

Role of a subdomain in the folding of bovine pancreatic trypsin inhibitor

Jonathan P. Staley*† & Peter S. Kim†‡

Departments of * Chemistry and ‡ Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

† Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, Massachusetts 02142, USA

THE disulphide-bonded intermediates that accumulate in the oxidative folding of bovine pancreatic trypsin inhibitor (BPTI) were characterized some time ago¹. Structural characterization of these intermediates would provide an explanation of the kinetically preferred pathways of folding for BPTI. When folding occurs under strongly oxidizing conditions, more than half the molecules become trapped in an intermediate, designated N*, which is similar to the native protein but lacks the 30–51 disulphide bond^{2–4}. We have tested the hypothesis⁵ that the precursor to N* is the one-disulphide intermediate [5–55], which contains the most stable disulphide in BPTI, and present evidence here that this is the case. A peptide model of [5–55], corresponding to a subdomain of BPTI, seems to fold into a native-like conformation, explaining why [5–55] does not lead to native protein and why it folds rapidly⁵ to N*. A native-like subdomain structure in a peptide model⁶ of [30–51], the other crucial one-disulphide intermediate, may explain the route by which [30–51] folds to native protein. Thus, much of the folding pathway⁵ of BPTI can be explained by the formation of a native-like subdomain in these two early intermediates. This suggests that a large part of the protein folding problem can be reduced to identifying and understanding subdomains of native proteins.

As part of an investigation of the BPTI folding pathway (Fig. 1a), we have designed a peptide model of [5–55] to determine whether it folds into a stable conformation in aqueous solution. Following a strategy used previously⁶ to design a peptide model of the intermediate [30–51], we designed our model, called P α P γ , by assuming that the structure in [5–55] would be native-like and localized to a region near the 5–55 disulphide.

In native BPTI, the 5–55 disulphide joins the 3₁₀-helix to the α -helix, both of which contact the central β -sheet. P α P γ includes residues from these three secondary structural units (Fig. 1b) and consists of two peptides, P α and P γ , joined by a disulphide bond corresponding to the 5–55 disulphide. P α includes residues corresponding to the α -helix and a short β -strand. The residues in P γ correspond to those in the 3₁₀-helix and the central β -sheet, which are joined by a linker of six glycines. The formation of the 30–51 disulphide was prevented by replacing Cys 30 and Cys 51 with alanines. To evaluate the importance of the central β -sheet region, we made a second peptide model, called P α P γ ($\Delta\beta$), which lacks the β -sheet and glycine linker sequences (Fig. 1c).

P α P γ folds in aqueous solution, as shown by the sigmoidal transition observed for the temperature dependence of the circular dichroism (CD) signal at 222 nm (Fig. 2a). The transition indicates that P α P γ is essentially 100% folded at 0 °C and that the melting temperature is \sim 28 °C. A sigmoidal transition is not observed for reduced P α P γ or for oxidized P α P γ ($\Delta\beta$), indicating that both the disulphide (Fig. 2a) and the β -sheet sequence (Fig. 2b) are required for structure in P α P γ .

The nuclear magnetic resonance (NMR) signals for protons in P α P γ are dispersed at low temperatures (Fig. 3a), providing further evidence that P α P γ is folded. The pattern of dispersion is similar (Fig. 3b) to that observed for a peptide model of [30–51], called P α P β , which contains many of the residues in P α P γ . As two-dimensional NMR showed that the structure of P α P β was native-like (ref. 6; and T. G. Oas and P.S.K., unpublished results) and as a given pattern of dispersion is charac-

teristic of a particular structure⁷, we came to the preliminary conclusion that the structure of P α P γ is also native-like. Two-dimensional NMR studies of P α P γ support this conclusion (unpublished results).

Four aromatic residues in native BPTI have reduced ring-flipping rates at low temperatures, as determined by NMR^{8,9}. Three of these residues are contained in P α P γ . At least one of these aromatic rings, probably Phe 45, shows reduced flipping rates at low temperatures in P α P γ (data not shown). Thus, the native-like structure in P α P γ probably includes close-packed tertiary structure.

Given our demonstration that P α P γ folds into a stable conformation, we concluded that [5–55] is the precursor to N* (Fig. 4). Several observations support this conclusion. Firstly, [5–55], produced by folding BPTI in denaturing conditions, interconverts relatively slowly with the other one-disulphide intermediates under native conditions¹⁰, suggesting that [5–55] does fold. Secondly, [5–55] accumulates during the folding of circular BPTI, in which the two termini of BPTI are joined by a peptide bond¹¹. Thirdly, a mutant of BPTI, with Cys 30 and Cys 51 replaced by alanines, folds readily¹² to the equivalent of N*, indicating that Cys 30 and Cys 51 are not required for the formation of N*. Fourthly, peptide models of two other possible precursors to N*, [14–38] (ref. 13) and [5–51] (unpublished results), are unfolded in aqueous solution at 0 °C. Finally, of the three disulphides in native BPTI, removal of the 5–55 disulphide results in the greatest loss of stability⁵.

The native-like structure of P α P γ provides a structural explanation for why [5–55] rapidly forms a second disulphide⁵ (Fig. 1a). The 14–38 disulphide lies at one end of the antiparallel β -sheet in native BPTI (Fig. 1b). Formation of the β -sheet would bring Cys 14 and Cys 38 close together and enhance disulphide formation. Moreover, linking the termini of the protein via the 5–55 disulphide and packing this region of the molecule against the β -sheet in a native-like way would constrain the loop regions containing Cys 14 and Cys 38 to be close together (Fig. 1b) and further enhance the rate of formation of the 14–38 disulphide.

Thus, whereas [30–51] is significant in the folding of BPTI because it is well populated^{10,14}, that is, thermodynamically stable, [5–55] is significant because in kinetic terms it favours the next step in folding⁵. A native-like structure does, however, appear to be responsible for both the thermodynamic stability of [30–51]⁶ and the rapid folding of [5–55] to N*.

The native-like structure in P α P γ also explains why [5–55] does not readily form the 30–51 disulphide to yield the intermediate [30–51, 5–55] (Fig. 4), as Cys 30 and Cys 51 would be buried. Similarly, these residues remain buried^{3,4} in N*, explaining why N* does not lead to native protein^{3,15}. Thus, the native-like structure in [5–55] seems to hinder the simple sequential formation of native disulphides in the branch of the BPTI folding pathway⁵ that leads to a dead-end intermediate. In the folding of proteins without disulphides, the burying of water molecules or ions may also hinder simple sequential acquisition of native structure; structural rearrangements may be required to release trapped water molecules or ions before folding to native protein can finish.

NMR studies of a peptide model for [30–51] indicate that the native-like structure in [30–51] leads to partial burying⁶ of Cys 55. This may explain why 5–55 does not form readily in [30–51] (Fig. 4) and why the branch of the BPTI folding pathway that leads to the native protein includes disulphide rearrangements through non-native intermediates (T. G. Oas and P.S.K., unpublished results; but see refs 16, 17).

It seems, therefore, that much of the apparently complex pathway for folding of BPTI can be explained by the formation of a native-like subdomain in the two early intermediates, [30–51] and [5–55]. We conclude that native-like subdomains are key determinants in the folding of BPTI. If this conclusion turns out to be a general one, then understanding the hierarchy of subdomains in native proteins^{18,19} will be an important part of understanding protein folding. □

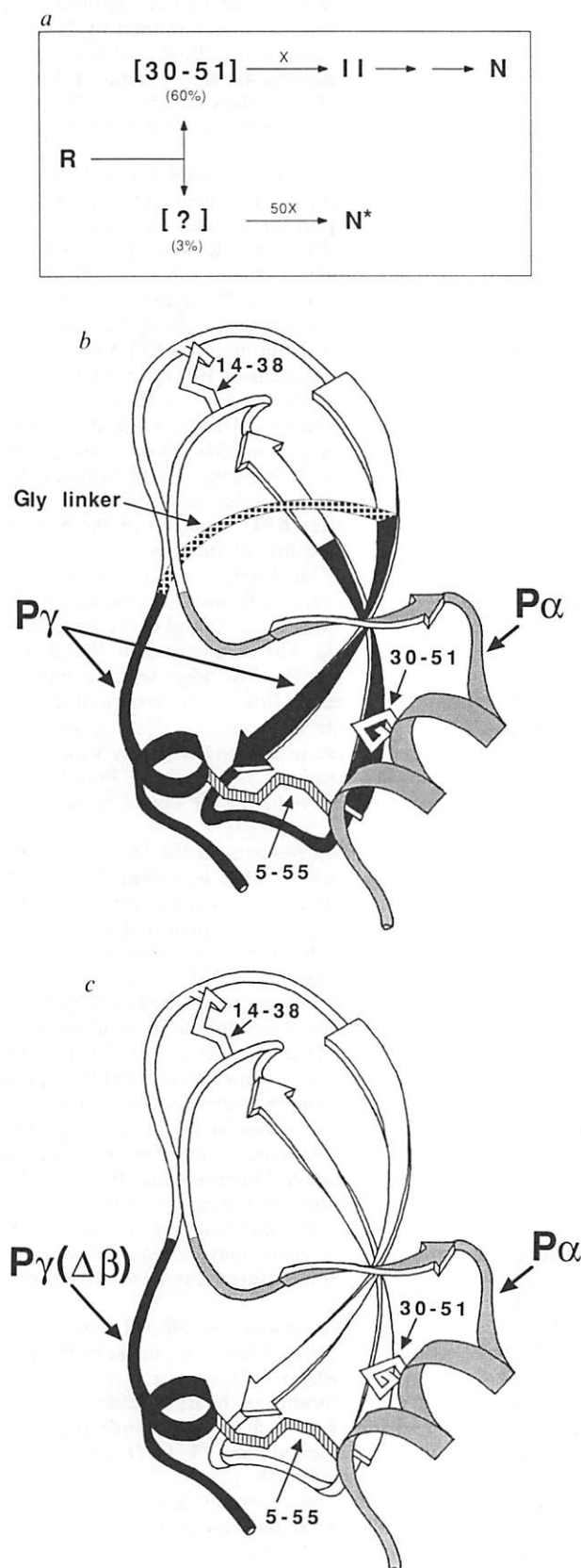


FIG. 1 Peptide models of the precursor to N^* . **a**, A simplified schematic of the folding pathway⁵ of BPTI at 25 °C, pH 8.7. Formation of both native BPTI and N^* results from a fork in the folding pathway at the one-disulphide stage of folding⁵. R refers to reduced BPTI; [30-51], to the one-disulphide intermediate containing a disulphide between Cys 30 and Cys 51; II, to the two-disulphide intermediates on the pathway to native BPTI; N, to native BPTI; N^* , to the native-like kinetically trapped intermediate; and [?], to the unidentified one-disulphide precursor to N^* . N^* is stable for days at 25 °C in the absence of reducing agents. Apart from reduction of the 30-51 disulphide, there are no major structural differences between N^* and native BPTI as judged by NMR³. Moreover, the X-ray crystal structure of a mutant of BPTI, with Cys 30 and Cys 51 replaced with alanines, reveals that the mutant has essentially a native fold⁴. The relative proportions of the one-disulphide intermediates are indicated⁵. As shown, [?] forms a second disulphide 50 times faster than [30-51]⁵, accounting for the significant accumulation of N^* from a poorly populated intermediate. **b**, Regions of BPTI included in $P\alpha P\gamma$ are indicated on a representation²⁰ of folded BPTI. The three native disulphides of BPTI are labelled. $P\alpha$, (shaded) includes residues 42-58: RNNFKSAEDAMRTCGGA (one-letter amino acid code); Cys 51 has been substituted with Ala. (Note: $P\alpha$ of $P\alpha P\gamma$ differs slightly from $P\alpha$ of $P\alpha P\beta$ in that R42 is not included⁵ in $P\alpha P\beta$.) $P\gamma$ (black) includes residues 1-9, a linker of six glycines and residues 20-33: RPDFCLEPPGGGGGGRYFY-NAKAGLAQTG; Cys 30 has been substituted with Ala. The glycine linker (spotted) was used to join covalently residues 1-9 to 20-33. The disulphide between $P\alpha$ and $P\gamma$ corresponding to the 5-55 disulphide is striped. **c**, Regions of BPTI included in $P\alpha P\gamma(\Delta\beta)$ are indicated on a representation²⁰ of folded BPTI. $P\alpha$ is shaded. $P\gamma(\Delta\beta)$ (black) corresponds to residues 1-9: RPDFCLEPP, lacking the β -sheet and glycine linker sequences. **METHODS.** Regions of BPTI important for stabilizing the 5-55 disulphide were defined by identifying residues involved in van der Waals' contacts and hydrogen bonds near the 5-55 disulphide in the structure²¹⁻²³ of native BPTI. Peptides containing these residues were synthesized on an Applied Biosystems model 430A peptide synthesizer using 4[oxymethyl]phenylacetamidomethyl resins and standard reaction cycles, modified to include acetic anhydride capping. All peptides were cleaved from the resins using HF (Immunodynamics Inc., San Diego) except for $P\gamma(\Delta\beta)$ which was cleaved using trifluoromethane sulphonic acid. Peptides were desalted on a Sephadex G-10 column in 5% acetic acid and purified by high-pressure liquid chromatography on a Vydac C18 column using a linear water/acetonitrile gradient containing 0.1% trifluoroacetic acid. The identity of each peptide was confirmed by fast-atom bombardment mass spectrometry (at the NIH Mass Spectrometry Facility at MIT), yielding relative molecular masses of 1,826.4 for $P\alpha$ (calculated, 1,828.0), 3,046.3 for $P\gamma$ (calculated, 3,047.4) and 1,072.6 for $P\gamma(\Delta\beta)$ (calculated, 1,073.2). The heterodimeric disulphides were formed by activating one thiol with Ellman's reagent²⁴ dithio-bis-(2-nitrobenzoic) acid (DTNB). One volume of $P\alpha$ (1 mM in 1 mM HCl) was added dropwise to nine volumes of 25 mM DTNB, 0.1 M phosphate, 1 mM EDTA at pH 7.3, 23 °C. After 5 min, an equal volume of diethyl ether was added and the pH of the aqueous layer dropped to pH 2.0 by addition of 1 M HCl. The aqueous layer was washed seven times with ether to extract the DTNB. (Saturating freshly opened ether with water prevents oxidation of Met 52; T. G. Oas and J. S. Fetrow, unpublished results.) One equivalent of $P\gamma$ or $P\gamma(\Delta\beta)$ in an equal volume of 0.5% acetic acid, 1 mM EDTA was then added to the aqueous layer. The reaction was initiated by titrating to pH 7.3 with 0.75 M Na_2HPO_4 and quenched after 15 min by dropping the pH to pH 2.0 with 1 M HCl. The peptides were purified as described above. The identity of each was confirmed by mass spectrometry yielding relative molecular masses of 4,872.7 for $P\alpha P\gamma$ (calculated, 4,873.4) and 2,900.0 for $P\alpha P\gamma(\Delta\beta)$ (calculated, 2,899.2). Overall yield was ~45%. Glycine was used for the linker in $P\alpha P\gamma$ for maximum flexibility and minimum perturbation. The length of the linker was based on computer modelling. Linkers of different lengths were energy-minimized using the steepest descents algorithm of Discover (version 2.4, Biosym Technologies Inc., San Diego) in the presence of residues 1-9, 20-33 and 42-58 of BPTI. The coordinates of these residues were fixed in their native conformation²¹. The linker was required to make peptide bonds with residues 9 and 20. After minimization, linkers were evaluated by inspection of their backbone bond lengths using Insight (version 2.4, Biosym Technologies Inc.). Five glycines were sufficient to bridge residues 9 and 20 without perturbing native structure. Six glycines were used for additional flexibility.

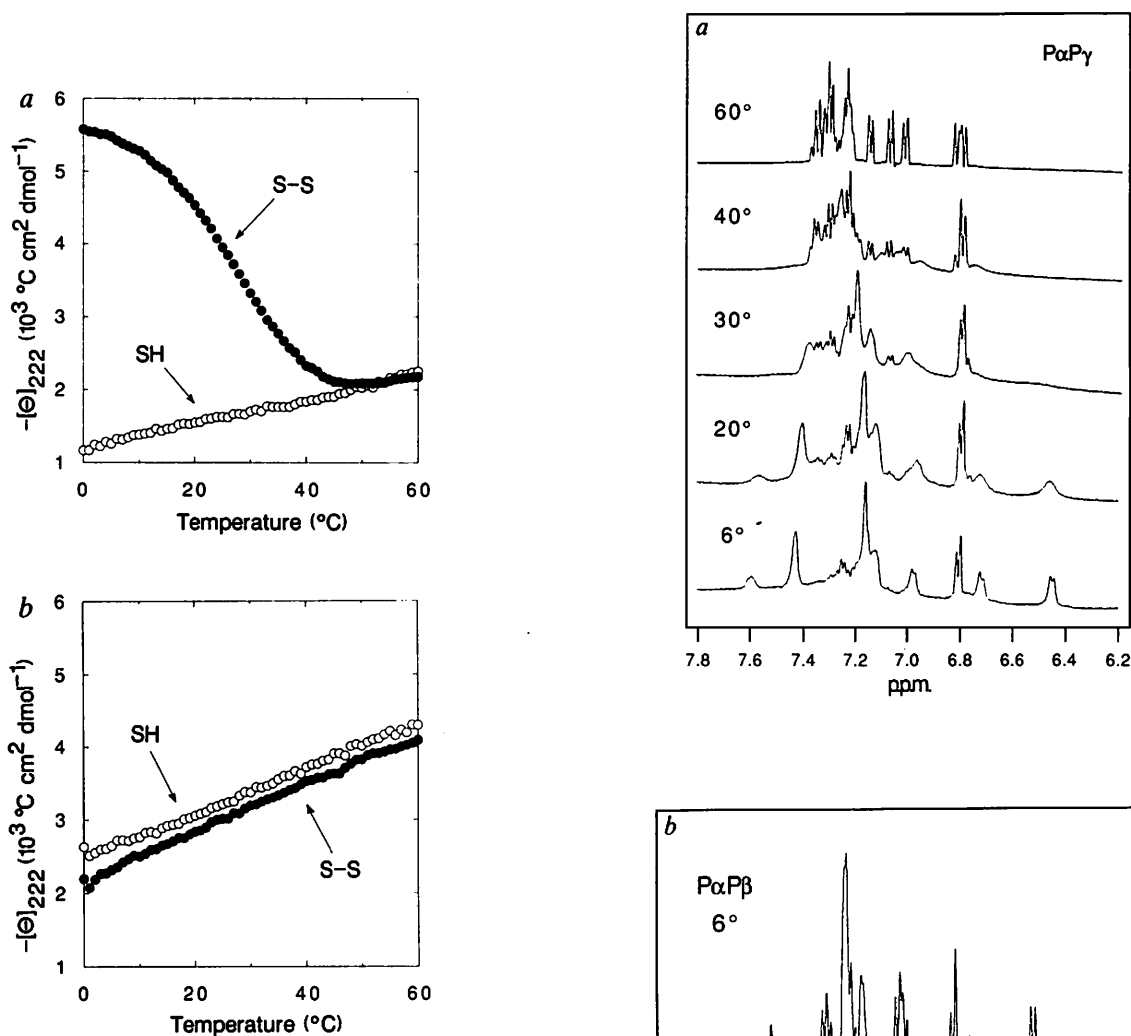


FIG. 2 $P\alpha P\gamma$ folds and requires the disulphide bond and the central β -sheet sequence for folding. *a*, Temperature dependence of $[\theta]_{222}$ for the oxidized (solid circles) and reduced (open circles) forms of $P\alpha P\gamma$. The melting temperature of oxidized $P\alpha P\gamma$ is $28 \pm 1^\circ\text{C}$, as determined by the maximum of the first derivative of the temperature dependence. Gel filtration studies indicate that $P\alpha P\gamma$ is a monomer under the conditions of these experiments. *b*, Temperature dependence of $[\theta]_{222}$ for the oxidized and reduced forms of $P\alpha P\gamma(\Delta\beta)$.

METHODS. Measurements were made with an Aviv Model 60DS circular dichroism spectrometer, equipped with an HP Model 89100A Peltier temperature control unit. A 10 mm pathlength cell was used, and all samples were degassed by vacuum. The concentration of peptide in all the experiments was $\sim 18 \mu\text{M}$. The concentrations of peptide stock solutions were determined by tyrosine and cystine absorbance²⁵ for $P\alpha P\gamma$ and by a ninhydrin assay²⁶ for $P\alpha P\gamma(\Delta\beta)$ (the ninhydrin colour yield of proline was determined as 7.3% at 570 nm). The sample buffer contained 40 mM sodium phosphate, 40 mM NaCl at pH 7.0. The reduced samples were generated in sample buffer, that also contained 0.2 mM dithiothreitol and 1 mM EDTA, by raising the pH temporarily to pH 8.7. Reduced samples were maintained under argon. Reverse-phase high-pressure liquid chromatography confirmed that each peptide was reduced before and after the circular dichroism experiments. The relative molecular mass (M_r) of $P\alpha P\gamma$ in sample buffer at 4°C was determined using a Sephadex G-25(f) column with the following standards: BPTI (M_r , 6,513), oxidized insulin B chain (M_r , 3,496), bombesin (M_r , 1,620), vitamin B_{12} (M_r , 1,355), human angiotensin I (M_r , 1,297) and oxytocin (M_r , 1,007). $P\alpha P\gamma$ was loaded at $20 \mu\text{M}$ and eluted at a peak concentration of $1 \mu\text{M}$. The apparent M_r of $P\alpha P\gamma$ was 4,900 (calculated 4,873) indicating that $P\alpha P\gamma$ is monomeric in these conditions. The apparent M_r was unchanged when $P\alpha P\gamma$ was loaded at 1 mM , eluting at a peak concentration of $50 \mu\text{M}$.

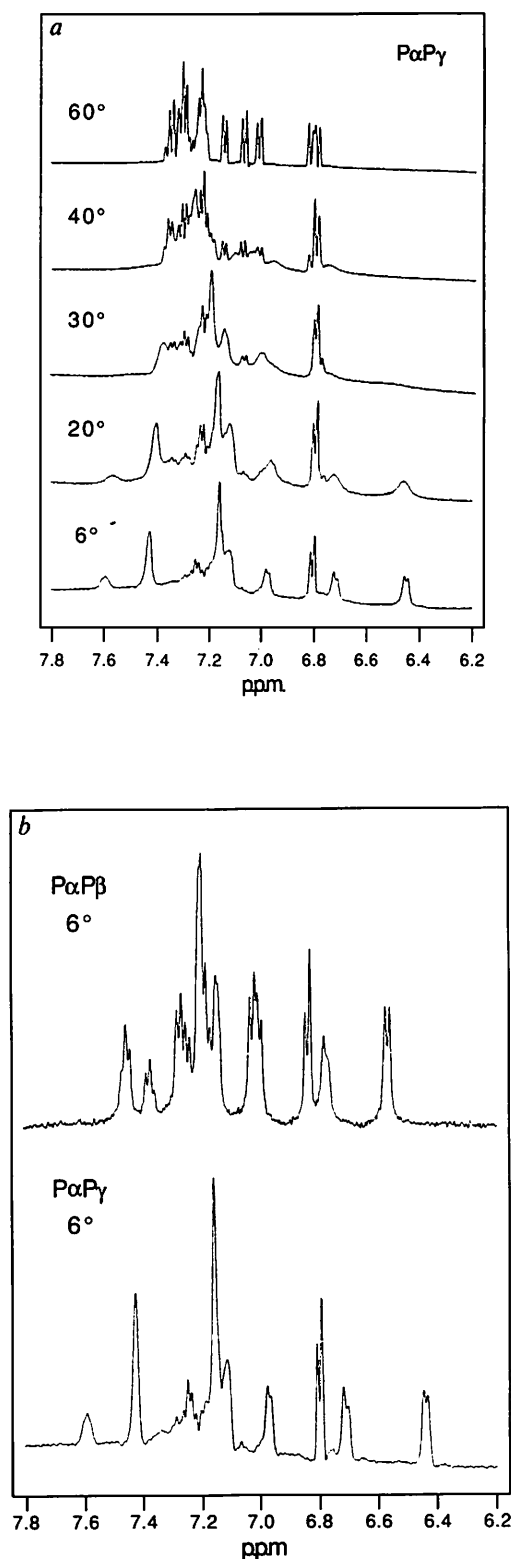


FIG. 3 Chemical shift dispersion in the aromatic region of the one-dimensional NMR spectrum of $P\alpha P\gamma$ depends on temperature and is similar to that of $P\alpha P\beta$. *a*, Temperature dependence of the aromatic region of $P\alpha P\gamma$. *b*, Comparison between the aromatic region of $P\alpha P\gamma$ and $P\alpha P\beta$ (ref. 6) at low temperatures.

METHODS. The spectra of $P\alpha P\gamma$ in 40 mM sodium phosphate, 40 mM NaCl at pH 7.0 in D_2O were collected, at a peptide concentration of 0.8 mM as determined by tyrosine and cystine absorbance²⁵. Trimethylsilylpropionic acid was used as a standard and its chemical shift at pH 7.0 was taken to be -0.018 p.p.m. (ref. 27). Data were collected on a 500 MHz spectrometer at the MIT Francis Bitter National Magnet Laboratory, presaturating residual HOD for 2 s. A shifted sine-bell apodization was used to enhance resolution.

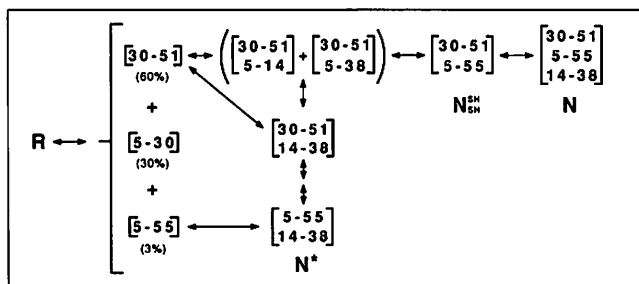


FIG. 4 Schematic diagram of the folding pathway⁵ of BPTI at 25 °C, pH 8.7. R, N and N* are as in Fig. 1. N^{SH} refers to the immediate precursor to native BPTI. Intermediates are designated by the disulphide bonds that they contain. The one-disulphide intermediates grouped together, are in rapid equilibrium on the minute timescale²⁸. It is interesting that [5-30], which contains a non-native disulphide, can serve as an intermediate in the rearrangement between [30-51] and [5-55]. The proportions of the one-disulphide intermediates are indicated⁵. The remaining 7% includes [30-55], [5-51] and other unidentified one-disulphide intermediates. The intermediates [30-51, 5-14] and [30-51, 5-38] are grouped together because both form from the one-disulphides, both form from [30-51, 14-38] and both rearrange to [30-51, 5-55] (ref. 16). Of the fifteen possible one-disulphide intermediates only [30-51], [5-30], [5-51] and [30-55] are observed in folding²⁸, but formation of N* from these intermediates would require both disulphide bond formation and rearrangement, involving intermediates that have not been observed. N* can form by rearrangement of a two-disulphide intermediate but this pathway accounts for only ~10% of N* formed⁵. Molecules that fold to N* are stable and kinetically trapped in strongly oxidizing conditions.

Received 10 November 1989; accepted 5 February 1990.

1. Creighton, T. E. *Prog. Biophys. molec. Biol.* **33**, 231-297 (1978).
2. States, D. J., Dobson, C. M., Karplus, M. & Creighton, T. E. *Nature* **288**, 630-632 (1980).
3. States, D. J., Dobson, C. M., Karplus, M. & Creighton, T. E. *J. molec. Biol.* **174**, 411-418 (1984).
4. Eigenbrot, C., Randal, M. & Kossiakoff, A. *Protein Engineering* (submitted).
5. Creighton, T. E. & Goldenberg, D. P. *J. molec. Biol.* **179**, 497-526 (1984).
6. Oas, T. G. & Kim, P. S. *Nature* **338**, 42-48 (1988).
7. Pardi, A., Wagner, G. & Wüthrich, K. *Eur. J. Biochem.* **137**, 445-454 (1983).
8. Wagner, G., DeMarco, A. & Wüthrich, K. *Biophys. struct. Mechanisms* **2**, 139-158 (1976).
9. Wagner, G., Bruhwiler, D. & Wüthrich, K. *J. molec. Biol.* **196**, 227-231 (1987).
10. Creighton, T. E. *J. molec. Biol.* **113**, 313-328 (1977).
11. Goldenberg, D. P. & Creighton, T. E. *J. molec. Biol.* **179**, 527-545 (1984).
12. Marks, C. B., Naderi, H., Kosen, P. A., Kuntz, I. D. & Anderson, S. in *Protein Structure, Folding and Design* Vol. 2 (ed. Oxender, D. L.) 335-340 (Liss, New York, 1987).
13. Goodman, E. M. & Kim, P. S. in *Current Research in Protein Chemistry* (ed. Villafranca, J. J.) (Academic, New York, in the press).
14. Creighton, T. E. *J. molec. Biol.* **87**, 579-602 (1974).
15. States, D. J., Creighton, T. E., Dobson, C. M. & Karplus, M. *J. molec. Biol.* **195**, 731-739 (1987).
16. Creighton, T. E. *J. molec. Biol.* **113**, 275-293 (1977).
17. Creighton, T. E. *J. molec. Biol.* **151**, 211-213 (1981).
18. Lesk, A. M. & Rose, G. D. *Biochemistry* **78**, 4304-4308 (1981).
19. Richardson, J. S. *Adv. Protein Chem.* **34**, 167-339 (1981).
20. Richardson, J. S. *Meth. Enzym.* **115**, 359-380 (1985).
21. Delsenhofer, J. & Steigemann, W. *Acta crystallogr.* **B31**, 238-250 (1975).
22. Wodawer, A., Walter, J., Huber, R. & Sjölin, L. *J. molec. Biol.* **180**, 301-329 (1984).
23. Wodawer, A., Nacham, J., Gilliland, G. L., Gallagher, W. & Woodward, C. *J. molec. Biol.* **198**, 469-480 (1987).
24. Eliman, G. *Arch. Biochem. Biophys.* **82**, 70-77 (1959).
25. Edelhoch, H. *Biochemistry* **6**, 1948-1954 (1967).
26. Rosen, H. *Arch. Biochem. Biophys.* **67**, 10-15 (1957).
27. DeMarco, A. *J. magn. Reson.* **26**, 527-528 (1977).
28. Creighton, T. E. *J. molec. Biol.* **87**, 603-624 (1974).

ACKNOWLEDGEMENTS. We thank R. Rutkowski for assistance with synthesis and purification; J. Fetrow for help with computer graphics; T. Oas for help in identifying van der Waals' contacts and hydrogen bonds in BPTI; W. DeGrado for suggesting the use of DTNB in the synthesis of peptide pairs and R. Baldwin, T. Creighton, D. Goldenberg, E. Goodman, T. Oas and E. O'Shea for discussion. J. P. S. is a Howard Hughes Medical Institute predoctoral fellow. This research was supported by grants from the NIH and the Lucille P. Markey Charitable Trust.