

## An 8-fold $\beta\alpha$ barrel protein with redundant folding possibilities

Karolin Luger, Halina Szadkowski and Kasper Kirschner<sup>1</sup>

Abteilung Biophysikalische Chemie, Biozentrum der Universität Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland

<sup>1</sup>To whom correspondence should be addressed

Protein sequences containing redundant segments of secondary structure at both termini have the choice *a priori* of folding into several possible circularly permuted variants of the wild-type tertiary structure. To test this hypothesis the gene of phosphoribosyl anthranilate isomerase from yeast, which is a single-domain 8-fold  $\beta\alpha$  barrel protein, was modified to produce a 10-fold  $\beta\alpha$  homologue in *Escherichia coli*. It contained a duplicate of the two C-terminal  $\beta\alpha$  units of supersecondary structure fused to its N-terminus. Most of the protein was recovered from the insoluble fraction of disrupted cells by dissolution in guanidinium chloride solutions and refolding. Pristine protein was purified from the soluble fraction. The purified  $(\beta\alpha)_{10}$  proteins were enzymically almost fully active. Absorbance, fluorescence and circular dichroism spectra as well as the reversible unfolding behaviour of both proteins were also very similar to the properties of the original  $(\beta\alpha)_8$  protein. Digestion with endopeptidases converted both the pristine and the refolded  $(\beta\alpha)_{10}$  variant to the same large fragment that had the N-terminal sequence and mol. wt of the wild-type  $(\beta\alpha)_8$  protein. The data suggest that the folding of the  $(\beta\alpha)_{10}$  variant is controlled thermodynamically both *in vivo* and *in vitro*.

**Key words:**  $\beta\alpha$  barrel/circular permutation/protein folding/terminal repeat/TIM barrel

### Introduction

It is still unknown how the amino acid sequence of a protein determines its three-dimensional structure (Anfinsen and Scheraga, 1975; Jaenicke, 1987). A novel experimental approach towards solving the protein folding problem is to study proteins whose sequence has been altered at the genetic level (Goldenberg, 1988; Alber, 1989). The substitution of one or more amino acids can in principle identify single residues that are crucial either for the pathway towards the folded structure (Yu and King, 1988) or for its stability (Pakula and Sauer, 1989; Lim and Sauer, 1989). However, more radical changes, such as insertions and deletions (Castagnoli *et al.*, 1989) or the exchange of surface loops (Wharton *et al.*, 1984; Riechmann *et al.*, 1988; Hynes *et al.*, 1989), are also tolerated in certain cases.

Another way of altering the protein sequence drastically is to permute it circularly. This approach was pioneered by Goldenberg and Creighton (1983), who used chemical condensation and limited proteolysis to permute bovine trypsin inhibitor. We have recently shown that two circularly permuted versions of the TRP-1 gene of yeast can be expressed in *Escherichia coli*

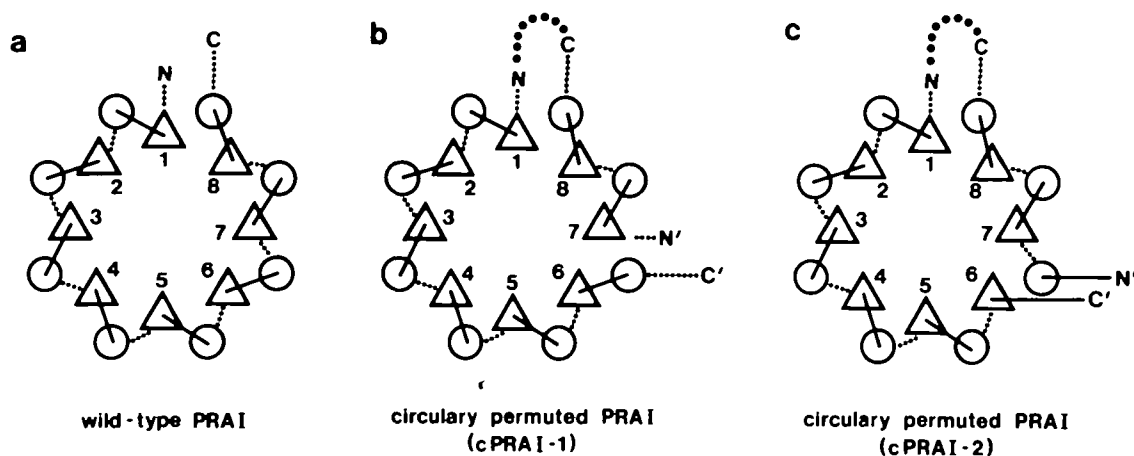
to produce two different, catalytically active variants of phosphoribosyl anthranilate isomerase (PRAI; Luger *et al.*, 1989). The ability of proteins with circularly permuted sequences to fold spontaneously to active molecules proves that both the original termini and the particular surface loops that were cleaved to create the new termini are unimportant for protein folding.

The structure of PRAI from *E. coli* is known (Priestle *et al.*, 1987), and modelling studies with the homologous yeast PRAI (M. Wilmanns, unpublished work) show that it probably has the same fold, namely that of an 8-fold  $\beta\alpha$  barrel. This symmetrical fold was first found in triosephosphate isomerase and has since been observed in at least 16 different enzymes (Chothia, 1988).

Circular permutation of PRAI at the three-dimensional level is illustrated schematically (Sternberg and Thornton, 1977) in Figure 1. The adjacent wild-type N- and C-termini (N and C in Figure 1A) are connected by an oligopeptide linker, and a surface loop is cleaved to create new termini. In the case of one of the circularly permuted variants of PRAI the new termini are between  $\alpha$  helix 6 and  $\beta$  strand 7 (cPRAI-1, N' and C' in Figure 1B). In the case of the other unorthodox variant, which begins with an  $\alpha$  helix rather than a  $\beta$  strand, the new termini are between  $\beta$  strand 6 and  $\alpha$  helix 6 (cPRAI-2, Figure 1C).

It occurred to us that addition of a duplicated segment of supersecondary structure (or 'folding unit'; Levitt and Chothia, 1976) from one terminus to the opposite terminus of PRAI would create a novel sequence with redundant folding information. Figure 2 illustrates this point with a two-dimensional representation of a terminally repeated  $(\beta\alpha)_{10}$  variant of PRAI. The eight sequential  $\beta\alpha$  units of the wild-type PRAI (see Figure 1A) have been 'rolled out' with the side chains of the internal hydrophobic core below the plane of the paper. The mutual orientation of  $\beta$  strands and  $\alpha$  helices is preserved. For illustration the  $\beta\alpha$  units 7 and 8 from the C-terminus of PRAI are duplicated and connected to its N-terminus, with apostrophes indicating the sequence identity.

Lasters *et al.* (1988) have argued that 10-fold  $\beta\alpha$  barrels are probably less stable than 8-fold  $\beta\alpha$  barrels. Detailed studies of the packing of side chains in the core of  $(\beta\alpha)_8$  proteins by Lesk *et al.* (1989) support this view. Therefore it is likely that the  $(\beta\alpha)_{10}$  variant can only fold to a  $(\beta\alpha)_8$  barrel. As shown by the brackets in Figure 2, this sequence of  $\beta$  strands and  $\alpha$  helices has, in principle at least, five different 'folding frames'. The core domain of each of these alternatives contains a circular permutation of the wild-type amino acid sequence. Frame E corresponds to the wild type (see Figure 1A) and frame A to the known circularly permuted variant cPRAI-1 (see Figure 1B). Therefore these two folding frames have a high probability of folding to a stable  $(\beta\alpha)_8$  core. No circular permutations corresponding to folding frames B, C and D are known. Whereas the  $(\beta\alpha)_8$  core of C represents a conventional barrel in that it begins with a  $\beta$  strand, the  $(\beta\alpha)_8$  cores of B and D are unorthodox. However, by beginning with an  $\alpha$  helix they resemble the unorthodox fold of cPRAI-2 (see Figure 1C). Therefore folding frames B, C and



**Fig. 1.** The  $(\beta\alpha)_8$  barrel structure of phosphoribosyl anthranilate isomerase (PRAI) and two circularly permuted variants (Luger *et al.*, 1989). Schematic view down the central axis of the barrel (Sternberg and Thornton, 1977) with the C-ends of the perpendicular  $\beta$  strands ( $\Delta$ ) pointing up. These are numbered sequentially from the wild-type N-terminus (N), and are connected by external perpendicular  $\alpha$  helices (O). The  $\alpha$  helices are numbered as the preceding  $\beta$  strands, and have their C-ends pointing down. (C) C-terminus of the wild-type sequence at the end of  $\alpha$  helix 8. Solid lines represent the loops connecting strands to helices at the top of the barrel and dashed lines represent loops connecting helices to strands at the bottom of the barrel. The active site of PRAI is above the plane of the paper. (A) Wild-type PRAI. (B) First circularly permuted variant (cPRAI-1). The original termini were connected by a new oligopeptide linker (●●●). The new N- (N') and C-termini (C') were generated by cleaving the loop between  $\alpha$  helix 6 and  $\beta$  strand 7. (C) Second circularly permuted variant cPRAI-2, with new termini between  $\beta$  strand 6 and  $\alpha$  helix 6. For partial sequences of critical loops see Figure 4.

D of the  $(\beta\alpha)_{10}$  variant also have a good chance *a priori* of folding to stable  $(\beta\alpha)_8$  cores.

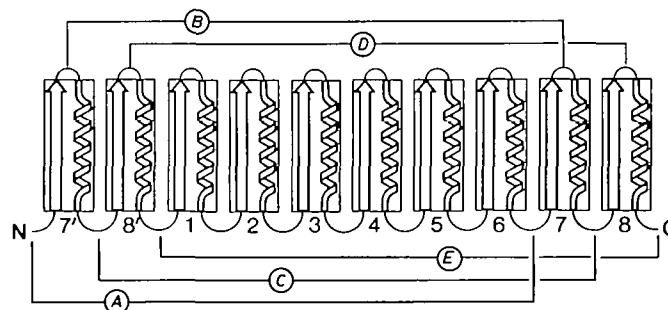
The work described here had the following objectives. (i) Does the presence of redundant segments of supersecondary structure in the protein sequence affect the overall rate and extent of protein folding? (ii) If folding occurs, which is the preferred folding frame? (iii) Would the products of folding *in vivo* and *in vitro* have the same folding frame? It is conceivable that the folding frame A is preferred *in vivo*, if folding proceeds from the N-terminus.

The gene corresponding to the  $(\beta\alpha)_{10}$  variant of PRAI shown schematically in Figure 2 was constructed, and expressed in *E. coli*. The gene product was purified to homogeneity from both the soluble and the insoluble cell fractions. Both the pristine protein and the preparation obtained after refolding from solutions of guanidinium chloride (GuCl) were almost fully active. Preliminary investigations suggest that both have the wild-type  $(\beta\alpha)_8$  folding frame, with the redundant sequence protruding from the N-terminus. These results were presented at the AAAS meeting in San Francisco in January 1989 and have been commented on previously by Goldenberg (1989).

## Materials and methods

### Materials

Restriction enzymes and other DNA-modification enzymes were purchased from Pharmacia, Staehelin or Boehringer. The DNA-sequencing kit was from United States Biochemical Corp.; acrylamide and urea for sequencing gels were from Biorad. Oligonucleotides for sequencing and for mutagenesis were synthesized on an Applied Biosystems ABI 380B oligonucleotide synthesizer and purified using the Beckman HPLC system with a Waters  $\mu$ Bondpak C-18 column (McLaughlin and Krusche, 1982). GuCl grade I, which was used for protein purification, was purchased from Sigma, and ultrapure GuCl, which was used for spectroscopic studies, was from Schwarz-Mann. Chromatography media were brought from Pharmacia. Hydroxylapatite, synthesized according to Atkinson *et al.* (1973), was a kind gift from Ciba-Geigy, Basel. Proteases were purchased from



**Fig. 2.** Schematic two-dimensional representation of the  $(\beta\alpha)_{10}$  variant of PRAI and its folding frames. The  $\beta\alpha$  supersecondary structure units (see Figure 1A) are rolled out in the plane. They are boxed and numbered sequentially from the wild-type N-terminus. Arrows represent  $\beta$  strands. Each  $\beta$  strand is connected by a semicircle that represents a surface loop to the following spiral that represents an  $\alpha$  helix. The N-terminal  $\beta\alpha$  units 7' and 8' are duplicates of the wild-type  $\beta\alpha$  units 7 and 8. Active sites at the top.  $(\beta\alpha)$  units are connected at the bottom by semicircles representing loops between  $\alpha$  helices and  $\beta$  strands. Brackets indicate different possible  $(\beta\alpha)_8$  folding frames. (A) Known circularly permuted  $(\beta\alpha)_8$  variant (cPRAI-1, see Figure 1B). (B, D) Hypothetical circularly permuted  $(\beta\alpha)_8$  variants, in which the core barrel begins with an  $\alpha$  helix, analogous to cPRAI-2 (see Figure 1C). (C) Hypothetical  $(\beta\alpha)_8$  variant. (E) Wild-type PRAI (see Figure 1A).

Boehringer, the corresponding model substrates from Sigma. Tryptone and yeast extract were from Oxoid. All other biochemicals were from Merck or Sigma, of the highest purity available.

### Plasmids and strains

The vector pDS56/RBSII-1 (Certa *et al.*, 1986) with the *E. coli* strain SG200-50 (strain collection by D. Stüber, personal communication) was used for heterologous gene expression, as described earlier (Luger *et al.*, 1989). SG-200-50 is a *lonA-* (*Tn5*) derivative of strain MC 4110 (*F-*, *delta LacU169*, *araD139*, *rpsL*, *relA*, *thiA*, *flbB*) (Casabadian, 1976). The expression-mutagenesis vectors pMa/c with the corresponding host strains WK6 and WK6mutS have been described by Stanssens *et al.* (1989). WK8 is the same as WK6 (in addition *71-18*, *dam-3*) and was used with the repressor plasmid pCI587.

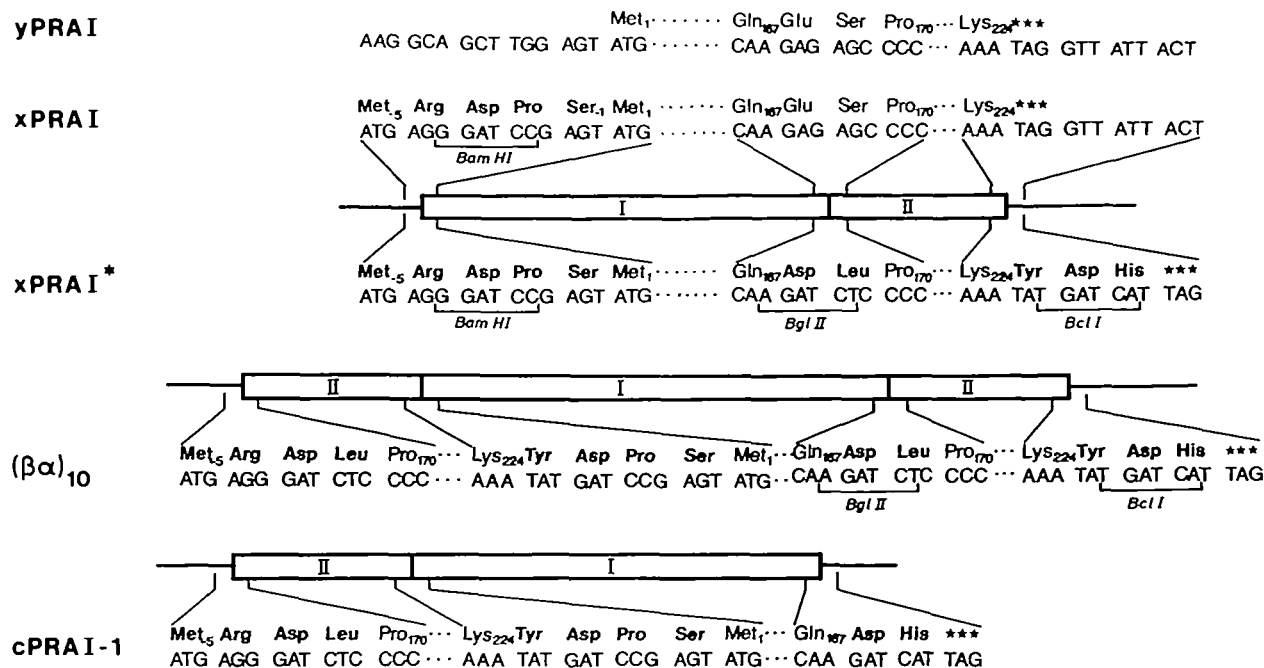


Fig. 3. Construction of the gene encoding the  $(\beta\alpha)_{10}$  variant of PRAI. Partial amino acid and nucleotide sequences; numbering of wild-type PRAI deduced from the TRP1 gene (Tschümper and Carbon, 1980). New amino acids are shown in bold type. New restriction sites are underlined. (yPRAI) Wild-type phosphoribosyl anthranilate isomerase from yeast and the TRP1 gene. (xPRAI) Extended wild-type reference protein. The sequence changes result from introduction of a new *Bam*HI restriction site into the TRP-1 gene. (xPRAI\*) Mutant of xPRAI. The sequence changes result from introduction of new *Bgl*II and *Bcl*I restriction sites.  $[(\beta\alpha)_{10}]$  Variant of PRAI containing an N-terminal copy of the two C-terminal  $\beta\alpha$  units. A copy of fragment II was inserted into the *Bam*HI site of plasmid xPRAI\* in the correct orientation. (cPRAI-1) Circularly permuted  $(\beta\alpha)_8$  variant of xPRAI (see Figure 1B).

#### Isolation of DNA and phages

Double-stranded plasmid DNA was isolated according to Humphreys *et al.* (1975; modified version for large-scale preparation) and Birnboim and Doly (1979; small scale). Single-stranded DNA was purified after infection with the helper phage M13K07 (Pharmacia) according to the procedure described by Stanssens *et al.* (1989). The ratio of single-stranded pMa/c to M13K07 was typically 90%.

#### Site-directed mutagenesis

The *Eco*RI–*Pst*I fragment of the TRP1 gene (Tschümper and Carbon, 1980) was subcloned into the *Eco*RI–*Pst*I sites of the polylinker of pMa. Site-directed mutagenesis was performed in pMa as described by Stanssens *et al.* (1989). The gapped duplex method was used to generate the three unique restriction sites in the TRP1 gene in three consecutive rounds of mutagenesis. In each case the size of the gap was 800 bp. The mutagenic oligonucleotides were 20–40 nucleotides long, depending on the size of the mismatch, with 10–20 matching nucleotides upstream and downstream of the mismatch. The gapped duplex was filled *in vitro* using Klenow DNA polymerase and T4 DNA ligase. The filled gapped duplexes were amplified in the strain WK6mutS and segregated in WK6 (for the mutations generating the unique *Bam*HI and *Bgl*II restriction sites), or in the *dam*<sup>–</sup> strain WK8 (for screening for the methylation-sensitive *Bcl*I restriction site). In all cases, transformants were selected for their resistance towards chloramphenicol. After the N-terminal *Bam*HI site had been introduced, the *Bam*HI–*Pst*I fragment was subcloned again in pMa cleaved with *Bam*HI and *Pst*I to enable expression of the corresponding insert after heat induction. The mutants were screened by checking the isolated plasmid DNA for the presence of the new restriction site. A typical yield of 15–50% positive mutants was routinely obtained by this procedure. Ligated nucleotide regions were confirmed by sequencing (Sanger *et al.*,

1977). All other cloning methods were performed according to standard protocols (Maniatis *et al.*, 1982). The following procedures were carried out as described previously (Luger *et al.*, 1989), unless stated otherwise.

#### Protein purification procedures

The  $(\beta\alpha)_{10}$ -variant (see Figures 2 and 3) and the mutant of PRAI (xPRAI\*, see Figure 3) were purified from inclusion bodies as described previously with slightly lower yields for the  $(\beta\alpha)_{10}$ -variant.

**Purification of the  $(\beta\alpha)_{10}$  variant from the soluble fraction.** All purification steps were performed at 4°C, unless stated otherwise. Transformed cells were cultivated and induced as described previously. After harvesting, the cells were washed with 50 mM Tris–HCl buffer, pH 8.5, containing 5 mM DTE, 2 mM EDTA and 0.3 mM PMSF (buffer A), and resuspended to a concentration of 0.2 g of wet cells per ml in buffer A containing 0.1% Triton X-100. The suspension was quickly frozen in liquid nitrogen and stored at –70°C (Miozzari *et al.*, 1978). A 100 ml quantity of the cell suspension was thawed in a water bath at 0–4°C. Cell debris was removed by centrifugation and after being adjusted with water to a conductivity of 1.8 mS, the clear supernatant was stirred gently with 50 ml of wet DEAE-Sephacel equilibrated with buffer A. This suspension was loaded on top of a prepacked column of DEAE-Sephacel (5 cm × 30 cm) equilibrated with buffer A containing 0.5% Triton X-100 (final column volume of 640 ml). The column was washed with 1 l of buffer A at a flow rate of 4 ml/min and was then eluted with 2.4 l of a linear gradient increasing from 0 to 250 mM NaCl in buffer A. The active fractions were pooled and dialysed against 10 mM potassium phosphate buffer, pH 6.5, containing 100 mM KCl, and 2 mM DTE (buffer B). The dialyzed fractions were applied to a column of hydroxylapatite (3.6 cm × 21 cm, 215 ml), previously equilibrated with buffer B, at a flow rate

of 34 ml/h. The column was eluted with six column volumes of a linear gradient increasing from 10 to 250 mM potassium phosphate pH 6.5, containing 100 mM KCl, at a flow rate of 68 ml/h. The active fractions were pooled and concentrated to ~0.3 mg protein per ml by ultrafiltration (Amicon PM10 filter). The protein solution was frozen rapidly by dripping into liquid nitrogen and stored at  $-70^{\circ}\text{C}$ . The yield was 32% after 49-fold purification from the crude-extract.

#### Determination of molecular weight

$M_r$  values of native proteins were determined by gel filtration on a column of Superose 12 (1 cm  $\times$  30 cm), using a Pharmacia FPLC system. The buffer (50 mM potassium phosphate, pH 7.5, containing 0.1 M NaCl, 1 mM EDTA, and 1 mM DTE; buffer C) was pumped at a flow rate of 0.12 ml/min, at  $20^{\circ}\text{C}$ . Samples (50  $\mu\text{l}$ ) of 1 mg protein per ml were injected and the elution profile was recorded on-line by monitoring the absorption at 280 nm. The column was calibrated by plotting the elution volumes of various standard native proteins versus the logarithm of the known  $M_r$ . The standard proteins and their  $M_r$  were: cytochrome *c* (12 300), horse heart myoglobin (16 950), bovine carbonic anhydrase (30 000), hen egg white ovalbumin (43 000), bovine serum albumin (66 300), rabbit muscle phosphorylase  $\zeta$  (194 800), pig heart lactate dehydrogenase (140 000) and bovine liver catalase (232 000). The apparent mol. wts of the unknown proteins were interpolated from this plot.

Ultracentrifugation was carried out using a Beckman Model E analytical ultracentrifuge equipped with a photo-electric scanner system. The sedimentation equilibrium was established at  $2 \times 10^4$  r.p.m. All measurements were performed at  $20^{\circ}\text{C}$  using a loading concentration of 0.5 mg protein per ml in buffer C.

Polyacrylamide gel electrophoresis in the presence of SDS was performed according to Lämmli (1970) and Studier (1973), or as described by Schägger and Jagow (1987). Routinely, gels with 12.5% polyacrylamide were used and stained with Coomassie Brilliant Blue. The following standard proteins were used: rabbit muscle phosphorylase (97 400), bovine serum albumin (66 300), *E. coli* tryptophan synthase  $\beta$ -subunit (43 000), rabbit muscle glyceraldehyde-3-phosphate dehydrogenase (35 700), *E. coli* tryptophan synthase  $\alpha$ -subunit (28 740), horse heart myoglobin (16 950), cytochrome *c* (12 300) and bovine pancreatic trypsin inhibitor (6500).

#### Determination of protein concentration

Protein concentrations were determined according to Bradford (1976) using the 'Biorad Protein Assay', with bovine serum albumin as a standard.

#### Unfolding and refolding studies

Equilibrium unfolding studies were carried out by monitoring the circular dichroism at 222 nm. Data from four runs were averaged, with the experimental error being ~5%.

The reversibility of unfolding was checked by means of enzyme activity as follows: solutions of 0.3 mg protein per ml in 50 mM potassium phosphate buffer, pH 7.5, containing 100 mM KCl, 2 mM DTE and 2 mM EDTA (buffer D) were dialysed for 90 min against buffer D containing 4 M GuCl (buffer E). Refolding was performed by dialysing the denatured protein solutions for 2 h intervals against three changes of 100 vols of buffer D. Activities were measured and compared to the activities of protein solutions that had not been denatured but had remained in buffer D for the same time interval. All steps were carried out at  $4^{\circ}\text{C}$ . Three independent measurements were averaged and the experimental error was within 3%.

**Refolding kinetics.** Due to the sensitivity of xPRAI and its variants towards irradiation by UV light, measurements were done at different times for periods of only 20 s, and the average relative fluorescence was recorded for the midpoint of the measuring period. The data of at least four independent measurements were averaged.

#### Limited proteolysis

**Proteolysis with trypsin.** Protein solutions (1.5 mg protein per ml) were dialyzed against 100 mM Tris-acetate buffer, pH 7.6, supplemented with 1 mM EDTA and 1 mM DTE (buffer F). Trypsin was dissolved in 1 mM HCl at a concentration of 0.5 mg/ml. A 200  $\mu\text{l}$  aliquot of the protein solution was mixed with 5  $\mu\text{l}$  of trypsin stock solution and incubated at room temperature. At the time intervals given in Figure 5, samples were removed from the incubation mixture, mixed with the same volume of sample buffer (Lämmli, 1970), and heated for 5 min at  $100^{\circ}\text{C}$ . Parallel samples were used to assay the activity of PRA isomerase as described previously. The heated samples were subsequently analyzed by gel electrophoresis in the presence of SDS as described above. The activity of the protease under the described conditions was checked by degradation of tryptophan synthase  $\beta_2$  subunit from *E. coli*. The characteristic fragments (Högberg-Raubaud and Goldberg, 1977) appeared within 10 min at room temperature.

**Proteolysis with LysC and ArgC endopeptidases.** A 250  $\mu\text{g}$  amount of protein in a final volume of 250  $\mu\text{l}$  of buffer F was mixed with either 3 milliunits of LysC or 5 units of ArgC protease dissolved in buffer F, and incubated at  $25^{\circ}\text{C}$ . Samples taken after different time intervals were treated as described above. The large proteolytic fragment ( $M_r$  ~25 000) was purified from a parallel batch by gel filtration on a column of Superose 12. The preparations were checked for homogeneity by gel electrophoresis in the presence of SDS and were then used for determining the amino-terminal sequence (Luger *et al.*, 1989).

**Proteolysis with factor  $X_a$ .** Activation of factor X was achieved by adding 25  $\mu\text{l}$  of Russel's Viper venom to a solution of 0.25 mg of factor X per ml of 8.3 mM Tris-HCl buffer, pH 8.0, containing 8.3 mM  $\text{CaCl}_2$  and 67 mM NaCl (Nagai and Thogersen, 1987). The mixture was incubated for 1 h at  $37^{\circ}\text{C}$ . The activated factor X (designated  $X_a$ ) was stored in aliquots at  $-20^{\circ}\text{C}$ . Protein samples were adjusted to a concentration of 2 mg protein per ml in 50 mM Tris-HCl buffer, pH 8.0, containing 100 mM NaCl and 1 mM  $\text{CaCl}_2$  (buffer G). Protein stock solution (20  $\mu\text{l}$ ) was mixed with 100  $\mu\text{l}$  of  $X_a$  stock solution, and the concentration of the buffer was adjusted to 50 mM by adding 10  $\mu\text{l}$  of 10-fold concentrated buffer G. After different incubation times, parallel aliquots were either used for enzyme assays or were heated for 5 min at  $100^{\circ}\text{C}$  in sample buffer and subjected to gel electrophoresis in the presence of SDS as described above. As a control for the activity of  $X_a$ , 42  $\mu\text{g}$  of prothrombin was digested with 3  $\mu\text{g}$  of  $X_a$  under the above conditions. Proteolysis of prothrombin to thrombin, as analyzed by gel electrophoresis in the presence of SDS, was completed after 1 h. The large proteolytic fragment was purified and sequenced as described above.

## Results

### Constructions

The construction of the coding sequence of the  $(\beta\alpha)_{10}$  variant of yeast PRAI that is shown schematically in Figure 2 is summarized in Figure 3. Introduction of a unique *Bam*HI site

**Table I.** Steady-state and refolding kinetics of  $(\beta\alpha)_8$  wild-type,  $(\beta\alpha)_8$  circularly permuted and  $(\beta\alpha)_{10}$  variants of yeast PRAI

Protein	Quality	Steady-state constants			Slow refolding rate constant $10^3 k_{\text{obs}}$ ( $s^{-1}$ )
		$k_{\text{cat}}$ ( $s^{-1}$ )	$K_M$ ( $\mu\text{M}$ )	$\frac{k_{\text{cat}}}{K_M}$ ( $\mu\text{M}^{-1}\text{s}^{-1}$ )	
Extended PRAI (xPRAI <sup>a</sup> )	refolded, fresh	79 ± 8	3.7 ± 0.6	21	3.5 ± 0.2
Mutant PRAI (xPRAI <sup>a,c</sup> )	refolded, stored	28 ± 9	3.4 ± 0.4	8	nd
Circularly permuted PRAI (cPRAI-1 <sup>a</sup> )	refolded, fresh	59 ± 3	5.5 ± 0.5	11	1.7 ± 0.1
$(\beta\alpha)_{10}$ variant <sup>b</sup>	pristine, stored	27 ± 5	2.7 ± 0.2	10	nd
$(\beta\alpha)_{10}$ variant <sup>b</sup>	refolded, fresh	39 ± 6	2.7 ± 0.2	14	1.7 ± 0.1
$(\beta\alpha)_{10}$ variant <sup>b</sup>	refolded, stored	22 ± 7	3.0 ± 0.2	7	nd

$k_{\text{cat}}$  and  $K_M$  values measured by monitoring disappearance of substrate as described (Luger *et al.*, 1989). 50 mM Tris-HCl buffer, pH 7.5, at 25°C.  $k_{\text{obs}}$  is the rate constant of the slow phase of refolding from 4 M GuCl solutions in 50 mM potassium phosphate buffer, pH 6.5, containing 0.1 M KCl, 20 mM DTE and 10 mM  $\text{K}_2\text{Mg-EDTA}$ , at 20°C. nd: not determined.

<sup>a</sup>See Figure 1.

<sup>b</sup>See Figure 2.

<sup>c</sup>See Figure 3.

was required to subclone the TRP-1 gene (Tschümper and Carbon, 1980) into a vector allowing high expression in *E. coli*. This mutation led to the extension of the wild-type PRAI (Figure 3, yPRAI) by five amino acids at the N-terminus (Figure 3, xPRAI). Because all variants discussed in this work begin with similar short extensions, we consider xPRAI to be the proper reference wild-type  $(\beta\alpha)_8$  protein.

Introduction of the new *Bgl*III and *Bcl*I sites that were required for the subsequent translocation step led to another mutant of yPRAI (Figure 3, xPRAI\*) that has the two substitutions Glu168Asp, Ser169Leu in the loop between  $\alpha$  helix 6 and  $\beta$  strand 7 (see Figure 1A) and three extra amino acid residues (Tyr, Asp and His) at the C-terminus. The properties of xPRAI\* are interesting because the folding frames B, C, D and E depicted in Figure 2 differ from the sequence of xPRAI at positions 168 and 169.

Insertion of a duplicate of the coding region II into the *Bam*HI site of the gene of xPRAI\* led to the continuous coding sequence of the  $(\beta\alpha)_{10}$  variant [Figure 3,  $(\beta\alpha)_{10}$ ], which has the same C-terminus as xPRAI\*. The partial sequence of cPRAI-1 (see Figure 1B) is shown for comparison in the bottom line of Figure 3.

#### Purifications

xPRAI\* and the  $(\beta\alpha)_{10}$  variant were expressed in *E. coli* as described previously for xPRAI and cPRAI-1. Again the major fraction of each overproduced protein was insoluble. After disruption of the cells and centrifugation, the pellets were dissolved in concentrated solutions of GuCl, refolded and purified by the procedure described previously. A pure protein was obtained with 60% yield after ~2-fold purification from the dissolved pellets (data not shown).

The cytoplasmic fraction contained ~5% of the total recoverable activity. It was used to purify pristine protein that had not passed through a cycle of aggregation, unfolding and refolding. All subsequent studies were done with both xPRAI\* (see Figure 3) and the refolded  $(\beta\alpha)_{10}$  variant, if not stated otherwise.

#### Determination of the molecular weight

To exclude the possibility that activity depends on the formation of dimers (Jackson and Yanofsky, 1969a,b) the mol. wts of xPRAI\* and of the pristine and refolded  $(\beta\alpha)_{10}$ -variants were determined by three independent methods. Gel electrophoresis in the presence of dodecyl sulphate showed good agreement with

the calculated  $M_r$  of 32 000 [ $(\beta\alpha)_{10}$  variant] and 25 000 (xPRAI\*; data not shown). Similar data were obtained by gel filtration of the native proteins on a column of Superose 12 and by equilibrium ultracentrifugation (data not shown). These results prove that xPRAI\* as well as both  $(\beta\alpha)_{10}$  variants from the cytoplasm and the insoluble fraction are monomeric under native conditions.

#### Steady-state kinetics

The active site of PRAI is formed by residues located in several of the  $\beta$  strands and their following loops (Priestle *et al.*, 1987). None of the alterations necessary for constructing either xPRAI\* or the  $(\beta\alpha)_{10}$  variant involve any of the active site residues or their immediate sequence neighbours (see Figures 2 and 3). Therefore,  $K_M$  and  $k_{\text{cat}}$  values should be sensitive indicators of the integrity of the overall structure of the central  $(\beta\alpha)_8$  core. The steady-state kinetic parameters of xPRAI, xPRAI\* and the  $(\beta\alpha)_{10}$  variants are shown in Table I.

The most important finding is that the  $K_M$  values of the pristine and the refolded  $(\beta\alpha)_{10}$  variants were quite similar to those of the reference (xPRAI) and mutant (xPRAI\*)  $(\beta\alpha)_8$  proteins. The  $k_{\text{cat}}$  values were ~2-fold smaller, but the effect of prolonged storage accounts for part of the decrease. In summary, these preliminary data indicate that xPRAI\* and both  $(\beta\alpha)_{10}$  variants have active sites that are very similar to that of the reference protein. However, since two of the folding frames depicted in Figure 2, namely those of xPRAI and cPRAI-1, have practically the same  $K_M$  and  $k_{\text{cat}}$  values (see Table I), these data cannot be used to distinguish between different folding frames. We therefore turned to the following qualitative method to make the distinction.

#### Limited proteolysis with endoproteases

If the tertiary structure of the  $(\beta\alpha)_{10}$  variant consists of a  $(\beta\alpha)_8$  barrel core (see Figure 2), the terminal extension(s) are probably joined by flexible linkers to the core, and therefore should be removable by endopeptidases. The reference  $(\beta\alpha)_8$  protein xPRAI and its two circularly permuted  $(\beta\alpha)_8$  variants cPRAI-1 and cPRAI-2 are resistant towards proteolysis by trypsin, ArgC and LysC proteases and factor  $X_a$ , despite the presence of numerous potential cleavage sites in loop regions as well as in external  $\alpha$  helices (Luger *et al.*, 1989). By treatment of the  $(\beta\alpha)_{10}$  variant with different endopeptidases under native conditions, and by sequencing of the N-termini of the resulting

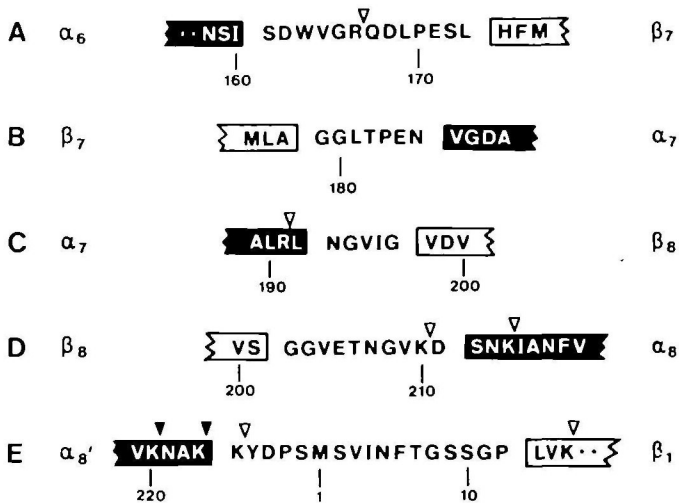


Fig. 4. Oligopeptide linkers of the  $(\beta\alpha)_{10}$  variant of PRAI between various  $(\beta\alpha)_8$  cores and their corresponding terminal extensions. Folding frames A, B, C, D and E as well as the numbering of segments of secondary structure as in Figure 2. Amino acid residues in single letter code are numbered as in Figure 3, and comprise the sequence of xPRAI\* from Asn158 to the C-terminal Lys224. Residues in secondary structures ( $\alpha$ :  $\alpha$  helix;  $\beta$ :  $\beta$  strand) are boxed. Boundaries of secondary structure segments are derived from model building of the sequence of yeast PRAI using the refined coordinates of PRAI from *E.coli* (Priestle *et al.*, 1987; M.Wilmanns and J.N.Jansonius, personal communication). ( $\nabla$ ) Potential and ( $\blacktriangledown$ ) actual cleavage sites for various endopeptidases described in the text.

peptidase-resistant cores, we attempted to gain information on the precise boundaries of the  $(\beta\alpha)_8$  barrel cores, that is, on the actual folding frame. Comparison of the folding frames of both the pristine and the refolded  $(\beta\alpha)_{10}$  variants should show whether the product of folding *in vivo* is different from that *in vitro*.

When the refolded  $(\beta\alpha)_{10}$  variant ( $M_r$  32 000) was treated with trypsin, it was degraded to a stable fragment with an  $M_r$  of  $\sim 25$  000, as judged by gel electrophoresis in the presence of SDS. The PRAI activity remained unchanged by comparison to the control without protease (data not shown). These results indicate that limited digestion with trypsin removes the terminal extension(s), wherever they are, and leaves an active  $(\beta\alpha)_8$  barrel core intact.

Figure 4 presents the various surface loops of the  $(\beta\alpha)_{10}$  variant that are potential linkers between the five different possible  $(\beta\alpha)_8$  cores and their respective terminal extensions (see Figure 2). As indicated by the triangles above the amino acid sequence, only the loops between  $\alpha_6$  and  $\beta_7$  (Arg166),  $\beta_8$  and  $\alpha_8$  (Lys210) and  $\alpha_8'$  and  $\beta_1$  (Lys223 and Lys224) contain potential cleavage sites for trypsin. The previous observation that the circularly permuted  $(\beta\alpha)_8$  variant cPRAI-1 (Figure 1B) is not cleaved by trypsin shows that the new link between  $\alpha_8'$  and  $\beta_1$  is inaccessible to proteases when it is part of a  $(\beta\alpha)_8$  core.

To identify the site of cleavage by trypsin we used separately the more specific endopeptidases ArgC and LysC, which cleave Arg-Xaa and Lys-Xaa peptide bonds, respectively. The loop between  $\alpha_8'$  and  $\beta_1$  contains only two sites for LysC (Lys223 and Lys224), whereas the loop connecting  $\alpha_6$  and  $\beta_7$  has only one potential cleavage site for ArgC (Arg166). Therefore, cleavage of the  $(\beta\alpha)_{10}$  variant with LysC to a fragment of  $M_r$  25 000 with retention to full PRAI activity would support the folding frame E, whereas complete cleavage with ArgC would support folding frame A. A mixture of two or more non-

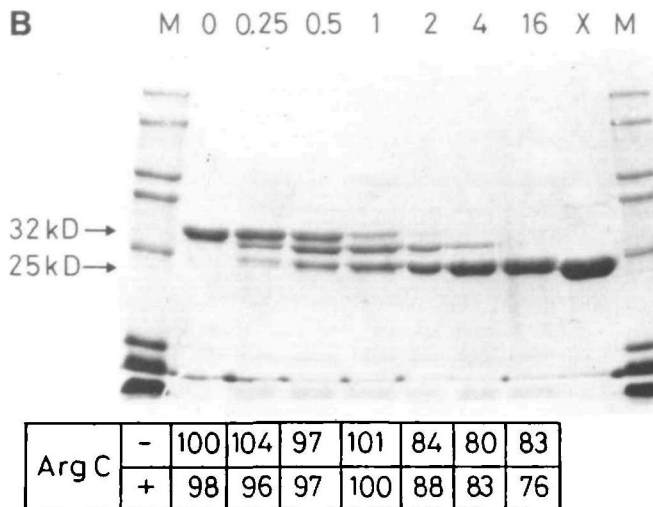
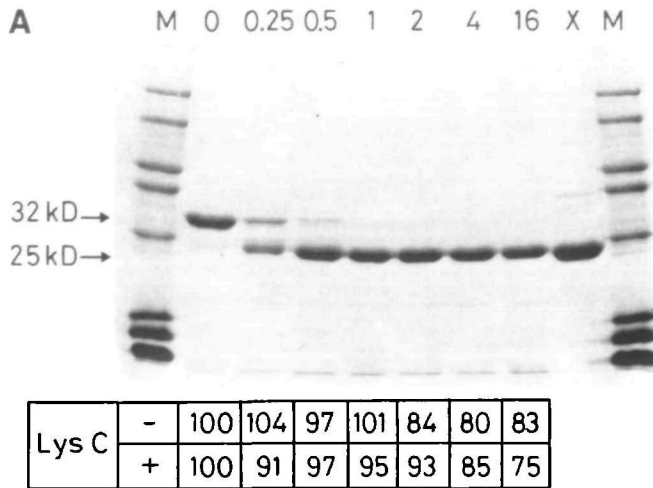


Fig. 5. The  $(\beta\alpha)_{10}$  variant is trimmed by LysC and ArgC endopeptidases. Kinetics of limited proteolysis measured by gel electrophoresis in presence of SDS. Lane M:  $M_r$  of marker proteins, from top to bottom: 97 000, 67 000, 43 000, 35 000, 29 000, 17 000, 12 000 and 6500. Lane X: wild-type reference protein xPRAI,  $M_r$  25 000. Lanes 0-16 indicate incubation period, in hours. The percentages below the shots show retention of PRAI activity at the indicated times in the absence (-) and presence (+) of the respective endopeptidase. (A) Action of LysC endopeptidase. (B) Action of ArgC endopeptidase.

equilibrating folding frames would be indicated by the inability of each protease to degrade all of the molecules to  $M_r$  25 000 fragments.

Contrary to expectation, both proteases led to complete degradation of the  $(\beta\alpha)_{10}$  variant to fragments of  $M_r$  25 000 (Figure 5) that retained  $\sim 80\%$  of the original enzymic activity. The large fragments were not degraded further, even after prolonged incubation. We used two different methods of gel electrophoresis for detecting polypeptide fragments in the  $M_r$  range 4000-10 000 (Lämmli, 1970; Schägger and Jagow, 1987), but obtained no evidence for the extraneous  $(\beta\alpha)_2$  region of  $M_r$  7000 expected as the other product of limited proteolysis of folding frame E (see Figure 2). The peptide of  $M_r$   $\sim 7000$  observed in Figure 5(B) is a contaminant of the  $(\beta\alpha)_{10}$  preparation present at zero time of incubation.

The  $M_r$  25 000 fragments obtained with each protease were purified and the respective N-terminal amino acid sequences were determined. LysC protease produced a unique fragment with the



N-terminal sequence KYDPSMSVINF. . . , and ArgC protease produced a fragment with NAKKYD. . . Thus both proteases cleaved peptide bonds that are in or at the end of  $\alpha$  helix 8' rather than in the loop connecting  $\alpha_8'$  with  $\beta_1$ , suggesting that the  $(\beta\alpha)_{10}$  variant possesses the folding frame E of the reference wild-type protein (see Figure 2).

Since no intermediate accumulated transiently in the course of proteolysis of  $(\beta\alpha)_{10}$  by LysC protease (Figure 5A), it appears that the Lys210'–Asp211' bond located in the loop connecting  $\beta_8'$  and  $\alpha_8'$  (see Figure 4) was cleaved more slowly than the Lys223–Lys224 bond. ArgC protease cleaved at one of the LysC-specific sites (Lys220'–Asn221'), for unknown reasons. The important observation is that none of the Arg–Xaa bonds that are located at the limits of the  $(\beta\alpha)_8$  cores of the other possible folding frames (e.g. Arg166–Gln in frame A, Arg191–Leu in frame C; see Figure 4) were cleaved by ArgC protease. The observed intermediate fragment of  $M_r$  30 000 (Figure 5B) could be due to specific cleavage at residue Arg191' at the end of  $\alpha_7'$ , before cleavage at Lys220' occurs.

In order to detect a minor fraction that possibly could possess the folding frame A, the  $(\beta\alpha)_{10}$  variant was digested with the blood plasma factor  $X_a$ . This extremely specific protease cleaves mainly at Ile–Glu–Gly–Arg–I–Xaa [the site of cleavage is indicated by an arrow (Coghlan and Vickery, 1989; Husten *et al.*, 1987)], but sometimes also at Gly–Arg–I–Xaa (Nagai and Thogersen, 1987). The only potential cleavage site for  $X_a$  is in the loop connecting  $\alpha_6$  with  $\beta_7$  (. . . Gly–Arg166–I–Gln. . . ; see Figure 4).

Even after long incubation times with large amounts of factor  $X_a$  endopeptidase, only partial degradation of the  $(\beta\alpha)_{10}$  variant to an  $M_r$  25 000 fragment was observed. Again, the PRAI activity remained practically constant (data not shown). N-Terminal sequencing of the purified  $M_r$  25 000 fragment revealed that it consisted of a mixture of peptides with at least two different N-termini. However, the two main sequences (NAKKYDP. . . and KYDPSM. . . ) prove that cleavage occurred at the sites attacked by LysC and ArgC proteinases (see Figure 4). The unexpected cleavage sites within the  $(\beta\alpha)_{10}$  variant cannot be due to thrombin, which is known to be the predominant contamination of preparations of factor  $X_a$  (Nagai and Thogersen, 1987), because thrombin is specific for Arg–Xaa bonds. It is unclear whether plasmin or some other protease is the responsible agent. Neither xPRAI nor cPRAI-1 were degraded by our preparation of factor  $X_a$ . Therefore the loops connecting  $\alpha_6$  with  $\beta_7$  in xPRAI and  $\alpha_8'$  with  $\beta_1$  in cPRAI-1 were inaccessible to both factor  $X_a$  and the unknown contaminating protease.

The  $(\beta\alpha)_{10}$  variant that was purified from the cytoplasm was also treated with both LysC and ArgC proteases. However, the SDS gel patterns were identical to those of the molecule refolded *in vitro*. The corresponding proteinase-resistant cores also retained full PRAI activity (data not shown), and the sequence of the LysC fragment (KYDPSM. . . ) agreed with the sequence obtained from the LysC fragment of the refolded molecule. Therefore, according to the criterion of susceptibility to endopeptidases, both the pristine and the refolded  $(\beta\alpha)_{10}$  variant must have the same folding frame.

#### Spectroscopic measurements

In order to obtain independent information on whether the N-terminal  $(\beta\alpha)_2$  extension is folded or not, we compared the circular dichroism (CD) spectra of the  $(\beta\alpha)_{10}$  variant to those of the reference protein xPRAI\* and the circularly permuted

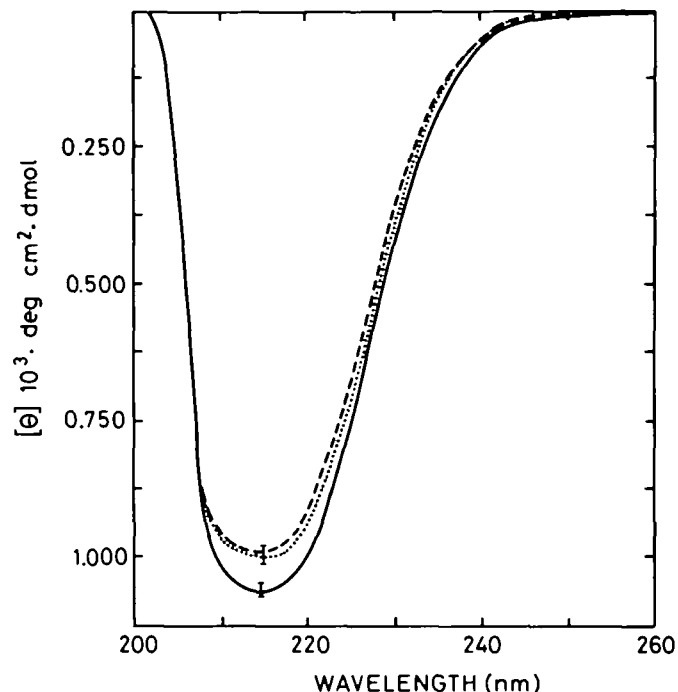


Fig. 6. The circular dichroism spectrum of the  $(\beta\alpha)_{10}$  variant is very similar to that of the  $(\beta\alpha)_8$  wild-type reference protein. Measurements in 0.05 M potassium phosphate buffer, pH 7.5, at 20°C in  $d = 0.1$  cm cuvettes. Solid line: 0.7 mg xPRAI/ml. Dashed (cPRAI-1) and dotted lines ( $(\beta\alpha)_{10}$  variant) are at concentrations giving the same mean residue concentration as xPRAI.

$(\beta\alpha)_8$  variant cPRAI-1. The protein solutions were adjusted to the same optical density at 230 nm, where the absorption of the peptide bonds is additive (Cantor and Schimmel, 1980), to ensure that each measurement was performed with an identical concentration of mean amino acid residues. The terminal extension of the  $(\beta\alpha)_{10}$  variant increases the number of residues and potential secondary structure segments of the sequence of the reference protein by 25%. Therefore, a completely random conformation of the extension would decrease the mean residue ellipticity by ~20%.

However, the normalized CD spectra were identical within the experimental error ( $\pm 5\%$ ) for all three proteins (Figure 6). It appears that the additional  $(\beta_7\alpha_7\beta_8\alpha_8)'$  fragment has a similar average content of secondary structure to that when it is part of the folded reference protein. The UV absorption and fluorescence emission spectra (excited at 280 nm), which were measured under the same conditions, were identical for the same three proteins within the experimental error. They provide little information on the secondary and tertiary structure of the extension and are therefore not shown here, because the 58 extra amino acids contain neither Tyr nor Trp and only two Phe residues (see Figures 3 and 4).

#### Equilibrium unfolding and refolding

To investigate the intrinsic stability of the  $(\beta\alpha)_{10}$  variant and to assess the influence of the terminal extension, we measured the equilibrium unfolding and refolding transitions and compared them to those of the reference protein (xPRAI) and one of the circularly permuted variants (cPRAI-1).

As shown in Figure 7, the overall GuCl-dependent unfolding transitions of both the  $(\beta\alpha)_{10}$  variant and xPRAI (as measured by CD at 222 nm) were practically indistinguishable, whereas that of cPRAI-1 was displaced more towards lower concentrations of GuCl.

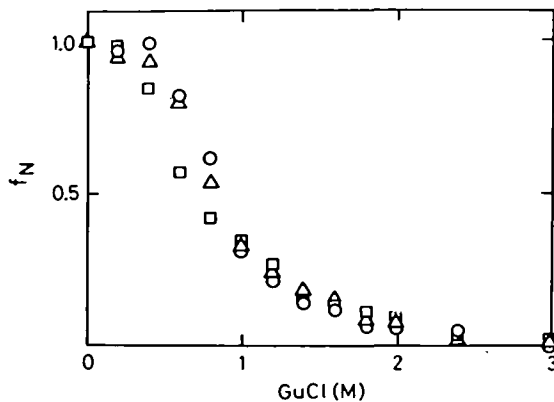


Fig. 7. The unfolding transitions of the  $(\beta\alpha)_{10}$  variant and the  $(\beta\alpha)_8$  wild-type reference protein at equilibrium are very similar. Averaged measurements are presented of circular dichroism at 222 nm in the presence of the indicated concentrations of GuCl in the buffer under Fig. 6, at 20°C. The linear change of mean residue ellipticity ( $\theta_{222}$ ) above 3 M GuCl was extrapolated to 0 M GuCl.  $f_N$  is the normalized deviation of  $\theta_{222}$  from this base line. (○) Wild-type  $(\beta\alpha)_8$ , xPRAI. (△)  $(\beta\alpha)_{10}$  variant. (□) Circularly permuted  $(\beta\alpha)_8$ , cPRAI-1.

The reversibility of unfolding was demonstrated by refolding experiments using dialysis. At a protein concentration of 0.3 mg/ml, the yield of enzymic activity after one round of unfolding and refolding was close to 100% for xPRAI, xPRAI\*, cPRAI-1 and the  $(\beta\alpha)_{10}$  variant (data not shown). Similar findings have also been reported for enolase (Teipel and Koshland, 1971), aldolase (Rudolph *et al.*, 1983) and the  $\alpha$  subunit of tryptophan synthase (Jackson and Yanofsky, 1969a; Yutani *et al.*, 1977; Matthews and Crisanti, 1981). High refolding yields may be an intrinsic property of  $(\beta\alpha)_8$  barrel proteins.

#### Refolding kinetics

As a preliminary check whether the N-terminal  $(\beta\alpha)_2$  region affects the kinetics of refolding of the  $(\beta\alpha)_{10}$  variant, refolding was initiated by rapid dilution of GuCl-containing solutions and monitored by tryptophan fluorescence. The progress curve consisted of both a rapid (half-time  $t_{1/2} < 1$  s) and a slow phase, as observed previously in the case of the reference protein xPRAI and one of the circularly permuted variants cPRAI-1 (Luger *et al.*, 1989). The latter phase, which has an amplitude of only 33% of the total amplitude, was fitted to a single exponential and the rate constants were compared to those of xPRAI and cPRAI-1 (see Table I). The rate constants of the slow refolding phase of cPRAI-1 and the  $(\beta\alpha)_{10}$  variant were approximately half that of xPRAI. Preliminary studies with cPRAI-1 involving CD measurements indicate that a large fraction of the total amplitude was attained immediately after dilution (K.Luger, unpublished data), and we assume that this is true for all three proteins. Therefore the presence of a duplicated region of the C-terminus fused to the N-terminus does not affect significantly the slow phase of folding of the  $(\beta\alpha)_8$  core.

#### Discussion

The initial investigations reported here show that the wild-type  $(\beta\alpha)_8$  region of the  $(\beta\alpha)_{10}$  variant of PRAI folds with the same overall rate and to the same extent as the wild-type reference  $(\beta\alpha)_8$  protein. Moreover, the product of folding *in vivo* is the same as that of folding *in vitro*.

It is pertinent to ask whether an equilibrium exists between the differing folding frames shown in Figure 2. If such were the case and only the wild-type conformer were susceptible to attack

by endopeptidases, the population would be shifted entirely towards that species, and the results of Figure 5 would be misleading.

The properties of another 8-fold  $\beta\alpha$  barrel protein, namely the  $\alpha$  subunit of tryptophan synthase (Hyde *et al.*, 1988), have some bearing on this question. The  $\alpha$  subunit forms non-covalent complementing dimers by mutual exchange of terminal regions of the polypeptide chain (Jackson and Yanofsky, 1969a). It also forms stoichiometric complexes with duplicate polypeptide fragments (Jackson and Yanofsky, 1969b). In that the  $(\beta\alpha)_{10}$  variant of PRAI contains redundant regions of polypeptide sequence at its termini, it represents a single-chain equivalent to the complexes of the  $\alpha$  subunit with duplicate fragments.

The important observations of Jackson and Yanofsky (1969a,b) in this regard is that these enzymically active complexes could only be formed by dissolving the protein (and the fragments) at high concentrations in 6 M urea and then removing the urea by dialysis. Moreover, these dimers and complexes could be dissociated by heating only at  $\sim 50^\circ\text{C}$ , which is close to the thermal unfolding transition of the  $\alpha$  subunit (Matthews *et al.*, 1982; Yutani *et al.*, 1982). We conclude that it is not very likely that the exchange of redundant segments that would be required during the transition from one folding frame of the  $(\beta\alpha)_{10}$  variant of PRAI to the other can take place under native conditions.

Another question is whether the final structure of  $(\beta\alpha)_{10}$  variant is preferred thermodynamically or rather is trapped kinetically in a metastable conformation. In this regard it is interesting that the direction of protein biosynthesis does not bias the folding of the  $(\beta\alpha)_{10}$  variant. If its folding were more rapid than its synthesis, the first stable  $(\beta\alpha)_8$  core to be completed on the ribosome, namely folding frame E (Figure 2), would be trapped kinetically. That is, even if alternative, more stable folding frames were to become available as the polypeptide chain is elongated further, they would have no opportunity of folding. By contrast, if protein folding were slower than protein synthesis, the processive nature of polypeptide chain elongation would not affect the final structure.

Although the half-life of the slowest phase of the fluorescence change observed during refolding from GuCl solutions is  $\sim 500$  s (see Table I,  $t_{1/2} = \ln 2/k_{\text{obs}}$ ), it is premature to conclude from these data that folding *in vivo* is slower than synthesis (elongation rate  $\sim 15$  amino acids added/s). The slow refolding steps observed *in vitro* with the analogous  $\alpha$  subunit of tryptophan synthase are probably due to the isomerization of Xaa-Pro bonds having non-native geometry (Cristanti and Matthews, 1981). Although it is not known which of the 11 Xaa-Pro bonds of pPRAI are generated in the wrong configuration during synthesis, none of these residues are located in  $\beta$  strands or  $\alpha$  helices (M.Wilmanns, unpublished work). By contrast, aldolase from *Staphylococcus aureus*, which is probably another monomeric  $(\beta\alpha)$  barrel protein (Matsuura *et al.*, 1984; Buisson *et al.*, 1987), refolds within 10 s (Rudolph *et al.*, 1983). Thus it is possible that the  $(\beta\alpha)_8$  cores of both the  $(\beta\alpha)_8$  and the  $(\beta\alpha)_{10}$  variants fold rapidly and that the slow final phase of folding, which is characterized by  $k_{\text{obs}}$  (see Table I), is due to the isomerization of peripheral Xaa-Pro bonds. In summary, the circumstantial evidence favours the view that the folding of the  $(\beta\alpha)_{10}$  variant to yield exclusively the wild-type  $(\beta\alpha)_8$  core is controlled thermodynamically (Anfinsen and Scheraga, 1975).

Why is folding frame E (see Figure 2) preferred over the other possible ones? Perhaps the different circularly permuted  $(\beta\alpha)$  variants of PRAI are intrinsically less stable than the wild-type.



Indeed, the concentration of GuCl at the midpoint of the unfolding transition of cPRAI-1 is slightly less than that of the reference ( $\beta\alpha$ )<sub>8</sub> protein (Figure 7). To test this hypothesis we have constructed and tested a terminally repeated variant of PRAI that carries the C-terminal region comprising  $\alpha_6\beta_7\alpha_7\beta_8\alpha_8$  joined to the N-terminus (K. Luger, unpublished experiments). This ( $\beta\alpha$ )<sub>10.5</sub> variant has the unorthodox ( $\alpha\beta$ )<sub>8</sub> core of cPRAI-2 (see Figure 1C) as the first folding frame to be synthesized on the ribosome. Furthermore, cPRAI-2 and the reference wild-type protein have superimposable unfolding transitions with GuCl as the denaturant (Luger *et al.*, 1989). However, the folding frames of both the pristine and the refolded forms of the ( $\beta\alpha$ )<sub>10.5</sub> variant were again the same, namely that of the wild-type ( $\beta\alpha$ )<sub>8</sub> core (data not shown). We conclude that the mid-point of the unfolding transition due to GuCl is not a sufficient criterion to predict the preferred folding frame of terminally duplicated variants.

Another explanation for the preference of the wild-type core is that, fortuitously or not, its interaction with the N-terminal extension represents the structure of minimal free energy. In this regard it is interesting that both the  $\alpha$  subunit of tryptophan synthase (Hyde *et al.*, 1988) and indoleglycerol phosphate synthase (Priestle *et al.*, 1987) possess structured extensions at their N-termini.

The CD spectra are consistent with the idea that the redundant segments of polypeptide chain have approximately the same content of  $\alpha$  helix as the same segment in the corresponding compact ( $\beta\alpha$ )<sub>8</sub> domain. However, the data could also be explained by assuming a collapsed, partially folded and rapidly fluctuating state of the extension (Ptitsyn, 1987; Ohgushi and Wada, 1983; Shortle and Meeker, 1989; Baum *et al.*, 1989). A more direct determination of the structure of the ( $\beta\alpha$ )<sub>10</sub> variant is needed to answer the important question of how the redundant chain segment interacts with the ( $\beta\alpha$ )<sub>8</sub> barrel core. Complementary information is to be expected from the properties of the intermediate stages of the folding pathway, where the final folding frame is selected. Experiments based on these considerations are in progress.

Proteins with terminally repeated sequences may represent transient intermediates during their evolution. About 30% of the presently known proteins have adjacent N- and C-terminal regions (Thornton and Sibanda, 1983). Therefore, at least in principle, they may have been subject to the same kind of strategy as used in this work on an 8-fold  $\beta\alpha$  barrel protein. Tandem duplications and transpositions at the genetic level have been shown to occur spontaneously. At a later stage, superfluous extensions could have been trimmed to generate circularly permuted sequences (Cunningham *et al.*, 1979).

More importantly, the new extensions or insertions might have been crucial for modulating the intrinsic properties of the wild-type protein. One advantage of this hypothetical mechanism of protein evolution is that each of the several new folding frames would probably retain partial activity. But, besides being more stable, the preferred configuration of ( $\beta\alpha$ )<sub>8</sub> core and terminal extensions could fortuitously be more active, or have some other new property, thus possibly affecting the generation time of the organism in a crucial manner.

### Acknowledgements

We thank M. Herold and U. Hommel for help in some of the experiments, M. Wilmanns and J. N. Jansonius for the refined secondary structure assignments of PRAI from yeast, and P. Stanssens and D. Stueber for providing procedures, plasmids and bacterial strains. We are grateful to J. Hofsteenge for performing protein sequence analyses, to A. Lustig for doing the hydrodynamic measurements and to H. Voser (Ciba-Geigy) for donating the hydroxylapatite. Critical comments

on the manuscript by E. W. Miles, J. Priestle and C. Yanofsky are gratefully acknowledged. This work was supported by the Swiss National Science Foundation, grant no. 3.255-0.85.

### References

- Alber, T. (1989) In Fasman, G. D. (ed.), *Prediction of Protein Structure and Principles of Protein Conformation*. Plenum Press, New York, pp. 161–192.
- Anfinsen, C. B. and Scheraga, H. A. (1975) *Adv. Prot. Chem.*, **29**, 205–300.
- Atkinson, A., Bradford, P. A. and Selmes, I. P. (1973) *J. Appl. Chem. Biotechnol.*, **23**, 517–529.
- Baum, J., Dobson, C. M., Evans, P. A. and Hanley, C. (1989) *Biochemistry*, **28**, 7–13.
- Bimboim, H. C. and Doly, J. (1979) *Nucl. Acids Res.*, **7**, 1513–1523.
- Bradford, M. (1976) *Anal. Biochem.*, **72**, 248–254.
- Buisson, G., Duée, E., Haser, R. and Payan, F. (1987) *EMBO J.*, **6**, 3909–3916.
- Cantor, C. R. and Schimmel, P. R. (1980) *Biophysical Chemistry II. Spectroscopic Analysis of Biopolymers*. W. H. Freeman, San Francisco, CA.
- Casabadian, M. J. (1976) *J. Mol. Biol.*, **104**, 541–555.
- Castagnoli, L., Scarpa, M., Kokkinidis, M., Banner, D. W., Tsernoglou, D. and Cesareni, G. (1989) *EMBO J.*, **8**, 621–629.
- Certa, U., Bannwarth, W., Stüber, D., Gentz, R., Lanzer, M., LeGrice, S., Guillot, F., Wendler, I., Hunsmann, G., Bujard, H. and Mous, J. (1986) *EMBO J.*, **5**, 3051–3056.
- Chothia, C. (1988) *Nature*, **333**, 598–599.
- Coghlan, V. M. and Vickery, L. E. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 835–839.
- Crisanti, M. M. and Matthews, C. R. (1981) *Biochemistry*, **20**, 2700–2706.
- Cunningham, B. A., Hemperly, J. J., Hopp, T. P. and Edelman, G. M. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 3218–3222.
- Goldenberg, D. P. (1988) *Annu. Rev. Biophys. Biophys. Chem.*, **17**, 481–507.
- Goldenberg, D. P. (1989) *Prot. Engng*, **2**, 493–495.
- Goldenberg, D. P. and Creighton, T. E. (1983) *J. Mol. Biol.*, **165**, 407–413.
- Högberg-Raubaud, A. and Goldberg, M. (1977) *Biochemistry*, **16**, 4014–4020.
- Hommel, U., Lustig, A. and Kirschner, K. (1989) *Eur. J. Biochem.*, **180**, 33–40.
- Humphreys, G. O., Willshaw, G. A. and Anderson, E. S. (1975) *BBA*, **383**, 457–463.
- Husten, E. J., Esmon, C. T. and Johnson, A. E. (1987) *J. Biol. Chem.*, **262**, 12953–12961.
- Hyde, C. C., Ahmed, S. A., Padlan, E. A., Miles, E. W. and Davies, D. R. (1988) *J. Biol. Chem.*, **263**, 17857–17871.
- Hynes, T. R., Kautz, R. A., Goodman, M. A., Gill, J. F. and Fox, R. O. (1989) *Nature*, **339**, 73–76.
- Jackson, D. A. and Yanofsky, C. (1969a) *J. Biol. Chem.*, **244**, 4526–4538.
- Jackson, D. A. and Yanofsky, C. (1969b) *J. Biol. Chem.*, **244**, 4539–4546.
- Jaenicke, R. (1987) *Prog. Biophys. Mol. Biol.*, **49**, 117–237.
- Lämmli, U. K. (1970) *Nature*, **227**, 681–685.
- Lasters, J., Wodak, S. J., Alard, P. and van Cutsem, E. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 3338–3342.
- Lesk, A. M., Brändén, C.-I. and Chothia, C. (1989) *Proteins*, **5**, 139–148.
- Levitt, M. and Chothia, C. (1976) *Nature*, **261**, 552–558.
- Lim, W. A. and Sauer, R. T. (1989) *Nature*, **339**, 31–36.
- Luger, K., Hommel, U., Herold, M., Hofsteenge, J. and Kirschner, K. (1989) *Science*, **243**, 206–210.
- Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Matsuura, Y., Kusunoki, M., Harada, W. and Kakudo, M. (1984) *J. Biochem. (Tokyo)*, **95**, 697–702.
- Matthews, C. R. and Crisanti, M. M. (1981) *Biochemistry*, **20**, 784–792.
- Matthews, C. R., Crisanti, M. M., Gepner, G. L., Velicelebi, G. and Sturtevant, J. M. (1982) *Biochemistry*, **19**, 1290–1293.
- McLaughlin, L. W. and Krusche, J. U. (1982) In Lange, A. and Gassen, H. G. (eds), *Chemical and Enzymatic Synthesis of Gene Fragments*. Verlag Chemie, Weinheim, pp. 177–198.
- Miozzari, G. F., Niederberger, P. and Hütter, R. (1978) *Anal. Biochem.*, **90**, 5220–5233.
- Nagai, K. and Thogersen, H. C. (1987) *Methods Enzymol.*, **153**, 461–481.
- Ohgushi, M. and Wada, A. (1983) *FEBS Lett.*, **164**, 21–24.
- Pakula, A. A. and Sauer, R. T. (1989) *Proteins*, **5**, 202–210.
- Priestle, J. P., Grütter, M. G., White, J. L., Vincent, M. G., Kania, M., Wilson, E., Jardtzyk, T. S., Kirschner, K. and Jansonius, J. N. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 5690–5694.
- Ptitsyn, O. (1987) *J. Prot. Chem.*, **6**, 273–293.
- Riechmann, L., Clark, M., Waldmann, H. and Winter, G. (1988) *Nature*, **332**, 323–327.
- Rudolph, R., Bohrer, M. and Fischer, S. (1983) *Eur. J. Biochem.*, **131**, 383–386.

- Sanger, F., Nicklen, S. and Coulson, A. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
- Schägger, H. and Jagow, G. (1987) *Anal. Biochem.*, **166**, 368–379.
- Shortle, D. and Meeker, A.K. (1989) *Biochemistry*, **28**, 936–944.
- Stanssens, P., Opsomer, C., McKeown, Y.M., Kramer, W., Zabeau, M. and Fritz, H.J. (1989) *Nucl. Acids Res.*, **17**, 4441–4454.
- Sternberg, M.J.E. and Thornton, J.M. (1977) *J. Mol. Biol.*, **110**, 269–283.
- Studier, W.F. (1973) *J. Mol. Biol.*, **79**, 237–248.
- Teipel, W.J. and Koshland, D.E. (1971) *Biochemistry*, **10**, 792–798.
- Thornton, J.M. and Sibanda, B.L. (1983) *J. Mol. Biol.*, **167**, 443–460.
- Tschümper, G. and Carbon, J. (1980) *Gene*, **10**, 157–166.
- Wharton, R.P., Brown, E.L. and Ptashne, M. (1984) *Cell*, **38**, 361–369.
- Yu, M.-H. and King, J. (1988) *J. Biol. Chem.*, **263**, 1412–1431.
- Yutani, K., Ogasahara, K., Sugino, Y. and Matsushiro, Y. (1977) *Nature*, **267**, 274–275.
- Yutani, K., Ogasahara, K., Kimura, A. and Sugino, Y. (1982) *J. Mol. Biol.*, **160**, 387–390.

Received on November 17, 1989; accepted on January 15, 1990