Protein Science (1994), *3*:1822–1832. Cambridge University Press. Printed in the USA.
Copyright © 1994 The Protein Society

Formation of a native-like subdomain in a partially folded intermediate of bovine pancreatic trypsin inhibitor

JONATHAN P. STALEY[1,3] AND PETER S. KIM[2]

[1] Department of Chemistry and [2] Department of Biology,
Howard Hughes Medical Institute, Whitehead Institute for Biomedical Research,
Massachusetts Institute of Technology, Cambridge, Massachusetts 02142

(RECEIVED May 11, 1994; ACCEPTED July 5, 1994)

Abstract

In the folding of bovine pancreatic trypsin inhibitor (BPTI), the single-disulfide intermediate [30–51] plays a key role. We have investigated a recombinant analog of [30–51] using 2-dimensional nuclear magnetic resonance (2D-NMR). This recombinant analog, named $[30–51]_{Ala}$, contains a disulfide bond between Cys-30 and Cys-51, but contains alanine in place of the other cysteines in BPTI to prevent the formation of other intermediates. By 2D-NMR, $[30–51]_{Ala}$ consists of 2 regions—one folded and one predominantly unfolded. The folded region resembles a previously characterized peptide model of [30–51], named PαPβ, that contains a native-like subdomain with tertiary packing. The unfolded region includes the first 14 N-terminal residues of [30–51] and is as unfolded as an isolated peptide containing these residues. Using protein dissection, we demonstrate that the folded and unfolded regions of $[30–51]_{Ala}$ are structurally independent. The partially folded structure of $[30–51]_{Ala}$ explains many of the properties of authentic [30–51] in the folding pathway of BPTI. Moreover, direct structural characterization of $[30–51]_{Ala}$ has revealed that a crucial step in the folding pathway of BPTI coincides with the formation of a native-like subdomain, supporting models for protein folding that emphasize the formation of cooperatively folded subdomains.

Keywords: BPTI; NMR; protein dissection; protein folding; subdomain

The transient nature of most protein folding intermediates precludes a detailed structural study of these intermediates (for review, see Matthews, 1993). Conveniently, however, intermediates in the folding pathway of bovine pancreatic trypsin inhibitor are distinguished covalently by the disulfide bonds that they contain (Creighton, 1977; Creighton & Goldenberg, 1984; Weissman & Kim, 1991, 1992a). The covalent nature of these intermediates (Fig. 1) allows them to be trapped readily and characterized in structural detail.

The single-disulfide intermediate [30–51], which contains a native disulfide bond between Cys-30 and Cys-51, is a prime target for detailed structural studies because of its importance early in the folding pathway of BPTI. This intermediate (Creighton, 1974) is the first well-populated intermediate on the pathway that leads productively to native protein (Fig. 1B). In addition, though there are 15 possible single-disulfide intermediates for BPTI, only [30–51] and 1 other single-disulfide intermediate, namely [5–55], accumulate substantially during folding at neutral pH (Fig. 1B; Weissman & Kim, 1991). Furthermore, [30–51] forms a second native disulfide bond at a much faster rate than the rate at which reduced, unfolded BPTI forms the same disulfide (Fig. 1B; Weissman & Kim, 1992b). Thus, [30–51] plays a crucial role in folding by restricting the available conformational space of the folding polypeptide and by facilitating subsequent formation of native structure.

During the folding of BPTI, [30–51] can be trapped by using iodoacetate to chemically modify free thiols and prevent formation of subsequent intermediates (Creighton, 1974). One-dimensional NMR studies suggest that the β-sheet and part of a hydrophobic core of native BPTI fold into a native-like conformation in iodoacetate-trapped [30–51] (States et al., 1987). The limited solubility of iodoacetate-trapped [30–51], however, has prevented detailed structural studies by 2-dimensional NMR.

The first detailed structural studies of [30–51] utilized a soluble peptide model of [30–51]. This peptide model, named

Reprint requests to: Peter S. Kim, Department of Biology, Howard Hughes Medical Institute, Whitehead Institute for Biomedical Research, Massachusetts Institute of Technology, Nine Cambridge Center, Cambridge, Massachusetts 02142.

[3] Present address: Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, California 94143-0448.

Abbreviations: BPTI, bovine pancreatic trypsin inhibitor; COSY, correlation spectroscopy; Gdn·HCl, guanidine hydrochloride; HMQC, heteronuclear multiple quantum coherence; HSMQC, heteronuclear single-multiple quantum coherence; HSQC, heteronuclear single quantum coherence; NOESY, nuclear Overhauser effect spectroscopy; ppm, parts per million; TOCSY, total correlation spectroscopy; 2D-NMR, 2-dimensional NMR.

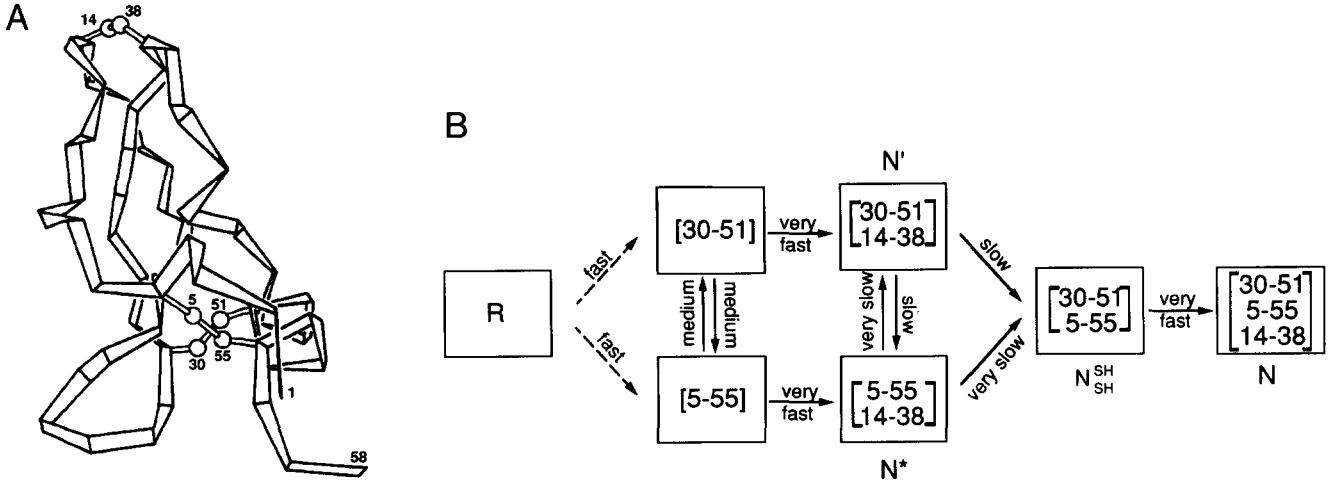

Fig. 1. A: Schematic representation of the structure for BPTI (Deisenhofer & Steigemann, 1975; Wlodawer et al., 1984; Berndt et al., 1992). The terminal residues and the cysteine residues involved in disulfide bonds are labeled. The α-helix includes residues 47–56, and the 2-stranded, antiparallel β-sheet includes residues 16–36. The disulfide bond 14–38 is accessible to solvent, exposing 50% of its total surface area, whereas the disulfide bonds 30–51 and 5–55 are inaccessible, exposing 0% of their total surface area (Lee & Richards, 1971). **B:** Schematic diagram of the folding pathway of BPTI at 25 °C, pH 7.3 (Weissman & Kim, 1991, 1992a). The upper branch of this bifurcated pathway leads productively to native protein. R refers to reduced BPTI; N, to native BPTI; N^{SH}_{SH}, to the precursor of native BPTI; N*, to a native-like, kinetically trapped intermediate; N′, to the major intermediate preceding the rate-limiting step in folding. The dashed arrows indicate that the major single-disulfide intermediates do not form directly from reduced protein but rather from rearrangement of other single-disulfide intermediates. Qualitative descriptions for the relative rates of intramolecular transitions at pH 7.3 are indicated; "very fast" rates are on the order of milliseconds, whereas "very slow" rates are on the order of weeks (Weissman & Kim, 1992b). N^{SH}_{SH}, N*, N′, and [5–55] fold into essentially the same conformation as native BPTI (Stassinopoulou et al., 1984; States et al., 1984; Eigenbrot et al., 1990; Naderi et al., 1991; van Mierlo et al., 1991a, 1991b; Staley & Kim, 1992).

PαPβ, includes the α-helical and β-sheet regions of native BPTI (Fig. 1A) and corresponds to roughly half of the residues in BPTI. PαPβ folds into a stable native-like structure that includes the α-helix and the β-sheet, as well as native-like tertiary packing between these structural elements (Oas & Kim, 1988). A second peptide model, named PαPβ-2, includes several more residues of the β-sheet and folds into a more stable structure (B. Odaert, J.M. Tasoff, D.Y. Kwon, & T.G. Oas, 1994, Structure and stability of an excised BPTI subdomain, in prep.). Detailed 2D-NMR studies of PαPβ-2 confirm that the α-helical and the β-sheet regions fold and pack in an essentially native-like manner (Odaert et al., in prep.).

To investigate the structure of full-length [30–51], we have produced a recombinant analog of [30–51], named $[30\text{–}51]_{Ala}$, that contains alanine in place of cysteines not involved in the 30–51 disulfide bond and the substitution Met-52 → Leu for purposes of expression. A different recombinant analog, named $[30\text{–}51]_{Ser}$, has been studied previously (van Mierlo et al., 1992, 1993). Unlike $[30\text{–}51]_{Ala}$, $[30\text{–}51]_{Ser}$ contains serine in place of cysteines not involved in the 30–51 disulfide bond and the substitution Met-52 → Arg. These 2 sets of mutations have dramatically different effects in the recombinant analogs $[5\text{–}55]_{Ala}$ (Staley & Kim, 1992) and $[5\text{–}55]_{Ser}$ (Darby et al., 1991; van Mierlo et al., 1991a). Most strikingly, $[5\text{–}55]_{Ser}$ is roughly 35 °C less stable than $[5\text{–}55]_{Ala}$. Moreover, $[5\text{–}55]_{Ser}$ is completely unfolded at 25 °C, a temperature at which [5–55] is observed to accumulate during folding at neutral pH (Weissman & Kim, 1991). Though the folded conformations of $[5\text{–}55]_{Ala}$ and $[5\text{–}55]_{Ser}$ are similar, the destabilizing effects of the mutations in $[5\text{–}55]_{Ser}$ suggest that $[5\text{–}55]_{Ala}$ is a better model for [5–55] at neutral pH and that $[30\text{–}51]_{Ala}$ may be the most accurate analog for authentic [30–51] at neutral pH.

Two-dimensional NMR studies (van Mierlo et al., 1993) indicate that $[30\text{–}51]_{Ser}$ folds into a structure qualitatively similar to that observed in PαPβ. In addition, the 15 N-terminal residues of $[30\text{–}51]_{Ser}$ appear to be essentially unfolded. It remains to be determined, however, whether or not the serine substitutions are responsible for the unfolded nature of these N-terminal residues in $[30\text{–}51]_{Ser}$. Furthermore, although broad lines were observed in NMR spectra of $[30\text{–}51]_{Ser}$ (van Mierlo et al., 1992, 1993), no attempt was reported to evaluate or control for protein aggregation.

Using 2D-NMR, we show that the native-like structure observed in PαPβ persists in $[30\text{–}51]_{Ala}$, as in $[30\text{–}51]_{Ser}$, and we confirm that the first 14 N-terminal residues are essentially unfolded. Though $[30\text{–}51]_{Ala}$ does aggregate under certain conditions, sedimentation equilibrium experiments demonstrate that $[30\text{–}51]_{Ala}$ does not aggregate under the conditions used for our NMR studies. Most importantly, we demonstrate that the folded and unfolded regions of $[30\text{–}51]_{Ala}$ are structurally independent, indicating that the folding of BPTI begins within restricted regions of the polypeptide.

Results

The CD signal of $[30\text{–}51]_{Ala}$ at 222 nm exhibits a sigmoidal dependence on temperature that is reversible, indicating that $[30\text{–}51]_{Ala}$ folds (Fig. 2A). The melting temperature of $[30\text{–}51]_{Ala}$ is ~35 °C, similar to the melting temperature of PαPβ-2 (Fig. 2B), which contains roughly two-thirds of the residues in $[30\text{–}51]_{Ala}$

(Odaert et al., in prep.). This similarity in stability suggests that the additional residues in $[30\text{–}51]_{Ala}$ do not fold. These additional residues consist of the first 17 N-terminal residues in BPTI and the loop residues 36–41. The stability of $[30\text{–}51]_{Ala}$ is also comparable to the stability of $[30\text{–}51]_{Ser}$ (van Mierlo et al., 1992), suggesting that $[30\text{–}51]_{Ala}$ is no more folded than $[30\text{–}51]_{Ser}$.

Two distinct sets of peaks are observed in 2D-NMR spectra of $[30\text{–}51]_{Ala}$ (Fig. 3), suggesting that $[30\text{–}51]_{Ala}$ consists of 2 structurally distinct regions. One set of peaks is broad (for example, Phe-22, Tyr-35, and Cys-51), roughly twice as broad as the corresponding peaks in native BPTI spectra obtained under identical conditions. Many of these broad peaks also have dispersed chemical shifts, indicating that they arise from a folded region of $[30\text{–}51]_{Ala}$. Strong similarity between spectra of $[30\text{–}51]_{Ala}$ and spectra of PαPβ-2 (Odaert et al., in prep.), $[30\text{–}51]_{Ser}$ (van Mierlo et al., 1993), and BPTI (Wagner & Wüthrich, 1982b; Tüchsen & Woodward, 1987; Wagner et al., 1987) facilitated the assignment of these peaks (Fig. 3; Table 1). The peaks in this first set were assigned primarily to residues that are contained in PαPβ-2 and that include those residues having amide protons protected from exchange in $[30\text{–}51]_{Ala}$ (Staley, 1993).

A second set of peaks is sharp and lacks chemical-shift dispersion (Fig. 3; for example, Asp-3, Glu-7, and Thr-11), suggesting that they arise from an unfolded region of $[30\text{–}51]_{Ala}$. Some of these peaks are strong in intensity, whereas others are weak. The strong peaks were readily assigned to the first 14 N-terminal residues of $[30\text{–}51]_{Ala}$ (Fig. 3; Table 1). The weak peaks were not assigned, though they most likely arise from the same N-terminal residues (see below).

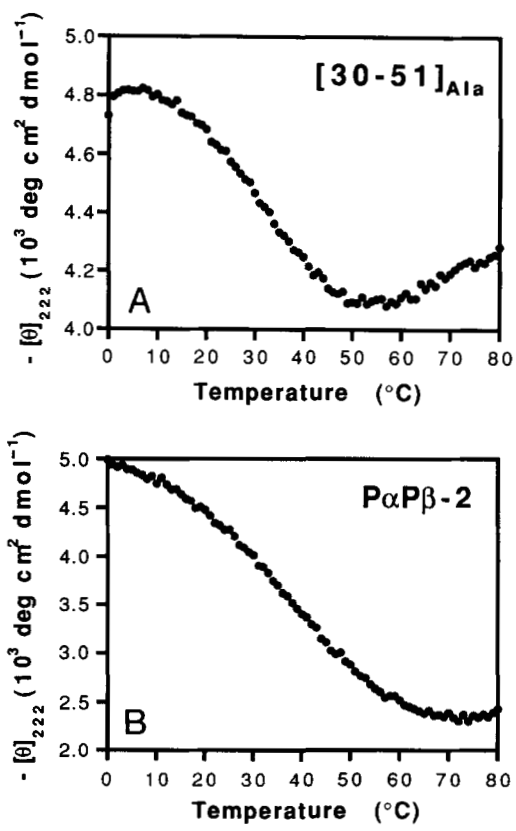

Fig. 2. $[30\text{–}51]_{Ala}$ folds into a stable structure, but no more stable than the peptide model PαPβ-2 (Odaert et al., in prep.). The temperature dependence of $[\theta]_{222}$ is shown for **(A)** $[30\text{–}51]_{Ala}$ and **(B)** PαPβ-2. Data were collected under identical conditions: 10 mM sodium phosphate, pH 6.0, 150 mM NaCl, 1 mM EDTA. The melting temperature for $[30\text{–}51]_{Ala}$ is ~35 °C and for PαPβ-2, ~40 °C, as estimated by the maximum of the first derivative of the temperature dependence of $[\theta]_{222}$.

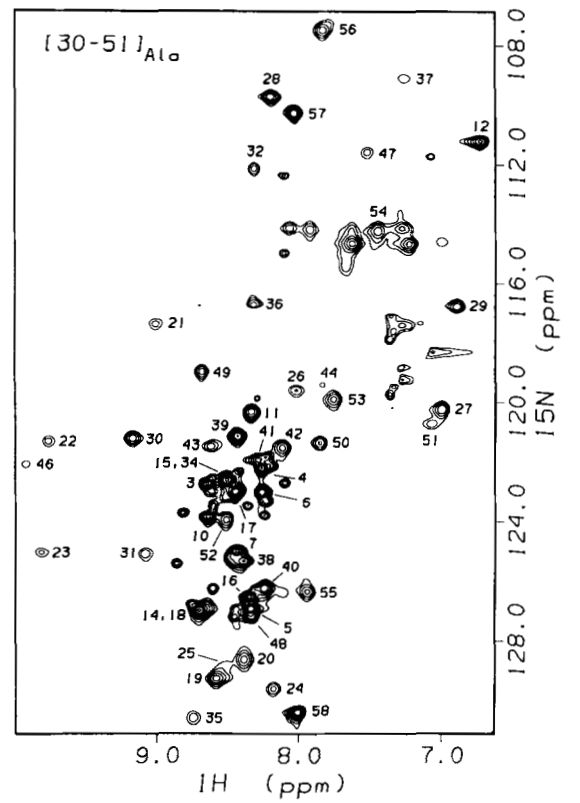

Fig. 3. $[30\text{–}51]_{Ala}$ contains 2 structurally distinct regions. An HSQC spectrum of ^{15}N-labeled $[30\text{–}51]_{Ala}$ in 90% H_2O/10% D_2O, pH 4.6, at 2 °C is shown. Crosspeaks are labeled with their residue assignments. (Crosspeaks from residues 33 and 45 cannot be seen at the contour level plotted.) Notice that crosspeaks from the first 14 N-terminal residues are sharp and lack chemical-shift dispersion, with the exception of Gly-12, whereas crosspeaks from other residues in the molecule tend to be dispersed and more broad.

If $[30\text{–}51]_{Ala}$ truly is only partially folded, then the folded and unfolded regions of $[30\text{–}51]_{Ala}$ should be structurally independent. To test for such structural independence, we dissected $[30\text{–}51]_{Ala}$ into 2 peptides. The first peptide, named $P_{1\text{–}15}^{[30\text{–}51]}$, consists of residues 1–15 of $[30\text{–}51]_{Ala}$ and the second peptide, named $P_{12\text{–}58}^{[30\text{–}51]}$, consists of residues 12–58. The set of peaks in HSQC spectra of $P_{12\text{–}58}^{[30\text{–}51]}$ and $P_{1\text{–}15}^{[30\text{–}51]}$ are similar to subsets of peaks in the HSQC spectrum of $[30\text{–}51]_{Ala}$ (cf. Fig. 4A with 4B and Fig. 4C with 4D, respectively).

A quantitative comparison of the backbone amide chemical shifts of $[30\text{–}51]_{Ala}$ with $P_{12\text{–}58}^{[30\text{–}51]}$ and $P_{1\text{–}17}^{[30\text{–}51]}$ is shown in Figure 5A. (Due to end effects, the peptide $P_{1\text{–}17}^{[30\text{–}51]}$, rather than $P_{1\text{–}15}^{[30\text{–}51]}$, represents most accurately the 14 N-terminal residues in $[30\text{–}51]_{Ala}$.) Though residues near the artificial terminus of $P_{12\text{–}58}^{[30\text{–}51]}$ have significantly different backbone amide proton chemical shifts from the corresponding residues in [30–51], residues 19–58 are strikingly similar (Fig. 5A), differing by 0.01 ppm on average, and by no more than 0.03 ppm (Fig. 5B). This strik-

Table 1. *^1H and ^{15}N chemical shifts (ppm) of $[30-51]_{Ala}$ at 2 °C, pH 4.6*[a]

Residue	$C^\alpha N$	NH	$C^\alpha H$
1	NA	NA	NA
2	NA	–	4.46[b]
3	122.74	8.65	4.52
4	122.21	8.27	4.61
5	126.93	8.34	4.29[c]
6	123.03	8.27	4.34
7	124.98	8.44	4.64
8	NA	–	NA
9	NA	–	4.42[b]
10	123.83	8.64	4.66
11	120.31	8.34	4.36
12	111.21	6.73	3.98
13	NA	–	4.44[b]
14	126.98	8.71	4.29
15	122.60	8.51	4.28
16	126.53	8.35	4.30
17	122.99	8.45	4.36
18	127.11	8.66	4.19
19	129.26	8.59	4.49
20	128.62	8.39	4.77
21	117.37	9.01	5.66
22	121.30	9.76	5.40
23	125.02	9.80	4.53
24	129.60	8.18	4.59
25	128.83	8.53[d]	3.77[d]
26	119.60	8.03	4.09[c]
27	120.24	7.00	4.28
28	109.72	8.20	3.92/3.75
29	116.85	6.89	4.77
30	121.20	9.17	5.50
31	125.06	9.09	4.89
32	112.14	8.32	4.93
33	120.00	8.80	4.85[c]
34	122.67	8.52[d]	4.06
35	130.61	8.75	4.42[c]
36	116.62	8.33	3.85
37	109.11	7.26	NA
38	125.33	8.39	4.28
39	121.14	8.44	4.23
40	126.24	8.25	4.27
41	121.93	8.29	4.27
42	121.53	8.12	4.30
43	121.49	8.63	4.68
44	119.39	7.84	5.28[c]
45	119.77	9.40	NA
46	122.07	9.90	4.43
47	111.57	7.52	4.59
48	127.20	8.33	3.08
49	118.97	8.68	3.84
50	121.36	7.86	4.24
51	120.72	7.07	NA
52	123.94	8.51	3.78
53	119.90	7.76	4.01
54	114.22	7.45	4.14
55	126.36	7.94	3.83
56	107.48	7.83	3.83
57	110.28	8.04	3.92/4.00
58	130.39	8.01	4.18

[a] Resonances not assigned are indicated by NA. A dash indicates that a resonance does not exist.
[b] Measured at 10 °C.
[c] Chemical shift is given for the corresponding resonance in $P_{12-58}^{[30-51]}$.
[d] Measured at 0 °C.

ing similarity in chemical shift demonstrates clearly that the structure contained in residues 19–58 of $[30-51]_{Ala}$ does not depend on the N-terminal residues of $[30-51]_{Ala}$.

A comparison of the chemical shifts of the backbone amide protons in $[30-51]_{Ala}$ with their corresponding protons in PαPβ-2 (Odaert et al., in prep.) indicates that these chemical shifts are also similar (Fig. 5C). In PαPβ-2, the α-helical and β-sheet regions of BPTI fold into a native-like subdomain with tertiary packing (Odaert et al., in prep.). The similarity in chemical shifts demonstrates that the region of $[30-51]_{Ala}$ corresponding to PαPβ-2 also folds into a native-like subdomain. Thus, the native-like subdomain contained in PαPβ-2 represents a truly autonomous folding unit within the context of full-length $[30-51]_{Ala}$.

Similarly, the amide proton chemical shifts of the first 14 residues of $[30-51]_{Ala}$ are very similar to those in $P_{1-17}^{[30-51]}$ (Fig. 5A), as has been observed for $[30-51]_{Ser}$ and a similar peptide (Kemmink et al., 1993). The chemical shifts differ on average by 0.01 ppm and by no more than 0.03 ppm (Fig. 5B). In contrast, these chemical shifts of $[30-51]_{Ala}$ differ, at 10 °C, from those of the native-like molecule $[5-55]_{Ala}$ (Staley & Kim, 1992) on average by 0.25 ppm, and by as much as 0.8 ppm (Staley, 1993). The striking similarity between $[30-51]_{Ala}$ and $P_{1-17}^{[30-51]}$ indicates clearly that the conformation of the first 14 N-terminal residues of $[30-51]_{Ala}$ does not depend on the rest of the molecule.

By 2D-NMR, there is little evidence for long-range structure in $P_{1-17}^{[30-51]}$ (data not shown). We do, however, observe a strong NOE from the amide proton of Gly-12, which is shifted upfield by ~1 ppm from the expected random coil shift (Bundi & Wüthrich, 1979) to aromatic protons of Tyr-10, as has been observed in a similar peptide (Kemmink et al., 1993), suggesting that the amide dipole of Gly-12 is interacting with the ring dipole of Tyr-10. These observations and others (Amir et al., 1992; Kemmink & Creighton, 1993; Lumb & Kim, 1994) have contributed to a picture of the denatured state of BPTI that includes local clusters of structure. Still, we conclude that $P_{1-17}^{[30-51]}$ is predominantly unfolded and that the 14 N-terminal residues in $[30-51]_{Ala}$ are also predominantly unfolded.

Many of the weak sharp peaks observed in spectra of $[30-51]_{Ala}$ (Fig. 3) are also observed in spectra of $P_{1-15}^{[30-51]}$ (Staley, 1993), indicating that these peaks arise predominantly from conformational heterogeneity of the N-terminal residues in $[30-51]_{Ala}$. The most likely source of the weak peaks is *cis/trans* isomerization about the 4 X-Pro imide bonds contained in the first 13 N-terminal residues. Observation of strong $C^\alpha H_{(i)}$ to $C^\delta H_{(i+1)}$ NOEs for X-Pro imide bonds in $P_{1-15}^{[30-51]}$ indicate (Wüthrich, 1986) that all of these bonds are predominantly *trans* (Staley, 1993), as has been observed in a similar peptide (Kemmink et al., 1993).

At 2 °C, pH 4.6, the dispersed peaks in $[30-51]_{Ala}$ are broader than the corresponding peaks in BPTI. Broad peaks might reflect aggregation or exchange between different conformations that is intermediate on the NMR time scale. Under certain conditions, our recombinant analog $[30-51]_{Ala}$ does not thermally denature reversibly, as monitored by CD (data not shown), suggesting that the broadening of peaks in $[30-51]_{Ala}$ may result from aggregation. Furthermore, in 20 mM acetate, pH 4.6, 100 mM NaCl at 2 °C, $[30-51]_{Ala}$ exhibits a molecular weight 50% higher than expected, as determined by sedimentation equilibrium (data not shown). In the absence of NaCl, however, $[30-51]_{Ala}$ is monomeric by sedimentation equilibrium over a

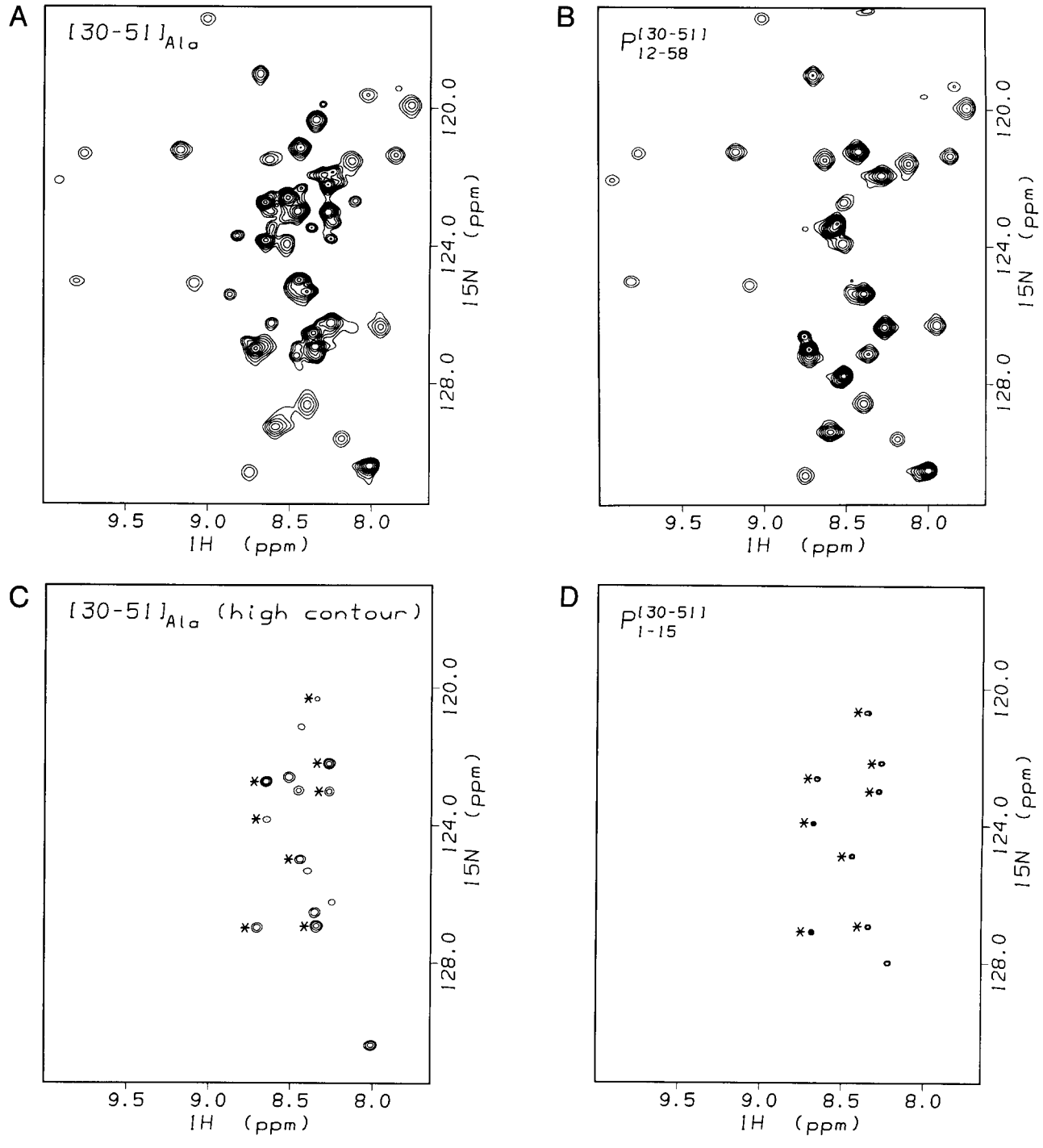

Fig. 4. $[30\text{–}51]_{Ala}$ can be dissected into 2 independent polypeptides without a significant effect on chemical shifts. HSQC spectra of ^{15}N-labeled polypeptides are shown. A comparison of **(A)** a spectrum of $[30\text{–}51]_{Ala}$ and **(B)** a spectrum of $P^{[30-51]}_{12-58}$ shows extensive similarities, demonstrating that the conformation of $P^{[30-51]}_{12-58}$ is largely conserved in $[30\text{–}51]_{Ala}$. A comparison of **(C)** a spectrum of $[30\text{–}51]_{Ala}$, plotted at a high contour level to display selectively sharp peaks, and **(D)** a spectrum of $P^{[30-51]}_{1-15}$ also shows extensive similarities, demonstrating that the conformation of $P^{[30-51]}_{1-15}$ is also largely conserved in $[30\text{–}51]_{Ala}$. To facilitate this comparison, peaks arising from analogous residues are labeled with an asterisk, except for Lys-15—the C-terminal residue of $P^{[30-51]}_{1-15}$. Peaks not labeled with an asterisk in C arise from residues 15, 16, 17, 38, 39, 40, and 58. Spectra were collected in 90% H_2O/10% D_2O, pH 4.6, at 2 °C, in the absence of buffer or salt.

concentration range of 25–400 μM, in 20 mM acetate, pH 4.6, at 2 °C (Fig. 6A, B). A proton 1-dimensional NMR spectrum collected under these conditions at a protein concentration of 200 μM (Fig. 6C, top) is essentially indistinguishable from a spectrum collected at a protein concentration of 4 mM under the conditions used for 2D-NMR experiments (Fig. 6C, bottom). Thus, we conclude that the broadening of peaks in $[30\text{–}51]_{Ala}$ does not result from aggregation. The broadening may, however,

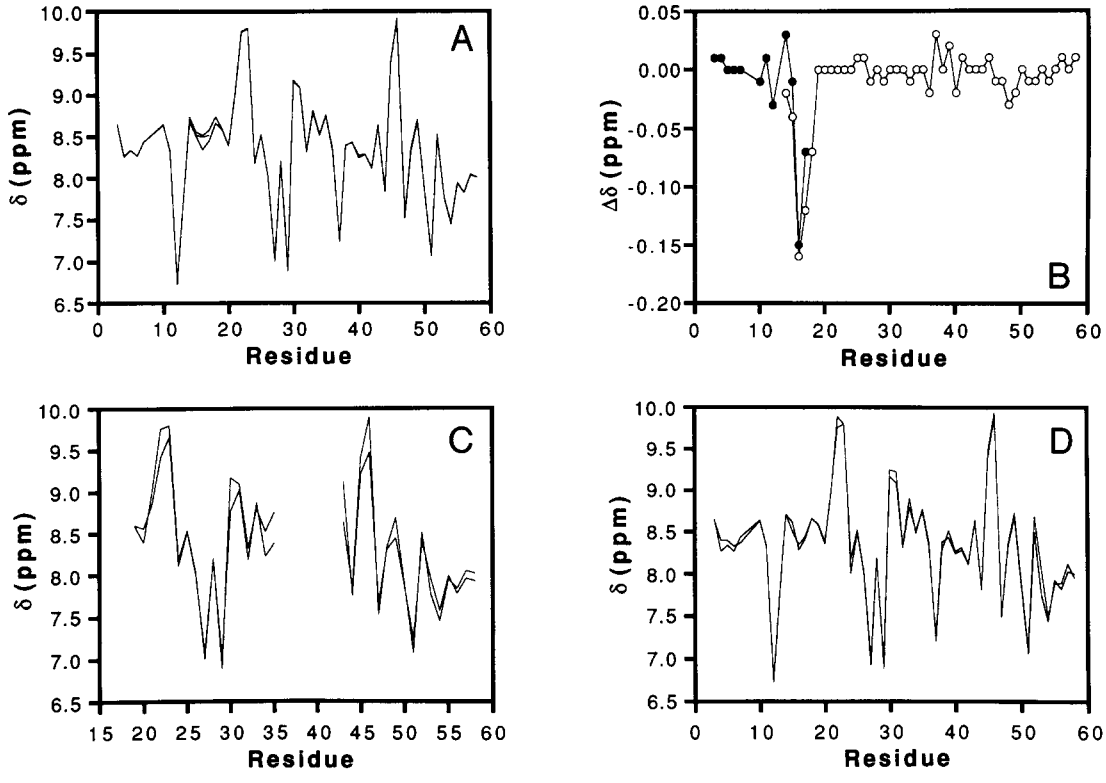

Fig. 5. The structured region of $[30\text{–}51]_{Ala}$ folds into a native-like conformation and the remaining region is as unfolded as an isolated peptide. **A:** The amide proton chemical shifts of residues 19–58 in $[30\text{–}51]_{Ala}$ and the corresponding residues in $P_{12-58}^{[30-51]}$ are strikingly similar, as are the amide chemical shifts of the first 14 N-terminal residues in $[30\text{–}51]_{Ala}$ and the corresponding residues in $P_{-17}^{[30-51]}$. All 3 sets of chemical shifts are plotted and overlaid. **B:** Chemical shift differences between $[30\text{–}51]_{Ala}$ and $P_{12-58}^{[30-51]}$ (open circles) and between $[30\text{–}51]_{Ala}$ and $P_{-17}^{[30-51]}$ (solid circles). **C:** The native-like subdomain of PαPβ-2, corresponding to residues 18–35 and 42–58, folds into a similar conformation in $[30\text{–}51]_{Ala}$. The amide chemical shifts of each are plotted for comparison. **D:** The amide chemical shifts of $[30\text{–}51]_{Ala}$ and $[30\text{–}51]_{Ser}$ are similar and are plotted together for comparison. The chemical shifts for $[30\text{–}51]_{Ala}$, $P_{12-58}^{[30-51]}$, and $P_{-17}^{[30-51]}$ were measured in the absence of buffer or salt at 2 °C, pH 4.6; the chemical shifts for $[30\text{–}51]_{Ser}$, in the absence of buffer or salt at −2 °C, pH 4.6 (van Mierlo et al., 1993); the chemical shifts for PαPβ-2, in 20 mM sodium phosphate, 200 mM sodium sulfate at −2 °C, pH 5.8 (Odaert et al., in prep.).

result from intermediate exchange because the extent of broadening increases with increasing temperature (data not shown).

Though 6 of the 59 residues in $[30\text{–}51]_{Ser}$ (van Mierlo et al., 1992) are different in $[30\text{–}51]_{Ala}$, the backbone amide chemical shifts in these 2 molecules are similar (Fig. 5D). The chemical shifts differ by 0.05 ppm on average and by as much as 0.32 ppm for R53, though this difference probably results from the different amino acids at position 52. These observations indicate that the specific mutations made in $[30\text{–}51]_{Ala}$ and $[30\text{–}51]_{Ser}$ are not responsible for their partially folded nature. Rather, their partially folded nature is an inherent characteristic of the BPTI polypeptide having only the 30–51 disulfide bond.

Discussion

Explanation for the properties of [30–51] in the folding pathway of BPTI

Our results indicate that $[30\text{–}51]_{Ala}$ adopts a partially folded structure in which the α-helical and β-sheet regions of BPTI form a native-like subdomain while the N-terminal residues remain essentially unfolded (Fig. 7). This conclusion is consistent with results observed for the peptide model PαPβ (Oas & Kim, 1988) and the recombinant analog $[30\text{–}51]_{Ser}$ (van Mierlo et al., 1993). This structural view provides an explanation for many of the thermodynamic and kinetic properties of [30–51] in the folding pathway of BPTI as discussed below.

1. Of the 15 single-disulfide intermediates, only the intermediates [30–51] and [5–55] accumulate to a significant extent at neutral pH (Fig. 1B; Weissman & Kim, 1991). The accumulation of [30–51] is explained, as in the case of [5–55] (Staley & Kim, 1990, 1992), by the formation of stable, native-like structure.
2. At neutral pH, [30–51] forms the 14–38 disulfide bond almost as fast as the native intermediate N_{SH}^{SH} forms the same disulfide bond (Fig. 1B; Weissman & Kim, 1992b). The presence of the β-sheet of BPTI in the native-like subdomain of [30–51] explains the rapid rate at which this intermediate forms the 14–38 disulfide bond (Fig. 7A).
3. The intermediate [30–51] does not readily form the 5–55 disulfide bond (Fig. 1B; Creighton, 1977; Weissman & Kim, 1991). The 5–55 disulfide bond is completely buried in native BPTI (Fig. 1A). Formation of a native-like subdomain in [30–51] probably results in the partial burial of

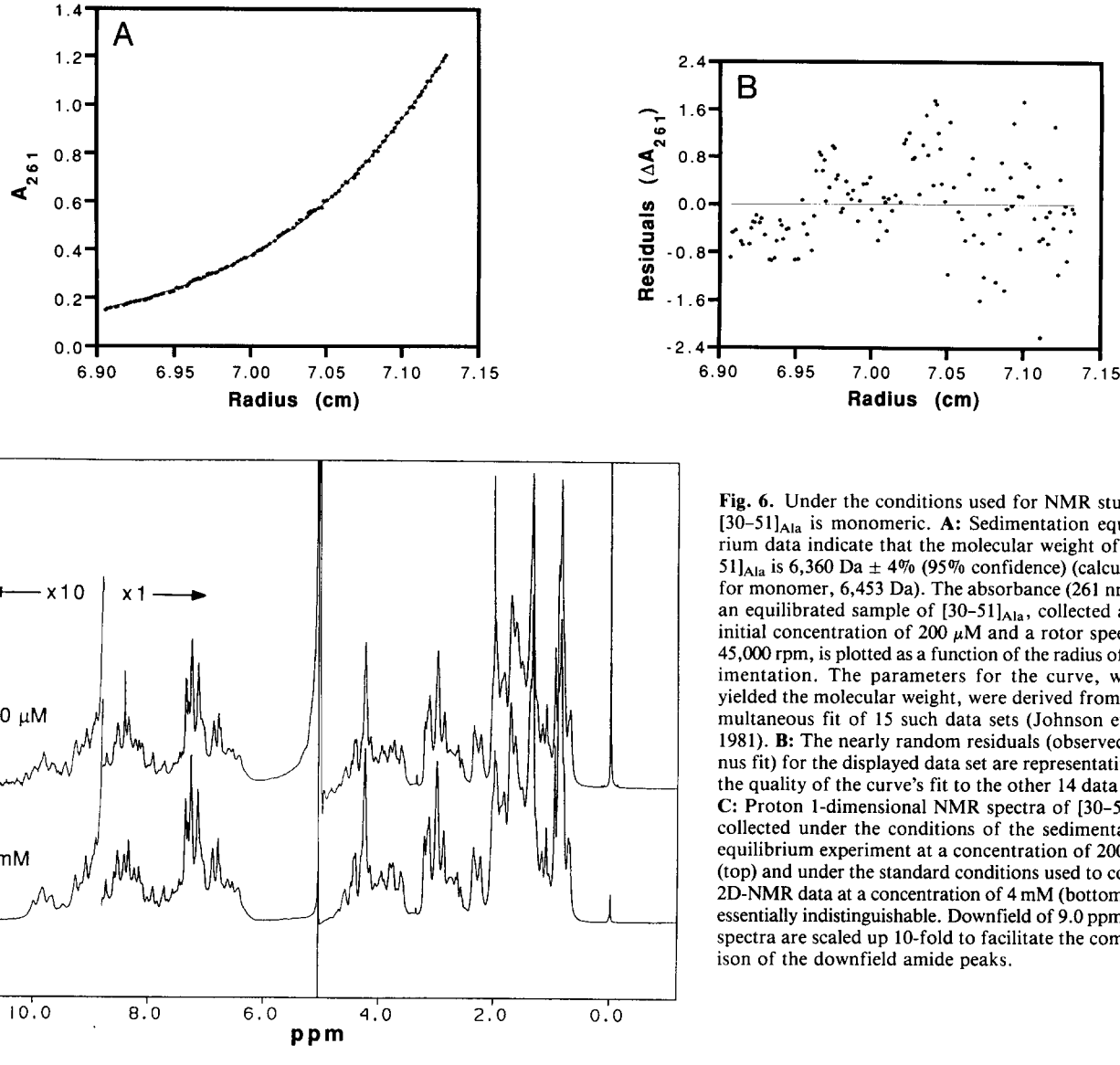

Fig. 6. Under the conditions used for NMR studies, $[30\text{–}51]_{Ala}$ is monomeric. **A:** Sedimentation equilibrium data indicate that the molecular weight of $[30\text{–}51]_{Ala}$ is 6,360 Da ± 4% (95% confidence) (calculated for monomer, 6,453 Da). The absorbance (261 nm) of an equilibrated sample of $[30\text{–}51]_{Ala}$, collected at an initial concentration of 200 μM and a rotor speed of 45,000 rpm, is plotted as a function of the radius of sedimentation. The parameters for the curve, which yielded the molecular weight, were derived from a simultaneous fit of 15 such data sets (Johnson et al., 1981). **B:** The nearly random residuals (observed minus fit) for the displayed data set are representative of the quality of the curve's fit to the other 14 data sets. **C:** Proton 1-dimensional NMR spectra of $[30\text{–}51]_{Ala}$ collected under the conditions of the sedimentation equilibrium experiment at a concentration of 200 μM (top) and under the standard conditions used to collect 2D-NMR data at a concentration of 4 mM (bottom) are essentially indistinguishable. Downfield of 9.0 ppm, the spectra are scaled up 10-fold to facilitate the comparison of the downfield amide peaks.

Cys-55, inhibiting the rate of disulfide bond formation with Cys-5 (cf. Goto & Hamaguchi, 1981).

4. Although [30–51] and [5–55] accumulate in roughly a 1:1 ratio at neutral pH, the relative amount of [5–55] is decreased substantially at pH 8.7 (Weissman & Kim, 1991). At pH 8.7 (i.e., near the pK_a for cysteine residues), approximately half of all cysteines free in solution are negatively charged. Thus, burial of both Cys-30 and Cys-51 in the hydrophobic core of [5–55] (van Mierlo et al., 1991a; Staley & Kim, 1992) will be unfavorable at alkaline pH.
5. At neutral pH, the rearrangement of [30–51] to [5–55] by thiol-disulfide exchange is slow compared to the rate of rearrangement for an unfolded single-disulfide intermediate (Fig. 1B; Weissman & Kim, 1992b). Formation of a native-like subdomain in [30–51] buries the 30–51 disulfide bond (Fig. 1A) and restricts the flexibility of residues Cys-14, Cys-38, and Cys-55 (Fig. 7A), thereby reducing the ability of [30–51] to rearrange via thiol-disulfide exchange.

Formation of the threaded loop in BPTI

Folding of the native-like subdomain in $[30\text{–}51]_{Ala}$ is expected to include formation of a threaded loop that is present in the native protein. The loop is defined by the 30–51 disulfide bond and the thread corresponds to a strand of the β-sheet that traverses through the loop (Fig. 7A; Creighton, 1974). Schellman (1976) has proposed that this threaded loop results from a twisting of the β-sheet prior to formation of the 30–51 disulfide bond. Such a mechanism has been proposed as a general mechanism for forming threaded loops in proteins (Benham & Jafri, 1993). Our finding that $[30\text{–}51]_{Ala}$ folds reversibly indicates, however, that it is possible for the loop to be threaded even in the presence of the 30–51 disulfide bond.

Nevertheless, forming the threaded loop in $[30\text{–}51]_{Ala}$ may be rate-limiting for folding. By NMR, $[30\text{–}51]_{Ala}$ interconverts between its folded and unfolded states with a slower relaxation time (τ = 1–5 ms; data not shown; see also States et al., 1987)

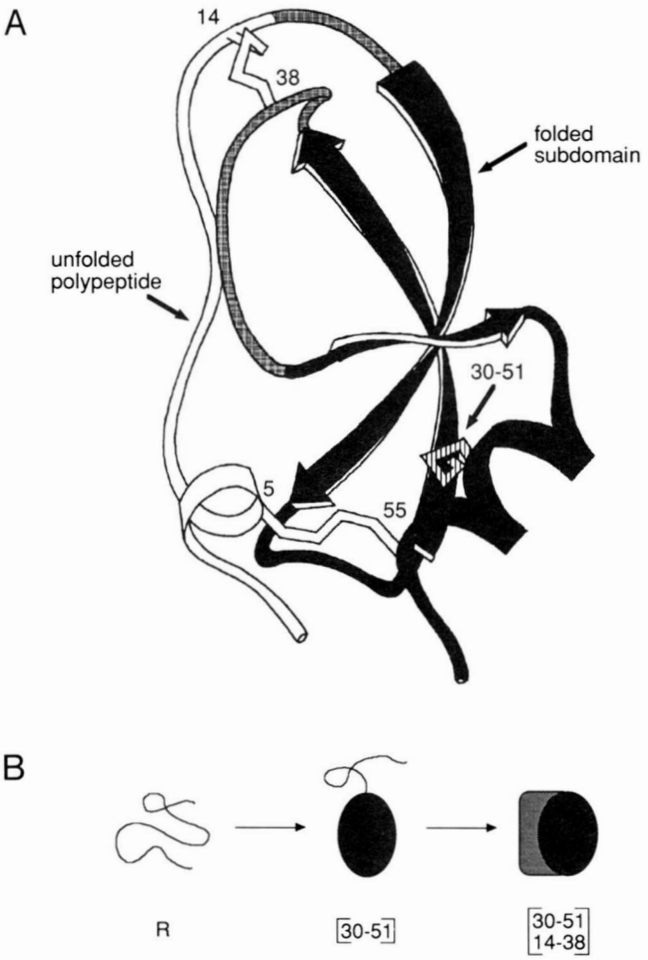

Fig. 7. Protein folding via a subdomain. **A:** Simplified representation of the partially folded structure of $[30\text{–}51]_{Ala}$ within the context of a schematic for the native structure of BPTI. Black indicates the folded region, which contains PαPβ-2; white indicates the essentially unfolded region; gray indicates ill-defined regions. The 3 disulfide bonds in native BPTI are included in the schematic; the 30–51 disulfide bond (striped) is the only one contained in $[30\text{–}51]_{Ala}$. The 4 free cysteines of authentic [30–51] are labeled. **B:** Schematic representation of structurally significant steps in the acquisition of native structure along the pathway for productive folding of BPTI. Lines represent largely unfolded structure; solid, shaded shapes represent folded, native-like structure. For [30–51], the line represents the first 14 N-terminal residues. Labels are as in Figure 1B.

than PαPβ or PαPβ-2 ($\tau \ll 1$ ms; Oas & Kim, 1988; Odaert et al., in prep.). Although the folded structures of $[30\text{–}51]_{Ala}$, PαPβ, and PαPβ-2 are similar (Fig. 5C; Odaert et al., in prep.), only $[30\text{–}51]_{Ala}$ can form the threaded loop. Thus, the need to form a threaded loop in $[30\text{–}51]_{Ala}$ may reduce the rate at which the native-like subdomain folds.

Implications for choosing mutations in designing analogs of disulfide-bonded folding intermediates

There has been some debate as to whether Ala or Ser is the appropriate substitution for a Cys residue not involved in a disulfide bond (Darby et al., 1991; Staley & Kim, 1992). In general, the following arguments favor Ala substitutions for modeling disulfide intermediates at neutral pH. The melting temperature of $[30\text{–}51]_{Ala}$ is comparable to the melting temperature of $[5\text{–}55]_{Ala}$ (Staley & Kim, 1992). This similarity in thermal stability reflects accurately the similarity in stability of the authentic intermediates [30–51] and [5–55] during folding at neutral pH (Weissman & Kim, 1991). In contrast, the stabilities of $[30\text{–}51]_{Ser}$ and $[5\text{–}55]_{Ser}$ differ by at least 20 °C (van Mierlo et al., 1991a, 1992). Moreover, whereas $[5\text{–}55]_{Ala}$ is folded, $[5\text{–}55]_{Ser}$ is essentially unfolded at 25 °C (van Mierlo et al., 1991a), the temperature at which the authentic intermediate is observed to accumulate at neutral pH (Weissman & Kim, 1991).

Although the cysteine substitutions Ala and Ser have very different effects in [5–55], these substitutions are essentially equivalent in [30–51]. The sensitivity of [5–55] to specific substitutions presumably results from the burial of the substitution sites, whereas the insensitivity of [30–51] presumably results from the solvent accessibility of these sites. Thus, without prior structural knowledge of an intermediate, we conclude that Ala is the preferred substitution for a Cys residue when modeling folding intermediates at neutral pH.

Implications for the folding of proteins in general

The structural characteristics of $[30\text{–}51]_{Ala}$ support a stepwise model for protein folding (Fig. 7B) in which folding proceeds by sequential formation of native-like subdomains (Rose, 1979; Richardson, 1981; Oas & Kim, 1988). In further support of a stepwise model for protein folding, native-like subdomains have been implicated in the folding of other intermediates. For example, a peptide model of the single-disulfide intermediate [5–55] (Fig. 1B), named PαPγ, folds into a native-like subdomain similar to the one found in [30–51] (Staley & Kim, 1990). A fragment of *trp* aporepressor also forms a native-like subdomain (Tasayco & Carey, 1992) and is a good candidate for an intermediate in the folding pathway of this protein (Gittelman & Matthews, 1990). In addition, a peptide model of an early intermediate in the folding of cytochrome *c* folds into a subdomain that appears native-like, though not close-packed (Wu et al., 1993).

Subdomains can simplify folding by serving as a template for the remaining unfolded polypeptide. In a similar manner, subdomains could simplify predicting the fold of a native protein. Furthermore, folding intermediates containing native-like subdomains provide an intriguing starting point for theoretical calculations.

It would be useful to identify native-like subdomains in other proteins. One might be able to identify subdomains by identifying the most stable region of a native protein (Woodward, 1994). For example, the β-sheet of BPTI, which is included in the subdomain, contains the slowest exchanging amide protons in native BPTI (Wagner & Wüthrich, 1982a). Thus, the most stable region of native BPTI is a key component of a crucial early folding intermediate. Furthermore, the structures formed in early folding intermediates of both cytochrome *c* (Roder et al., 1988; Elöve & Roder, 1991) and apomyoglobin (Hughson et al., 1990; Jennings & Wright, 1993) also correspond to a region of the native protein that contains the slowest exchanging amide protons. Formation of the most stable subdomain of a native protein early in folding may be a general property of protein folding (Oas & Kim, 1988; Staley & Kim, 1990).

Materials and methods

Plasmid construction

All plasmids were constructed using standard cloning procedures (Sambrook et al., 1989), and their sequences were confirmed by DNA sequencing (Sanger et al., 1977). A plasmid has been constructed previously (Staley & Kim, 1992; Weissman & Kim, 1992b) to produce wild-type BPTI utilizing a T7 expression system (Studier et al., 1990) and includes an F1 replication origin to allow for single-stranded mutagenesis. In order to increase yields of purified protein and to enable the expression of small BPTI peptides, this plasmid has been modified to express a fusion protein of BPTI and a portion of the trp ΔLE 1413 polypeptide (Miozzari & Yanofsky, 1978; Kleid et al., 1981). The trp ΔLE sequence starts after the *Nde* I site of pAED4 (Doering, 1992) with the first residue of the trp L gene and continues through codon Ala-105. This codon is followed by a *Hin*d III site, a Trp codon, and the gene for BPTI. To allow for cyanogen-bromide cleavage of the trp ΔLE polypeptide from BPTI polypeptide, the gene for BPTI was mutated with the substitution Met-52 → Leu, which destabilizes $[5\text{–}55]_{Ala}$ by only 2 °C (Staley & Kim, 1992)—in contrast to the substitution Met-52 → Arg, contained in $[30\text{–}51]_{Ser}$ (van Mierlo et al., 1992), which destabilizes $[5\text{–}55]_{Ala}$ by 10 °C (Staley & Kim, 1992). The 6 methionines in the trp ΔLE sequence were replaced with leucines to simplify the cyanogen-bromide cleavage reaction. The 2 cysteines in the trp ΔLE sequence were replaced with alanines to simplify the oxidation of BPTI mutants while fused to the trp ΔLE polypeptide. The plasmid containing the gene for $[30\text{–}51]_{Ala}$ was constructed by mutating Cys-5, Cys-14, Cys-38, and Cys-55 to Ala, and is called $p[30\text{–}51]_{Ala}$. The gene for producing the peptide $P_{1\text{–}15}^{[30\text{–}51]}$ was constructed by introducing a stop codon after the 15th codon of $[30\text{–}51]_{Ala}$; the gene for the peptide $P_{1\text{–}17}^{[30\text{–}51]}$, by introducing a Met codon after the 17th codon (as a result, this peptide ends in homoserine); the gene for the peptide $P_{12\text{–}58}^{[30\text{–}51]}$, by introducing a Met codon before the 12th codon.

Protein expression

Expression of unlabeled $[30\text{–}51]_{Ala}$ was achieved as described previously for the expression of $[5\text{–}55]_{Ala}$ (Staley & Kim, 1992), except that rifampicin was omitted. Expression of ^{15}N-labeled $[30\text{–}51]_{Ala}$ was achieved using the same procedure, except that minimal M9 media containing 1 g/L ^{15}N-ammonium sulfate was used in place of Luria broth and cells were induced at an OD_{590} of ~0.8 and harvested 3 h later.

Purification of recombinant protein

Inclusion bodies were purified as follows (Marston & Hartley, 1990). Cell pellets were sonicated on ice in 50 mM Tris·HCl, pH 8.7, 1 mM EDTA, and spun at 35,000 × *g* for 20 min. The resulting pellet was sonicated in 50 mM Tris·HCl, pH 8.7, 1 mM EDTA, 1% Nonidet P-40, 1% deoxycholic acid, and spun at 35,000 × *g* for 20 min. The resulting pellet was dissolved in 0.2 M Tris·HCl, pH 8.7, 6 M Gdn·HCl, 0.1 M dithiothreitol, and dialyzed against 5% (vol/vol) HOAc. The entire dialysate (soluble and insoluble material) was lyophilized. To oxidize $[30\text{–}51]_{Ala}$ and $P_{12\text{–}58}^{[30\text{–}51]}$, lyophilized material was dissolved in 0.2 M Tris·HCl, pH 8.7, 6 M Gdn·HCl to a final protein concentration of roughly 100 μM, and oxidized by air overnight with stirring. After oxidation, reactions were dialyzed against 5% (v/v) HOAc and lyophilized. For all fusion proteins, the trp ΔLE polypeptide and, in some cases, a portion of the BPTI polypeptide were cleaved from the polypeptide of interest by resuspending lyophilized material in 5 mL 70% formic acid/L cells and roughly 200 mg cyanogen bromide/L cells for 1–2 h. Reactions were dialyzed against 5% (v/v) HOAc, dialysate was spun at 35,000 × *g* for 20 min, and products, in the supernatant, were purified by reversed-phase HPLC. Expression of $[30\text{–}51]_{Ala}$ in minimal M9 media resulted in final yields of ~10 mg/L. All preparations were >95–99% pure as judged by reversed-phase HPLC. The molecular mass of each gene product was confirmed by laser desorption mass spectrometry (Finnigan MAT Lasermat): ^{15}N-labeled $[30\text{–}51]_{Ala}$: observed, 6,453 Da; calculated, 6,453 Da; ^{15}N-labeled $P_{1\text{–}15}^{[30\text{–}51]}$: observed, 1,678 Da; calculated, 1,678 Da; ^{15}N-labeled $P_{1\text{–}17}^{[30\text{–}51]}$: observed, 2,013 Da; calculated, 2,013 Da; ^{15}N-labeled $P_{12\text{–}58}^{[30\text{–}51]}$: observed, 5,155 Da; calculated, 5,152 Da.

CD

Measurements were made using a 10-mm-pathlength cell and an Aviv model 62DS CD spectrometer equipped with a temperature-control unit. Samples, degassed by reduced pressure, contained protein at a concentration of 15 μM as determined by tyrosine and cystine absorbance of stock solutions (Edelhoch, 1967). Sample buffer contained 10 mM sodium phosphate, pH 6.0, 150 mM NaCl, 1 mM EDTA.

NMR

Strong similarity between spectra of $[30\text{–}51]_{Ala}$ and spectra of BPTI (Wagner & Wüthrich, 1982b; Tüchsen & Woodward, 1987; Wagner et al., 1987), PαPβ-2 (Odaert et al., in prep.), and $[30\text{–}51]_{Ser}$ (van Mierlo et al., 1993) facilitated the assignment of resonances in $[30\text{–}51]_{Ala}$. Assignments were checked by identifying unambiguous, sequential, NOE connectivities (Wüthrich, 1986). Each residue assignment was supported by at least 1 unambiguous sequential NOE connectivity. Spectra of $[30\text{–}51]_{Ala}$ were collected at a protein concentration of 4 mM, whereas spectra of $P_{1\text{–}15}^{[30\text{–}51]}$, $P_{1\text{–}17}^{[30\text{–}51]}$, and $P_{12\text{–}58}^{[30\text{–}51]}$ were collected at 10 mM. All samples, which lacked both buffer and salt, were prepared and adjusted to pH 4.6 at 4 °C. Trimethylsilylpropionic acid was used as an internal standard for measuring ^{1}H chemical shifts (DeMarco, 1977); $^{15}NH_4Cl$, as an external standard for measuring ^{15}N chemical shifts (Levy & Lichter, 1979). Data were collected on a Bruker AMX 500-MHz spectrometer. Two-dimensional homonuclear experiments included COSY (Piantini et al., 1982; Rance et al., 1983; Shaka & Freeman, 1983), NOESY (Jeener et al., 1979; Kumar et al., 1980; Macura et al., 1981), and TOCSY (Rance, 1987). Two-dimensional heteronuclear experiments included the HSQC experiment (Bodenhausen & Ruben, 1980; Bax et al., 1990; Norwood et al., 1990), as well as the HMQC (Bax et al., 1983) and HSMQC (Zuiderweg, 1990) experiments in combination with the COSY and NOESY experiments, respectively, to yield the HMQC-COSY and HSMQC-NOESY experiments (Gronenborn et al., 1989; Norwood et al., 1990). Water was presaturated for 1 s and 1,024 data points were collected in the t_2 dimension. In general, 256 increments were used in the t_1 dimension for homonuclear experiments and 128

increments for heteronuclear experiments, although 256 increments were used in the experiments displayed in Figures 3 and 4. Data from NOESY experiments were collected with a mixing time of 150 ms; data from TOCSY, with mixing times of 45 or 110 ms. In all data sets, the first t_1 was multiplied by 0.5 (Otting et al., 1986). The proton 1-dimensional NMR data shown in Figure 6C were collected with a NOESY pulse sequence using a mixing time of 150 ms.

Equilibrium sedimentation

Measurements were made in Beckman 12-mm-pathlength, 6-sector cells in a Beckman XL-A 90 analytical ultracentrifuge. Samples were dialyzed against 20 mM acetic acid, pH 4.6. Data were collected at 2 °C using 3 rotor speeds (37, 41, and 45 krpm), 3 wavelengths (249, 261, and 272 nm), and 3 initial concentrations (50, 100, and 200 μM), as determined by tyrosine and cystine absorbance of stock solutions (Edelhoch, 1967). Final concentrations varied continuously from 25 μM to 400 μM. Molecular weights were calculated using the program HID4000 (Johnson et al., 1981) by simultaneously fitting 15 data sets using a single molecular weight, 15 intercepts, 14 offsets, and the second virial coefficient as parameters. No systematic variation in the residuals was observed. The second virial coefficient was used as a fitting parameter because nonideality was observed as a *decrease* in the apparent molecular weight as a function of increasing protein concentration. This nonideality most likely results from the high net charge to molecular weight ratio (Williams et al., 1958) of ~1:1,000 for $[30\text{–}51]_{Ala}$ at the low ionic strength of the buffer required to prevent aggregation. An apparent net charge of +3, calculated from the second virial coefficient (Williams et al., 1958), is consistent with this interpretation; a net charge of +7 is expected in the absence of counterions at pH 4.6 given average pK_a's of amino acid side chains and termini (Creighton, 1993). Values for the partial specific volumes of protein species and the density of solutions were calculated using values from Laue et al. (1992). An analysis of ^{15}N-labeled $[30\text{–}51]_{Ala}$ yields a molecular weight of 6,360 Da ± 4% (95% confidence); calculated for monomer, 6,453 Da. A similar analysis of native BPTI yields a molecular weight of 6,480 Da ± 6% (95% confidence); calculated, 6,514 Da.

Acknowledgments

We thank M.G. Milla, M.A. Milhollen, and D.M. Nguyen for plasmid construction and help with protein expression and purification. We also thank T.G. Oas for providing the peptide PαPβ-2 and sharing data prior to publication. In addition, we thank M.L. Johnson and J. Lary for providing the program HID4000 and A.D Frankel for use of his Silicon Graphics computer. Finally, we thank L.P. McIntosh, K.J. Lumb, and B.A. Schulman for help with NMR spectroscopy and Z.-Y. Peng and J.S. Weissman for many helpful discussions. J.P.S. was a Howard Hughes Medical Institute predoctoral fellow. P.S.K. was a Pew Scholar in the Biomedical Sciences. This research was supported by a grant from the National Institutes of Health (GM 41307).

References

Amir D, Krausz S, Haas E. 1992. Detection of local structures in reduced unfolded bovine pancreatic trypsin inhibitor. *Proteins Struct Funct Genet 13*:162–173.

Bax A, Griffey RH, Hawkins BL. 1983. Correlation of proton and nitrogen-15 chemical shifts by multiple quantum NMR. *J Magn Reson 55*:301–315.

Bax A, Ikura M, Kay LE, Torchia DA, Tschudin R. 1990. Comparison of different modes of two-dimensional reverse-correlation NMR for the study of proteins. *J Magn Reson 86*:304–318.

Benham CJ, Jafri MS. 1993. Disulfide bonding patterns and protein topologies. *Protein Sci 2*:41–54.

Berndt KD, Guntert P, Orbons LP, Wüthrich K. 1992. Determination of a high-quality nuclear magnetic resonance solution structure of the bovine pancreatic trypsin inhibitor and comparison with three crystal structures. *J Mol Biol 227*:757–775.

Bodenhausen G, Ruben DJ. 1980. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem Phys Lett 69*:185–189.

Bundi A, Wüthrich K. 1979. ^{1}H-NMR parameters of the common amino acids residues measured in aqueous solutions of the linear tetrapeptides H-Gly-Gly-*X*-L-Ala-OH. *Biopolymers 18*:285–297.

Creighton TE. 1974. The single-disulphide intermediates in the refolding of reduced pancreatic trypsin inhibitor. *J Mol Biol 87*:603–624.

Creighton TE. 1977. Conformational restrictions on the pathway of folding and unfolding of the pancreatic trypsin inhibitor. *J Mol Biol 113*:275–293.

Creighton TE. 1993. *Proteins*. New York: W.H. Freeman and Company.

Creighton TE, Goldenberg DP. 1984. Kinetic role of a meta-stable native-like two-disulphide species in the folding transition of bovine pancreatic trypsin inhibitor. *J Mol Biol 179*:497–526.

Darby NJ, van Mierlo CPM, Creighton TE. 1991. The [5–55] single-disulphide intermediate in folding of bovine pancreatic trypsin inhibitor. *FEBS Lett 279*:61–64.

Deisenhofer J, Steigemann W. 1975. Crystallographic refinement of the structure of bovine pancreatic trypsin inhibitor at 1.5 Å resolution. *Acta Crystallogr B 31*:238–250.

DeMarco A. 1977. pH dependence of internal references. *J Magn Reson 26*:527–528.

Doering DS. 1992. Functional and structural studies of a small F-actin binding domain [thesis]. Cambridge, Massachusetts: Massachusetts Institute of Technology.

Edelhoch H. 1967. Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry 6*:1948–1954.

Eigenbrot C, Randal M, Kossiakoff AA. 1990. Structural effects induced by removal of a disulfide-bridge: The X-ray structure of the C30A/C51A mutant of basic pancreatic trypsin inhibitor at 1.6 Å. *Protein Eng 3*:591–598.

Elöve GA, Roder H. 1991. Structure and stability of cytochrome *c* folding intermediates. In: Georgiou G, De Bernardez-Clark E, eds. *Protein refolding*. Washington, D.C.: American Chemical Society. pp 50–63.

Gittelman MS, Matthews CR. 1990. Folding and stability of *trp* aporepressor from *Escherichia coli*. *Biochemistry 29*:7011–7020.

Goto Y, Hamaguchi K. 1981. Formation of the intrachain disulfide bond in the constant fragment of the immunoglobulin light chain. *J Mol Biol 146*:321–340.

Gronenborn AM, Bax A, Wingfield PT, Clore GM. 1989. A powerful method of sequential proton resonance assignment in proteins using relayed ^{15}N-^{1}H multiple quantum coherence spectroscopy. *FEBS Lett 243*:93–98.

Hughson FM, Wright PE, Baldwin RL. 1990. Structural characterization of a partly folded apomyoglobin intermediate. *Science 249*:1544–1548.

Jeener J, Meier BH, Bachmann P, Ernst RR. 1979. Investigation of exchange processes by two-dimensional NMR spectroscopy. *J Chem Phys 71*: 4546–4553.

Jennings PA, Wright PE. 1993. Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science 262*:892–896.

Johnson ML, Correia JJ, Yphantis DA, Halvorson HR. 1981. Analysis of data from the analytical ultracentrifuge by nonlinear least-squares techniques. *Biophys J 36*:575–588.

Kemmink J, Creighton TE. 1993. Local conformations of peptides representing the entire sequence of bovine pancreatic trypsin inhibitor and their roles in folding. *J Mol Biol 234*:861–878.

Kemmink J, van Mierlo CPM, Scheek RM, Creighton TE. 1993. Local structure due to an aromatic-amide interaction observed by ^{1}H-nuclear magnetic resonance spectroscopy in peptides related to the N terminus of bovine pancreatic trypsin inhibitor. *J Mol Biol 230*:312–322.

Kleid DG, Yansura D, Small B, Dowbenko D, Moore DM, Grubman MJ, McKercher PD, Morgan DO, Robertson BH, Bachrach HL. 1981. Cloned viral protein vaccine for foot-and-mouth disease: Responses in cattle and swine. *Science 214*:1125–1129.

Kumar A, Ernst RR, Wüthrich K. 1980. A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochem Biophys Res Commun 95*:1–6.

Laue TM, Shah BD, Ridgeway TM, Pelletier SL. 1992. Computer-aided interpretation of analytical sedimentation data for proteins. In: Harding SE, Rowe AJ, Horton JC, eds. *Analytical ultracentrifugation in biochemistry and polymer science*. Cambridge, UK: Royal Society of Chemistry. pp 90–125.

Lee B, Richards FM. 1971. The interpretation of protein structures: Estimation of static accessibility. *J Mol Biol 55*:379–400.

Levy GC, Lichter RL. 1979. *Nitrogen-15 nuclear magnetic resonance spectroscopy*. New York: John Wiley & Sons.

Lumb KJ, Kim PS. 1994. Formation of a hydrophobic cluster in denatured bovine pancreatic trypsin inhibitor. *J Mol Biol 236*:412–420.

Macura S, Huang Y, Suter D, Ernst RR. 1981. Two-dimensional chemical exchange and cross-relaxation spectroscopy of coupled nuclear spins. *J Magn Reson 43*:259–281.

Marston AO, Hartley DL. 1990. Solubilization of protein aggregates. *Methods Enzymol 182*:264–276.

Matthews CR. 1993. Pathways of protein folding. *Annu Rev Biochem 62*:653–683.

Miozzari GF, Yanofsky C. 1978. Translation of the leader region of the *Escherichia coli* tryptophan operon. *J Bacteriol 133*:1457–1466.

Naderi HM, Thomason JF, Borgias BA, Anderson S, James TL, Kuntz ID. 1991. ^{1}H NMR assignments and three-dimensional structure of Ala 14/Ala 38 bovine pancreatic trypsin inhibitor based on two-dimensional NMR and distance geometry. In: Nall BT, Dill KA, eds. *Conformations and forces in protein folding*. Washington, D.C.: American Association for the Advancement of Science. pp 86–114.

Norwood TJ, Boyd J, Heritage JE, Soffe N, Campbell ID. 1990. Comparison of techniques for ^{1}H-detected heteronuclear ^{1}H-^{15}N spectroscopy. *J Magn Reson 87*:488–501.

Oas TG, Kim PS. 1988. A peptide model of a protein folding intermediate. *Nature 336*:42–48.

Otting G, Widmer H, Wagner G, Wüthrich K. 1986. Origin of t_1 and t_2 ridges in 2D NMR spectra and procedures for suppression. *J Magn Reson 66*:187–193.

Piantini U, Sørensen OW, Ernst RR. 1982. Multiple quantum filters for elucidating NMR coupling networks. *J Am Chem Soc 104*:6800–6801.

Rance M. 1987. Improved techniques for homonuclear rotating-frame and isotropic mixing experiments. *J Magn Reson 74*:557–564.

Rance M, Sørensen OW, Bodenhausen G, Wagner G, Ernst RR, Wüthrich K. 1983. Improved spectral resolution in COSY ^{1}H NMR spectra of proteins via double quantum filtering. *Biochem Biophys Res Commun 117*:479–485.

Richardson JS. 1981. The anatomy and taxonomy of protein structure. *Adv Protein Chem 34*:167–339.

Roder H, Elöve GA, Englander SW. 1988. Structural characterization of folding intermediates in cytochrome *c* by H-exchange labeling and proton NMR. *Nature 335*:700–704.

Rose GD. 1979. Hierarchic organization of domains in globular proteins. *J Mol Biol 134*:447–470.

Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: A laboratory manual*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA 74*:5463–5467.

Schellman CG. 1976. A proposed folding path for pancreatic trypsin inhibitor. *Fed Proc 35*:1716.

Shaka AJ, Freeman R. 1983. Simplification of NMR spectra by filtration through multiple-quantum coherence. *J Magn Reson 51*:169–173.

Staley JP. 1993. Structural studies of early intermediates in the folding pathway of bovine pancreatic trypsin inhibitor [thesis]. Cambridge, Massachusetts: Massachusetts Institute of Technology.

Staley JP, Kim PS. 1990. Role of a subdomain in the folding of bovine pancreatic trypsin inhibitor. *Nature 344*:685–688.

Staley JP, Kim PS. 1992. Complete folding of bovine pancreatic trypsin inhibitor with only a single disulfide bond. *Proc Natl Acad Sci USA 89*:1519–1523.

Stassinopoulou CI, Wagner G, Wüthrich K. 1984. Two-dimensional ^{1}H NMR of two chemically modified analogs of the basic pancreatic trypsin inhibitor. Sequence-specific resonance assignments and sequence location of conformation changes relative to the native protein. *Eur J Biochem 145*:423–430.

States DJ, Creighton TE, Dobson CM, Karplus M. 1987. Conformations of intermediates in the folding of the pancreatic trypsin inhibitor. *J Mol Biol 195*:731–739.

States DJ, Dobson CM, Karplus M, Creighton TE. 1984. A new two-disulphide intermediate in the refolding of reduced bovine pancreatic trypsin inhibitor. *J Mol Biol 174*:411–418.

Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW. 1990. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol 185*:60–89.

Tasayco ML, Carey J. 1992. Ordered self-assembly of polypeptide fragments to form native-like dimeric *trp* repressor. *Science 255*:594–597.

Tüchsen E, Woodward C. 1987. Assignment of asparagine-44 side-chain primary amide ^{1}H NMR resonances and the peptide amide N1H resonance of glycine-37 in basic pancreatic trypsin inhibitor. *Biochemistry 26*:1918–1925.

van Mierlo CPM, Darby NJ, Creighton TE. 1992. The partially folded conformation of the [30–51] intermediate in the disulfide folding pathway of bovine pancreatic trypsin inhibitor. *Proc Natl Acad Sci USA 89*:6775–6779.

van Mierlo CPM, Darby NJ, Keeler J, Neuhaus D, Creighton TE. 1993. Partially folded conformation of the [30–51] intermediate in the disulphide folding pathway of bovine pancreatic trypsin inhibitor. ^{1}H and ^{15}N resonance assignments and determination of backbone dynamics from ^{15}N relaxation measurements. *J Mol Biol 229*:1125–1146.

van Mierlo CPM, Darby NJ, Neuhaus D, Creighton TE. 1991a. Two-dimensional ^{1}H nuclear magnetic resonance study of the [5–55] single-disulphide folding intermediate of bovine pancreatic trypsin inhibitor. *J Mol Biol 222*:373–390.

van Mierlo CPM, Darby NJ, Neuhaus D, Creighton TE. 1991b. [14–38, 30–51] Double-disulphide intermediate in folding of bovine pancreatic trypsin inhibitor: A two-dimensional ^{1}H nuclear magnetic resonance study. *J Mol Biol 222*:353–371.

Wagner G, Braun W, Havel TF, Schaumann T, Gō N, Wüthrich K. 1987. Protein structures in solution by nuclear magnetic resonance and distance geometry: The polypeptide fold of the basic pancreatic trypsin inhibitor determined using two different algorithms, DISGEO and DISMAN. *J Mol Biol 196*:611–639.

Wagner G, Wüthrich K. 1982a. Amide protein exchange and surface conformation of the basic pancreatic trypsin inhibitor in solution. Studies with two-dimensional nuclear magnetic resonance. *J Mol Biol 160*:343–361.

Wagner G, Wüthrich K. 1982b. Sequential resonance assignments in protein ^{1}H nuclear magnetic resonance spectra. Basic pancreatic trypsin inhibitor. *J Mol Biol 155*:347–366.

Weissman JS, Kim PS. 1991. Reexamination of the folding of BPTI: Predominance of native intermediates. *Science 253*:1386–1393.

Weissman JS, Kim PS. 1992a. Kinetic role of nonnative species in the folding of bovine pancreatic trypsin inhibitor. *Proc Natl Acad Sci USA 89*:9900–9904.

Weissman JS, Kim PS. 1992b. The pro region of BPTI facilitates folding. *Cell 71*:841–851.

Williams JW, van Holde KE, Baldwin RL, Fujita H. 1958. The theory of sedimentation analysis. *Chem Rev 58*:715–806.

Wlodawer A, Walter J, Huber R, Sjölin L. 1984. Structure of bovine pancreatic trypsin inhibitor. Results of joint neutron and X-ray refinement of crystal form II. *J Mol Biol 180*:301–329.

Woodward CK. 1994. Hydrogen exchange rates and protein folding. *Curr Opin Struct Biol 4*:112–116.

Wu LC, Laub PB, Elöve GA, Carey J, Roder H. 1993. A noncovalent peptide complex as a model for an early folding intermediate of cytochrome *c*. *Biochemistry 32*:10271–10276.

Wüthrich K. 1986. *NMR of proteins and nucleic acids*. New York: John Wiley & Sons.

Zuiderweg ERP. 1990. A proton-detected heteronuclear chemical-shift correlation experiment with improved resolution and sensitivity. *J Magn Reson 86*:346–357.