Is the Continuity of the Domains Required for the Correct Folding of a Two-Domain Protein?†

Monica Ritco-Vonsovici, Philippe Minard, Michel Desmadril,* and Jeannine M. Yon

Laboratoire d'enzymologie physicochimique et moléculaire, Unité de Recherche du Centre National de la Recherche Scientifique, Université de Paris-Sud, 91405 F Orsay, France

Received March 28, 1995; Revised Manuscript Received October 19, 1995⊗

ABSTRACT: The role of domains in protein folding has been widely studied and discussed. Nevertheless, it is not clear whether the continuity of the domains in a protein is an essential requirement in determining the folding pathway. Previous studies have shown that the isolated structural domains of the two-domain monomeric enzyme, yeast phosphoglycerate kinase (yPGK), are able to fold independently into a quasi-native structure, but they neither reassociate nor generate a functional enzyme [Minard, P., Hall, L., Betton, J. M., Missiakas, D., & Yon, J. M. (1989) *Protein Eng. 3*, 55−60; Fairbrother, W. J., Bowen, D., Hall, L., Williams, R. J. P. (1989) *Eur. J. Biochem. 184*, 617−625; Missiakas, D., Betton, J. M., Minard, P., & Yon, J. M. (1990) *Biochemistry 29*, 8683−8689]. In the present work, two circularly permuted variants of the yPGK gene were constructed. The natural adjacent chain termini were directly connected and the new extremities were created within the N-domain (at residues 71 and 72) or the C-domain (at residues 291 and 292), respectively. These two proteins were overexpressed and purified. They exhibit 14% and 23% of the wild-type enzyme activity, respectively. The two mutants fold in a compact conformation with slight changes in the secondary and tertiary structure probably related to the presence of a heterogeneous population of molecules. The unfolding studies reveal a large decrease in stability. From the present data it appears that, although the circular permutations induce some perturbations in the structure and stability of the protein, the continuity of the domains is not required for the protein to reach a native-like and functional structure.

The mechanism by which the amino acid sequence of a protein directs its folding into the unique native and functional conformation remains an unsolved problem. According to Kim and Baldwin (1990) the folding pathway can be considered as a sequential process in which the unstructured polypeptide chain first forms local structures that interact to produce domains stabilized by long-range interactions which finally generate the native protein. Wetlaufer (1973) has proposed that the domains are independent folding units that assemble and produce the native molecule. This concept of autonomous folding units has been also introduced for fragments smaller than structural domains but larger than individual secondary structure elements, referred to as subdomains (Levitt & Clothia, 1976). These structural regions (subdomains and domains) are expected to fold independently when they are isolated. Several studies on the isolation and folding of structural domains and subdomains have been reported (Oas & Kim, 1988; Minard *et al.*, 1989; Missiakas *et al.*, 1990; Sharma *et al.*, 1990; Herold *et al.*, 1991; Williams & Shoelson, 1993; Jecht *et al.*, 1994).

This hypothesis was also confirmed for yeast phosphoglycerate kinase, a monomeric protein of 415 amino acids that folds in two continuous globular domains of approximately the same size (Watson *et al.*, 1982), called the N- and C-domains (Figure 1). In the native enzyme, domain interactions involve helix V, which links the two domains

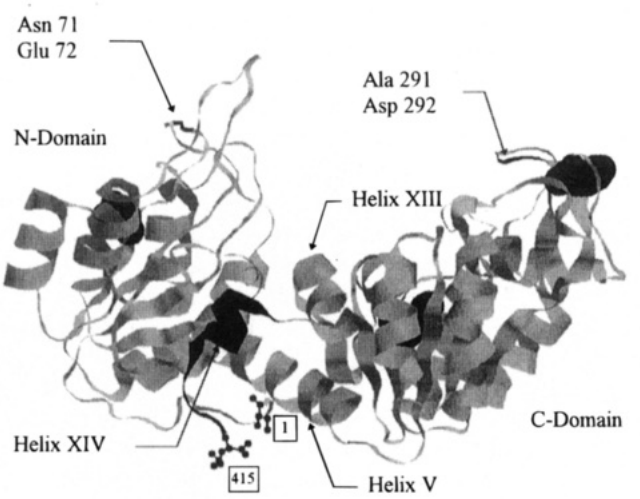

FIGURE 1: Three-dimensional structure of yeast PGK (Watson *et al.*, 1982). The cleavage sites and the conformational probes in the N- and C-domains are indicated.

and C-terminal helices XIII and XIV interacting with the N-domain. Helix XIV acts to lock the C-domain on the N-domain, insuring the expression of the biological function (Ritco-Vonsovici *et al.*, 1995). The two domains of yeast phosphoglycerate kinase were produced by recombinant techniques (Minard *et al.*, 1989). These isolated structural domains are able to fold independently into a quasi-native structure, but they neither reassociate nor generate a functional enzyme (Minard *et al.*, 1989; Fairbrother *et al.*, 1989; Missiakas *et al.*, 1990). These results have suggested that the isolated domains are stable, but the interdomain inter-

† This work was supported by the Centre National de la Recherche Scientifique, the Ministère de l'Enseignement Supérieur et de la Recherche, the Fondation pour la Recherche Médicale Française, and by a contract with the Centre National de la Transfusion Sanguine.

* To whom all correspondence should be addressed.

⊗ Abstract published in *Advance ACS Abstracts,* December 1, 1995.

0006-2960/95/0434-16543$09.00/0

actions are not strong enough to produce a fully complemented enzyme. Another study on several pairs of contiguous yeast phosphoglycerate kinase (yPGK)[1] fragments obtained by chemical cleavage has been reported (Pecorari *et al.*, 1993). Some of these fragments have a low degree of structure but are able to complement the rest of the molecule to generate an active enzyme. From these data, it appears that the correct folding of the isolated fragments is not a prerequisite for their complementation.

The role of the domains as folding intermediates can intuitively be understood when they are made by continuous stretches of the amino acid sequence. Indeed, it is widely believed that folding starts independently in several parts of the polypeptide chain and that local interactions dominate the first folding steps. A network of short-range interactions could result in the prefolding of the individual domains which would then be able to interact with each other. This intuitive concept, which associates local interactions and early steps, does not simply explain the folding of proteins made of discontinuous domains. This leads us to address the following questions: Is the continuity of the domains an obligatory requirement for the correct folding of the proteins made of continuous domains? Is the natural topology of a protein critical for its stability and folding pathway? An answer to these questions could emerge from studies of circularly permuted proteins. The sequence of bovine pancreatic trypsin inhibitor was the first to be permuted circularly by chemical condensation and limited proteolysis (Goldenberg & Creighton, 1983). However, a method applicable to other proteins is the circular permutation of the corresponding structural gene that was first introduced by Luger *et al.* (1989) and subsequently used by several authors (Buchwalder *et al.*, 1992; Horlick *et al.*, 1992; Yang *et al.*, 1993; Protasova *et al.*, 1994; Hahn *et al.*, 1994). Phage T4 lysozyme, in particular, a two-domain protein, has been shown to fold correctly even when a circular permutation of its sequence was constructed (Zhang *et al.*, 1993).

In this paper we describe the construction of two circularly permuted variants of yeast phosphoglycerate kinase by a simple polymerase chain reaction methodology. In one of the mutants, new C- and N-extremities were created in the N-domain. In the other mutant, new C- and N-extremities were created in the C-domain. The major goal of this approach was to study the importance of the continuity of yeast phosphoglycerate kinase domains for the correct folding of this enzyme. Structural and functional properties of these mutants, as well as their stability, were characterized.

MATERIALS AND METHODS

Site-Directed Mutagenesis. The 3.1 kb PGK fragment was subcloned into the pBluescript II bacterial vector (pBS), and all of the mutagenesis experiments were carried out within the resultant plasmid. pBluescript II bacterial vector, T7 primer, and Reverse primer were purchased from Stratagene.

The first step was the insertion of the *Bam*HI/*Xho*I digestion fragment of pYE-PGK vector (Figure 2) in the multiple cloning site of pBS plasmid (pBS-PGK). The pYE-PGK plasmid was constructed by Minard *et al.* (1990). The *Bam*HI site, which is located downstream from the 3.1 kb PGK fragment, was converted to an *Xho*I site by linker addition. In a second step, unique *Nde*I and *Aat*II restriction sites was inserted in pBS-PGK using the PCR method in order to facilitate the cloning of the circularly permuted coding sequences of yeast phosphoglycerate kinase (yPGK) in this vector (Figure 3). Two overlapping fragments were synthesized in separate primary amplifications and the fragments were then combined and amplified in a second

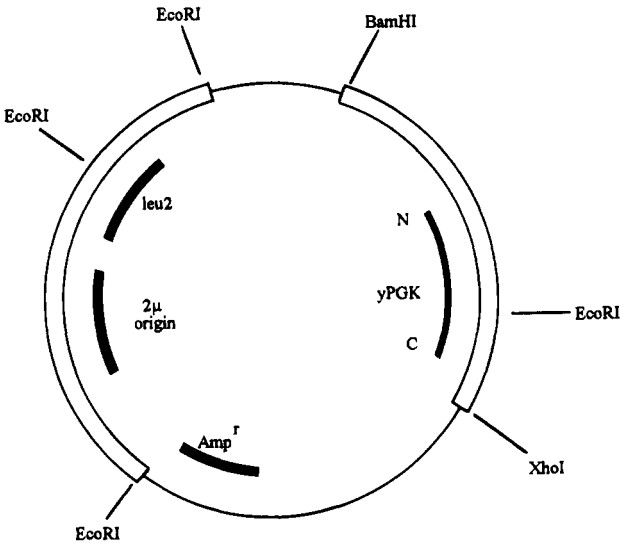

FIGURE 2: The pYE-PGK vector for a high expression level of yPGK in *S. cerevisiae*.

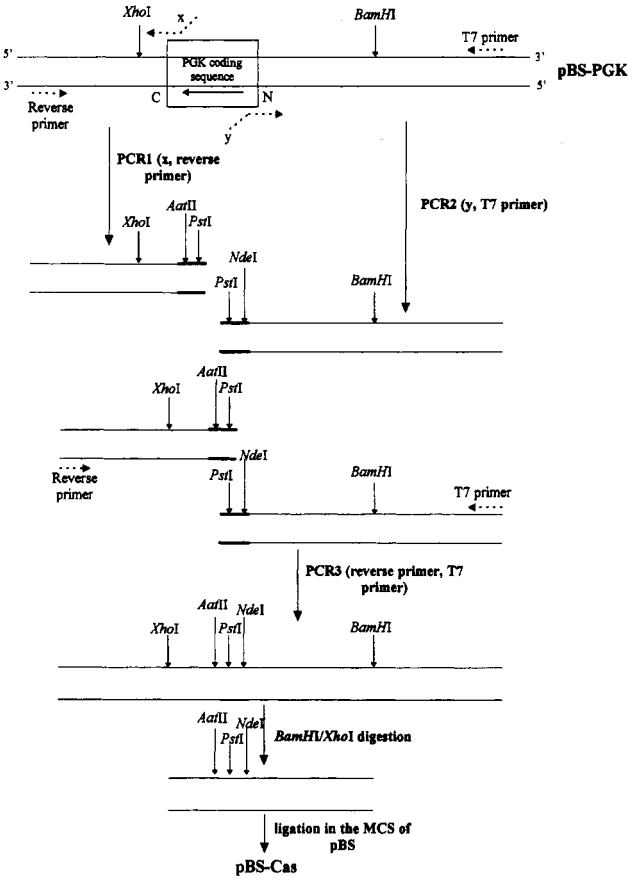

FIGURE 3: Schematic illustration of the construction of pBS-Cas vector by the PCR methodology.

[1] Abbreviations: CD, circular dichroism; EDTA, ethylenediaminetetraacetate; Gdn-HCl, guanidine hydrochloride; NbS_2, 5.5′-dithio-bis-(2-nitrobenzoate); yPGK, yeast phosphoglycerate kinase (E.C. 2.7.2.3); PAGE, polyacrylamide gel electrophoresis; PCR, polymerase chain reaction; 3PG, 3-phosphoglycerate; SDS, sodium dodecyl sulfate; Tris, Tris[hydroxymethyl]aminomethane; UV, ultraviolet.

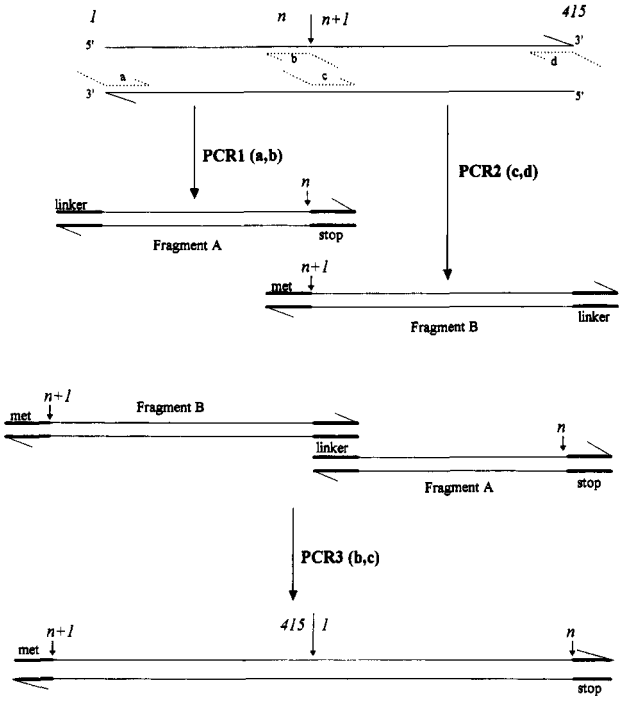

FIGURE 4: Construction of circularly permuted genes of yPGK.

reaction. Primer *x* includes the *Aat*II restriction site and a region that is complementary to the 3′ noncoding flanking sequence. Primer *y* incorporates the *Nde*I restriction site, a region complementary to the 5′ noncoding flanking sequence, and a region that overlaps primer *x*. This overlapping region contains a *Pst*I recognition site, which is very useful for the screening of the recombinant clones. The resulting fragment was restricted with *Bam*HI and *Xho*I and cloned into the pBS plasmid, creating the new plasmid pBS-Cas. The pBS-Cas vector inserts the circularly permuted variants. The modified region of the yPGK gene in pBS-Cas plasmid was sequenced. All the restriction enzymes and the synthetic primers were purchased from Eurogentec (Seraing, Belgium).

A similar strategy was employed in a third step for the construction of the permuted coding sequences (Figure 4). Two sets of oligonucleotide primers were used as follows. The first primer set, designated *a* and *b*, was used in the first PCR step to create and amplify the DNA fragment A. Primer *a* included the codons of the last five amino acids of the sequence (residues 411−415) joined to the codons of the first amino acids of the sequence (residues 1−4). Primer *b* contained the natural TAA stop codon and the *Aat*II unique restriction site. The second primer set, designated *c* and *d*, was used in the second PCR step to create and amplify the DNA fragment B. Primer *c* incorporated the *Nde*I restriction site with the ATG start codon, and primer *d* was the reverse of primer *a* (named "linker" region). Fragments A and B that encode the same "linker" region were joined in a third PCR step with *b* and *c* primers. The first and second amplifications used the pYE-PGK or pBS-PGK vectors as templates. The amplified A and B fragments were purified from agarose gels and precipitated with ethanol. The final amplified product was digested with *Nde*I and *Aat*II and subcloned in pBS-Cas to give pBS-PGK72p or pBS-PGK292p. Finally, the digestion fragment *Bam*HI/*Xho*I, including the new coding sequence, was cloned and expressed into pYE vector.

DNA amplifications were performed using *pfu* DNA polymerase from Stratagene. Twenty-five PCR cycles (95 °C for 30 s, 35 °C for 30 s or 1 min, and 72 °C for 1 min) were carried out.

The presence of the correct circularly permuted sequences was first checked by appropriate restriction analyses and then confirmed by DNA sequencing.

Protein Expression and Purification. The bacterial vectors pBS-PGK72p and pBS-PGK292p were restricted with *Bam*HI and *Xho*I, the resultant gene fragments were ligated into the yeast expression vector pYE and subsequently transformed into *pgk Saccharomyces cerevisiae* strain BC3 (Piper & Curran, 1990). Yeast cells were transformed by the spheroplast method (Hinnen *et al.*, 1978) and grown on a defined minimal medium without leucine, YOM leu(−), for leucine selection (Hawthorne & Mortimer, 1960). The protein extracts were subjected to SDS−PAGE (Laemmli, 1970) to estimate the overexpression of permuted yPGK.

Protein purification was performed according to Ritco-Vonsovici *et al.* (1995), and the purity was determined by SDS−PAGE.

Enzyme Activity Measurements. Enzyme activity was measured by a coupled assay adapted from Bücher (1955) as described by Betton *et al.* (1985). The protein concentration was determined by the absorbance at 278 nm using the same extinction coefficient as for the wild-type protein, $\epsilon = 0.49$ mL mg^{-1} cm^{-1} (Watson *et al.*, 1982).

Mass Spectrometry. Electrospray mass spectrometry was performed after injection of 10 μL of 10 μM protein samples with a solvent containing CH_3CN (50%)/H_2O (49%)/HCOOH (1%). Protein samples were desalted by reverse phase HPLC on an Aquapore BU300 column by using a linear gradient of CH_3CN/H_2O containing 0.1% trifluoroethanol.

Amino-Terminal Sequencing. The amino-termini were checked by the automated Edman method on an Applied Biosystems 473 A sequencer using samples of 200 pmol of purified proteins.

Gel Filtration. The elution volumes of the permuted and wild-type proteins were determined by HPLC on a Superose 12 column equilibrated with a 20 mM Tris-HCl buffer, pH 7.5, containing 0.2 M NaCl. The detection was performed at 278 nm with an LKB spectrophotometer.

Reactivity of the Thiol Groups. The reactivity of the thiol groups toward 5,5′-dithio-bis(2-nitrobenzoate) (NbS_2) was measured at 22 °C in a 20 mM Tris-HCl buffer, pH 7.5. The reaction was started by addition of a small volume of 6 mM NbS_2 (250 μM final concentration) to 1 mL of a 10 μM PGK solution. The reaction was followed by monitoring the absorbance of the NbS^- at 412 nm ($\epsilon = 14\,150$ M^{-1} cm^{-1}, Riddles *et al.*, 1979) with a Cary spectrophotometer equipped with a thermostated cell holder. The apparent first-order rate constants were calculated from the whole data set using a nonlinear regression (Marquardt algorithm; Press *et al.*, 1986).

Circular Dichroism and Fluorescence. The circular dichroic spectra were recorded from 260 to 200 nm in far-UV and from 340 to 250 nm in near-UV on a Mark V dichrograph (Jobin-Yvon) equipped with a thermostated cell holder and connected to a microcomputer for data acquisition.

Steady state fluorescence emission spectra were recorded between 300 and 400 nm after excitation at 293 nm with a

SLM 8000C Aminco spectrofluorimeter and a 10 mm light-path cell.

Unfolding Studies. The unfolding experiments were carried out with a 2 μM protein concentration in 20 mM Tris-HCl buffer, pH 7.5, containing 0.5 mM EDTA and 1 mM 2-mercaptoethanol at 22 °C. The circular dichroism and fluorescence measurements were performed after 12 h of incubation at 22 °C in different concentrations of Gdn-HCl.

Ultrapure Gdn-HCl was obtained from Pierce; the denaturant concentrations were checked by refractometry, using the relationship provided by Nozaki (1970).

The CD transition curves were constructed by plotting the variation in ellipticity at 220 nm as a function of denaturant concentration.

The transition curves as assessed by the enzymatic activity were analyzed by using the model of linear dependency of ΔG_x upon denaturant concentration, x, according to Pace (1986)

$$\Delta G_x = \Delta G_0 - mx \quad (1)$$

ΔG_0 being the variation of free energy in the absence of denaturant and m a constant proportional to the increase in the degree of exposure of the protein on denaturation. The following equation was derived from eq 1 in order to fit the experimental data:

$$y_x = y_n + A\left\{\frac{e^{[(\Delta G_0 - mx)/RT]}}{1 + e^{[(\Delta G_0 - mx)/RT]}}\right\} \quad (2)$$

where y_x is the enzymatic activity in the presence of x molar Gdn-HCl, y_n the enzymatic activity of the native form, and A the amplitude of the transition.

The transition curves obtained by spectroscopic measurements were fitted to an equation in which both the baselines and the transition regions of the curve are fitted simultaneously. The transition was analyzed following an equation derived from the denaturant binding model, as described by Garcia *et al.* (1995). For fluorescence data, the transition region was described by a linear combination of two single-transition curves of opposite amplitude.

All data were fitted by using a simplex procedure based on the Nelder and Mead algorithm (Press *et al.*, 1986).

RESULTS

Construction of the Mutants. The aim of this work was to "break" the continuity of the domains joining the natural extremities of the protein and creating new chain termini within each domain. The direct junction of wild-type N- and C-ends was possible because the distance between the N- and C-termini in yPGK is very small, about 5.33 Å, no linker being necessary. The positions of the new extremities were chosen in order to be compatible with the structure and function of the protein: the substrate binding regions and the conserved regions in different species were not modified, and the new extremities were introduced on surface loops.

While there are several such nonconserved surface loops in the N-domain, there are only a few in the C-domain, this domain being involved in ATP binding. Considering that flexible regions of the protein exposed to the solvent are less important for structural stability (Alber & Matthews, 1987), residue 72 was chosen as the new N-terminus in the N-domain and residue 292 was chosen as the new N-terminus in the C-domain. Glu 72 is in a surface loop joining the βB strand and α-helix II in the N-domain, and Asp 292 is in a β-turn joining the βJ and βK strands in the C-domain (Figure 1). These two new terminal residues were replaced by a serine, an amino acid that, in the N-terminal position, is favorable for the stability of the proteins *in vivo* (Bachmair *et al.*, 1986). The orientation of Glu 72 and Asp 292 side chains in the 3D structure of yPGK suggested that their replacement will not affect protein stability. The two mutants were designated as PGK72p and PGK292p.

Table 1: Molecular Weights of Permuted and Wild-Type Yeast Phosphoglycerate Kinase as Determined by Elestrospray Mass Spectrometry

	expected mol wt	exptl mol wt
wtPGK	44650	44662
PGK72p	44608	44616
PGK292p	44622	44623

In order to start the translation of the permuted proteins, a methionine codon was placed before Ser 72 and Ser 292 in engineered *Nde*I restriction sites; to stop translation, a TAA codon was introduced at the end of the permuted coding sequence.

The vector pYE-PGK is suitable for expressing yPGK to high levels (Minard *et al.*, 1990). A 3.1 kb *Bam*HI/*Xho*I fragment containing the entire yeast PGK gene, together with 5′ and 3′ flanking regions may be easily inserted and removed by cleavage of these unique restriction sites (Figure 2). The insertion of the permuted sequences in this fragment makes necessary the introduction of the unique *Nde*I restriction site, which contains the ATG start codon, at the beginning of the coding sequence of yPGK and another at the end of this sequence. The pYE vector already contains an *Nde*I restriction site within a 2 μ region, and its replacement may affect expression. Therefore, for the construction of permuted mutants, the *Bam*HI/*Xho*I fragment from pYE-PGK was first subcloned into the multiple cloning site (MCS) of the bacterial vector pBluescript II (pBS-PGK; see Materials and Methods).

Expression. The *Bam*HI/*Xho*I fragment contains 1.5 kb of sequence upstream and 0.3 kb downstream from the PGK coding sequence. These sequences possess elements necessary for a high expression level (Mellor *et al.*, 1985; Chen *et al.*, 1984). They were slightly modified by the addition of 10 bases, not affecting the natural PGK promotor. Under these conditions, the two mutants were overexpressed at a level similar to that of the wild-type protein using a yeast strain devoid of the PGK chromosomal gene.

Characterization of the Circularly Permuted Variants. After purification, each mutant gave a single band on SDS–PAGE and migrated as the wild-type protein (data not shown). In gel filtration, a single symmetric peak was observed with an elution volume corresponding to that of wtPGK, indicating that the permuted proteins possess the same Stokes' radii and, therefore, have a compact globular structure. The observed molecular weights of these proteins as determined by electrospray mass spectrometry are fully consistent with the theoretical values expected from the amino acid composition (Table 1). The amino termini was examined by the automated Edman method, but no N-terminus was detectable, indicating that they are blocked as

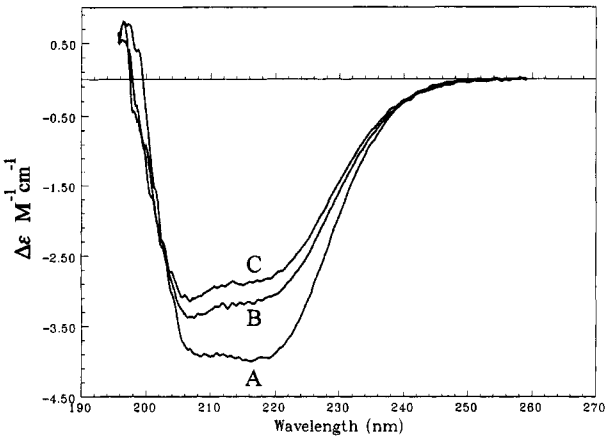

FIGURE 5: Far-UV circular dichroism spectra of wild-type PGK (A), PGK292p (B), and PGK72p (C). Spectra were recorded at 22 °C in a 20 mM Tris-HCl buffer, pH 7.5, containing 0.5 mM EDTA and 1 mM 2-mercaptoethanol. Protein concentration was 2.5 μM in a 2 mm path cell.

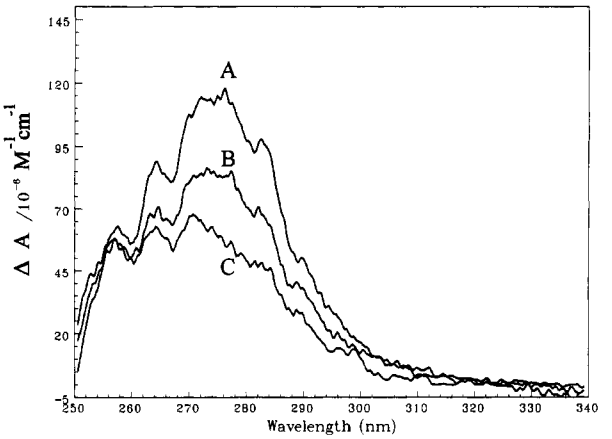

FIGURE 6: Near-UV CD spectra of wild-type PGK (A), PGK292p (B), and PGK72p (C). Spectra were recorded at 22 °C in a 20 mM Tris-HCl buffer, pH 7.5, containing 0.5 mM EDTA and 1 mM 2-mercaptoethanol. Protein concentration was 10 μM in a 10 mm path cell.

in the wild-type enzyme. Since the two permuted sequences have a serine as the penultimate amino acid, it is likely that the final products start with an N-acetylated serine residue after methionine removal.

Enzymatic Activity. The specific activities of PGK72p and PGK292p were about 300 IU/mg and 480 IU/mg, respectively, which represent 14% and 23% of wild-type specific activity (2100 IU/mg). The C-domain possesses the ATP binding site. It seems, therefore, that the permutation of the sequence in this region can modify the enzyme affinity for this substrate. This would have implications on the specific activity of PGK292p. The K_m for ATP was found to be 0.5 mM, and, therefore, was of the same order of magnitude as that of wtPGK (K_m = 0.3 mM). This result indicates that the structural changes mainly affect the catalytic constant and not the binding constant.

The reversibility of the unfolding process was tested by recording the reactivation kinetics for the two mutants. Proteins were completely unfolded in 2 M Gdn-HCl and then refolded. A maximum level of specific activity of 84% for PGK292p and 85% for PGK72p with respect to the original specific activity was recovered after 70 min of refolding in 0.2 M Gdn-HCl. This suggests that the two mutants are able to refold to a large extent *in vitro*.

Structure of the Permuted Proteins. The secondary structures of the permuted and wild-type proteins were compared by CD spectroscopy (Figure 5). The far-UV CD spectra of the mutants have about 80% of the ellipticity of the wild-type protein, the minimum at 220 nm being more affected. The near-UV CD spectra in the region 340−250 nm are presented in Figure 6. The ellipticity of the permuted variants is lower than that of wtPGK according to the far-UV spectra. The CD spectrum of wtPGK at 277 nm results mainly from contributions of tyrosine residues located in the N-terminal domain, while the CD spectrum at 295 nm originates essentially from the two tryptophans of the protein located within the C-terminal domain (Gast *et al.*, 1995). The aromatic spectrum of PGK292p is slightly modified in the 290−295 nm region, indicating the occurrence of some perturbations in the environment of the two tryptophans. The spectrum of PGK72p is modified in the 270−280 nm region, in agreement with the introduction of important changes within the N-domain. The perturbations in the polarity of the tryptophan environment in PGK292p were further supported by the shift of the maximum emission fluorescence wavelength (336 nm for the mutant protein instead of 333 nm for the wild-type protein; data not shown).

A specific conformational probe for the N-domain is the unique cysteine residue, Cys 97. In wild-type PGK, this residue is buried and the rate constant of the reaction with NbS_2 is independent of reagent concentration, the first-order rate constant being $k = 6 \times 10^{-5}$ s^{-1}. In the PGK72p mutant, the thiol group is more reactive, the apparent first-order rate constant of labeling being $k = 6.5 \times 10^{-2}$ s^{-1} (at 250 μM NbS_2 concentration). This indicates that the cysteine side chain is more accessible in the permuted protein than in the wild-type protein. In PGK292p mutant, the circular permutation within the C-domain would not significantly affect the accessibility of Cys 97 residue. In this case, the reaction of the thiol group with NbS_2 displayed biphasic kinetics: a rapid phase with an apparent first-order rate constant $k = 7.7 \times 10^{-2}$ s^{-1} was followed by a slower one with $k = 1 \times 10^{-4}$ s^{-1}. The first phase (about 20% of the total amplitude) corresponds to an accessible cysteine residue, and the slow phase corresponds to a buried cysteine residue. This indicates the presence of two different populations of enzyme molecules.

Stability of the Permuted Proteins. Three different signals were used to compare the stabilities of the permuted and wild-type proteins. The Gdn-HCl denaturation curves of these variants were obtained by measuring the variations in ellipticity at 220 nm and the decrease in activity for increasing concentrations of denaturant and by following the effects of denaturant on the fluorescence properties of the mutant enzymes.

The unfolding transitions of the permuted proteins as assessed by the variations in enzyme activity are symmetrical, and their quantitative analysis shows that the c_m values (0.31 M for PGK72p and 0.29 M for PGK292p) and cooperativity are lower than that of wtPGK (Table 2). Since there are some indications for the presence of two different populations, the two-state equation for unfolding is probably not valid. Therefore, the thermodynamic ΔG_0 values obtained from eq 2 are only apparent $\Delta G^{app}{}_0$, which refer only to the equilibrium between active and inactive populations. The permuted proteins are less stable than wild-type PGK, $\Delta G^{app}{}_0$

Table 2: Unfolding Parameters of Permuted and Wild-Type Yeast Phosphoglycerate Kinase

	signal	wild-type	PGK292p	PGK72p
c_m[b] (M Gdn-HCl)	CD	0.80 ± 0.02[a]	0.26 ± 0.04	0.41 ± 0.08
	$\Delta\lambda_{max}c_{m1}$	0.73 ± 0.02[a]	0.25 ± 0.02	0.36 ± 0.01
	c_{m2}		1.80 ± 0.15	1.59 ± 0.03
	$F_{Int}c_{m1}$	0.73 ± 0.02[a]	0.30 ± 0.03	0.37 ± 0.01
	c_{m2}	1.38 ± 0.02[a]	1.02 ± 0.06	
	activity	0.70 ± 0.02	0.29 ± 0.01	0.31 ± 0.04
n[c]	CD	12.8 ± 0.1[a]	1.09 ± 0.02	4.4 ± 1
	$\Delta\lambda_{max}n_1$	9.0 ± 0.1[a]	2.80 ± 0.55	5.8 ± 0.1
	n_2		2.8 ± 0.7	3 ± 1
	$F_{Int}n_1$	13.0 ± 0.1[a]	1.02 ± 0.08	9.7 ± 2.6
	n_2	3.8 ± 0.3[a]	2.2 ± 0.2	
	activity	19.0 ± 0.4	7.2 ± 3	4 ± 2
$\Delta G^{app}{}_0$ (kcal/mol)	activity	7.9 ± 0.5	4.2 ± 0.2	3.3 ± 0.2
m (kcal mol^{-1} M^{-1})	activity	9.9 ± 1.5	14.5 ± 0.8	10.5 ± 0.6

[a] From Missiakas *et al.*, 1990. [b] c_m, transition midpoints. [c] n, cooperativity indexes of transition.

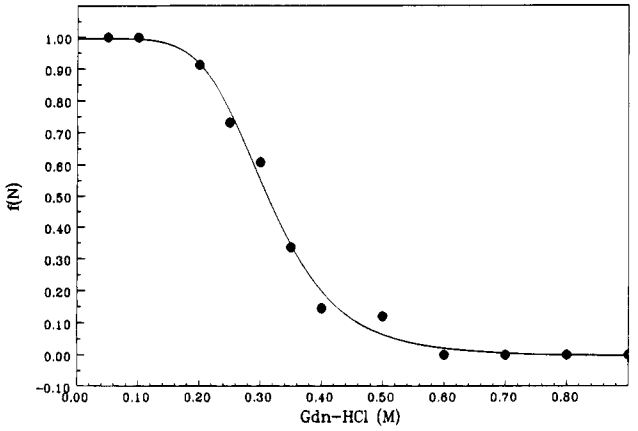

FIGURE 7: Unfolding transition of PGK72p mutant as assessed by the variations in enzyme activity. Experimental data were fitted according to eq 2.

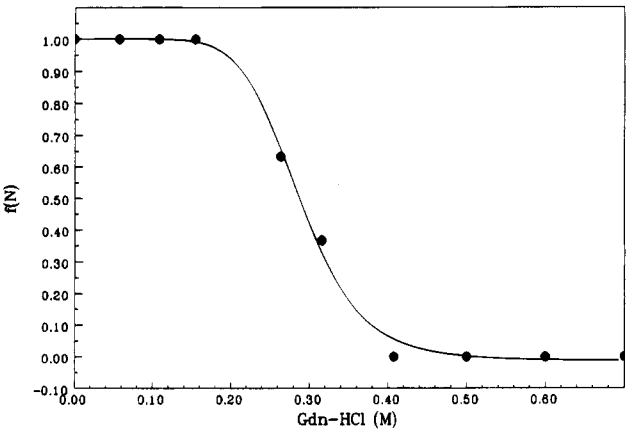

FIGURE 8: Unfolding transition of PGK292p mutant as assessed by the variations in enzyme activity. Experimental data were fitted according to eq 2.

values being 3.3 kcal/mol for PGK72p and 4.2 kcal/mol for PGK292p compared to 7.9 kcal/mol for wtPGK. Similar values for the midpoints were obtained for the unfolding CD transitions (Table 2).

The analysis of fluorescence emission spectra at different denaturant concentrations shows the existence of two different events during the denaturation: a shift in the maximum emission fluorescence wavelength and a change in the fluorescence intensity. These variations are different for the two permuted proteins. In PGK72p, the increase in fluorescence intensity for concentrations of denaturant lower than 0.2 M Gdn-HCl is not accompanied by a change in the position of the maximum emission fluorescence wavelength; in this range of concentrations the enzyme is still active (Figure 7), the shift being associated with the decrease in fluorescence intensity for concentrations of Gdn-HCl greater than 0.2 M. In contrast, in PGK292p, the increase in fluorescence intensity for concentrations of denaturant lower than 0.9 M Gdn-HCl is accompanied by a shift in the maximum emission fluorescence wavelength and by the disappearance of the secondary structure and enzyme activity (Figure 8).

The transitions obtained, on the one hand, from the shift in the maximum fluorescence emission wavelength and, on the other hand, from the change in the fluorescence intensity are reported in Figures 9 and 10 and in Figures 11 and 12, respectively. All transition curves are asymmetrical; the midpoints are reported in Table 2. The unfolding transition

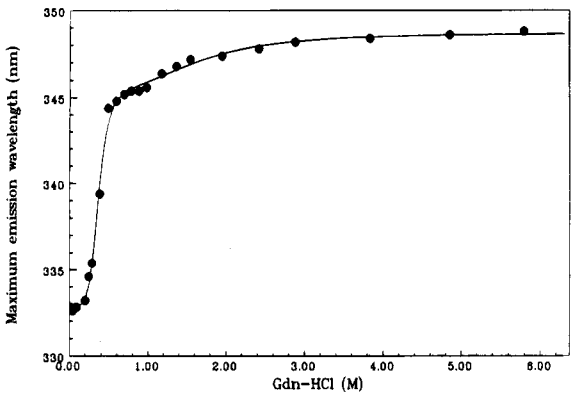

FIGURE 9: Unfolding transition of PGK72p as assessed by the variations in the maximum fluorescence emission wavelength (excitation wavelength: 293 nm).

of PGK72p as assessed by the change in the fluorescence intensity is comparable to that observed for the isolated C-domain and for a mutant deleted of the 12 C-terminal residues (Ritco-Vonsovici *et al.*, 1995). The unfolding transition of PGK292p is comparable to that of wtPGK; it can be deconvoluted in two resolved transitions of opposite amplitude (Missiakas *et al.*, 1990). The first transition corresponds to the unfolding of the native protein leading to a species with a fluorescence higher than those of both native and completely unfolded proteins. The second transition corresponds to the disappearance of the residual

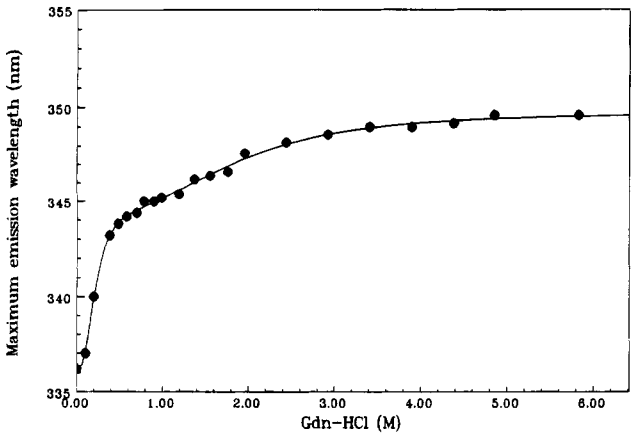

FIGURE 10: Unfolding transition of PGK292p as assessed by the variations in the maximum fluorescence emission wavelength (excitation wavelength: 293 nm).

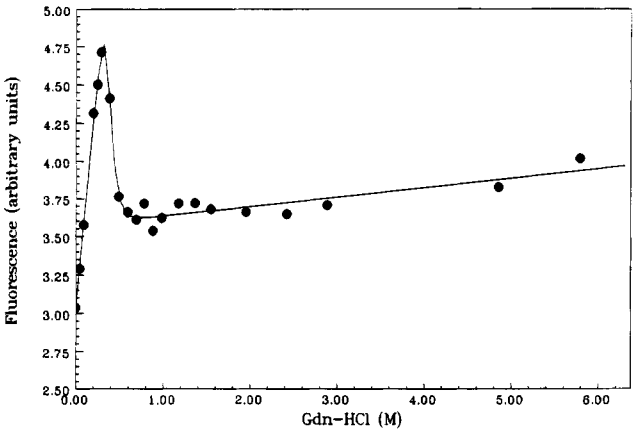

FIGURE 11: Unfolding transition of PGK72p as assessed by the variations in fluorescence emission at 330 nm (excitation wavelength: 293 nm).

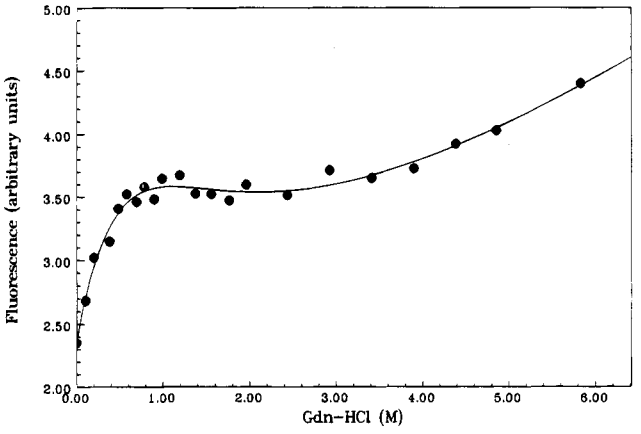

FIGURE 12: Unfolding transition of PGK292p as assessed by the variations in fluorescence emission at 350 nm (excitation wavelength: 293 nm).

microstructures located around tryptophans 308 and 333 in the C-domain (Garcia *et al.*, 1995). The two transitions are also detected in the unfolding curves obtained from the plot of the change in the maximum fluorescence emission wavelength upon denaturant concentration.

The denaturation studies consistently indicated that the permutations drastically destabilize yPGK. Since denaturation is clearly not a two-state process and under native conditions there is some degree of heterogeneity, it is not possible to determine accurate unfolding ΔG_0 values for the mutants from transition curves as assessed by structural signals.

DISCUSSION

In order to investigate the importance of the continuity of the domains in the correct folding and stability of yPGK, two circularly permuted proteins were constructed by a simple PCR methodology. Such permutations were possible because yeast phosphoglycerate kinase belongs to that category of proteins which have the N- and C-termini in relatively close proximity (Thornton & Sibanda, 1983).

To construct these variants, the natural extremities of yPGK were directly connected. Indeed, in the refined three-dimensional structure of yPGK the N- and C- ends are only 5.33 Å apart, suggesting that the introduction of a linker such as polyglycine would be not necessary and possibly could induce an entropic destabilization. These permuted variants can be overexpressed in *S. cerevisiae*. This first remarkable result immediately shows that the permuted proteins fold *in vivo*. Furthermore, the growth of the yeast cells devoid of the PGK chromosomal gene in glucose medium, which requires a functional PGK, indicates that the permuted proteins are enzymatically active.

Gel chromatography, gel electrophoresis, and N-terminal amino acid analysis as well as electrospray mass spectrometry showed that the purified proteins had the expected molecular weights and were not proteolyzed. They possess a compact, globular form as wild-type PGK. These proteins also underwent the same processing as yPGK: removal of the N-terminal Met and acetylation of Ser 1.

The two permuted mutants display a significant specific activity: 14% for PGK72p and 23% for PGK292p with respect to wild-type enzyme activity. This indicates that the overall tertiary structure of yPGK is not strongly affected by the circular permutations. Furthermore, the permuted proteins are able to recover all the original level of activity following the *in vitro* unfolding−refolding cycle. Therefore, the continuity of the domains is not a prerequisite for the folding of enzymatically active yPGK *in vitro* or *in vivo*.

The active site of yPGK includes the cleft between the two domains (Figure 1), 3PG interacting mainly with the N-domain and ATP mainly with the C-domain (Banks *et al.*, 1979; Watson *et al.*, 1982; Harlos *et al.*, 1992). The catalytic reaction requires a hinge-bending motion of the domains bringing the substrates close together in a water-free environment favorable for catalysis. The circular permutation within the C-domain did not affect the enzyme affinity for ATP (K_m for ATP was not greatly affected), but it induced a conformational change that contributes to a lower expression of the catalytic activity.

The direct comparisons of the structures of the wild-type and permuted proteins by spectroscopic methods reveal, as expected, the existence of conformational changes mainly within the permuted domain. Thus, PGK292p mutant presents an increased accessibility of the two tryptophan residues located in the C-domain. The red shift of the maximum emission fluorescence wavelength indicates an increase in polarity of this environment. In this mutant, the reactivity of the cysteine residue toward NbS_2 is slightly increased compared to that of the buried Cys 97 residue in wtPGK: $k_{wtPGK} = 6 \times 10^{-5}$ s^{-1} and $k_{PGK292p} = 1 \times 10^{-4}$ s^{-1}. PGK72p mutant presents structural modifications

mainly in the N-domain. As revealed by the near-UV CD spectrum, the environment of the tyrosine residues is perturbed. The accessibility of the cysteine side chain to solvent is also largely increased. Surprisingly, the circular permutation within this domain also has an effect on the C-domain. The two tryptophans are in the same environment as in wild-type PGK; the position of the maximum emission fluorescence wavelength of native PGK72p is 333 nm as for wtPGK. However, the fluorescence intensity of these residues is strongly affected by low denaturant concentrations. This phenomenon is similar to that described for the mutant deleted of the 12 C-terminal residues, PGK Δ404−415 (Ritco-Vonsovici *et al.*, 1995). In denaturant concentrations lower than 0.2 M Gdn-HCl, the increase in fluorescence intensity is not the consequence of an important change in protein structure, the protein being active and the tryptophans being in the same environment as in the native enzyme. As observed for PGK Δ404−415, under these low denaturant concentrations conditions, the increase in fluorescence intensity seems to correspond to a solvent effect on the native protein due to a higher flexibility induced by the denaturant. The conclusion of these observations is that the permuted domains become, as expected, more "floppy" in the constructed variants; moreover, the permutation in the N-domain also induces some degree of flexibility in the C-domain.

In addition to the structural changes induced in each permuted domain, the circular permutations introduce a heterogeneity of populations. It is well-known that the catalytic properties of an enzyme are very sensitive to any small structural change in the structure of the catalytic site. The significant specific activities of the permuted proteins (14% and 23%) suggest that there are no drastic changes in the structure of these variants. In contrast, the far-UV CD spectra reveal a secondary structure content of 80% with respect to wtPGK. It seems, therefore, that there is a contradiction between the significant level of activity and the significant loss of secondary structure. This discrepancy can be simply explained by the existence of two forms of the protein with a different secondary structure content. The presence of this heterogeneity of populations is confirmed for PGK292p by the biphasic labeling kinetics of Cys 97 with NbS_2, which reveal that 20% of the molecules contain an accessible cysteine residue and 80% of the molecules a buried one. This is consistent with the global secondary structure content resulting from 20% of misfolded molecules and 80% of folded molecules. However, gel filtration data indicate that these heterogeneous populations have similar hydrodynamic properties. For the PGK72p variant, the labeling of the thiol group with NbS_2 is very fast, and therefore a possible heterogeneity cannot be observed by this method.

An important difference in the properties of the permuted and wild-type proteins is the large decrease in stability of the mutants (Table 2). Are these changes in stability attributable to possible strains imposed by the linkage between residues 415 and 1? Zhang *et al.* (1993) reported that a protein with two domains can be permuted with little change in stability when the connection is performed by a linker. This result addresses the question about a possible destabilizing effect of this connection. However, if the absence of linker has some destabilizing effect, we have no evidence that all the effects of the permutations can be simply accounted for by the direct connection. The properties of the two proteins are different even though they have the same connection. Furthermore, the destabilizations are mainly located in the domain where the new discontinuities were introduced. Thus, even if the absence of linker contributes to the destabilization, it probably does not explain all the observed effects of the permutations. Although some strains could be introduced by the junction, the discontinuity of the domains probably introduces a larger contribution to the destabilization of the protein.

The equilibrium unfolding transitions as assessed by the variations in fluorescence emission intensity reveal, in both mutants, the presence of an "hyperfluorescent" state corresponding to different phenomena. For PGK292p, the increase in fluorescence intensity for concentrations lower than 0.9 M Gdn-HCl corresponds to the unfolding of the protein. The second transition corresponds to the disruption of residual structures (Garcia *et al.*, 1995), as also revealed by the asymmetry of the transitions as assessed by the change in the maximum emission wavelength (Figure 9). In contrast, for PGK72p, the larger increase in fluorescence intensity for concentrations lower than 0.2 M Gdn-HCl is due to a strong solvent effect on the fluorescence of the native protein triggered by these low denaturant concentrations (Ritco-Vonsovici *et al.*, 1995). This effect is followed by a decrease in fluorescence intensity between 0.2 and 0.6 M Gdn-HCl which corresponds to the unfolding of the protein. It seems likely that the process of disruption of residual microstructures also occurs for this mutant as indicated by the asymmetry in the transitions as assessed by the change in the maximum emission wavelength (Figure 10).

In summary, the circularly permuted proteins are able to fold efficiently in an active conformation with slight changes in the secondary and tertiary structures related to some heterogeneity in the native state. Although these proteins have a lower stability, their ability to fold is not fundamentally affected, suggesting that the circular permutations have not disrupted any structural units that are essential for folding. Thus, the continuity of the domains might be important for the stability of the folded protein but it is not necessary for this protein to find a folding pathway and to achieve a native-like and active structure.

It is widely believed that protein folding is a "hierarchical process gradually proceeding from local secondary structure via subdomains and domains to the complete tertiary structure" (Jaenicke, 1991). However, the present data indicate that a discontinuity in a domain might change the sequence of events in the folding pathway but does not prevent the protein from achieving a native and active conformation. It remains to be shown how the sequence of folding events is modified in order to accommodate the discontinuity in the domain structure. Previous studies on the folding of yPGK have shown that the N- and C-domains are able to fold independently (Minard *et al.*, 1989). Nevertheless, they interact and fold simultaneously during the folding of the wild-type protein, suggesting that they associate at a stage when they are not completely prefolded (Ballery *et al.*, 1993). Furthermore, studies on yPGK fragments have shown that fragments 1−96 or 249−415 display a low degree of secondary structure but are able to associate functionally with the complementary fragments (Pecorari *et al.*, 1993). This suggests that the association process does not correspond to the association of prestruc-

tured elements but involves a mutual induction of structure.

Taken together, our results suggest that a strictly sequential and hierarchical description of the folding process is not appropriate for yPGK.

ACKNOWLEDGMENT

The authors thank Dr. C. Ghélis for the amino-terminal sequencing of the proteins and Dr. J. P. Lecaer from the Institut Alfred Fessard (C.N.R.S.) who provided expert assistance with the electrospray mass spectrometry. We thank Dr. V. Liepkalns for carefully reading the manuscript and revising the English.

REFERENCES

Alber, T., & Matthews, B. W. (1987) *Methods Enzymol. 154*, 511−533.

Bachmair, A., Finley, D., & Varshavsky, A. (1986) *Science 234*, 179−186.

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, H. M., & Phillips, A. W. (1979) *Nature 279*, 773−777.

Betton, J. M., Desmadril, M., Mitraki, A., & Yon, J. M. (1985) *Biochemistry 24*, 4570−4576.

Bücher, T. (1955) *Methods Enzymol. 1*, 415−422.

Buchwalder, A., Szadkowski, H., & Kirschner, K. (1992) *Biochemistry 31*, 1621−1630.

Chen, C. Y., Opperman, H., & Hitzeman, R. A. (1984) *Nucleic Acids Res. 12*, 8951−8970.

Fairbrother, W. J., Bowen, D., Hall, L., & Williams, R. J. P. (1989) *Eur. J. Biochem. 184*, 617−625.

Garcia, P., Desmadril, M., Minard, P., & Yon, J. M. (1995) *Biochemistry 34*, 397−404.

Gast, K., Damaschun, G., Desmadril, M., Minard, P., Müller-Frohne, M., Pfeil, W., & Zirwer, D. (1995) *FEBS Lett. 358*, 247−250.

Goldenberg, D. P., & Creighton, T. E. (1983) *J. Mol. Biol. 165*, 407−413.

Hahn, M., Piotukh, K., Borriss, R., & Heinemann, U. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 10417−10421.

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

Hawthorne, D. C., & Mortimer, R. K. (1960) *Genetics 45*, 1085−1110.

Herold, M., Leitsler, M., Hage, A., Luger, K., & Kirschner, K. (1991) *Biochemistry 30*, 3612−3620.

Hinnen, A., Hicks, J. B., & Fink, G. R. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 1929−1933.

Horlick, R. A., George, H. J., Cooke, G. M., Tritch, R. J., Newton, R. C., Dwivedi, A., Lischwe, M., Salemne, F. R., Weber, P. C., & Horuk, R. (1992) *Protein Eng. 5*, 427−431.

Jaenicke, R. (1991) *Biochemistry 30*, 3147−3161.

Jecht, M., Tomschy, A., Kirschner, K., & Jaenicke, R. (1994) *Protein Sci. 3*, 411−418.

Kim, P. S., & Baldwin, R. L. (1990) *Annu. Rev. Biochem. 59*, 631−660.

Laemmli, U. K. (1970) *Nature 227*, 680−685.

Levitt, M., & Clothia, C. (1976) *Nature 261*, 552−558.

Luger, K., Hommel, U., Herold, M., Hofsteenge, J., & Kirschner, K. (1989) *Science 243*, 206−210.

Mellor, J., Dobson, M. J., Roberts, N. A., Kingsman, A. J., & Kingsman, S. M. (1985) *Gene 33*, 215−226.

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

Minard, P., Bowen, D. J., Hall, L., Littlechild, J. A., & Watson, H. C. (1990) *Protein Eng. 3*, 515−521.

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

Nozaki, Y. (1970) *Methods Enzymol. 26*, 43−50.

Oas, T. G., & Kim, P. S. (1988) *Nature 336*, 42−48.

Pace, C. N. (1986) *Methods Enzymol. 131*, 266−280.

Pecorari, F., Minard, P., Desmadril, M., & Yon, J. M. (1993) *Protein Eng. 6*, 313−325.

Piper, P. W., & Curran, B. P. G. (1990) *Curr. Genet. 17*, 119−123.

Press, W. H., Flannery, B. P., Teukolsky, S. K., & Vetterling, W. T. (1986) *Numerical Recipes*, Cambridge University Press, Cambridge.

Protasova, N. Y., Kireeva, M. L., Murzina, N. V., Murzin, A. G., Uversky, V. N., Gryaznova, O. I., & Gudkov, A. T. (1994) *Protein Eng. 7*, 1373−1377.

Riddles, P. W., Blakeley, R. L., & Zerner, B. (1979) *Anal. Biochem. 94*, 75−81.

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

Sharma, A. K., Minke-Gogl, V., Gohl, P., Siebendritt, R., Jaenicke, R., & Rudolph, R. (1990) *Eur. J. Biochem. 194*, 603−609.

Thornton, J. M., & Sibanda, B. L. (1983) *J. Mol. Biol. 167*, 443−460.

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.

Wetlaufer, D. B. (1973) *Proc. Natl. Acad. Sci. U.S.A. 70*, 697−701.

Williams, K. P., & Shoelson, S. E. (1993) *Biochemistry 32*, 11279−11284.

Yang, Y. R., & Schachman, H. K. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 11980−11984.

Zhang, T., Bertelsen, E., Benvegnu, D., & Alber, T. (1993) *Biochemistry 32*, 12311−12318.

BI950699Z