Domain Behavior during the Folding of a Thermostable Phosphoglycerate Kinase

Martin J. Parker,*,‡ James Spencer,‡ Graham S. Jackson,‡ Steven G. Burston,‡,§ Laszlo L. P. Hosszu,|| C. Jeremy Craven,|| Jonathan P. Waltho,|| and Anthony R. Clarke‡

Department of Biochemistry, School of Medical Sciences, University of Bristol, University Walk, Bristol BS8 1TD, U.K. and Krebs Institute for Biomolecular Research, Department of Molecular Biology and Biotechnology, University of Sheffield, P.O. Box 594, Sheffield S10 2UH, U.K.

Received June 4, 1996; Revised Manuscript Received September 19, 1996[⊗]

ABSTRACT: *Bacillus stearothermophilus* phosphoglycerate kinase (bsPGK) is a monomeric enzyme of 394 residues comprising two globular domains (N and C), covalently linked by an interdomain α-helix (residues 170−185). The molecule folds to the native state in three stages. In the first, each domain rapidly and independently collapses to form an intermediate in which the N-domain is stabilized by 5.1 kcal mol^{-1} and the C-domain by 3.3 kcal mol^{-1} over their respective unfolded conformations. The N-domain then converts to a folded state at a rate of 1.2 s^{-1} (ΔG_{I-F} = 3.8 kcal mol^{-1}), followed by the C-domain at 0.032 s^{-1} (ΔG_{I-F} = 12.1 kcal mol^{-1}). It is this last step that limits the rate of acquisition of enzyme activity. In the dynamics of unfolding in water, the N-domain converts to the intermediate state at a rate of 8×10^{-4} s^{-1}, some 10^7 times faster than the C-domain. Consequently, the most populated intermediate in the folding reaction has a native-like N-domain, while that in the unfolding direction has a native-like C-domain. In a conventional sense, therefore, the folding/unfolding kinetics of bsPGK can be described as random order. Consistent with these observations, cutting the molecule in the interdomain helix produces two, independently stable units comprising residues 1−175 and 180−394. A detailed comparison of their folding behavior with that of the whole molecule reveals that true interdomain contacts are relatively weak, contributing ∼1.4 kcal mol^{-1} to the stability of the active enzyme. The only interactions which contribute to the stability of rapidly formed intermediates or to transition states along the productive folding pathways are those within domain cores. Contacts formed either between domains or with the interdomain helix are made only in the folded ground state, but do not constitute a separate step in the folding mechanism. Intriguingly, the most pronounced effect of interdomain contacts on the kinetics of folding is inhibitory; the presence of the C-domain appearing to reduce the effective rate of acquisition of native structure within the N-domain.

For methodological reasons and because they serve as the simplest model objects, small single-domain proteins have formed the focus of most folding studies (Kim & Baldwin, 1990; Creighton, 1992). There is less information, however, about the development of higher levels of organization in large proteins which are composed of well-defined structural domains. Although these proteins require specific interdomain interactions to maintain the native state, little is known of the strength of interdomain contacts at each stage in folding and of the part these interactions play in the folding mechanism.

Addressing this problem produces an immediate difficulty, that of defining the term "domain". A domain is usually described as a substructure within a protein with one or more of the following properties (Garel, 1992). (i) When isolated, it forms the same well-defined folded conformation as it does in the intact molecule. In this respect it has a high degree of structural autonomy, the energetic definition. (ii) It acts as a discrete genetic unit which can be identified in different proteins, the evolutionary definition. (iii) On examination of the global, three-dimensional fold, it appears as a distinct substructure, the morphological or topological definition. (iv) It has a particular mechanistic function within the protein, e.g., a unit within an enzyme which binds one of the reactants in a multisubstrate reaction, the functional definition.

At one extreme, there are proteins which fulfill all four criteria, but are composed of domains with no intimate noncovalent interactions necessary for their individual biological activities. In these, chain connectivity is used to maintain crude proximity (e.g., the type κ-immunological light chain (Tsunenga et al., 1987)). In other proteins, the domains interact more extensively and cannot fold as separate units, making it impossible to examine them in isolation (e.g., α-lactalbumin (Peng & Kim, 1994)).

Phosphoglycerate kinase (PGK)[1] is a more promising paradigm for studies of the folding of multi-domain proteins (Betton et al., 1984, 1985; Yon et al., 1988, 1990). This

* Author to whom correspondence should be addressed.
‡ University of Bristol.
|| University of Sheffield.
§ Present address: Howard Hughes Medical Institute Research Laboratories, Boyer Center for Molecular Medicine, 295 Congress Avenue, New Haven, CT 06510.
⊗ Abstract published in *Advance ACS Abstracts,* November 15, 1996.

[1] Abbreviations: 3-PGA, 3-phosphoglycerate; bsPGK, phosphoglycerate kinase from *Bacillus stearothermophilus*; bsPGK′, W290Y bsPGK; CD, circular dichroism; DTT, dithiothreitol; EDTA, ethylenediaminetetraacetic acid; GAPDH, glyceraldehyde-3-phosphate dehydrogenase from horse muscle; GuHCl, guanidinium hydrochloride; IPTG, isopropyl β-D-thiogalactoside; NMR, nuclear magnetic resonance; PCR, polymerase chain reaction; PGK, 3-phosphoglycerate kinase; PMSF, phenylmethanesulfonyl fluoride; SDS−PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; TEA, triethanolamine hydrochloride; Tris·HCl, tris(hydroxymethyl)methylamine hydrochloride.

S0006-2960(96)01330-X CCC: $12.00

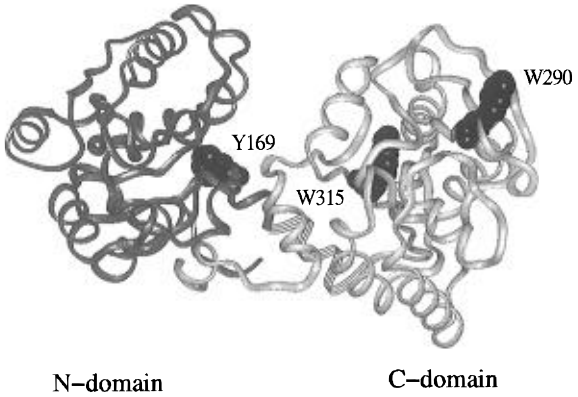

FIGURE 1: Domain structure of bsPGK. The structure of bsPGK (Davies et al., 1993) is represented here as a ribbon diagram (created using the program Insight II, Biosym Technologies Inc.). Residues 1−175 (isolated N-domain) and 186−394 (isolated C-domain) are rendered solid and colored dark and light gray, respectively. The ribbon encompassing residues 176−185, which forms part of the interdomain bridging helix (residues 170−185), is shown as threads. The side chains of the two wild-type tryptophans (W290 and W315; located in the C-domain) and the interdomain-located tyrosine Y169 are shown as space-filling structures.

protein adopts a bilobed structure in which the two domains (N- and C-domains) are separated by a "hinge" region (see Figure 1). Crystal structures are available for the horse muscle (Banks et al., 1979), yeast (Watson et al., 1982), pig muscle (Harlos et al., 1992), and *Bacillus stearothermophilus* (Davies et al., 1993) proteins and reveal that the structure is highly conserved (Watson & Littlechild, 1990). Essentially, the folds adopted by each domain are rather similar and may be outlined as a central, six-stranded parallel β-sheet flanked by amphipathic α-helices (the C-domain has two short additional β-strands on the face most distal from the N-domain). The sites for binding of the three-carbon substrate and the coenzyme are in the N- and C-domains, respectively. It has been proposed (Banks et al., 1979) that catalysis takes place by a "hinge-bending" mechanism, requiring movements of the domains about the interdomain region to facilitate an exact juxtaposition of reactants in the phosphoryl transfer reaction.

This protein satisfies all four of the above criteria but has a small interface between the domains (see Figure 1). Interdomain interactions are provided by two α-helices; a long helix which covalently connects the N- and C-domains (residues 170−185) and a short helix at the extreme C-terminus (residues 385−391). This last helix recrosses the domain interface to make contact with the N-domain. Despite these contacts, the molecule can be cleaved in the interdomain region to yield native-like fragments (Minard et al., 1989; Fairbrother et al., 1989; Missiakas et al., 1990, 1992; Hosszu et al., 1996).

In this study, we use the heat-stable PGK from *B. stearothermophilus* to examine the folding behavior of its N- and C-domains in the context of the intact protein and as isolated units. Equilibrium and kinetic studies are used to elucidate the development of inter- and intradomain interactions during folding.

METHODS

Experimental

Source Of Proteins. Mutants of the *B. stearothermophilus* PGK gene were made on the single tryptophan template,

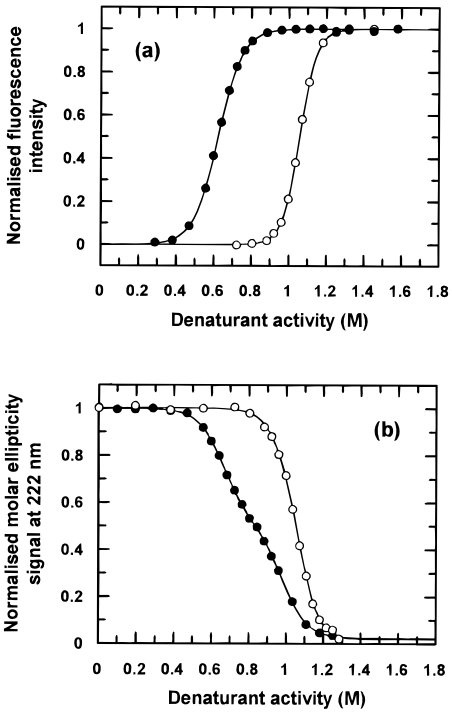

FIGURE 2: Equilibrium unfolding of intact bsPGK′. Shown in panel a are equilibrium unfolding profiles of bsPGK′ (open circles) and I217A+L221A bsPGK′ (closed circles), following the fluorescence of W315 as a function of the molar denaturant activity (eq 1; Analytical Procedures). For comparative purposes, the amplitudes have been normalized. Both sets of data have been fitted to a two-state equilibrium process (eq 2, Analytical Procedures), to yield values of $\Delta G_{F-U(w)}$ and Δm_{U-F} (see Table 1). Shown in panel b are equilibrium unfolding profiles for bsPGK′ (open circles) and I217A+L221A bsPGK′ (closed circles) using circular dichroism (CD). The fraction of the folded state [θ_{222nm}] is plotted as a function of the molar denaturant activity (see Analytical Procedures). Two two-state transitions can be resolved for I217A+L221A bsPGK′, resulting from the selective destabilization of the C-domain. The CD equilibrium unfolding profile for bsPGK′ does not fit satisfactorily to a single two-state process, suggesting that there are two, almost coincident transitions (in agreement with the kinetic results in Figure 5, which show that the folding/unfolding of the N- and C-domains are very weakly coupled and of random order). Therefore, both sets of data have been fitted to a relationship involving two independent, two-state equilibria (eq 3, Analytical Procedures) to yield values of $\Delta G_{F-U(w)}$ and Δm_{U-F} for the N- and C-domains (see Table 1). While the data for I217A+L221A bsPGK′ has been freely fitted, that for bsPGK′ has been constrained using $\Delta G_{F-U(w)}$ and Δm_{U-F} values for the C-domain derived from its fluorescence data in panel a. All fits to the data are shown as continuous curves.

W290Y (bsPGK′), and the proteins were expressed and purified as previously described (Staniforth et al., 1993). The isolated N-domain of bsPGK′ was produced by replacing the Leu 176 codon with a stop codon and the protein expressed and purified as previously described (Parker et al., 1995). The isolated C-domain of bsPGK′ was created using the polymerase chain reaction (PCR). An N-terminal oligonucleotide, corresponding to codons 180−190 of the bsPGK′ template, was designed to substitute codons 181 and 182 with an *Eco*RI restriction site and codon 184 with a start codon. A C-terminal oligonucleotide was designed so as to incorporate a *Hin*dIII restriction site downstream from the stop codon. These two oligonucleotides were used to prime 20 cycles of PCR off the bsPGK′ template DNA using *Taq* polymerase (Boehringer Mannheim) in a Perkin Elmer DNA thermal cycler. Cycle temperature steps used were 94 °C for 1 min (melting); 48 °C for 1 min (annealing); 72 °C for

2 min (polymerization). A single major band of expected size was obtained on a 1.0% low melting point agarose gel. This was excised, purified using the Pharmacia "Bandprep" kit, and digested for 1 h at 37 °C with 5 units each of *Eco*RI and *Hin*dIII (New England Biolabs) in manufacturer's buffer. After gel purification, the digested DNA was ligated into the fusion protein expression vector pMAL-c2 (New England Biolabs; *Eco*RI/*Hin*dIII cut and purified). XL1-Blue cells (Stratagene) were transformed by electroporation and pMAL-c2 clones selected for growth on media containing 0.5 mg mL^{-1} ampicillin (Beecham Research). Expression gels of recombinant clones, grown in NZCYM broth plus ampicillin at 37 °C and induced with 1 mM isopropyl β-D-thiogalactoside (IPTG) (Sigma Chemical Co.) at mid-log phase, revealed a band with an apparent molecular weight of 67 kDa (as expected from the predicted molecular weight for the C-domain of 24.4 kDa). This band comprises ~15% of total cell protein. The entire DNA of the bsPGK′ C-domain construct was sequenced to check for fidelity. For protein purification, IPTG-induced cells in NZCYM broth plus ampicillin were grown at 37 °C overnight with shaking. The cells were harvested by centrifugation, resuspended in a minimal volume of column buffer (20 mM tris(hydroxymethyl)methylamine hydrochloride (Tris·HCl) (pH 7.4), 200 mM NaCl, 1 mM ethylenediaminetetraacetic acid (EDTA), and 1 mM dithiothreitol (DTT); Boehringer Mannheim)), and lysed by sonication, and phenylmethanesulfonyl fluoride (PMSF) (Sigma) was added to 0.5 mM, to prevent proteolysis. Ammonium sulfate (Boehringer Mannheim) was then added to 50% saturation, and contaminating precipitants were removed by centrifugation. The ammonium sulfate concentration was then increased to 80% saturation and the protein pelleted by centrifugation. Protein pellets were resuspended in a minimal volume of column buffer and dialyzed overnight against 200 volumes of this buffer. The 67 kDa fusion protein was then purified by affinity chromatography using amylose resin (New England Biolabs), the fusion protein eluting to high purity with 10 mM maltose (Sigma), as assessed by SDS−PAGE. Fusion protein at 1.25 mg mL^{-1} was digested for 48 h at 25 °C with 0.1% w/w factor Xa (New England Biolabs) and then PMSF added to 0.5 mM to quench the reaction. The digest mixture was then dialyzed overnight against 200 volumes of 20 mM Tris·HCl (pH 7.5), 1 mM EDTA, and 1 mM DTT and loaded onto a Q-sepharose (Pharmacia) column equilibrated in the same buffer, and the protein eluted with a 0−0.4 M NaCl gradient. Those fractions containing pure C-domain protein, as assessed by SDS−PAGE, were pooled, precipitated in 80% ammonium sulfate, and stored at 4 °C. Gas phase amino acid sequencing of the N-terminal eight amino acids was performed to confirm that cleavage had taken place at the single expected site.

Oligonucleotides used for PCR and sequencing were synthesized on a Du Pont Coder 300 DNA synthesizer. Sequencing of DNA was performed by the chain termination procedure (Sanger et al., 1980), using a Du Pont Genesis 2000 automated sequencer.

Protein concentrations were estimated by UV absorption of aromatic residues at 280 nm (ϵ = 5500 M^{-1} cm^{-1} for tryptophan and 1100 M^{-1} cm^{-1} for tyrosine). Enzymatic activity was determined in a linked assay with horse muscle glyceraldehyde-3-phosphate dehydrogenase (GAPDH), as described by Bücher (1955), monitoring the conversion of NADH to NAD^+ by the change in absorbance at 340 nm in a Perkin Elmer spectrophotometer. The assay solution contained 50 mM triethanolamine hydrochloride (TEA) (pH 7.2), 20 mM 3-phosphoglycerate (3-PGA), 20 mM ATP, 20 mM $MgCl_2$, 0.2 mM NADH, and 0.1 mg mL^{-1} GAPDH. Assay reagents were obtained from Boehringer Mannheim.

Equilibrium Unfolding Measurements. For fluorescence studies, 2−4 μM protein was incubated in 50 mM TEA (pH 7.2) and increasing concentrations of ultrapure guanidinium hydrochloride (GuHCl) (Sigma) at 25 °C. Denaturation concentrations were checked by refractometry, using the method of Nozaki (1970). The unfolding reactions were allowed to equilibrate for approximately 4 h before fluorescence emission intensity was recorded in a Spex Fluoromax spectrofluorometer. For tryptophan fluorescence, an excitation wavelength of 295 nm was selected by a single excitation monochromator (slit width 2.5 nm) and emission intensity recorded at 345 nm using a single emission monochromator (slit width 5.0 nm). For tyrosine fluorescence, an excitation wavelength of 276 nm was selected (slit width 2.5 nm) and the emission intensity recorded at 340 nm (slit width 10.0 nm).

For circular dichroism (CD) measurements, 2−4 μM protein was incubated in 50 mM K_2HPO_4 (Boehringer Mannheim) (pH 7.2) and increasing concentrations of GuHCl at 25 °C. The unfolding reactions were allowed to equilibrate for 4 h before molar ellipticity ([θ], deg M^{-1} cm^{-1}) was recorded at 222 nm, using a xenon light source in a Jobin-Yvon CD6 spectrometer (cell path length 10 mm; slit width 1.0 nm; integration time 20 s).

Folding and Unfolding Rates. For the folding rates of the isolated C-domain and the C-domain in the intact PGK molecules, 20 μL of 200−300 μM protein in 50 mM TEA (pH 7.2) and 3 M GuHCl was injected into a rapidly stirred quartz cuvette, containing 2 mL of 50 mM TEA (pH 7.2) plus an appropriate concentration of GuHCl, in a Spex Fluoromax spectrofluorometer. Fluorescence emission intensity at 345 nm was recorded (slit width 5.0 nm) as a function of time (integration time 1−2 s) using an excitation wavelength of 295 nm (slit width 2.5 nm). For the unfolding rates, 20 μL of 200−300 μM protein in 50 mM TEA (pH 7.2) was injected into a rapidly stirred cuvette containing 2 mL of 50 mM TEA (pH 7.2) plus an appropriate concentration of GuHCl and fluorescence emission intensity recorded as described for the folding rates.

For the folding rates of the N-domain in the intact PGK molecules, 20−30 μM protein in 50 mM TEA (pH 7.2) and 3 M GuHCl was mixed against 10 volumes of 50 mM TEA (pH 7.2) plus an appropriate concentration of GuHCl in an Applied Photophysics stopped-flow apparatus. An excitation wavelength of 270 nm was selected by a single excitation monochromator (slit width 5.0 nm) and the total fluorescence emission intensity recorded as a function of time (time constant 100 μs). For unfolding rates, 2−3 μM protein in 50 mM TEA (pH 7.2) was mixed with an equal volume of 50 mM TEA (pH 7.2) plus an appropriate concentration of GuHCl. Determination of the folding and unfolding rates for the isolated N-domain has been described elsewhere (Parker et al., 1995).

Rate of Regain of Catalytic Activity. Five microliters of 2−5 μM PGK in 50 mM TEA (pH 7.2) and 3 M GuHCl was added to 1 mL of assay solution (see above) at 25 °C.

Table 1: Summary of Folding Parameters

	C-domain			N-domain		
	intact			intact		
parameter[a]	WT	MUT	isolated	WT	MUT	isolated
$K_{F/U(w)}^{EQ}$	5.5×10^9	1.5×10^4	2.7×10^5 [b]	6.7×10^6 [c]	1.1×10^7	9.3×10^6
	(5.0×10^8)	(6.5×10^3)	(1.2×10^5)	(9.8×10^5)	(2.0×10^6)	(1.4×10^6)
$\Delta G_{F-U(w)}^{EQ}$	−13.2	−5.7	−7.4[b]	−9.3[c]	−9.6	−9.5
(kcal mol^{-1})	(−11.8)	(−5.2)	(−6.9)	(−8.1)	(−8.6)	(−8.3)
Δm_{U-F}^{EQ}	−20.9	−15.5	−19.6[b]	−15.7[c]	−16.7	−16.6
(M^{-1})	(−15.2)	(−12.9)	(−15.9)	(−12.0)	(−13.0)	(−12.0)
$K_{F/U(w)}^{KIN}$ [d]	1.3×10^{10}	1.5×10^4		5.0×10^6	5.4×10^6	3.3×10^6
	(2.5×10^8)	(5.0×10^3)		(2.7×10^6)	(2.6×10^6)	(2.0×10^6)
$\Delta G_{F-U(w)}^{KIN}$ [d]	−13.7	−5.7		−9.1	−9.1	−8.9
(kcal mol^{-1})	(−11.4)	(−5.0)		(−8.7)	(−8.7)	(−8.6)
$k_{I-F(w)}$	0.028	0.009	0.019	1.10	0.98	9.50
(s^{-1})	(0.029)	(0.009)	(0.019)	(1.10)	(0.98)	(9.50)
$k_{F-I(w)}$	7.6×10^{-10}	2.0×10^{-6}	3.6×10^{-5}	0.0013	0.0014	0.015
(s^{-1})	(2.0×10^{-8})	(5.4×10^{-6})	(9.6×10^{-5})	(0.0027)	(0.0030)	(0.030)
$K_{I/U(w)}$	300	3.4	240[e]	5.9×10^3	7.7×10^3	5.3×10^3
	(170)	(3.0)	(130)	(6.7×10^3)	(8.0×10^3)	(6.3×10^3)
$\Delta G_{I-U(w)}$	−3.4	−0.7	−3.2[e]	−5.1	−5.3	−5.1
(kcal mol^{-1})	(−3.0)	(−0.6)	(−2.9)	(−5.2)	(−5.3)	(−5.2)
Δm_{F-tu}	12.0	6.9	10.2	5.0	4.6	3.1
(M^{-1})	(8.0)	(5.1)	(7.8)	(3.7)	(3.8)	(2.1)
Δm_{U-I}	−8.8	−5.1	−7.4[e]	−10.5	−11.3	−11.1
(M^{-1})	(−7.1)	(−5.0)	(−6.0)	(−9.2)	(−10.0)	(−9.7)
Δm_{I-tf}	−0.13	−3.4	−2.0	−0.3	−0.5	−1.6
(M^{-1})	(−0.10)	(−3.5)	(−2.1)	(−0.3)	(−0.5)	(−1.6)

[a] EQ, values obtained from equilibrium measurements; KIN, values obtained from kinetic measurements; WT, bsPGK′; MUT, I217A+L221A bsPGK′; isolated N-domain, residues 1−175 of bsPGK′; isolated C-domain, residues 185−394 of bsPGK′. Values given in parentheses were calculated using raw GuHCl concentration, as opposed to a linearized activity scale (see Analytical Procedures). [b] $K_{F/U(w)} = K_{F/I(w)}K_{I/U(w)}$; $\Delta m_{U-F} = \Delta m_{I-F} + \Delta m_{U-I}$ (obtained from the three-state, equilibrium fluorescence unfolding data (see Figure 6, panel a)): $K_{F/I(w)} = 1125$ (923) and $\Delta m_{I-F} = -12.2$ (−9.9) M^{-1}. The kinetic parameters for the intact N-domains of bsPGK′ and I217A+L221A bsPGK′ were calculated by fitting the rate profiles to eq 5, Analytical Procedures. The kinetic data for the intact N-domain of bsPGK′ has also been fitted to eq 9, Results, which involves an additional off-pathway, inhibitory equilibrium (see Results for discussion and revised, calculated values). Average errors in equilibium constants, rate constants and Δm values do not exceed 20%, 10%, and 15%, respectively. [c] Fit constrained using $K_{F/U(w)}$ and Δm_{U-F} derived from the equilibrium fluorescence data for the intact C-domain of bsPGK′ (see legend to Figure 2). [d] $K_{F/U(w)}^{KIN} = [k_{I-F(w)}/k_{F-I(w)}]K_{I/U(w)}$. [e] Values obtained from equilibrium CD data (see Figure 6, panel a).

After rapid and vigorous mixing, the absorption at 340 nm was recorded as a function of time in a Perkin Elmer spectrophotometer.

For all measurements, solutions were maintained at 25 °C using thermostated circulating water baths.

Analytical Procedures

It is well established that the change in solvation free energy of protein hydrocarbon moieties (Tanford, 1970; Pace, 1975) and the free energy change of protein unfolding (Staniforth et al., 1993; Johnson & Fersht, 1995) have a nonlinear dependence on denaturant concentration. For the equilibrium and kinetic analyses described in this paper, GuHCl concentration ([GuHCl]) is converted to molar denaturant activity (D), according to the relationship

$$D = C_{0.5}[\text{GuHCl}]/(C_{0.5} + [\text{GuHCl}]) \qquad (1)$$

where $C_{0.5}$ is an empirically derived denaturation constant with the value 7.5 M (Parker et al., 1995). Where high denaturant concentrations are required to unfold a protein, this linearized scale allows a more reliable extrapolation of data to a condition where $D = 0$. This is particularly important for the evaluation of unfolding rates and equilibrium constants in water and for determining proper m values (see below). It is worth pointing out that a direct calorimetric evaluation of the dependence of the free energy of protein unfolding on denaturant concentration, performed by Johnson and Fersht (1995), provides a molar denaturation scale in very close agreement to ours, derived using solubility data. Nonetheless, for the sake of comparison, we provide values for equilibrium and kinetic parameters calculated using GuHCl concentration as well as denaturant activity (see Table 1).

Treatment Of Equilibrium Data. Molar ellipticity at 222 nm ($[\theta_{222}]$) is used qualitatively as a measure of the degree of secondary structure (primarily α-helix) in the folded state. Equilibrium unfolding data measured by CD are presented here as the fraction of the fully folded state $[\theta_{222}]$ versus denaturant activity.

For single equilibrium unfolding transitions (i.e., a two-state, folded (F) to unfolded (U) transition), data were fitted to the equation

$$I = \alpha_F I_F + \alpha_U I_U \qquad (2)$$

where α_F and α_U are the fractional proportion of molecules in the folded and unfolded states, respectively, and I, I_F, and I_U are signals (measured, folded, and unfolded, respectively). The following relations (temporary variables) were used in the fitting procedure:

$$K_{F/U} = K_{F/U(w)} \exp(\Delta m_{U-F} D)$$

$$\alpha_F = K_{F/U}/(1 + K_{F/U})$$

$$\alpha_U = 1 - \alpha_F$$

where $K_{F/U}$ and $K_{F/U(w)}$ are equilibrium constants (F/U) at a

given denaturant activity (D) and in water, respectively, and Δm_{U-F} describes the sensitivity of the equilibrium to the denaturant activity.

For two, parallel and independent unfolding reactions (as in the unfolding of independent domains, A and B, with distinct optical changes for each), data were fitted to the equation

$$I = \alpha_{F(A)}I_{F(A)} + \alpha_{U(A)}I_{U(A)} + \alpha_{F(B)}I_{F(B)} + \alpha_{U(B)}I_{U(B)} \quad (3)$$

where $\alpha_{F(A)}$ and $\alpha_{U(A)}$ are the fractional proportion of molecules with folded and unfolded A, respectively, and $\alpha_{F(B)}$ and $\alpha_{U(B)}$ the fractional proportion of molecules with folded and unfolded B, respectively. Temporary variables used in the fitting procedure were as described for single equilibrium transitions, but relate to each structural unit.

For a three-state mechanism (i.e., U = I = F; where I is an equilibrium intermediate), data were fitted to the equation

$$I = I_F\alpha_F + I_I\alpha_I + I_U\alpha_U \quad (4)$$

The following temporary variables were used in the fitting procedure:

$$K_{F/I} = K_{F/I(w)} \exp(\Delta m_{I-F}D)$$

$$K_{I/U} = K_{I/U(w)} \exp(\Delta m_{U-I}D)$$

$$\alpha_F = 1/(1 + (1/K_{F/I}) + (1/K_{F/I}K_{I/U}))$$

$$\alpha_I = 1/(1 + K_{F/I} + (1/K_{I/U}))$$

$$\alpha_U = 1/(1 + K_{F/I}K_{I/U} + K_{I/U})$$

where $K_{F/I}$ and $K_{I/U}$ are equilibrium constants (F/I and I/U, respectively) and Δm_{I-F} and Δm_{U-I} describe the sensitivity of these equilibrium processes to the denaturant activity (D), respectively.

Treatment of Kinetic Data. For transients of fluorescence intensity (I) versus time, for single relaxation processes, data were fitted to the equation $I = I_a(1 - \exp(-kt)) + I_0$ for rising intensities (where I_a is the fluorescence amplitude of the reaction, k is the observed rate constant for the relaxation, and I_0 is the initial intensity) and $I = I_a \exp(-kt) + I_f$ for decreasing intensities (where I_f is the final fluorescence intensity). For two independent, random-order processes, as in the folding and unfolding reactions monitored by tyrosine fluorescence in the intact PGK molecules (see Figure 3, panel c), transients were fitted to the equation $I = I_{a1}(1 - \exp(-k_1t)) + I_{a2}(1 - \exp(-k_2t)) + I_0$ for rising intensities (where I_{a1} and I_{a2} are the fluorescence amplitudes of reactions 1 and 2, respectively, and k_1 and k_2 are the observed rate constants for the relaxation processes of reactions 1 and 2, respectively) and $I = I_{a1} \exp(-k_1t) + I_{a2} \exp(-k_2t) + I_f$ for decreasing intensities.

Linear free energy profiles (observed rate constant (k_{obs}) versus denaturant activity (D)) were fitted to the equation

$$k_{obs} = k_{F-I} + k_{I-F}/(1 + 1/(K_{I/U})) \quad (5)$$

(Parker et al., 1995), where k_{I-F} and k_{F-I} are rate constants describing the forward and reverse transitions between the folded and intermediate states, respectively, and $K_{I/U}$ is the equilibrium constant for the rapid interconversion of the intermediate and unfolded state. The following temporary variables were used in the fitting procedure:

$$k_{F-I} = k_{F-I(w)} \exp(\Delta m_{F-tu}D)$$

$$k_{I-F} = k_{I-F(w)} \exp(\Delta m_{I-tf}D)$$

$$K_{I/U} = K_{I/U(w)} \exp(\Delta m_{U-I}D)$$

where the subscript "w" denotes the value of the constant in water and Δm values describe how the relative free energies of states associated with a particular transition (denoted by the subscripts) vary as a function of the molar denaturant activity (tu, transition state for unfolding; tf, transition state for folding). As discussed in the Results, this form of the equation, unlike that previously described by Parker et al. (1995), treats the forward and reverse reactions separately.

Rate of Regain of Enzyme Activity. For a single relaxation folding process (U → F; where U and F are catalytically inactive and active forms of the protein, respectively) with a rate constant k, the amount of catalytically active protein at time t (F_t) is given by

$$F_t = F_{FIN}(1 - \exp(-kt))$$

where F_{FIN} is the amount of catalytically active protein present when the folding reaction has reached completion. Similarly, the catalytic activity at time t (A_t) is given by

$$A_t = A_{FIN}(1 - \exp(-kt))$$

where A_{FIN} is the activity present in the U → F reaction when it has come to completion. Activity is expressed as dOD_{340}/dt, (where OD_{340} is the optical density at 340 nm); therefore

$$(dOD_{340}/dt)_{(t)} = (dOD_{340}/dt)_{(FIN)}(1 - \exp(-kt))$$

where $(dOD_{340}/dt)_{(FIN)}$ is the final activity. Integrating this expression between the limits of $t = 0$ and $t = t'$ gives

$$OD_{340(t=t')} - OD_{340(t=0)} = (dOD_{340}/dt)_{(FIN)}(t' - (1/k) + (1/k) \exp(-kt')) \quad (6)$$

where $OD_{340(t=t')}$ and $OD_{340(t=0)}$ are the absorbance at time t and time 0, respectively. For progress curves of OD_{340} versus time, data are fitted directly to this expression.

All data were fitted using the Grafit analysis software (Erithracus software, U.K.). When kinetic data were fitted to eq 5, proportional weighting was used, so that the fitted values took account of rate constants equally across the whole range.

RESULTS

Origin of Spectroscopic Signals in the Intact Molecule. The parent molecule used in these studies is designated bsPGK′ and has one of the two wild-type tryptophans, at position 290 on the surface of the C-domain, replaced by tyrosine. This leaves the buried tryptophan, W315, as the sole indole group for reporting structural changes within the C-domain (see Figure 1). This conservative mutation has little influence on the stability of the folded state ($\Delta\Delta G_{F-U}$ = 0.6 kcal mol^{-1}), and the resultant protein has a catalytic activity indistinguishable from that of the wild-type (data not shown). Complete unfolding of bsPGK′, induced by high concentrations of the denaturant GuHCl, results in a 2.5-fold increase in the indole fluorescence intensity. The

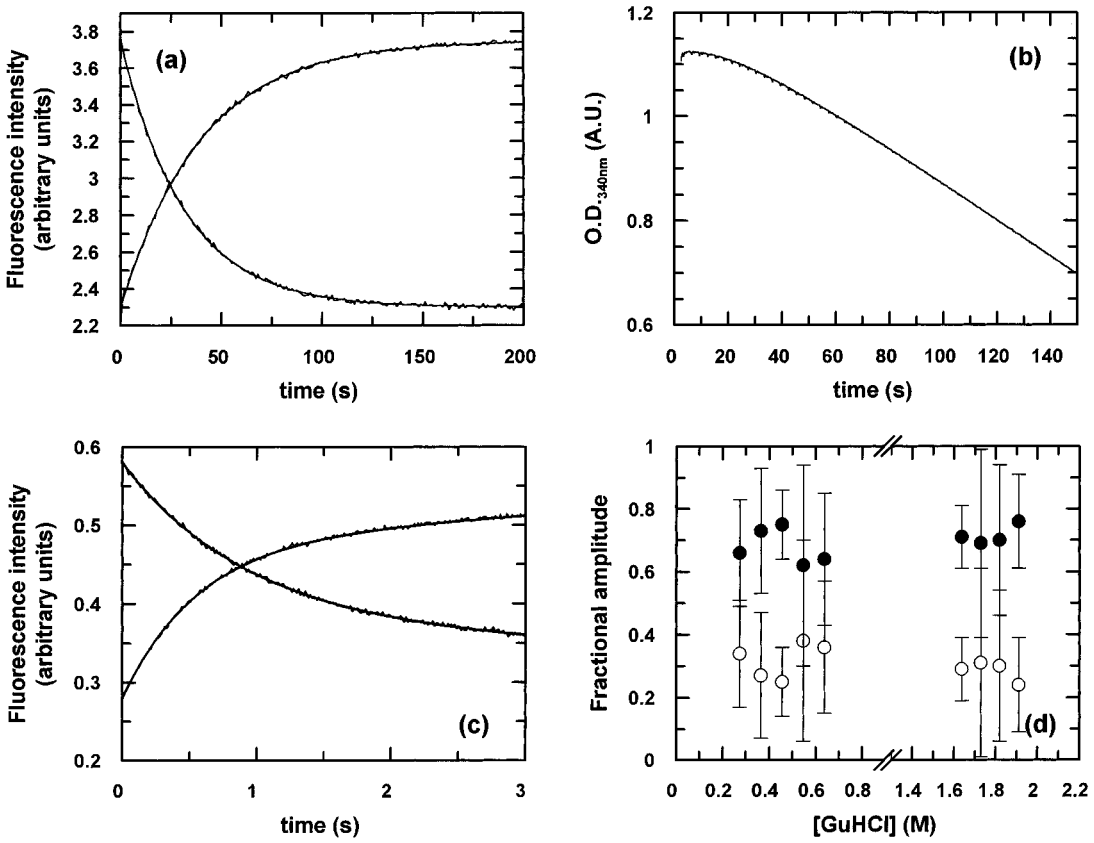

FIGURE 3: Folding and unfolding kinetics of intact bsPGK′. Shown in panel a are representative folding (fluorescence decrease) and unfolding (fluorescence increase) kinetic transients for bsPGK′ in 0.03 and 1.80 M GuHCl, respectively, following the fluorescence emission intensity of W315 at 345 nm, using an excitation wavelength of 295 nm. The calculated first-order rate constant (see Analytical Procedures) for folding is 0.028 s^{-1}, while that for unfolding is 0.025 s^{-1}. In panel b, bsPGK′ is refolded in an activity assay solution (see Experimental Procedures) and the absorbance at 340 nm (OD_{340nm}) measured as a function of time. The data has been fitted to eq 6 (see Analytical Procedures) to yield a first-order rate constant for folding of 0.032 s^{-1}. In panel c, representative folding (0.27 M GuHCl) and unfolding (1.80 M GuHCl) kinetic transients for bsPGK′ are shown, following total fluorescence emission intensity, using an excitation wavelength of 270 nm. The folding transient represents a double-exponential decrease in intensity (see Analytical Procedures) and yields first-order rate constants of 1.2 and 0.034 s^{-1}. The unfolding transient represents a double-exponential increase in intensity and yields first-order rate constants of 1.8 and 0.025 s^{-1}. These fast and slow kinetic transients are used to construct the rate profiles for the N- and C-domains, respectively, shown in Figure 5. Fits to the data are shown as continuous curves. In panel d the fractional amplitudes of the transients in panel c have been plotted against final GuHCl concentration, well above and below the transition zone (open circles, fast transient; closed circles, slow transient). That these fractional amplitudes remain constant is consistent with the folding of the N- and C-domains occurring in a random order (Ikai & Tanford, 1973).

equilibrium unfolding profile of bsPGK′, monitored by the fluorescence of W315, is effectively two-state (see Figure 2, panel a), while that measured by circular dichroism (CD) is complicated by the near coincident unfolding transition of the N-domain (see Figure 2, panel b).

Both folding and unfolding kinetics of the C-domain of bsPGK′, reported by the fluorescence of W315, are single-exponential processes (see Figure 3, panel a). The folding rate at 0.03 M GuHCl is 0.028 s^{-1}, while at 1.80 M GuHCl the unfolding rate is 0.025 s^{-1}. This folding rate is identical to the rate of regain of enzyme activity (0.028 s^{-1}; see Figure 3, panel b), demonstrating that the folding of the C-domain limits the rate of acquisition of the fully native enzyme. Similar conclusions have been made for both yeast and pig muscle PGK (Betton et al., 1985; Missiakas et al., 1992). In double-jump experiments (Brandts et al., 1975), where the molecule is unfolded at low temperature (0 °C) and then refolded at 25 °C after a series of delay times from 20 to 5000 s, no change is seen in the amplitude or rate associated with this refolding transient, showing that this process is not limited by cis−trans isomerization of proline residues (data not shown).

If the same folding/unfolding reactions are monitored using an excitation wavelength of 270 nm, with no emission filter, the additional contribution of tyrosine to the total fluorescence emission intensity can be observed. The folding reaction (at 0.27 M GuHCl) now produces two exponential decreases in intensity with rates of 1.2 and 0.034 s^{-1}, whereas the unfolding reaction (at 1.8 M GuHCl) produces two exponential increases in intensity with rates of 1.8 and 0.025 s^{-1} (see Figure 3, panel c). The slow folding and unfolding rates are identical to those obtained by monitoring solely the indole fluorescence of W315 in the C-domain. We interpret the fast rates as arising from the N-domain, where, presumably, the tyrosine fluorescence (arising from four residues in the N-domain) is not as heavily quenched by the close proximity of a tryptophan. This interpretation is confirmed by comparing the equilibrium and kinetic folding behavior of the isolated domains (see below). In addition, when W315 is replaced by tyrosine, to produce a protein with no indole groups, the equilibrium unfolding profile measured by tyrosine fluorescence reveals two transitions (see Figure 4, panel a). The fluorescence enhancement caused by unfolding the N-domain, which previously was only seen in the kinetic measurements, is now visible at equilibrium: replacing the internal tryptophan with tyrosine selectively destabilizes the C-domain, which now unfolds (with a resultant decrease in tyrosine fluorescence) at a lower

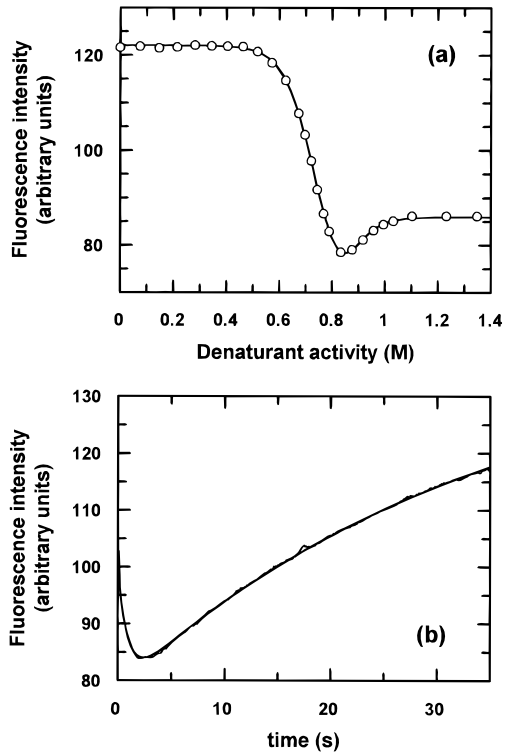

FIGURE 4: Equilibrium and kinetic folding behavior of a tryptophanless PGK. The two wild-type tryptophans (see Figure 1) have been replaced with tyrosine residues to produce a tryptophanless PGK molecule (W^- bsPGK). The equilibrium unfolding profile of W^- bsPGK is shown in panel a, where tyrosine fluorescence intensity (see Experimental Procedures) is plotted as a function of the calculated molar denaturant activity (eq 1, Analytical Procedures). Analogous to the effect of the I217A+L221A mutation on bsPGK′ (see Figure 2), mutation of the wild-type tryptophans selectively destabilizes the C-domain, such that two distinct unfolding transitions can be resolved. The calculated values of $\Delta G_{F-U(W)}$ and m_U, using eq 3 (see Analytical Procedures), are -7.3 ± 0.1 kcal mol^{-1} and -16.7 ± 0.2 M^{-1} for the C-domain and -9.9 ± 0.2 kcal mol^{-1} and -19.5 ± 0.3 M^{-1} for the N-domain. A kinetic transient for the folding of W^- PGK, following tyrosine fluorescence (see Experimental Procedures), in 0.03 M GuHCl is shown in panel b. This transient fits to a double-exponential process (see Analytical Procedures) to yield first-order rate constants of 1.1 (fluorescence decrease) and 0.029 s^{-1} (fluorescence increase). The relative amplitudes of the fast and slow transients correlate, respectively, with the relative amplitudes associated with the equilibrium unfolding processes of the N- and C-domains shown in panel a. These rates correspond to those measured by following the total fluorescence emission intensity of bsPGK′ (see Figure 3). Fits to the data are shown as continuous curves.

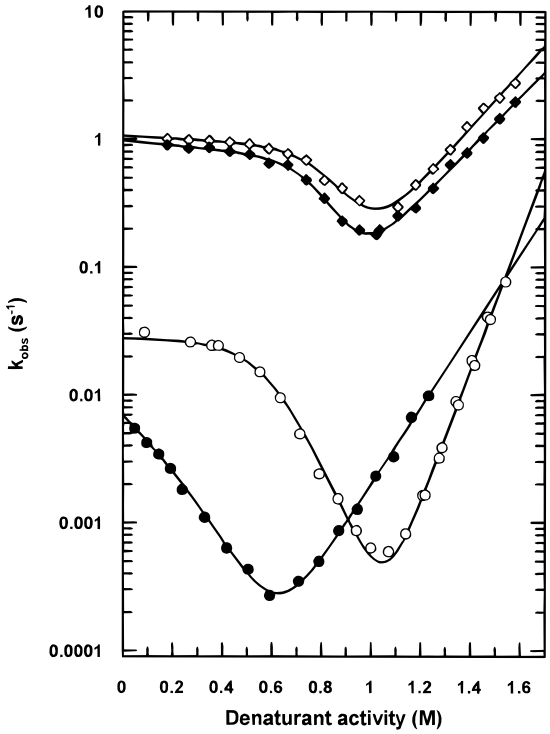

FIGURE 5: Reaction rate profiles for the folding/unfolding processes of the N- and C-domains. The observed rate constants (k_{obs}) for the folding/unfolding processes of the N- and C-domains (see Figure 3) are plotted as a function of molar denaturant activity (eq 1, Analytical Procedures): open circles, bsPGK′ C-domain; closed circles, I217A+L221A bsPGK′ C-domain; open diamonds, bsPGK′ N-domain; closed diamonds, I217A+L221A bsPGK′ N-domain. The data have been fitted to a three-state relaxation system (illustrated in Scheme 2, Results) using eq 5, Analytical Procedures. Fits are shown as continuous curves, and the calculated values of $k_{I-F(w)}$, $k_{F-I(w)}$, $K_{I/U(w)}$, Δm_{U-I}, Δm_{I-tf}, and Δm_{F-tu} (see Analytical Procedures and Results) are given in Table 1. The equilibrium constant for the full F-to-U process ($K_{F/U(w)}$) and its linear free energy slope (Δm_{U-F}) can be calculated from the following kinetic data: $K_{F/U(w)} = [k_{(I-F(W))}/k_{(F-I(w))}]K_{I/U(w)}$; $\Delta m_{U-F} = \Delta m_{U-I} + \Delta m_{I-tf} - \Delta m_{F-tu}$. The close agreement of $K_{F/U(w)}$ and Δm_{U-F} derived from the kinetic and equilibrium analyses for the N- and C-domains in both proteins (see Table 1) demonstrates that the kinetic and equilibrium data are mutually consistent.

denaturant concentration. The folding of this tryptophanless mutant in 0.27 M GuHCl reveals two reaction phases; an intensity decrease at a rate of 1.1 s^{-1} and an increase at 0.027 s^{-1} (see Figure 4, panel b). Plotting the amplitudes of these transients as a function of GuHCl concentration produces an unfolding profile identical to that obtained at equilibrium (data not shown). These fast and slow rates match those obtained by monitoring the total fluorescence emission intensity of bsPGK′, corresponding to the folding processes of the N- and C-domains, respectively.

Domain Behavior in the Intact Molecule. Although the results in Figure 2 reveal that the N- and C-domains of bsPGK′ have near coincident unfolding transitions at equilibrium, the kinetic results described above show that folding/unfolding steps in the two domains do not occur in a compulsory order (see Figure 3, panels a and c, and Figure 5). In the absence of denaturant, the folding rate of the N-domain (1.2 s^{-1}) is some 40 times faster than the C-domain (0.028 s^{-1}). In the unfolding process, at 1.8 M GuHCl (activity = 1.45 M), structure is lost in the N-domain at a rate of 1.8 s^{-1}, some 70 times faster than in the C-domain (0.025 s^{-1}). Thus, the most long-lived and stable intermediate in the folding reaction has a native-like N-domain, while that in the unfolding reaction has a native-like C-domain. In a conventional sense, therefore, the folding/unfolding kinetics of the N- and C-domains in this molecule can be described as random-order, as shown below:

Scheme 1

unfolding:

$$N_{folded}C_{folded} \xrightarrow{1} N_{denatured}C_{folded} \xrightarrow{2} N_{denatured}C_{denatured}$$

folding:

$$N_{denatured}C_{denatured} \xrightarrow{3} N_{folded}C_{denatured} \xrightarrow{4} N_{folded}C_{folded}$$

Note that the subscript "denatured" is used to describe both the unfolded state and rapidly formed, partially folded intermediate states.

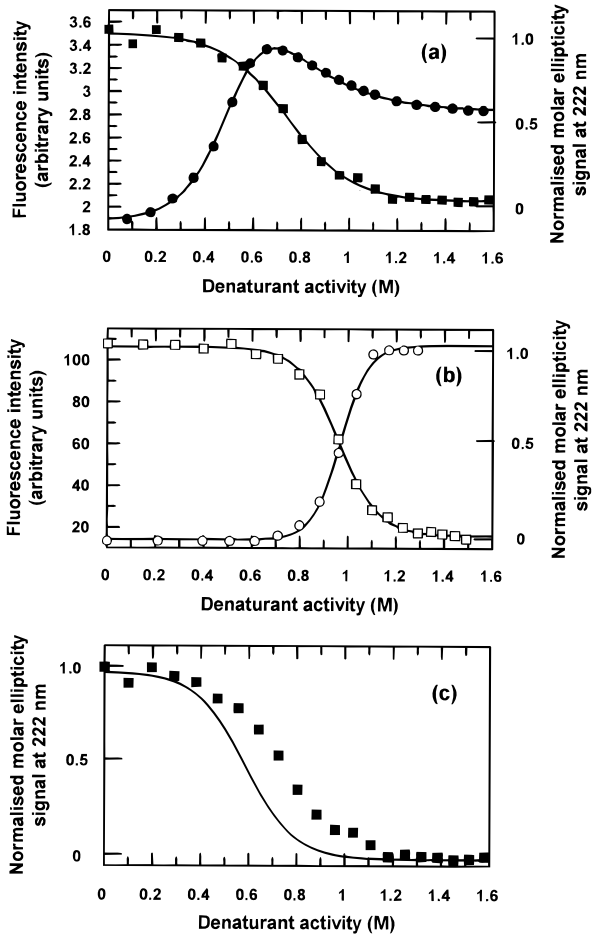

FIGURE 6: Equilibrium unfolding of the isolated N- and C-domains of bsPGK′. Shown in panel a are equilibrium unfolding profiles for the isolated C-domain of bsPGK′ (residues 185−394), monitoring the fluorescence of W315 (closed circles) and the fraction of the folded state [θ_{222nm}] (closed squares; see Analytical Procedures). The fluorescence data have been fitted to a process involving two coupled equilibria (U = I = F; eq 4, Analytical Procedures), whereas the CD data has been fitted to a two-state process (U = I; eq 2, Analytical Procedures). The calculated values of $\Delta G_{F-I(w)}$, $\Delta G_{I-U(w)}$, Δm_{I-F}, and Δm_{U-I} are given in Table 1. The biphasic fluorescence equilibrium unfolding profile for the isolated C-domain is insensitive to protein concentration, and light-scattering measurements also suggest that neither of the two fluorescence transitions involve aggregation. Shown in panel b are equilibrium unfolding profiles for the isolated N-domain (residues 1−175) monitoring tyrosine fluorescence (open circles) and the fraction of the folded state [θ_{222nm}] (open squares). Both sets of data have been fitted to a two-state process (eq 2, Analytical Procedures), and the values of ΔG_{F-U} and Δm_{U-F} are given in Table 1. Fits to the data are shown as continuous curves. In panel c, the CD data for the isolated C-domain is compared to an unfolding profile for the kinetic intermediate of the intact C-domain. This latter profile was constructed from the values of $\Delta G_{I-U(w)}$ and Δm_{U-I}, derived from the analysis of the intact C-domain kinetic data (see Figure 5 and Table 1). The correspondence of these two curves illustrates the similarity in the energetic properties of the isolated C-domain equilibrium folding intermediate and the intact C-domain kinetic folding intermediate.

The observed folding/unfolding kinetics for the N- and C-domains of bsPGK′, as a function of denaturant activity (see Analytical Procedures), are plotted in Figure 5. The relaxation rates for the two domains, particularly in the transition zone, differ roughly by 2 orders of magnitude so, although the system relaxes to a mixture of species, the folding and unfolding processes of the two domains are kinetically uncoupled. These rate profiles show distinct "kinks" in their folding limbs, demonstrating that each

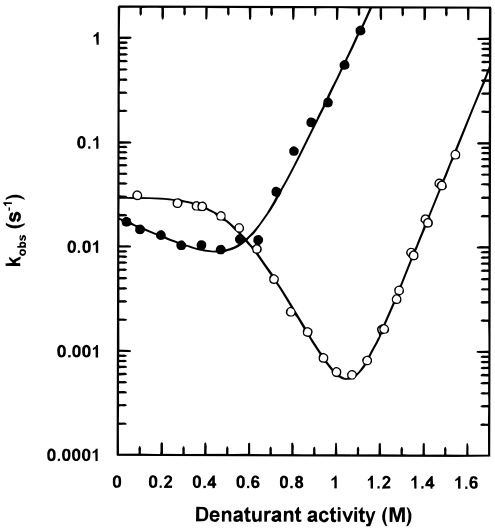

FIGURE 7: Reaction rate profile for the isolated C-domain. The observed rate constant (k_{obs}) for the rate-limiting folding/unfolding kinetics of the isolated C-domain (closed circles), following the fluorescence of W315 (see Experimental Procedures), is plotted as a function of denaturant activity (eq 1, Analytical Procedures). For comparison, the rate profile for the intact C-domain of bsPGK′ from Figure 5 is also shown (open circles). In accordance with the equilibrium unfolding results (see Figure 6 panel a and Table 1), the intermediate species is fully populated throughout the folding limb of the rate profile, such that the associated ground states sampled in this kinetic relaxation system are I and F. Thus, the data has been fitted to the two-state kinetic equation $k_{obs} = k_{F-I(w)} \exp(\Delta m_{F-tu}D) + k_{I-F(w)} \exp(\Delta m_{I-tf}D)$. The fit to the data is shown as a continuous curve and the calculated values of $k_{F-I(w)}$, $k_{I-F(w)}$, Δm_{F-t}, and Δm_{I-t} are given in Table 1.

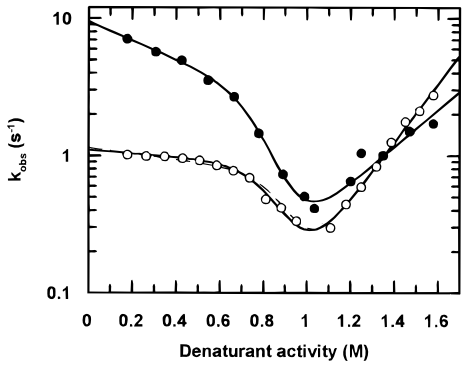

FIGURE 8: Reaction rate profile for the isolated N-domain. The observed rate constant (k_{obs}) for the rate-limiting folding/unfolding kinetics of the isolated N-domain, following tyrosine fluorescence, is plotted as a function of denaturant activity (closed circles; taken from Parker et al., 1995). Also shown is the rate profile for the intact N-domain of bsPGK′ from Figure 5, for comparison (open circles). Both sets of data have been fitted to eq 5, Analytical Procedures, describing a three-state relaxation system (see Scheme 2, Results). The fits to the data are shown as continuous curves, and the calculated values are given in Table 1. The value of $k_{I-F(W)}$ for the N-domain in the intact molecule is approximately 8 times lower than that for the isolated N-domain. As discussed in the text, the most likely explanation for this is the reversible formation of inhibitory contacts with the partially folded C-domain, such that the population available to fold is effectively reduced. Therefore, the data for the intact N-domain has also been fitted to eq 3 in Results, using the values of $k_{I-F(w)}$, $K_{I/U(w)}$, Δm_{I-tf}, and Δm_{U-I} derived from the kinetic data for the isolated N-domain, as a constraint. The fit to the data using this equation is shown as a dashed line. See Results for calculated values.

domain folds through a rapidly formed intermediate state that is populated only in strong folding conditions (Parker et al., 1995, 1996; Baldwin, 1996). The folding processes for each domain, therefore, can be described by the following

reaction scheme:

Scheme 2

$$\mathrm{U} \underset{}{\overset{K_{\mathrm{I/U}}}{=}} \mathrm{I} \underset{k_{\mathrm{F-I}}}{\overset{k_{\mathrm{I-F}}}{=}} \mathrm{F}$$

where U, I, and F represent the unfolded, intermediate, and folded states, respectively. The observed relaxation rate (k_{obs}) is then given by

$$k_{obs} = k_{F-I} + k_{I-F}/(1 + 1/K_{I/U}) \quad (7)$$

(Parker et al., 1995), where k_{F-I} and k_{I-F} are the unfolding and folding rates, respectively, and $K_{I/U}$ describes the rapid equilibrium ratio I/U. Although such data cannot *a priori* distinguish between a mechanism involving an on-pathway and an off-pathway intermediate (Baldwin, 1996), two lines of experimental evidence support the contention that the intermediate states of the N- and C-domains are productive. Firstly, $^1H/^2H$ pulsed protection studies of the isolated N-domain reveal amide proton protection patterns in the intermediate state that are remarkably native-like (J. P. Waltho and A. R. Clarke, unpublished data). Secondly, thermodynamic and kinetic data collected for a series of bsPGK mutants suggest that the intermediate state of the C-domain in the intact molecule has the properties expected for an on-pathway species, i.e., contacts formed in the intermediate are native-like (Parker et al., 1996).

From the random-order reaction mechanism illustrated in Scheme 1, it is clear that an analysis of the kinetics is more complicated than that applicable either to a single-domain protein or to a compulsory-order two-domain mechanism. For instance, the experimentally measured folding reaction of the N-domain in Scheme 1 (step 3; $N_{denatured}C_{denatured} \rightarrow N_{folded}C_{denatured}$) is not the reverse of the measured unfolding reaction in denaturing conditions (step 1; $N_{folded}C_{folded} \rightarrow N_{denatured}C_{folded}$). The same lack of equivalence applies also to the C-domain in the intact molecule. Thus, although both sets of data have been fitted to the relationship described in eq 7, the values of k_{F-I} and k_{I-F} do not necessarily reflect forward and backward rates in a common equilibrium relaxation process, so they must be interpreted separately. The sensitivity of these rates to the denaturant activity is therefore described by

$$k_{obs} = k_{F-I(w)} \exp(\Delta m_{F-tu} D) + k_{I-F(w)} \exp(\Delta m_{I-tf} D)/(1 + (1/(K_{I/U(w)} \exp(\Delta m_{U-I} D)))) \quad (8)$$

where the subscript (w) represents the rate or equilibrium constant in water and D is the denaturant activity which modifies these values. Subscripts U, I, and F represent the unfolded, intermediate, and folded states of the domain and tf and tu the rate-limiting transition states for the I-to-F transition in the folding and unfolding directions, respectively. Values of Δm show how the stabilities of the given states vary as a function of the denaturant activity (Parker et al., 1995). These values provide a qualitative measure of the change in the exposure of nonpolar (i.e., hydrocarbon) groups between particular states in the folding reaction; a positive value signifies an increase in solvent exposure (Tanford, 1970; Shortle et al., 1988; Staniforth et al., 1993; Parker et al., 1996). Hence, Δm_{F-tu} describes this relationship for the activation barrier for unfolding while Δm_{I-tf} describes this relationship for folding.

Values derived for both domains, by fitting their rate profiles to eq 8, using both GuHCl concentration and the calculated denaturant activity (see Analytical Procedures), are given in Table 1 and demonstrate that the intermediate states formed by each domain are distinct from each other, the N-domain intermediate being 1.7 kcal mol^{-1} more stable than that formed by the C-domain intermediate. The formation of these states, therefore, is unlikely to occur by a uniform and random collapse of the polypeptide chain.

Given that each domain folds via a distinct intermediate species and that the folding/unfolding of the two domains is of random-order, then one must consider a thermodynamic cycle containing nine possible free energy states (see Figure 9). In order to establish the development of interdomain interactions during folding, we need to know to what extent the domains are energetically coupled in each of these states.

Mutational Effects Do Not Cross between Domains. A core mutation which significantly destabilizes the C-domain of bsPGK′ (I217A+L221A) splits the normally coincident equilibrium unfolding transitions of the two domains (see Figure 2). In this core mutant, the C-domain unfolds at a much lower GuHCl concentration, but the unfolding transition associated with the N-domain is not perceptibly moved. Indeed, the values of $K_{F/U(w)}$ and Δm_{U-F} derived from the CD equilibrium data for the N-domain in this mutant correlate well with those derived for the wild-type (bsPGK′; see Figure 2 and Table 1). This suggests that interdomain contacts do not markedly couple the observed equilibrium folding/unfolding processes of the N- and C-domains. The rate profiles for the kinetic relaxation processes of the N- and C-domains in this mutant are shown in Figure 5. The mutation affects both the stabilities and solvent exposures (Δm values) of the I, t, and F states of the C-domain (see Table 1). The relevance of these effects has been described elsewhere (Parker et al., 1996). In stark contrast, the parameters derived for the N-domain of this mutant are not noticeably different from those derived for the wild-type (bsPGK′; see Table 1), showing that the domains are effectively independent in their intermediate and rate-limiting transitions states.

For each domain, in bsPGK′ and the I217A+L221A mutant, the values for $K_{F/U(w)}$ and Δm_{U-F} calculated from the equilibrium data (Figure 2) are close to those derived from the kinetic data (see Table 1 and legend to Figure 5), despite the folding and unfolding processes in the intact molecule populating different species (as illustrated in Scheme 1). Taken together, these results imply that there is little interdomain interaction in bsPGK and that the domains form highly autonomous units of structure.

It is interesting to compare this folding mechanism with that for hen lysozyme (Itzhaki et al., 1994; Parker et al., 1995). Unlike bsPGK, the folding of the α- and β-domains of hen lysozyme is, predominantly, a compulsory-order process, i.e., the folding units are energetically tightly coupled. This is likely to be a consequence of the fact that the α- and β-domains are not contiguous in sequence, so that neither represents a distinct, nonoverlapping folding unit. The studies of Peng and Kim (1994), on the structurally related α-lactalbumin, support this basic argument.

Folding of the Isolated Domains: Equilibrium Properties. The conclusion that interdomain contacts are weak is confirmed by expressing the domains individually. The

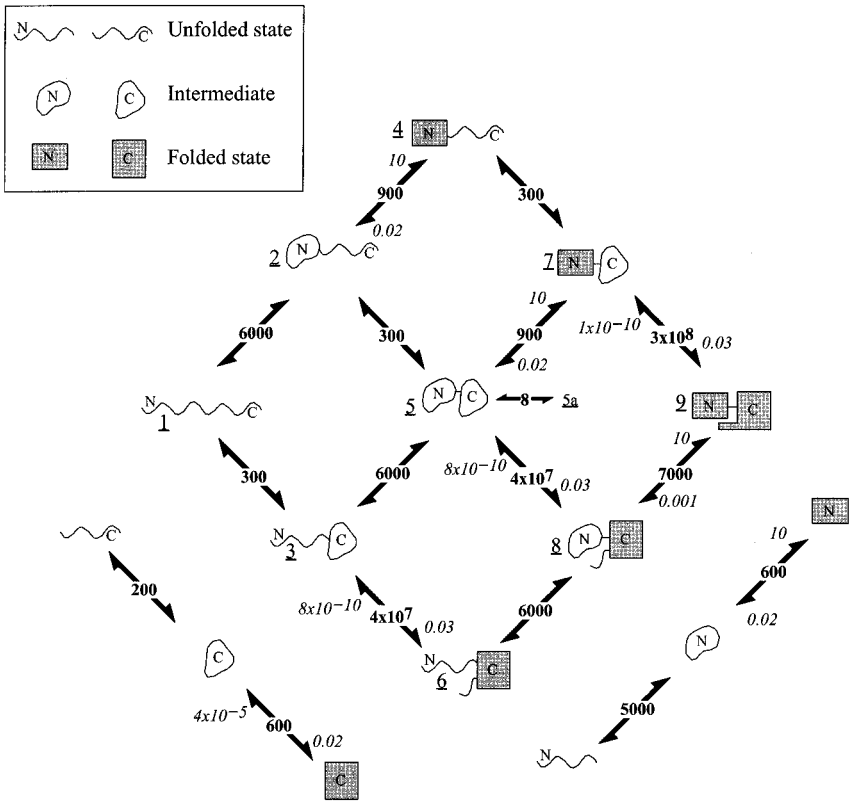

FIGURE 9: Summary of the folding pathway of bsPGK′ in the absence of denaturant. The unfolded, intermediate, and folded domain states are represented by open chains, circles, and squares, respectively. Individual states are given an underscored number. Transitions between states are annotated with forward and backward rate constants, where known, while rapid steps are represented by equilibrium constants (higher numbered state/lower numbered), which break the equilibrium symbol. Also shown is the reversible, productive intermediate (state 5) to abortive intermediate (state 5A) transition (see Results for discussion). The major pathway for folding in water proceeds through states 1−[2/3]−5−7−9, while the major unfolding pathway when the system is transferred to high denaturant concentrations proceeds through states 9−[8]−6−[3]−1.

division of the gene was made at the ends of the bridging interdomain helix to produce an N-domain comprising codons 1−174 and a C-domain comprising codons 185−394. The equilibrium unfolding transitions, measured by CD and intrinsic fluorescence, for the isolated domains of bsPGK′ are shown in Figure 6. The most striking result is the small effect on the N-domain stability, compared to that in the intact molecule, whereas the effect on the C-domain is large (see Table 1). This implies an uneven division of domains, i.e., the helical interdomain segment (176−185), which has been deleted in producing the isolated chains, makes stronger contacts with the C-domain than with the N. This is borne out by inspection of the X-ray structure (Figure 1), which shows this helix to make several hydrophobic contacts with the C-domain, but few with the N-domain.

While the isolated N-domain shows only a single equilibrium unfolding transition, coincident in the CD and tyrosine fluorescence measurements (Figure 6, panel b), the isolated C-domain (Figure 6, panel a) shows two. In the first phase the fluorescence increases, as it does in the unfolding of the C-domain in the whole molecule (Figure 2, panel a), but this transition is shifted to much lower denaturant concentrations. In the second phase, the fluorescence decreases in a process coincident with the loss of the amide CD signal at 222 nm (see Figure 6, panel a). These observations demonstrate the existence of an equilibrium intermediate in the folding pathway of the destabilized, isolated C-domain, which must have extensive secondary structure. When the energetic properties of the kinetic intermediate for the intact C-domain are compared with those of the isolated C-domain equilibrium state, the two are found to be rather similar (see Figure 6, panel c, and Table 1). In Figure 6, panel c, the unfolding profile for the kinetic intermediate of the intact C-domain has been reconstructed from kinetically derived parameters and superimposed on the equilibrium CD curve for the isolated C-domain.

Folding of the Isolated Domains: Kinetic Pathways. Figure 7 shows a direct comparison of the rate profiles obtained for the isolated and intact C-domains. For the unfolding limb in the intact molecule it must be remembered that the N-domain has already unfolded, so that interdomain contacts are abolished. Nonetheless, there is a large difference between the extrapolated value of $k_{F-I(w)}$ for the intact molecule (7.6×10^{-10} s^{-1}) and for the isolated domain (3.6×10^{-5} s^{-1}). Thus, relative to the transition state for unfolding, removing residues 1−184 has destabilized the C-domain by 6.3 kcal mol^{-1} ($=RT \ln(3.6 \times 10^{-5}/7.6 \times 10^{-10})$). This must result from removing contacts with the interdomain helix (in the region 176−184) rather than the remainder of the N-domain, which is in the unfolded state in this comparison.

While there is a distinct, kinetically populated intermediate in the folding of the C-domain in the whole molecule, the folding kinetics of the isolated domain appear two-state (see Figure 7). The reason for the simpler two-state kinetic profile is that the folded state, but not the intermediate, is rendered much less stable by the deletion. As a result, at the cusp point of the kinetic profile for the isolated C-domain (denaturant activity $\approx$ 0.5 M) the intermediate is populated, while in the intact molecule (cusp point $\approx$ 1.1 M) it is not (see Figure 7). Hence, in the case of the isolated C-domain, only the I−F transition is being kinetically sampled. The

free energy change for the I to F transition for the isolated C-domain calculated from rates ($-RT \ln(k_{I-F(w)}/k_{F-I(w)}) = -3.7$ kcal mol^{-1}) is close to that measured for the I to F equilibrium fluorescence transition shown in Figure 6 (-3.4 kcal mol^{-1}; see legend to Table 1).

Despite the instability of the folded state of the isolated C-domain, the rate of folding ($k_{I-F(w)}$) is similar to that in the intact C-domain (these values are 0.019 and 0.028 s^{-1}, respectively; see Table 1). Because both the stability of the intermediate ($\Delta G_{I-U(w)}$) and the height of the activation barrier for folding are not noticeably altered by the deletion, we can conclude that contacts made with the N-domain in general, or with the more intimately associated 175−184 portion of the interdomain helix, contribute little to the dynamics of folding. This contrasts with the removal of interactions from the core (e.g., the I217A+L221A mutation; see Figure 5 and Table 1), which have been shown to have a much greater effect on the folding dynamics of the C-domain (Parker et al., 1996). However, the large increase in the unfolding rate for the isolated C-domain shows that contacts formed with the 175−184 helix are required to sustain the fully native ground state. In other words, relatively long sequence range, interdomain interactions become important only when the rate-limiting transition state has been traversed.

In keeping with this role for the interdomain helix, the removal of residues 176 onward has only a subtle influence on the folding/unfolding kinetics of the N-domain (see Figure 8). In fact, while loss of interdomain contacts marginally increases the unfolding rate of the N-domain, it allows its folding to occur more rapidly (see below).

Strength of Interdomain Contacts. Owing to the fact that the N-domain unfolds first when bsPGK′ is denatured, a comparison of its unfolding rate as an isolated unit and as part of the intact molecule offers the best measure of the strength of interdomain contacts in the fully native ground state. That is, interactions with the folded C-domain must be broken in the unfolding transition. The results in Figure 8 show that removal of residues 176−394 increases the rate of unfolding by a factor of ∼11 (see Table 1). Thus, the contribution of interdomain contacts to the native state unfolding barrier is only 1.4 kcal mol^{-1}. Contacts between domains are not formed in a step which is separate from domain folding and unfolding, since a tryptophan probe inserted into the domain boundary (mutation Y169W on the tryptophanless bsPGK molecule) reports only the two domain folding transitions (data not shown). Studies on yeast PGK by Ballery et al. (1993) also suggest that domain pairing and formation of the native structure occur simultaneously in the slow phase of refolding (i.e., the folding of the C-domain).

Unproductive Interactions between Domains: Evidence for Transitory Misfolding. One aspect of the domain behavior observed in bsPGK contains an energetic inconsistency when interpreted in light of an orthodox folding pathway. The N-domain folds faster as an isolated unit than as part of the intact molecule, i.e., $k_{I-F(w)}$ for the isolated N-domain is about 8-fold greater than for the intact N-domain (see Figure 8 and Table 1). Such behavior has also been observed in the yeast protein (Missiakas et al., 1992). This 8-fold difference in $k_{I-F(w)}$ implies that the activation barrier to folding from the intermediate to the transition state (ΔG_{I-t}) is increased by 1.3 kcal mol^{-1} when interactions with the C-domain are present, while the barrier to unfolding (ΔG_{F-t}) is increased by 1.4 kcal mol^{-1} (see above). Using the values for $k_{I-F(w)}$ and $k_{F-I(w)}$ (see Table 1) then, with respect to I, the stability of the N-domain folded state (ΔG_{F-I}) appears to be reduced by only 0.1 kcal mol^{-1} by this deletion. This prompts the rather uncomfortable conclusion that the transition state in the N-domain is stabilized by 1.3 kcal mol^{-1} upon removal of interdomain contacts. However, the relaxation rates at the cusps of the rate profiles for the intact and isolated N-domains (where the F and U states are isoenergetic) are the same, within experimental error (see Figure 8). Also, no such phenomenon is seen in the C-domain; the removal of the N-domain causes little change in its rate of folding ($k_{I-F(w)}$) despite the much greater effect on its ground state stability.

The most compelling explanation for this phenomenon is that the partially structured C-domain forms unproductive contacts with the partially structured N-domain, which reduces the effective population available to fold. The formation of these abortive contacts must be reversible, since the protein regains full activity upon refolding. This effect is absent in the folding of the C-domain, because the N-domain reaches the folded state some 40 times faster (see Table 1); hence the C-domain can fold unencumbered.

The effect can be described by the following relationship:

$$k_{I-F(obs)} = k_{I-F}/(1 + K_{ab} + 1/K_{I/U}) \quad (9)$$

where $k_{I-F(obs)}$ and k_{I-F} are the observed and actual (i.e., on-pathway) folding rates for the I to F transition, respectively, and K_{ab} describes the equilibrium ratio between the productive (I) and abortive (I_{ab}) intermediate states ($K_{ab} = I_{ab}/I$). At high denaturant concentrations (near the cusp point of the rate profiles in figure 8) this relationship reduces to

$$k_{I-F(obs)} = k_{I-F}K_{I/U} \quad (10)$$

In these conditions, the measured rate depends only on the stability of the productive, obligatory intermediate and on the transition state between this state and the folded structure. As mentioned above, $k_{I-F(obs)}$ for the intact and isolated N-domains are essentially the same. In accordance with the results for the C-domain, this demonstrates that productive interdomain or helix V contacts contribute little to the dynamics of folding.

As the denaturant concentration is reduced, both K_{ab} and $K_{I/U}$ increase, so that the abortive and productive interactions compete in the case of the intact molecule. This reduces the observed N-domain folding rate ($k_{I-F(obs)}$) over that seen in the isolated species (k_{I-F}). As we know that k_{I-F} and $K_{I/U}$ are not measurably influenced by the presence of the C-domain (see above), we can use k_{I-F}, $K_{I/U(w)}$, Δm_{I-tf}, and Δm_{U-I} calculated for the isolated N-domain (see Table 1) to constrain a fit to the intact N-domain kinetic data to eq 9 (see Figure 8). This produces values for $K_{ab(w)}$ and Δm_{I-Iab} (a measure of the difference in the solvation of hydrocarbon between the productive and abortive intermediate) of 8 and 1.3 M^{-1}, respectively. In light of the ΔG and Δm values associated with the on-pathway transitions (see Table 1), these values imply that the abortive contacts are relatively weak and do not involve a substantial area of contact.

DISCUSSION

A scheme for the folding reaction of bsPGK′ derived from our equilibrium and kinetic results is summarized in Figure 9. One of the major conclusions to arise from the study is

that interdomain contacts are weak in this molecule and have a measurable influence only on the kinetics of unfolding. For each domain, we see that a productive intermediate rapidly accumulates (see Figure 9, structures 2, 3, and 5). These states have the same stability when the domains fold as isolated units as they do when they are integrated into the intact molecule, implying that the initial collapse is domain-specific rather than random. As the on-pathway activation barriers to folding from these intermediate states are also unaffected by the removal of these contacts, we may conclude that productive interdomain contacts do not become established until both domains have passed through their respective rate-limiting (I-to-F) transition states.

Prior to the formation of the fully folded state, the only interactions between domains are weak, unproductive contacts in the rapidly formed intermediate (see Figure 9, step 5 to 5a). These occur before the N-domain folds to its native state and inhibit this process. Given the observed propensity of folding intermediates to aggregate and retard folding rates (Ranson et al., 1995) and the close proximity of the independent domains during folding, it is surprising that there is not a greater extent of inhibitory interaction at this stage of folding.

Looking at these reactions in the unfolding direction provides information on interdomain contacts in the folded ground state. The C-domain unfolds 5×10^4 times faster, when isolated, than when integrated into the whole molecule; even though, in the latter case, the N-domain is already in the unfolded state when the C-domain unfolding kinetics are being measured. In contrast, the N-domain unfolds only 11 times faster when isolated, despite the fact that contacts with the folded C-domain have to be broken in this step (Figure 9, steps 9 to 8). This unequal effect must result from the way in which the domains were divided for the experiment, i.e., to make the isolated units, we effectively removed residues 176−185 from the interdomain helix (see Figure 1). In the folded ground state, this bridging helix must make much stronger interactions with the C-domain than with the N, and thus it acts as an integral part of the former. As a result of this, the strength of interdomain contacts in this PGK is best estimated by comparing the unfolding rates ($k_{F-I(w)}$) for the intact and isolated N-domains. These measurements (see Table 1) show that the contribution of residues 176−394 to the stability of the N-domain folded ground state ($\Delta\Delta G_{F-t}$) is only 1.4 kcal mol^{-1}. This energetic autonomy is further supported by recent NMR studies on the isolated N-domain from bsPGK (Hosszu et al., 1996). These studies demonstrate a very close similarity between the NMR-derived structure of the isolated N-domain and the X-ray structure of this domain in the whole molecule. The integrity and high exchange protection of this structure is maintained right through the domain, i.e., residues 1−175.

By contrast, the contribution of residues 1−185 to the stability of the folded ground state of the C-domain ($\Delta\Delta G_{F-t}$) is 6.4 kcal mol^{-1}, a value which includes contacts with the N-domain, proper, and with the bridging helix; the latter arising from residues 176−185. Even with this marked loss of stability in the folded ground state, the stabilities of the intermediate and transition states are not measurably altered, showing that interactions with both the N-domain and the bridging helix are made only after the transition state has been traversed. The lack of interdomain contacts in the rapidly formed intermediate states is supported by studies on the relative accessibility (reactivity) of single cysteine residues genetically inserted in yeast PGK (Ballery et al., 1990, 1993). Thiol groups located within the domains of yeast PGK show a very rapid phase of protection followed by a slower phase of protection, while an interdomain-located thiol group shows only a slow phase of protection (Ballery et al., 1993).

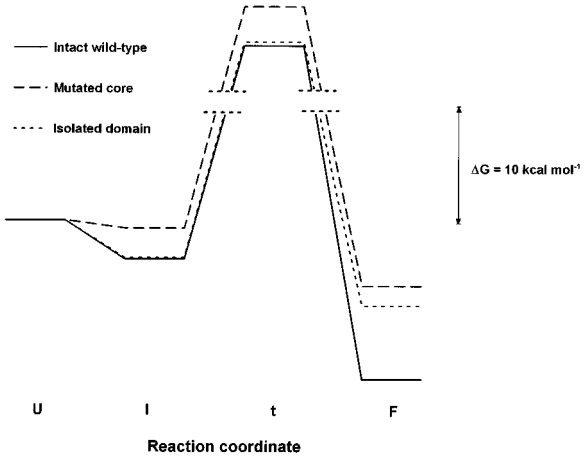

FIGURE 10: Reaction profiles for folding of the C-domain. The free energies of states in the folding reaction of the C-domain are shown using the unfolded state (U) as a reference. The profile marked "intact wild-type" represents the behavior of the C-domain in whole bsPGK′, while that marked "mutated core" carries the mutations I217A+L221A. The profile marked "isolated domain" represents the behavior of residues 186−394 in the truncated protein.

Our results imply that the critical steps in the folding pathway of bsPGK are dictated by two, separate cores: one in the N-domain and one in the C-domain. Contacts between these centers are formed only late in the folding process, but do not constitute a separate step in the folding pathway. The formation of these contacts are rather like those which consolidate the outer regions of small, single-domain proteins. For example, mutagenesis and kinetic studies on barnase suggest that the latest interactions to form on the folding pathway are those involved in the outer regions of hydrophobic cores, the N-termini of helices and surface loops (Matouschek et al., 1992; Serrano et al., 1992). In this respect, it is important to distinguish between those interactions which influence the rate of folding and those which influence the rate of unfolding; the former are required to direct the search for the folded state, and the latter merely serve to maintain it. This is illustrated by comparing the folding pathways of the intact C-domain core mutant (I217A+L221A) and the isolated C-domain with the intact C-domain in the wild-type (bsPGK′; see Figure 10). The removal of contacts from the hydrophobic core of the C-domain significantly affects the dynamics of folding and unfolding, while the removal of interdomain contacts affects only the dynamics of unfolding.

Taken in conjunction with our previous studies on bsPGK (Parker et al., 1996), it is possible to propose a general pattern of interactions which direct the folding of this large protein. The fastest contacts to form are between sequence-local, hydrophobic side chains which ultimately form core interactions in the folded molecule. These early contacts are native-like (i.e., nonrandom) and stabilize adjacent secondary structural elements in the rapidly collapsed intermediate states (Parker et al., 1996). The additional formation of structure within these domain nuclei then requires the development of statistically less likely interactions between more distant

parts of the chain. For both domains, the transition state species is highly hydrated, with respect to the folded ground state, i.e., Δm_{I-t} is significantly smaller than Δm_{F-t} (see Table 1). This observation implies that the transition state species are not uniformly distorted, high-energy forms of the folded ground states, as described by the "critical distortion" model (Segawa & Sugihara, 1984). Rather, they are more plausibly described by the "critical substructure" model developed by Kuwajima (1989).

The contacts made between domains in this molecule are the most distant in sequence and do not involve parts of the polypeptide chain which are integral to either of the domain cores. It is this combination of spatial remoteness and sequence separation of the folding cores which provides the best explanation for the observed autonomy of the two domains in the dynamics of folding.

ACKNOWLEDGMENT

We thank the BBSRC (U.K.) for project funding and the award of studentships to M.J.P., J.S., L.L.P.H., G.S.J., C.J.C., and S.G.B. We also thank the Wellcome Trust for an equipment grant. A.R.C. is a Lister Institute research fellow.

REFERENCES

Baldwin, R. L. (1996) *Folding Des. 1*, R1−R8.
Ballery, N., Minard, P., Desmadril, M., Betton, J. M., Perahia, D., Mouawad, L., Hall, L., & Yon, J. M. (1990) *Protein Eng. 3*, 199−204.
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. W., Merrett, M., & Phillips, A. W. (1979) *Nature 279*, 773−777.
Betton, J. M., Desmadril, M., Mitraki, A., & Yon, J. M. (1984) *Biochemistry 23*, 6654−6661.
Betton, J. M., Desmadril, M., Mitraki, A., & Yon, J. M. (1985) *Biochemistry 24*, 4570−4577.
Brandts, J. F., Halvorson, H. R., & Brennan, M. (1975) *Biochemistry 14*, 4953−4963.
Bücher, T. (1955) *Methods Enzymol. 1*, 415−422.
Creighton, T. E., Ed. (1992) *Protein Folding*, W. H. Freeman and Co., New York.
Davies, G. J., Gamblin, S. J., Littlechild, J. A., & Watson, H. C. (1993) *Proteins: Struct., Funct., Genet. 15*, 283−289.
Fairbrother, W. J., Bowen, D., Hall, L., & Williams, R. J. P. (1989) *Eur. J. Biochem. 184*, 617−625.
Garel, J. R. (1992) in *Protein Folding* (Creighton, T. E., Ed.; Chapter 9, pp 405−454, W. H. Freeman and co., New York.
Harlos, K., Vas, M., & Blake C. F. (1992) *Proteins: Struct., Funct., Genet. 12*, 133−144.
Hosszu, L. L. P., Craven, C. J., Spencer, J., Parker, M. J., Clarke, A. R., Broadhurst, R. B., Laue, E. D., & Waltho, J. P. (1996) *Biochemistry* (in press).
Ikai, A., & Tanford, C. (1973) *J. Mol. Biol. 73*, 145−163.
Itzhaki, L. S., Evans, P. A., Dobson, C. M., & Radford, S. E. (1994) *Biochemistry 33*, 5212−5220.
Johnson, C. M., & Fersht, A. R. (1995) *Biochemistry 34*, 6975−6804.
Kim, P. S., & Baldwin, R. L. (1990) *Annu. Rev. Biochem. 59*, 631−666.
Kuwajima, K. (1989) *Proteins 6*, 87−103.
Matouschek, A., Serrano, L., & Fersht, A. R. (1992) *J. Mol. Biol. 224*, 819−835.
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.
Missiakas, D., Betton, J. M., Chaffotte, A., Minard, P., & Yon, J. M. (1992) *Protein Sci. 1*, 1485−1493.
Nozaki, Y. (1970) *Methods Enzymol. 26*, 43−50.
Pace, C. N. (1975) *CRC Crit. Rev. Biochem. 3*, 1−43.
Parker, M. J., Spencer, J., & Clarke, A. R. (1995) *J. Mol. Biol. 253*, 771−786.
Parker, M. J., Sessions, R. B., Badcoe, I. G., & Clarke, A. R. (1996) *Folding Des. 1*, 145−156.
Peng, Z., & Kim, P. S. (1994) *Biochemistry 33*, 2136−2141.
Ranson, N. A., Dunster, N. J., Burston, S. G., & Clarke, A. R. (1995) *J. Mol. Biol. 250*, 581−586.
Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H., & Roe, B. A. (1980) *J. Mol. Biol. 143*, 161−178.
Segawa, S., & Sugihara, M. (1984) *Biopolymers 23*, 2473−2488.
Serrano, L., Matouschek, A., & Fersht, A. R. (1992) *J. Mol. Biol. 224*, 805−818.
Shortle, D., Meeker, A. K., & Freire, E. (1988) *Biochemistry 27*, 4761−4768.
Staniforth, R. A., Burston, S. G., Smith, C. J., Jackson, G. S., Badcoe, I. G., Atkinson, T., Holbrook, J. J., & Clarke, A. R. (1993) *Biochemistry 32*, 3842−3851.
Tanford, C. (1970) *Adv. Protein Chem. 24*, 1−95.
Tsunenaga, M., Goto, Y., Katawa, Y., & Hamaguchi, K. (1987) *Biochemistry 26*, 6044−6051.
Watson, H. C., & Littlechild, J. A. (1990) *Biochem. Soc. Trans. 18*, 187−190.
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.
Yon, J. M., Betton, J. M., Desmadril, M., Mitraki, A., Minard, P., Gaillard, S., Ballery, N., & Missiakas, D. (1988) *J. Chromatogr. 440*, 421−437.
Yon, J. M., Desmadril, M., Betton, J. M., Minard, P., Ballery, N., Missiakas, D., Gaillard, S., Perhia, D., & Mouawad, L. (1990) *Biochemie 72*, 417−429.

BI961330S