8. Garoff, H. & Simons, K. *Proc. natn. Acad. Sci. U.S.A.* **71**, 3988–3992 (1974).
9. Helenius, A. & Kartenbeck, J. *Eur. J. Biochem.* **106**, 613–618 (1980).
10. Fuller, S. D. *Cell* **48**, 923–934 (1987).
11. Gaulton, G. N. & Greene, M. L. *A. Rev. Immun.* **4**, 253–280 (1986).
12. Schreier, M. H. & Lefkovits, I. *Immunology* **36**, 743–752 (1979).
13. Reading, C. L. *J. Immun. Meth.* **53**, 261–291 (1982).
14. Cutler, D. F., Melancon, P. & Garoff, H. *J. cell. Biol.* **102**, 902–910 (1986).
15. Scheele, C. H. & Pfefferkorn, E. R. *J. Virol.* **3**, 369–375 (1969).
16. Rice, C. M. *et al. Science* **229**, 726–733 (1985).
17. Burge, B. W. & Pfefferkorn, E. R. *Nature* **210**, 1937–1940 (1966).
18. Lagwinska, E., Stewart, C. C., Adles, C. & Schlesinger, S. *Virology* **65**, 204–214 (1975).
19. Zavadova, Z., Zavada J. & Weiss, R. *J. gen. Virol.* **37**, 557–567 (1977).
20. Perez, L. G., Davis, G. L. & Hunter, E. *J. Virol.* **61**, 2981–2988 (1987).
21. Sege, K. & Peterson, P. A. *Nature* **271**, 167–169 (1978).
22. Sung Co, M., Gaulton, G. N., Fields, B. N. & Greene, M. I. *Proc. natn. Acad. Sci. U.S.A.* **82**, 1494–1498 (1985).
23. Sung Co, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 5315–5318 (1985).
24. Erlanger, B. F. *et al. Immun. Rev.* **94**, 23–37 (1986).
25. Elias D., Cohen, I. R., Schechter, Y., Spirer, Z. & Golander, A. *Diabetes* **36**, 348–354 (1987).
26. Farid, N. R. & Lo, T. C. Y. *Endocr. Rev.* **6**, 1–23 (1985).
27. Ardman, B., Khiroya, R. H. & Schwartz, R. S. *J. exp. Med.* **161**, 669–678 (1985).
28. Knigge, K. M., Piekut, D. T. & Berlove, D. J. *Cell Tissue Res.* **246**, 509–513 (1986).
29. Pain, D., Kanwar, Y. S. & Blobel, G. *Nature* **331**, 232–237 (1988).
30. Stuart, S. G. *et al. J. exp. Med.* **166**, 1668–1684 (1987).
31. Borrebaeck, C. A. K. & Moller, S. A. *J. Immun.* **136**, 3710–3715 (1986).
32. Shulman, M., Wilde, C. & Kohler, G. *Nature* **276**, 269–270 (1978).
33. Mellman, I., Steinman, R. M., Unkeless, J. C. & Cohn, Z. A. *J. cell. Biol.* **86**, 712–722 (1980).

A peptide model of a protein folding intermediate

Terrence G. Oas & Peter S. Kim

Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, Massachusetts 02142, USA
Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

It is difficult to determine the structures of protein folding intermediates because folding is a highly cooperative process. A disulphide-bonded peptide pair, designed to mimic the first crucial intermediate in the folding of bovine pancreatic trypsin inhibitor, contains secondary and tertiary structure similar to that found in the native protein. Peptide models like this circumvent the problem of cooperativity and permit characterization of structures of folding intermediates.

ONE of the basic goals of protein folding studies is to determine the structures of folding intermediates, ultimately at atomic resolution, and to understand the thermodynamic basis of their formation. It is difficult to obtain folding intermediates of single-domain proteins at equilibrium because the cooperativity of folding is high. Folding intermediates can be populated kinetically but they are short-lived, so detailed structural analyses (using nuclear magnetic resonance spectroscopy or X-ray crystallography for example) are difficult. Amide proton labelling methods have provided insights into the structures of kinetic folding intermediates[1,2], but a high-resolution structure has not yet been determined for any folding intermediate of a single-domain protein.

The most well characterized protein folding pathway is described in terms of disulphide-bond formation in the oxidative refolding of bovine pancreatic trypsin inhibitor[3,4] (BPTI). Reduction of the three disulphide bonds in native BPTI unfolds the protein in the absence of denaturants. The pathway of disulphide-bond formation in the refolding of reduced BPTI has been determined by Creighton (for review, see ref. 5). The term 'pathway' has led to some confusion and we use it here in the same sense as Creighton: a pathway describes the kinetically most accessible routes for folding. Other routes for folding are possible but are not depicted because they are less probable.

A simplified version of the pathway for folding of BPTI (ref. 3) is shown in Fig. 1*a*. A crucial intermediate in this pathway contains a single disulphide bond between residues 30 and 51 (ref. 6) and is called [30–51]. Because disulphide-bond formation and protein folding are thermodynamically linked processes, the conformations that favour particular disulphide bonds in folding are stabilized by those disulphides[3]. Thus, structural characterization of intermediates trapped by blocking the remaining thiols may allow one to understand the pathway in structural terms, provided that the blocking groups do not alter the conformations of these intermediates. Unfortunately, solubility limitations and the presence of substantial unfolded regions have hindered detailed structural characterization of trapped BPTI folding intermediates. Nevertheless, several trapped intermediates have been surveyed using NMR[7]: spectra of the intermediates were found to resemble spectra of BPTI (ref. 8) more closely as the number of native disulphide bonds in the intermediates increased. A comparison of the resonance pattern in the NMR spectrum of [30–51] with that of native BPTI suggested[7] that at low temperatures [30–51] contains part of the central β-sheet found in BPTI.

We have designed and synthesized a small (30 residues) synthetic analogue of [30–51]. The analogue (called PαPβ) is a disulphide-bonded peptide pair in which two short peptides are connected by a disulphide bond corresponding to the 30–51 disulphide of BPTI (Fig. 1*b*). PαPβ is very soluble in aqueous solution and does not aggregate. As judged by circular dichroism (CD) and NMR, PαPβ is >90% folded in aqueous solution (pH 6) at 4 °C. The structure unfolds when the disulphide bond is reduced. Two-dimensional nuclear Overhauser effect spectroscopy (NOESY) indicates that PαPβ contains much of the secondary and tertiary structure present in the corresponding region of native BPTI. These results indicate that native-like structure can form early in protein folding, and that peptide models can be used to characterize the structures of transient folding intermediates. In future work, it may be possible to crystallize peptide models like PαPβ.

Design of a synthetic analogue of [30–51]

Creighton's pathway[3] for the folding of BPTI, also referred to as the rearrangement pathway, is shown in Fig. 1*a*. BPTI can refold into a native-like structure without the 30–51 disulphide[9], but the resultant species is a non-productive intermediate that must undergo unfavourable disulphide rearrangements to complete folding[4]. This non-productive intermediate [N(30SH, 51SH)] and minor species have been omitted from Fig. 1*a* (see Fig. 6 of ref. 4 for a more complete description). Although there are fifteen possible one-disulphide combinations, [30–51] accounts for 60% of the one-disulphide species observed. The importance of [30–51] is clear from the observation that all further intermediates in the pathway retain the 30–51 disulphide (Fig. 1*a*).

Of the four remaining cysteines in [30–51], residues 5, 14 and 38 participate in formation of the second disulphide bond with approximately equal probability. In contrast, Cys 55 does not participate to an appreciable extent[3,10]. None of the pre-

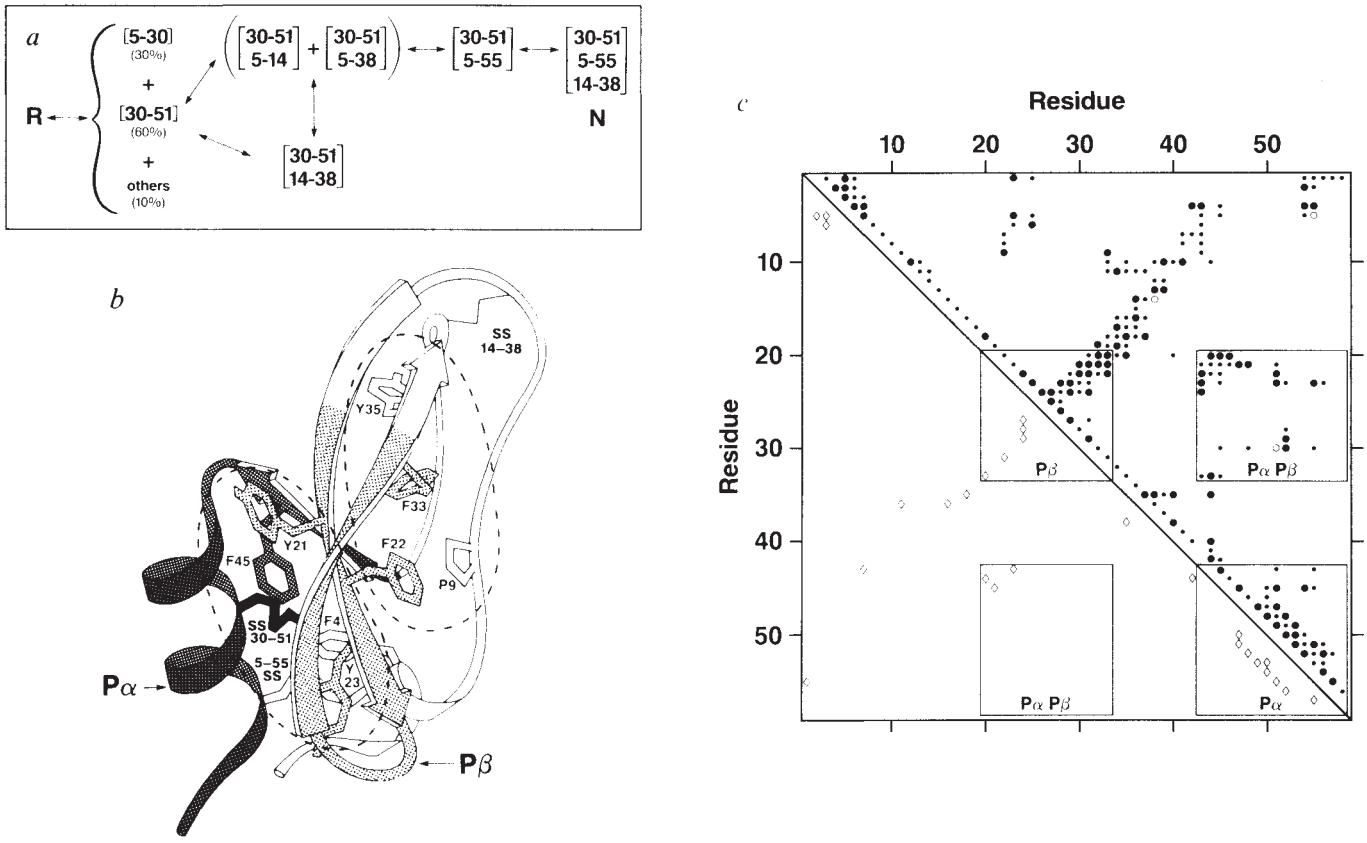


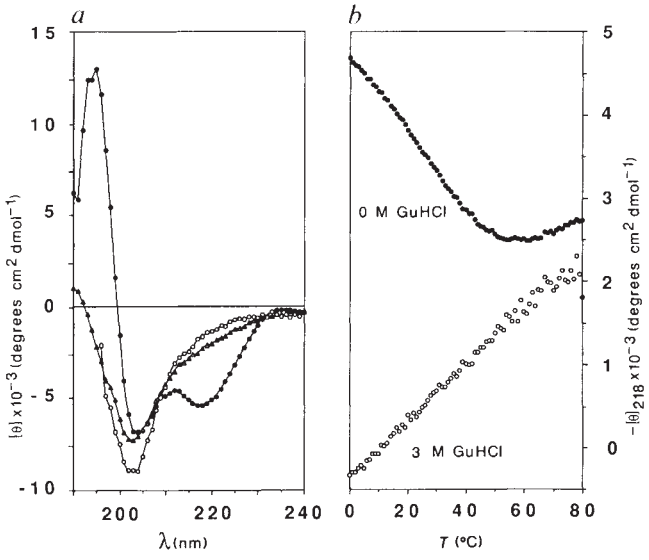
Fig. 1 Design of a peptide model for the [30–51] folding intermediate of BPTI. *a*, Schematic diagram of the folding pathway of BPTI (25 °C, 0.1 M Tris-HCl, 0.2 M KCl, 1 mM EDTA, *p*H 8.7)[3]. R designates the reduced, unfolded protein. N refers to the native, folded protein. Only the most populated intermediates are depicted in this diagram, which indicates the disulphide bonds formed at various stages of folding. The non-productive intermediate, N(30SH, 51SH), is not shown. The one-disulphide intermediates are grouped together because they are in rapid equilibrium; the two most populated intermediates are indicated along with their relative proportions. The plus sign between intermediates [30–51, 5–14] and [30–51, 5–38] indicates that they are both formed directly from the one-disulphide intermediates, that they are both converted directly to [30–51, 5–55], and that both are intermediates in the rearrangement of [30–51, 14–38] to [30–51, 5–55]. *b*, Schematic drawing[52] showing the regions of BPTI corresponding to the peptides Pα and Pβ. Most of one of the hydrophobic cores in BPTI (indicated with dotted lines) is contained in the region corresponding to PαPβ. *c*, A contact plot of BPTI based on X-ray diffraction data[17,18]. Van der Waals interactions between any atom of a residue and the atoms of other residues are indicated above the diagonal. If the distance between two atoms is less than 150% of the sum of their van der Waal's radii, a small dot is shown; less than 100% is indicated by a large dot. Radii were used as previously described[53]. Disulphide bonds are designated with open circles. Intra-residue van der Waal's interactions, and those between residues *i* and *i*+1 have been omitted for clarity. Hydrogen bonds between residues are indicated below the diagonal.

Methods. The peptides were synthesized on an Applied Biosystems Model 430A peptide synthesizer using standard reaction cycles, modified to include acetic anhydride capping. PAM (4[oxymethyl]phenylacetomidomethyl) resins were used. The peptides were cleaved from the resin using TFMSA (trifluoromethane sulphonic acid) and desalted on a Sephadex G-10 column in 5% acetic acid. The peptide complex (PαPβ) was formed by air oxidation in 5 M guanidine hydrochloride (GuHCl), 0.2 M Tris-HCl, *p*H 8, 25 °C for 20 h. GuHCl was present to prevent precipitation of the Pβ homodimer. PαPβ was purified from other products by HPLC on a Vydac C18 semipreparative column using a water/acetonitrile gradient in the presence of 0.1% trifluoracetic acid. The peptide complex was lyophilized from 5% acetic acid and stored in a desiccator. PαPβ synthesized by this procedure was indistinguishable, by analytical HPLC, from the product obtained under argon using glutathione as an oxidant. The identity of PαPβ was confirmed by analysis at the MIT Mass Spectrometry NIH Facility which gave the expected MH^+ relative molecular mass of 3,353.

dominant two-disulphide intermediates formed from [30–51] can easily form a third disulphide bond, even in the case of [30–51, 14–38] which contains two native disulphides[3]. Instead, an intramolecular disulphide exchange reaction leads to the intermediate [30–51, 5–55]. This last intermediate has a native-like structure and the remaining disulphide bond readily forms between cysteines 14 and 38 (Fig. 1*a*).

There are two central questions about Creighton's pathway[3,11]: (1) What is the basis for the accumulation of [30–51] among the one-disulphide intermediates? (2) Why does folding occur preferentially by rearrangements of incorrect two-disulphide intermediates, instead of direct formation of correct disulphides? It is probable that the answer to the first question is structural because [30–51] does not accumulate appreciably when denaturants are present during oxidation[12], and the amount of [30–51] increases both in the presence of stabilizing Hofmeister salts and at low temperatures[13]. Moreover, hydrodynamic[14], immunochemical[15], ultraviolet absorbance and CD studies[16] indicate that [30–51] has a non-random conformation. An answer to the second question requires an explanation of why Cys 55 does not participate significantly in the second step of the folding pathway (Fig. 1*a*).

In the design of a synthetic analogue of [30–51], we made the following working assumptions: (1) the structure detected[7] in [30–51] is localized to regions near the disulphide bond and (2) the structure in [30–51] involves native-like interactions. In BPTI, the 30–51 disulphide connects the C-terminal α-helical region (containing Cys 51) to a strand of the central antiparallel β-sheet (containing Cys 30). The core region of PαPβ therefore includes residues corresponding to the α-helix and to a substantial part of the central antiparallel β-sheet. Residues outside this core region were also included, using a contact map based



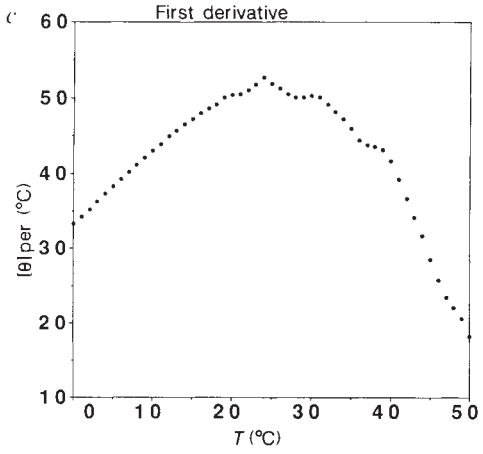


Fig. 2 PαPβ unfolds in denaturing conditions or when the disulphide bond is reduced. *a*, CD spectra of PαPβ (disulphide-bonded peptide pair) at 0 °C (●) or 60 °C (△); and the reduced complex (Pα + Pβ) at 0 °C (○). *b*, Temperature dependence of the CD signal at 218 nm for PαPβ, in the presence or absence of GuHCl. *c*, The first derivative of the temperature dependence in the absence of GuHCl. PαPβ was found to be a monomer under the conditions of these experiments as determined by Sephadex G-25 gel filtration studies. In addition, $[\theta]_{218}$ at 0 °C was independent (±8%) of peptide concentration over the range of 15μM to 0.3 mM.

Methods. All spectra were obtained in standard buffer (0.2 M Na_2SO_4, 10 mM Na_2HPO_4, *p*H 6.0) on an Aviv Model 60DS CD spectrometer using a 1-mm pathlength cell at a sample concentration of 0.06 mM (as determined by tyrosine and cystine absorbance at 275.5 nm in 6 M GuHCl; ref. 54). PαPβ was reduced in 2% mercaptoethanol, 0.5% NH_4HCO_3 and then lyophilized to form reduced Pα + Pβ, as determined by HPLC. The CD spectrum of reduced Pα + Pβ was obtained in the presence of 0.5 mM reduced dithiothreitol and 1 mM EDTA. For the temperature-dependence studies, a 10-mm pathlength cell was used and the sample concentration was 0.02 mM. Temperature was maintained with a HP Model 89100A Peltier temperature control unit. All samples were degassed under vacuum before data collection. A Sephadex G-25 fine column (1.6 cm × 40 cm) in standard buffer at 4 °C was used for the gel filtration experiments. The calibration standards were, in order of increasing elution volume: reduced, carboxymethylated BPTI; native BPTI; oxidized insulin B chain; Ac-SAEDAMRTAGGA; $(PNPA)_3$; and Ac-AKFERQHMDS. PαPβ eluting from the column (at a peak concentration of 0.9 mM) had an apparent relative molecular mass of 2,980, indicating that it was monomeric.

on the crystal structure of BPTI (ref. 17 and 18) as a guide (Fig. 1*c*). Our aim was to include residues in one peptide that interact with residues in the second peptide (that is, in the native structure) while keeping the peptides as short as possible. PαPβ represents approximately one-half of BPTI (30 of 58 residues) and it includes most of one of the hydrophobic cores of the protein (Fig. 1*b*).

The sequences of the individual peptides (called Pα and Pβ) are:

Asn-Asn-Phe-Lys-Ser-Ala-Glu-Asp-**Cys**
-Met-Arg-Thr-Ala-Gly-Gly-Ala Pα: [43–58]

Arg-Tyr-Phe-Tyr-Asn-Ala-Lys-Ala-Gly-
Leu-**Cys**-Gln-Thr-Phe Pβ: [20–33]

Pα contains 16 residues and corresponds to residues 43–58, which include the C-terminal α-helix and a short β-strand in BPTI. Cys 55 has been replaced with an alanine, so Pα contains only one thiol group (corresponding to Cys 51). Pβ contains 14 residues and corresponds to part of the antiparallel β-sheet in BPTI (residues 20–33) including Cys 30.

The peptides were synthesized chemically using solid-phase methods and the peptide pair was formed by oxidation as described in the Fig. 1 legend. After oxidation, PαPβ was purified using reverse-phase HPLC. The identity of PαPβ was confirmed using fast atom bambardment (FAB) mass spectrometry.

PαPβ folds into a native-like conformation

CD spectra indicate that PαPβ folds in aqueous solution and that this structure can be unfolded with increasing temperature, whereas there is no evidence for substantial folded structure when the disulphide bond is reduced (Fig. 2*a*). As monitored by CD, the thermal unfolding transition for PαPβ is broad, spanning over 60 °C, and a fully folded baseline is not reached even at 0 °C (Fig. 2*b*). The transition is completely reversible at temperatures up to 80 °C provided that the sample is degassed before use. Adding denaturant (guanidine hydrochloride) eliminates the transition, and a linear temperature-dependence of the CD signal, often seen with unfolded peptides[19] is observed instead (Fig. 2*b*). PαPβ does not aggregate, as judged by gel filtration, and there is no significant dependence of the CD signal on peptide concentration over a 20-fold range (see Fig. 2 legend).

Proton NMR spectra (aromatic region) of PαPβ at various temperatures are shown in Fig. 3*a*. There are not separate peaks for folded and unfolded species, indicating that interconversion between folded and unfolded molecules is fast on the NMR timescale. Fast-exchange allows the folding transitions for individual protons to be monitored by chemical shift measurements (Fig. 3*b*). The temperature-dependence curves are sigmoidal and broad, and different resonances may show qualitatively different transitions, although it is important first to establish proper baseline values for the transition curves.

The folded structure in PαPβ is similar to that in [30–51] as judged by NMR (compare Fig. 3*a* with Fig. 3 of ref. 7). The structure in [30–51] seems to be more stable[7] than that in PαPβ. An interesting difference between the two species is that the interconversion between folded and unfolded molecules of [30–51] shows intermediate-exchange kinetics on the NMR timescale[7], whereas PαPβ shows fast-exchange kinetics.

To characterize the structure in PαPβ, we have used two-dimensional NMR spectroscopy. Sequential strategies[20] have been used to assign 90% of the resonances in PαPβ. Unequivocal starting assignments were obtained as described in the Fig. 4 legend. A more complete description of the 2D-NMR analysis will be published elsewhere. Figure 4*a* summarizes experimental data used for backbone resonance assignments. Although the NMR assignments for PαPβ were made independently of the assignments for BPTI, the relative chemical shifts in the two

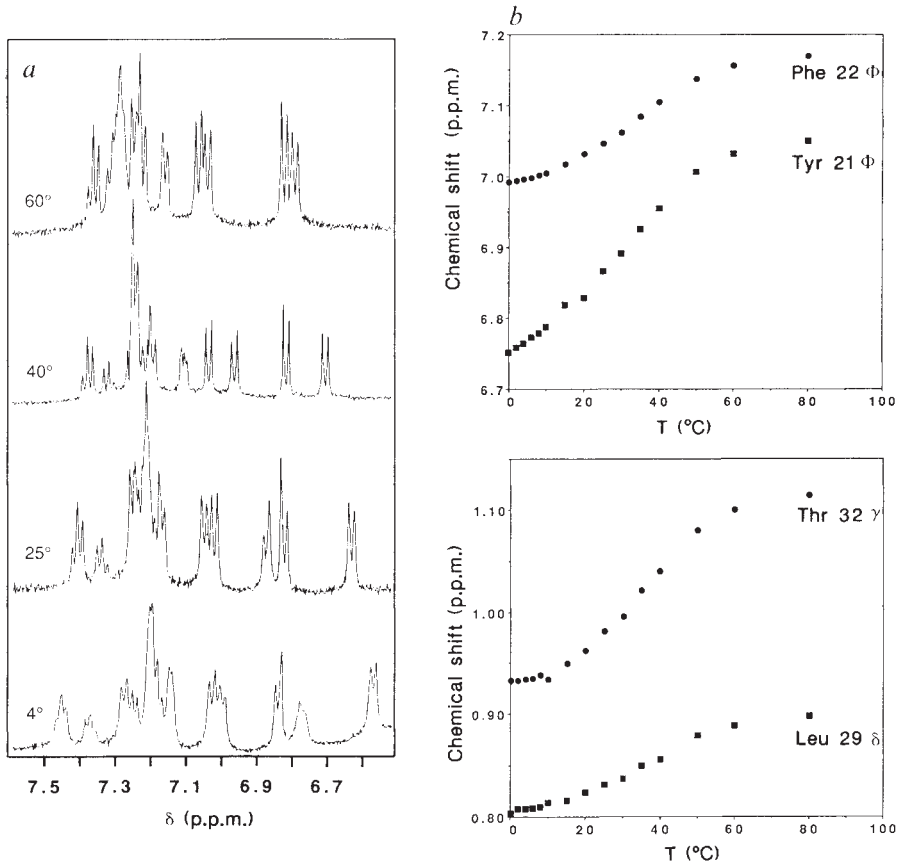


Fig. 3 Because the folding of PαPβ is fast on the NMR timescale, the folding transitions of individual residues can be monitored by chemical shift changes. *a*, Aromatic region of ^{1}H-NMR spectra for PαPβ at various temperatures. *b*, Temperature dependence of chemical shift changes for selected resonances of PαPβ. Resonances are indicated using nomenclature for the corresponding residue in intact BPTI. p.p.m., Parts per million.
Methods. The sample concentration was 5 mM in standard buffer made with D_2O. The chemical shift standard was TMSP (trimethylsilylpropionic acid), whose chemical shift at *p*H 6.0 was assumed to be −0.017 p.p.m. (ref. 55). The data were collected on a 500 MHz spectrometer at the MIT Francis Bitter National Magnet Laboratory, using pre-saturation of residual HOD for 2.0 s. A shifted sine-bell apodization was used to enhance resolution.

species are remarkably similar (Fig. 4*b*). For comparison, the range of chemical shifts for amino acid residues in model 'random-coil' peptides[21] are indicated with squares in Fig. 4*b*. These results indicate that the structures in PαPβ and native BPTI are similar and that the population of folded PαPβ molecules at 4 °C is high.

Figure 5*a* shows a region of a NOESY spectrum of PαPβ. Analysis of the NOESY data indicates that most, if not all, of the secondary structure present in the corresponding region of BPTI is also present in the folded conformation of PαPβ. In particular, the C-terminal α-helix and the central antiparallel β-sheet seem to be intact (Table 1). Tertiary interactions between these regions of secondary structure are also present in PαPβ (Table 1). These tertiary interactions are similar to those found in intact BPTI, as indicated in Fig. 5*b*.

Mechanism of protein folding

The finding that PαPβ folds into a conformation with native-like secondary and tertiary structure supports the initial assumptions made in the design of the analogue (that structure in [30-51] involves native-like interactions and that many of these interactions are localized near the disulphide). Our results indicate that most of the C-terminal α-helix, central antiparallel β-sheet and the hydrophobic core between them is a relatively stable entity in the presence of the 30-51 disulphide. This provides a structural explanation for why [30-51] is populated at high levels in early folding intermediates of BPTI[6].

The structure seen in [30-51] is thermodynamically coupled to disulphide bond formation. One would like to know if development of secondary and tertiary structure kinetically precedes disulphide bond formation, or vice versa. Creighton's experiments suggest that disulphide bond formation occurs first because all six cysteine residues in BPTI participate equally in the formation of the first disulphide bond[22]; rapid disulphide shuffling allows accumulation of the thermodynamically most stable intermediates. Kinetic studies of the association between Pα and Pβ may provide a direct evaluation of this conclusion.

Why doesn't Cys 55 participate in the second step of the rearrangement pathway (Fig. 1*a*)? The structure of PαPβ could explain why the 14-55 and 38-55 disulphide bonds are not formed readily from [30-51], because the antiparallel β-sheet places Cys 14 and Cys 38 away from Cys 55. The observation that the 5-55 disulphide bond does not form requires another

Table 1 Selected NOEs observed in NMR spectra of PαPβ

Structure	Observed NOESY cross-peak	Distance in crystal structure of BPTI
β-sheet	21α-32α	2.4 Å
	23α-30N	4.4 Å
	24α-29N	3.4 Å
	26α-27N	3.5 Å
α-helix	47β-49N	2.6 Å
	48β-52N	4.7 Å
	52α-55β	4.2 Å
	54N-55N	2.6 Å
Tertiary	21δ-46α	4.1 Å
	21ε-48N	3.2 Å
	23ε-55β	4.1 Å
	23β-55β	2.5 Å
	30β-48β	3.5 Å
	30N-52γ	3.4 Å
	45δ-50α	4.8 Å
	45ε-54γ	2.8 Å
	45ε-55β	4.5 Å

All nuclear Overhauser effects (NOEs) listed were observed as cross-peaks that could be unambigously assigned (that is, definitive assignments of resonances to the associated protons with no possibility of overlap with other resonances) in NOESY spectra of PαPβ at 4 °C (Fig. 5*a*). Cross-peaks were observed both in spectra accumulated using 250 ms and 350 ms mixing times. For comparison, the distances between the corresponding protons in the crystal structure of BPTI (ref. 18) are indicated.

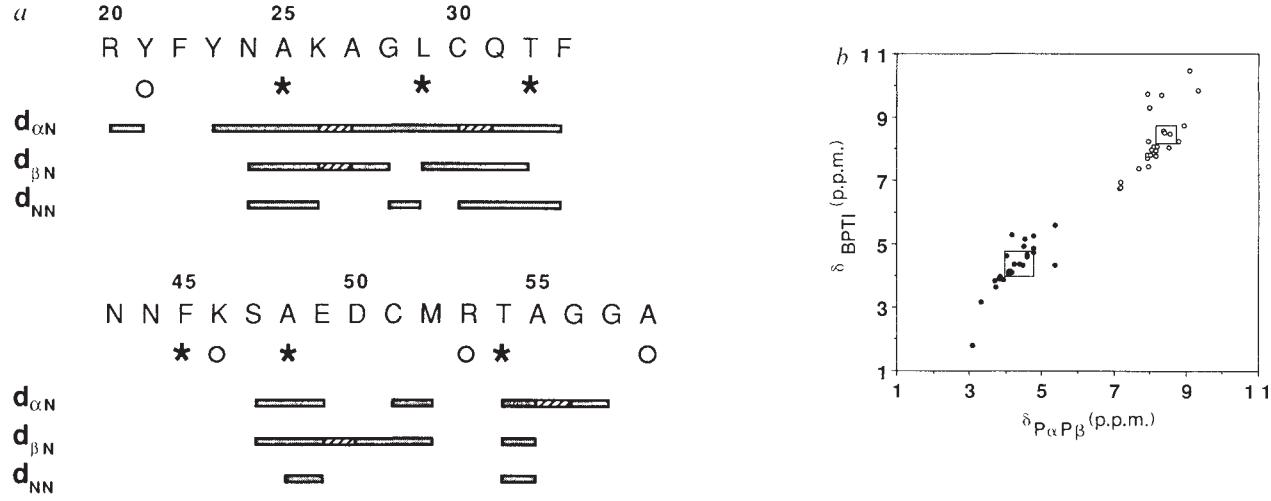


Fig. 4 Independent NMR assignments for PαPβ give relative chemical shifts that are remarkably similar to those for native BPTI. *a*, Summary of experimental data used for sequential assignments. The amino-acid sequence of PαPβ (using nomenclature for the corresponding residues in BPTI) is shown and the following symbols are used[20]: *Definitive assignments made independently from 2D-NMR, as described below; $d_{\alpha N}$, $d_{\beta N}$ and d_{NN} represent NOE connectivities between residues i and $i+1$; dotted regions indicate that an unambiguous NOE was observed; cross-hatched regions that the i to $i+1$ NOE was ambiguous, but that connectivities from this NOE could be unambiguously connected to the appropriate amide and alpha protons on the same residue; ○, assignments made on the basis of spin systems after elimination of other possibilities. Essentially complete side-chain assignments have been obtained for all residues except 22, 43 and 44. *b*, The chemical shifts of all assigned amide and alpha protons in PαPβ against those observed for the corresponding residues in BPTI (ref. 8). The range of chemical shifts values for amide and alpha protons in model 'random-coil' peptides[21] are indicated with squares.

Methods. DQF-COSY (double quantum filtered two-dimensional J-correlated spectroscopy), RELAY (relayed coherence transfer spectroscopy), TOCSY (total correlation spectroscopy) and NOESY (two-dimensional nuclear Overhauser enhancement spectroscopy)[20] spectra (800 $t_1 \times 2{,}048$ t_2) were collected on a 500 MHz Bruker spectrometer at the Fox Chase Cancer Center (Philadelphia) with a recycle delay of 2.0 s. The mixing times for the NOESY spectra were 250 or 350 ms. Peptide concentration was 15 mM. A sweep width of 5 kHz was used in both dimensions, and a phase-shifted sine-bell function was used to enhance resolution. Definitive assignments that served as starting points for the sequential assignments were made as follows: the methyl resonances of Ala 25 and Ala 48 were assigned using PαPβ variants which contained deuterated alanine (introduced by peptide synthesis) at those sites. The methyl resonances of the only leucine, Leu 29, were assigned on the basis of the unique COSY pattern. The methyl resonances of Thr 54 and Thr 32, the only threonine residues found in Pα and Pβ respectively, were assigned first in spectra of the isolated peptides at high temperatures (60 °C) on the basis of their COSY patterns and intrinsic chemical shift values[21]. These assignments were then transferred to spectra of PαPβ at 60 °C (that is, unfolding conditions) and the resonances were followed as a function of temperature to 4 °C. From these starting points, sequential assignment strategies[20] were used. Details of the assignment procedure will be published elsewhere.

explanation. Previous studies indicate that native-like structure in partially folded intermediates needs to be distorted significantly in the transition state to form the 5-55 disulphide[23,24]. The native-like structure observed in PαPβ supports this view, although more work is needed to define the nature of disulphide bond rearrangements in the folding of BPTI.

It has been suggested[25] that 'direct' formation of the 5-55 disulphide from [30-51] becomes more significant when folding occurs at higher temperatures. Our results are consistent with this because the structure in PαPβ and [30-51] (ref. 7) is thermally labile: higher temperatures would be expected to decrease the amount of structure in the vicinity of Cys 55 in [30-51]. The black mamba toxins, which are homologous to BPTI but significantly less stable, follow a more direct folding pathway[26]. The disulphide rearrangements in the folding pathway of BPTI may be a result of the unusual stability of the protein (see refs 23, 24), although disulphide rearrangements are also observed for less stable proteins such as ribonuclease A (ref. 27) and hen lysozyme[28].

In the simple framework model of protein folding[29-32], marginally stable units of isolated secondary structure form early and then coalesce at a later stage in folding. The interaction between secondary structures is mutually stabilizing. Earlier reports that isolated α-helices and β-turns can form in aqueous solution (see refs 33-35 for examples) provide evidence for the first step of the framework model. CD experiments suggest that a peptide containing the C-terminal α-helix of BPTI can also form an α-helix in isolation (E. M. Goodman and P.S.K., to be published). Our results indicate that interaction of this α-helix with β-sheet regions of BPTI provides a substantial increase in stability, supporting the second step of the framework model. Moreover, when six residues are removed from the C-terminus of BPTI (that is, including part of the helical region), the amount of [30-51] produced is reduced significantly and the molecule cannot refold[36].

Nevertheless, PαPβ does include most of one of the hydrophobic cores of BPTI (Fig. 1*a*). Our results do not rule out a folding mechanism in which a non-specific hydrophobic collapse of the polypeptide chain is followed by formation of interacting secondary structure units[37], nor have we addressed the possibility of a 'molten globule'[38,39] intermediate. The roles of specific secondary-structure interactions and non-specific hydrophobic interactions in protein folding remain to be determined.

PαPβ provides a new model system to study β-sheet formation. It has been difficult to obtain models of β-sheets because intermolecular aggregation often occurs. Where it has been studied in model systems (for example polylysine or polytyrosine), β-sheet formation occurs in the second-to-minute time range[40,41]. In contrast, the direct folding reactions (that is, uncomplicated by proline isomerization) of several β-sheet-containing proteins, including BPTI (ref. 42), can occur faster than this ($\tau < 50$ ms). Our results indicate that β-sheets can form in less than 1 ms because the interconversion between folded and unfolded PαPβ molecules is in fast-exchange on the NMR timescale; the frequency difference for these two states can exceed 300 Hz.

Prospects

PαPβ can be considered as a 'synthetic subdomain' of BPTI, where the term subdomain is used to refer to a unit of folded structure larger than an isolated helix or sheet, but smaller than

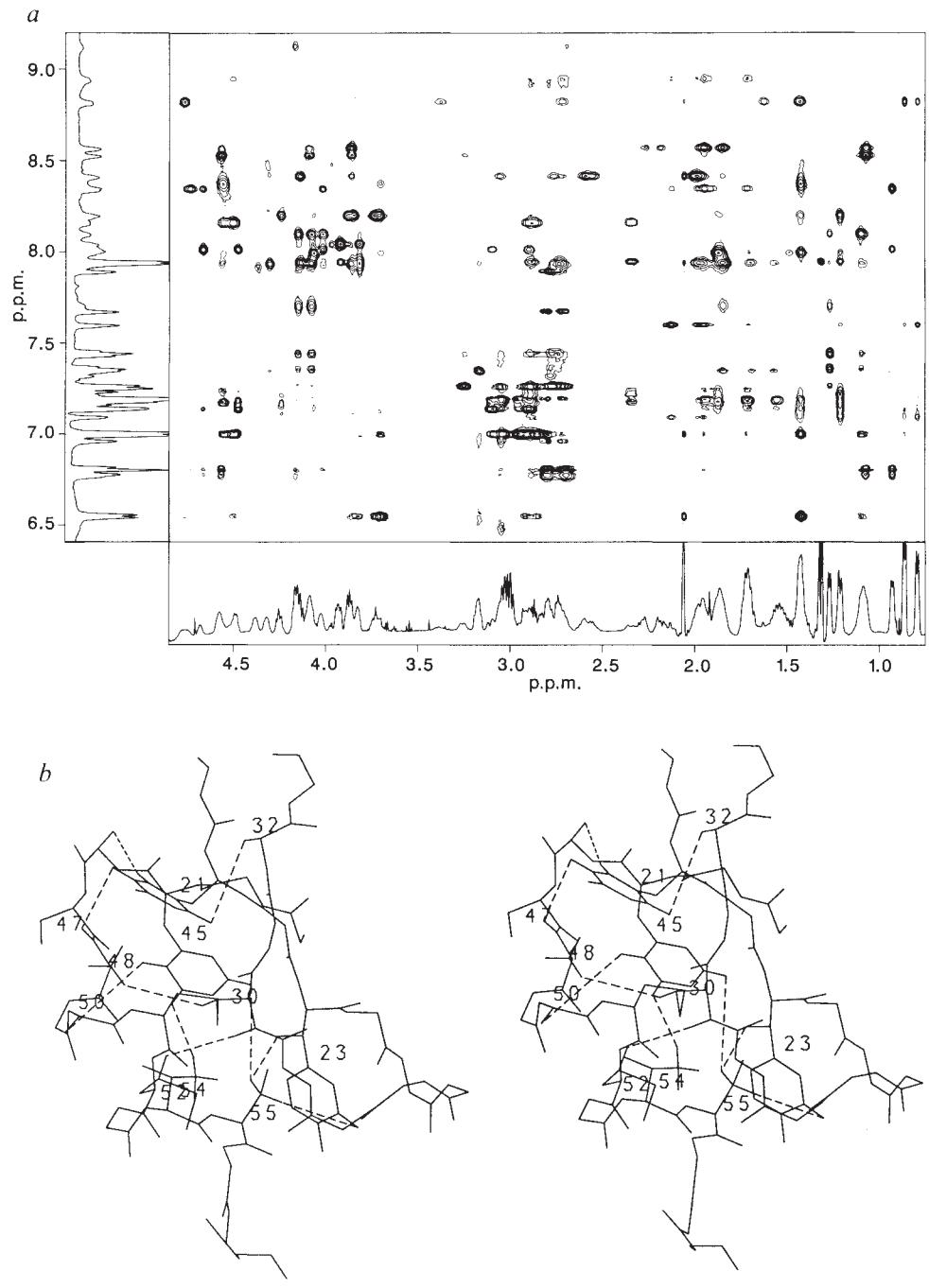


Fig. 5 NOESY spectra indicate that the structure of PαPβ is native-like. *a*, Region of a NOESY spectrum (unsymmetrized) of PαPβ, in 90% H_2O, 10% D_2O, standard buffer, 4 °C. Mixing time was 250 ms. *b*, Stereo drawing of the crystal structure[18] for the region of BPTI corresponding to PαPβ. Selected unambiguous NOEs observed in spectra of PαPβ which result from native-like tertiary interactions are indicated with dotted lines.

an entire domain[43]. Our demonstration that an isolated subdomain can have substantial and stable native-like structure in aqueous solution supports the validity of hierarchical analyses of protein structure and folding[43,44]. Our results indicate that protein regions like PαPβ, that are noncontiguous in primary amino acid sequence but which would otherwise correspond to compact units[45], should be considered in protein structure and folding hierarchies.

The approach used here is best suited for regions of a protein that include a disulphide bond, although in future work it may be possible to connect regions with a non-natural disulphide. Another potential limitation is that the method introduces artificial chain termini. These limitations can be compared with those of other methods for making synthetic subdomains. Proteolytic fragments corresponding to subdomains that fold in aqueous solution have been identified[46,47], but typically they are large by standards of normal peptide synthesis. Template-based synthetic strategies have been developed[48] that permit studies of the interaction between secondary structure units, but the template is non-native.

The ability to make autonomously folding protein fragments using disulphide-bonded peptide pairs may also have practical applications. Anti-peptide antibodies[49,50] often have low affinity for native proteins, possibly because the antibodies recognize unfolded conformations. Synthetic subdomains that fold might be useful immunogens. Similarly, synthetic subdomains could be used to map discontinuous epitopes[51] and to design analogues of protein ligands (for example, hormones). Synthetic subdomains from enzymes may prove to be catalytic.

The approach described here for making peptide models of protein folding intermediates and/or subdomains is simple. In particular, large quantities of product can be obtained using standard peptide synthesis methods, and interacting regions that are non-contiguous in primary sequence can be studied without

introducing a non-native linker. By removing a substantial part of the protein, the cooperativity of protein folding is circumvented and structural analysis is simplified. In theory, the free energy of interaction between different protein regions is represented in the equilibrium constant for forming the peptide pair, and the kinetics of association can give information about the mechanism of interaction. The ability to make amino acid changes using peptide synthesis, combined with the ability to carry out detailed NMR analyses will facilitate studies of the structure, stability and folding of peptide models like $P\alpha P\beta$.

Received 20 July; accepted 20 September 1988.

1. Roder, H., Elöve, G. A. & Englander, S. W. *Nature* **335,** 700–704 (1988).
2. Udgaonkar, J. B. & Baldwin, R. L. *Nature* **335,** 694–699 (1988).
3. Creighton, T. E. *J. molec. Biol.* **113,** 275–293 (1977).
4. Creighton, T. E. & Goldenberg, D. P. *J. molec. Biol.* **179,** 497–526 (1984).
5. Creighton, T. E. *Prog. Biophys. molec. Biol.* **33,** 231–297 (1978).
6. Creighton, T. E. *J. molec. Biol.* **87,** 603–624 (1974).
7. States, D. J., Creighton, T. E., Dobson, C. M. & Karplus, M. *J. molec. Biol.* **195,** 731–739 (1987).
8. Wagner, G. & Wüthrich, K. *J. molec. Biol.* **155,** 347–366 (1982).
9. States, D. J., Dobson, C. M., Karplus, M. & Creighton, T. E. *J. molec. Biol.* **174,** 411–418 (1984).
10. Goldenberg, D. P. *Biochemistry* **27,** 2481–2489 (1988).
11. Creighton, T. E. in *Protein Folding* (ed. Jaenicke, R.) 427–446 (Elsevier/North-Holland Biomedical, Amsterdam, 1980).
12. Creighton, T. E. *J. molec. Biol.* **113,** 313–328 (1977).
13. Creighton, T. E. *J. molec. Biol.* **144,** 521–550 (1980).
14. Creighton, T. E. *J. molec. Biol.* **87,** 603–624 (1974).
15. Creighton, T. E., Kalef, E. & Arnon, R. *J. molec. Biol.* **123,** 129–147 (1978).
16. Kosen, P. A., Creighton, T. E. & Blout, E. R. *Biochemistry* **22,** 2433–2440 (1983).
17. Deisenhofer, J. & Steigemann, W. *Acta crystallogr.* **B31,** 238–250 (1975).
18. Wlodawer, A., Walter, J., Huber, R. & Sjölin, L. *J. molec. Biol.* **180,** 301–329 (1984).
19. Shoemaker, K. R., Fairman, R., York, E. J., Stewart, J. M. & Baldwin, R. L. in *Proc. 10th Am. Peptide Symp.* (ed. Marshall, G. R.) 15–20 (ESCOM, Leiden, 1988).
20. Wüthrich, K. *NMR of Proteins and Nucleic Acids* (Wiley-Interscience, New York, 1986).
21. Bundi, A. & Wüthrich, K. *Biopolymers* **18,** 285–297 (1979).
22. Creighton, T. E. *J. molec. Biol.* **96,** 777–782 (1975).
23. Creighton, T. E. *J. phys. Chem.* **89,** 2452–2459 (1985).
24. Goldenberg, D. P. & Creighton, T. E. *Biopolymers* **24,** 167–182 (1985).
25. Marks, C. B., Naderi, H., Kosen, P. A., Kuntz, I. D. & Anderson, S. *Science* **235,** 1370–1373 (1987).

We thank R. Rutkowski for assistance with synthesis and purification, S. Almo and J. Fetrow for help with graphics, N. Kallenbach for NMR time, E. Goodman and E. O'Shea for help with NMR analysis, H. Roder and J. Wand for assitance in obtaining NMR data and T. Lin and J. Staley for help with disulphide chemistry. We also thank many colleagues, especially D. Goldenberg, E. Haas, P. Matsudaira and T. Creighton for discussions and T. Alber and C. R. Matthews for comments on the manuscript. T.G.O. is an American Cancer Society Postdoctoral Fellow. This research was supported in part by grants from the NSF and the Lucille P. Markey Charitable Trust.

26. Hollecker, M. & Creighton, T. E. *J. molec. Biol.* **168,** 409–437 (1983).
27. Anfinsen, C. B., Haber, E., Sela, M. & White, F. H. Jr *Proc. natn. Acad. Sci. U.S.A.* **47,** 1309–1314 (1961).
28. Saxena, V. P. & Wetlaufer, D. B. *Biochemistry* **9,** 5015–5022 (1970).
29. Ptitsyn, O. B. & Rashin, A. A. *Biophys. Chem.* **3,** 1–20 (1975).
30. Karplus, M. & Weaver, D. L. *Nature* **260,** 404–406 (1976).
31. Richmond, T. J. & Richards, F. M. *J. molec. Biol.* **119,** 537–555 (1978).
32. Kim, P. S. & Baldwin, R. L. *A. Rev. Biochem.* **51,** 459–489 (1982).
33. Kim, P. S. & Baldwin, R. L. *Nature* **307,** 329–334 (1984).
34. Dyson, H. J. *et al. Nature* **318,** 480–483 (1985).
35. Shoemaker, K. R., Kim, P. S., York, E. J., Stewart, J. M. & Baldwin, R. L. *Nature* **326,** 563–567 (1987).
36. Creighton, T. E. *J. molec. Biol.* **119,** 507–518 (1978).
37. Rose, G. D. & Roy, S. *Proc. natn. Acad. Sci. U.S.A.* **77,** 4643–4647 (1980).
38. Kuwajima, K. *J. molec. Biol.* **114,** 241–258 (1977).
39. Ptitsyn, O. B. *J. Protein Chem.* **6,** 273–293 (1987).
40. Hartman, R., Schwaner, R. C. & Hermans, J. *J. molec. Biol.* **90,** 415–429 (1974).
41. Auer, H. E. & Patton, E. *Biophys. Chem.* **4,** 15–21 (1976).
42. Jullien, M. & Baldwin, R. L. *J. molec. Biol.* **145,** 265–280 (1981).
43. Rose, G. D. *J. molec. Biol.* **134,** 447–470 (1979).
44. Go, M. & Nosaka, M. *Cold Spring Harb. Symp. quant. Biol.* **52,** 915–924 (1987).
45. Zehfus, M. H. & Rose, G. D. *Biochemistry* **25,** 5759–5765 (1986).
46. Fronticelli-Bucci, C. & Bucci, E. *Biochemistry* **14,** 4451–4458 (1975).
47. Reutimann, H., Luisi, P. L. & Holmgren, A. *Biopolymers* **22,** 107–111 (1983).
48. Mutter, M. in *Proc. 10th Am. Peptide Symp.* (ed. Marshall, G. R.) 349–353 (ESCOM, Leiden, 1988).
49. Lerner, R. A. *Nature* **299,** 592–596 (1982).
50. Walter, G. & Doolittle, R. F. *Genet. Eng.* **5,** 61–91 (1983).
51. Berzofsky, J. A. *Science* **229,** 932–940 (1985).
52. Richardson, J. S. *Meth. Enzym.* **115,** 359–380 (1985).
53. Lee, B. & Richards, F. M. *J. molec. Biol.* **55,** 379–400 (1971).
54. Edelhoch, H. *Biochemistry* **6,** 1948–1954 (1967).
55. DeMarco, A. *J. magn. Reson.* **26,** 527–528 (1977).

LETTERS TO NATURE

Birth of millisecond pulsars in globular clusters

J. E. Grindlay & C. D. Bailyn

Harvard/Smithsonian Center for Astrophysics,
Harvard College Observatory, Cambridge, Massachusetts 02138, USA

The recent report that two millisecond pulsars were found in the globular cluster 47 Tuc[1], together with the previous discoveries in the clusters M28[2], M4[3] and M15[4], indicates that formation of millisecond pulsars in globular clusters is greatly enhanced over the field, where a fourth example has recently been found[5]. This excess appears to be even larger than that for low-mass X-ray binaries (LMXBs), which show a factor of 100 excess (per unit mass) in globular clusters as opposed to the Galaxy[6,7]. If the birthrate of millisecond pulsars is indeed greater than that of LMXBs in clusters, then, contrary to the common assumption[8], LMXBs cannot be the sole precursors of millisecond pulsars. Here we suggest instead that accretion-induced collapse of white dwarfs in binaries[9] can form millisecond pulsars directly, without requiring a precursor LMXB stage[10]. Ablation of the pre-collapse binary companion by the millisecond pulsar's radiation field[11], a process invoked to explain some of the characteristics of the recently discovered eclipsing millisecond pulsar[5,12,13], can then yield isolated neutron stars without requiring an additional stellar encounter.

A neutron star can be formed by accretion-induced collapse (AIC) of a white dwarf which is driven past its Chandrasekhar limit by accretion from a binary companion. This can occur for a range of both white dwarf initial masses and of accretion rates[14]. An alternative mechanism for the formation of cluster neutron stars is one in which a fraction of the neutron stars formed by the collapse of the cores of massive stars is retained by the shallow potential well of the cluster[15]. In view of the large number of millisecond pulsars (MSPs) observed, we suggest that AIC must be occurring at present in globular clusters, because, as we show below, AIC can produce these objects without recourse to a precursor LMXB stage. Ablation of the secondary[11] eliminates the necessity for an additional stellar encounter[9] in the production of isolated neutron stars. It is interesting to note that two of the globular cluster MSPs have long ($\gtrsim$100-day) orbital periods; these are similar to pulsars in the field, for which an AIC origin has been suggested[16].

AIC will result from a C–O or O–Mg–Ne white dwarf of mass $M \gtrsim 1.0$–$1.2\ M_\odot$ accreting at a rate of either $\dot{M} < 10^{-9}$ or $\dot{M} > 4\times10^{-8}\ M_\odot\ \mathrm{yr}^{-1}$. Such massive white dwarfs can arise from the early generation of massive stars (up to $\sim 8\ M_\odot$; more massive stars which can produce neutron stars directly must be far fewer in number, to avoid disrupting the cluster and polluting it with metals from supernova explosions[9]) or could be produced by the dynamical merging of He white dwarfs[17] from tidal capture or collisions of $0.5\text{-}M_\odot$ He white dwarfs on cluster giants[9]. Although the low-accretion-rate case may lead to mass loss through nova explosions which inhibit the growth of white dwarf mass[18], the nova mass loss rates are uncertain and in a globular cluster the large number of white dwarf–main sequence star binaries that is expected from tidal capture may statistically favour this possibility. The high-accretion-rate case can be due