doi:10.1006/jmbi.2000.3622 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2000) 298, 493–501

JMB

Folding of Bovine Pancreatic Trypsin Inhibitor (BPTI) Variants in which Almost Half the Residues are Alanine

Yutaka Kuroda and Peter S. Kim*

Howard Hughes Medical Institute, Whitehead Institute for Biomedical Research Department of Biology, MIT Cambridge, MA 02142, USA

Recent studies indicate that a fraction of the information contained in an amino acid sequence may be sufficient for specifying a native protein structure. An earlier alanine-scanning experiment conducted on bovine pancreatic trypsin inhibitor (BPTI; 58 residues) suggested that if cumulative mutations have additive effects on protein stability, a native protein structure could be built from BPTI sequences that contained many alanine residues distributed throughout the protein. To test this hypothesis, we designed and produced six BPTI mutants containing from 21 to 29 alanine residues. We found that the melting temperature of mutants containing up to 27 alanine residues (48% of the total number of residues) could be predicted quite well by the sum of the change in melting temperature for the single mutations. Additionally, these same mutants folded into a native-like structure, as judged by their cooperative thermal denaturation curves and heteronuclear multiple quantum correlation (HMQC) NMR spectra. A BPTI mutant containing 22 alanine residues was further shown by 2D and 3D-NMR to fold into a structure very similar to that of native BPTI, and to be a functional trypsin inhibitor. These results provide insight into the extent to which native protein structure and function can be achieved with a highly simplified amino acid sequence.

Keywords: protein structure; alanine-scanning; information degeneracy; NMR

*Corresponding author

Introduction

Information degeneracy in an amino acid sequence (Bowie *et al.*, 1990; Cordes *et al.*, 1996) can be broken down into two types, relating to the base and the exponent of the equation used to calculate the number of possible amino acid sequences of a given length. The equation is: number of possible sequences = (residue type)$^{\text{number of residues}}$. The base in this equation is usually taken to be 20, for the 20 naturally occurring amino acids, but there is degeneracy inherent in the overlapping structural characteristics of these 20 amino acids. The exponent is given by the length of the polypeptide, but there is degeneracy here as well, since some sites along the amino acid sequence do not specify the protein structure. For the purposes of this discussion, we define these two types of information degeneracy as class redundancy and site redundancy, respectively.

Examples of substantial class redundancy have been reported by Gassner *et al.* (1996), Riddle *et al.* (1997) and Schafmeister *et al.* (1997). CD and NMR studies indicate that the small β-sheet SH3 domain can be built largely from five representative amino acid residues (Riddle *et al.*, 1997). In another study, the X-ray crystal structure of a *de novo* designed protein containing a reduced alphabet of seven amino acid residues was shown to adopt a native four-helix bundle structure (Schafmeister *et al.*, 1997). Examples of site redundancy are provided in a report by Brown & Sauer (1999) wherein the researchers designed Arc-repressor mutants in

Present address: Y. Kuroda, Genomic Sciences Center, Protein Research Group, RIKEN, The Institute for Physical and Chemical Research, Hirosawa, 2-1, Wako, Saitama, 351-0198 Japan.

Abbreviations used: BAPA, N^{α}-benzoyl-D,L-arginine *p*-nitroanilide; BPTI, bovine pancreatic trypsin inhibitor; CD, circular dichroism; DQF-COSY, double quantum filtered correlation spectroscopy; HMQC, heteronuclear multiple quantum correlation; HSQC, heteronuclear single quantum correlation; NOESY, nuclear Overhauser effect spectroscopy; T_m, midpoint temperature; TOCSY, total correlation spectroscopy.

E-mail address of the corresponding author: tocio@wi.mit.edu

0022-2836/00/030493–9 $35.00/0

which three to 15 residues (out of 53 residues) were simultaneously changed to alanine residues. The mutant proteins had native properties as judged by cooperative CD melts and DNA binding activity. These reports indicate that a significant amount of information degeneracy is present in an amino acid sequence, and stress the need for further study of both class and site redundancy.

Site redundancy, however, is less accessible to study than class redundancy, as it is difficult to remove structural information from specific sites in an amino acid sequence. The ideal procedure for doing this would be to replace residues at target positions with a ''structurally neutral'' building block, one that maintains the amino acid sequence length without directly assisting in determining the protein structure. In many reports, the alanine residue has emerged as the best representative of such a building block (Cunningham & Wells, 1989). The reasons underlying the choice of alanine for this purpose are the same as those invoked in alanine-scanning experiments: the alanine side-chain is small, which avoids steric clashes, and alanine does not introduce the flexibility of a glycine residue.

The conceptual classification of class and site redundancy may sometimes be equivocal, since some studies of information degeneracy in amino acid sequences may concern both types of degeneracy. For example, a stretch of ten residues corresponding to an α-helix in phage T4 lysozyme can be changed to alanine residues and the protein will retain the same overall fold as the wild-type protein (Heinz *et al*., 1992; Matthews, 1995). In another experiment, ten residues in the hydrophobic core of T4 lysozyme were replaced by methionine residues (Gassner *et al*., 1996). Although the fraction of amino acids changed to methionine residues is only 6 %, this experiment is particularly interesting because the replaced residues are located in the protein core, yet the protein's backbone structure remains essentially the same as that of native T4 lysozyme.

Here, we investigate the amount of site redundancy present in the BPTI sequence. BPTI has served as a model system for the analysis of many aspects of protein structure, and our mutants are derived from a previously constructed single disulfide bond variant of BPTI, called $[5\text{-}55]_{BPTI}$ (previously referred as $[5\text{-}55]_{Ala}\Delta M0/M52L$; Yu *et al*., 1995). This variant has a single disulfide bond between residues 5 and 55, and four cysteine residues mutated to alanine residues. A closely related protein called $[5\text{-}55]_{Ala}$, which differs from $[5\text{-}55]_{BPTI}$ in that it has an additional N-terminal Met residue and contains $Me^{+}52$ instead of Leu, has been shown to have essentially the same structure as wild-type BPTI (Staley & Kim, 1992). Based on the assumption that alanine is a good model for a structurally neutral building block, and on the results of a previous alanine-scanning experiment (Yu *et al*., 1995), we made and analyzed six variants of $[5\text{-}55]_{BPTI}$ containing from 21 to 29 alanine residues. We found that the effect of multiple substitutions on the midpoint temperature is additive over a large number of substitutions, and that the level of site redundancy is remarkably high: we observed a native BPTI-like fold for a sequence containing up to 27 alanine residues.

Results and Discussion

Design and characterization of mutant sequences

Site redundancy within the BPTI sequence was assessed by constructing six BPTI mutants containing 21, 22, 26, 27, 28 and 29 alanine residues, referred to as BPTI-21, -22, -26, -27, -28 and -29, respectively. These variant BPTI sequences (Figure 1) were designed based on the results of a previous alanine-scanning experiment for $[5\text{-}55]_{BPTI}$ (Yu *et al*., 1995). As $[5\text{-}55]_{BPTI}$ already contains six natural alanine residues plus four cysteine-replacing alanine residues, we designed the BPTI-22 sequence by simultaneously changing all

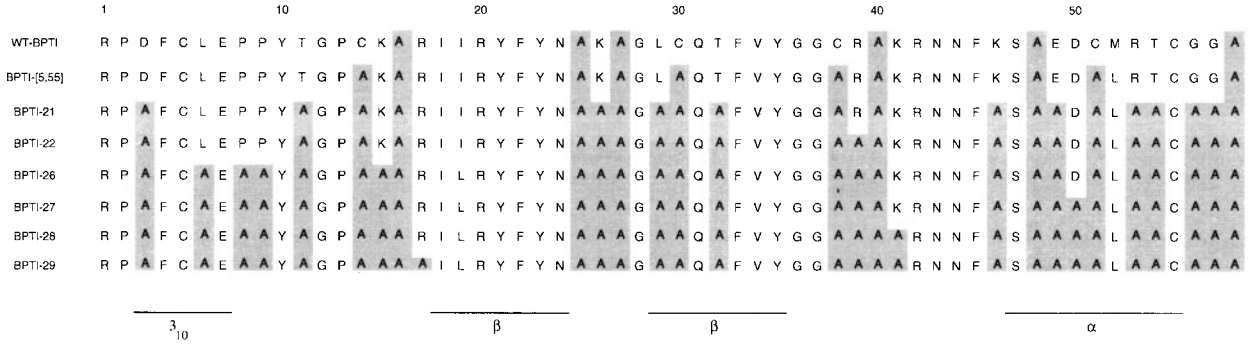

Figure 1. Sequences of variant BPTIs. The BPTIs are named after the number of alanine residues in their sequences. The alanine residues are emphasized by a shaded pattern. The wild-type BPTI contains six alanine residues and three disulfide bonds (5-55, 14-38, and 30-51). $[5\text{-}55]_{BPTI}$ is a single disulfide bond BPTI variant in which cysteine residues 14, 30, 38 and 51 are replaced by alanine and, in order to avoid cleavage by cyanogen bromide, M52 is replaced by a leucine residue. The secondary structures were determined by DSSP (Kabsch & Sander, 1983) for the native BPTI (PDB file; 5pti).

12 of the least destabilizing $[5\text{-}55]_{BPTI}$ residues to alanine. BPTI-21 was later derived from BPTI-22 by restoring R39 in the secondary active-site loop (Perona *et al*., 1993) in order to regain the protein's full trypsin-inhibitory activity. For BPTI-26, in addition to exploiting the results of the alanine-scanning experiment (Yu *et al*., 1995), we constrained the number of alanine residues placed in a row to three, since alanine has a strong tendency to form helices (Heinz *et al*., 1992; Padmanabhan *et al*., 1990).

The final BPTI-29 sequence, in which half of the residues are alanine, was designed according to the following three criteria: (i) no more than four alanine residues are present in a row; (ii) at least four charged residues are left to maintain the protein's solubility; and (iii) the least destabilizing alanine replacements are used. The most destabilizing replacements in this final sequence are P9A and L6A, which decrease the melting temperature of $[5\text{-}55]_{BPTI}$ by 5.5 and 4.6 deg. C, respectively (Yu *et al*., 1995). BPTI-27 and -28 were generated as a result of this strategy for constructing the final BPTI-29 sequence, and are included here for completeness.

Sedimentation equilibrium data indicate that all of the mutants are monomeric (Table 1). For each of the four mutants that contain up to 27 alanine residues, we observe a cooperative and reversible (>95%) thermal denaturation curve (Figure 2(a)), and a sharp and dispersed 1D-NMR spectrum (data not shown). Also, the HMQC spectra show characteristic peaks that strongly suggest an overall native-like fold for BPTI-21 through to BPTI-27 (Figure 2(b)).

BPTI-28 and BPTI-29, on the other hand, do not produce native-like conformations. BPTI-28 has characteristics intermediate between a native state and a molten globule state, while BPTI-29 exemplifies a molten globule (Ohgushi & Wada, 1983; Kuwajima, 1989; Ptitsyn, 1992; Peng *et al*., 1995). Although the 1D-NMR spectrum of BPTI-28 has many features reminiscent of a folded protein, and many cross peaks in the HMQC spectrum are sharp and well dispersed (Figure 2(b)), the thermal denaturation curve is not cooperative (Figure 2(a)). Furthermore, some minor peaks, probably originating from the presence of multiple backbone conformations, are observed in the HMQC spectrum. The thermal denaturation curve of BPTI-29 is also not cooperative (Figure 2(a)), and both the ^{1}H-NMR and HMQC (Figure 2(b)) spectra are qualitatively similar to those of a denatured protein.

As a further test of a native-like BPTI structure, we assessed the trypsin-inhibitory activity of the mutants. BPTI-21, which maintains all of the key residues important for trypsin binding, is fully inhibitory (100% compared to the wt-BPTI activity). BPTI-22, which lacks R39, an arginine residue involved in two hydrogen-bonding interactions with the trypsin residue backbone (Perona *et al*., 1993), has 60% activity at a 1:1 stoichiometry and a protein concentration of 280 nM; however, at a 1:2 stoichiometry, BPTI-22 completely blocks trypsin activity. Although we cannot rule out an induced-fit conformational change in the BPTI variants upon binding to trypsin, these inhibition studies provide additional support for the conclusion that BPTI-21 has a native-like BPTI structure. We speculate that the removal of the R39 side-chain results in a lower binding affinity due to the loss of a hydrogen bond between R39 and the main-chain carbonyl group of the trypsin residue N97 (Rühlmann *et al*., 1973). BPTI-26 has no trypsin-inhibitory activity, presumably because this mutant lacks not only R39 but also K15, which is buried in the trypsin-specificity pocket of the trypsin/BPTI complex (Rühlmann *et al*., 1973).

Additivity of the melting temperature changes

Remarkably, the melting temperatures of the $[5\text{-}55]_{BPTI}$ variants containing multiple alanine residue substitutions can be predicted by the sum of the change in melting temperatures for the single mutations (Table 1). While melting temperatures do not represent a direct quantitation of the unfolding free energy, additivity of melting temperature changes is expected when the unfolding entropy changes (ΔS) of the variants are the same,

Table 1. Melting temperatures and oligomerization states of the BPTI mutants

	Mol mass ratio [a] (M_{meas}/M_{exp})	T_m (°C)[b]	Predicted T_m (°C)[c]
$[5\text{-}55]_{BPTI}$	0.96	40	-
BPTI-21	1.00	38	36.7
BPTI-22	0.87	39	36.8
BPTI-26	0.96	20	20.6
BPTI-27	1.02	20	18.0
BPTI-28	1.07	-	15.0
BPTI-29	1.10	-	13.1

[a] The molecular masses in solution (M_{meas}) were determined by analytical ultra-centrifugation. M_{exp} is the expected molecular mass as calculated from the amino acid sequence.

[b] The midpoint temperature, T_m, is determined as the maximum of the derivative of the denaturation curve monitored by CD at 222 nm. The error is estimated to be ±1 deg. C. No T_m values were determined for BPTI-28 and BPTI-29 because their melts were not cooperative.

[c] Predicted T_m values were calculated from the melting temperature of $[5\text{-}55]_{BPTI}$ by assuming an additive effect for the simultaneous replacements. The melting temperatures for the single mutations are taken from Yu *et al*. (1995) and the error is estimated to be ±0.2 deg. C.

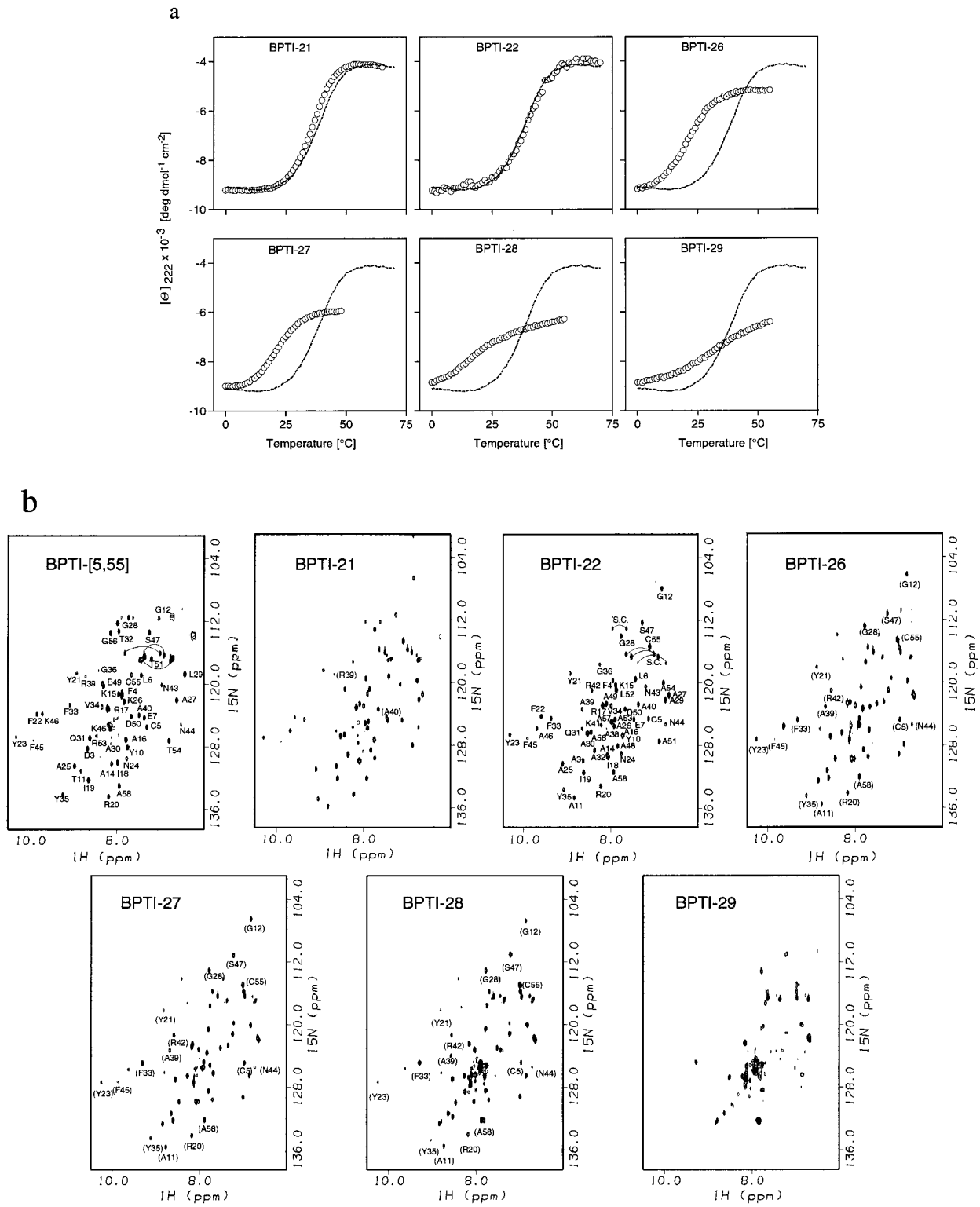

Figure 2. (a) Denaturation curves monitored by CD at 222 nm. In each panel, the open circles represent the denaturation curves of the respective mutant. For comparison, the denaturation curve of the reference protein, $[5\text{-}55]_{BPTI}$, is shown with a broken line. (b) HMQC spectra of the variant BPTIs at pH 4.7, 4 °C. The assignments for $[5\text{-}55]_{BPTI}$ are transferred from those of $[5\text{-}55]_{Ala}$ (B.A. Schulman & P.S.K., unpublished results). The assignments for BPTI-22 were carried out without reference to the NMR data of $[5\text{-}55]_{Ala}$ or wt-BPTI. All residues have corresponding cross peaks visible in the HMQC spectrum of BPTI-22, except for G37, whose peak lies outside of the plotted region. R39 and A40 are the only peaks that are shifted by the A39R mutation, and they are indicated in parentheses in the BPTI-21 spectrum. Tentative assignments of peaks readily transferred from BPTI-22 to BPTI-26 and BPTI-27 are shown in parentheses.

and the change in unfolding enthalpy changes ($\Delta\Delta H$) introduced by the combined mutations is equal to the sum of that for the single mutations. Such additivity has been observed previously for a smaller number of combined mutations (Kimura *et al.*, 1992; Blaber *et al.*, 1995; Zhang *et al.*, 1995), and additivity might be expected for multiple mutations exposed on the protein surface. The three most buried residues replaced in our study have relative accessibilities (Ooi *et al.*, 1987) above 37 % (P9, T54 and G56). Deviations from additivity have been observed in multiple-alanine mutants of Arc repressor in which some of the mutated residues are over 85 % buried in the native protein (Brown & Sauer, 1999).

BPTI-like structure of BPTI-22

We carried out an NMR study of BPTI-22, which is well structured, as demonstrated by its HMQC spectrum, and is sufficiently soluble for thorough NMR experiments. Resonance assignments for BPTI-22 were obtained using 2D-NMR and ^{15}N-edited 3D-NMR experiments, without reference to the NMR assignments of $[5\text{-}55]_{Ala}$ or wt-BPTI. A strong correlation between both the ^{15}N and the ^{1}H chemical shifts of the non-mutated residues of BPTI-22 and $[5\text{-}55]_{BPTI}$ suggests that these two proteins have very similar backbone structures (Figure 3(a)). The sequential and medium-range NOEs indicate that BPTI-22 has the same secondary structure elements as wild-type BPTI (Figure 3(b)). The structural similarity is further substantiated by the identification of 40 medium and long-range NOEs in BPTI-22 that are consistent with a BPTI-like structure (Figure 3(c) and Table 2). The NMR spectra, partial trypsin-inhibitory activity in BPTI-22 and full trypsin-inhibitory activity in the R39-containing BPTI-21, are strong indications for BPTI-like native structure in BPTI-22.

Native to molten globule state threshold

Although mutants with up to 27 alanine residues appear to be native-like, the structures of BPTI-26 and BPTI-27 could not be directly assessed by NMR because of their low melting temperatures, limited solubility, and the large number of alanine residues which introduce ambiguities in the sequential assignment procedure (Wüthrich, 1986). However, the significant overlap of the peaks in the HMQC spectra of BPTI-26 and BPTI-27 with those of BPTI-22 (Figure 2(b)) suggests that the overall structures of these molecules are similar.

By examining a series of mutants containing an increasing number of alanine residue replacements, we were able to detect the fraction of alanine content at which the characteristics for the native BPTI state disappear. The transition from a structurally unique state to a molten globule state occurs in BPTI-28. The variant with one additional alanine residue, BPTI-29, has a 50 % alanine content: this variant has characteristics typical of a molten globule state (Ohgushi & Wada, 1983; Kuwajima, 1989; Ptitsyn, 1992; Peng *et al.*, 1995). In our studies, both the relatively low solubility and high helical propensity of alanine were considered when designing sequences, and these considerations probably decrease the maximum number of alanine residues that can be contained in a BPTI-like foldable sequence. Thus the native to molten globule threshold occurs at approximately 50 % alanine content. More generally, the threshold number will depend on the protein's shape and size. For example, as larger proteins have a greater proportion of their residues buried in their core, they are therefore expected to have a lower threshold.

Information degeneracy experiments with molten globules provide insight into how the low-resolution backbone fold is specified. The observation of a molten globule state in BPTI-29 is similar to previous findings which indicate that the overall fold of a molten globule is defined by a small fraction of the information contained in the amino acid sequence (Kamtekar *et al.*, 1993; Davidson & Sauer, 1994; Davidson *et al.*, 1995; Wu & Kim, 1997). For example, the helical domain of α-lactalbumin, in which all the hydrophobic residues were replaced by leucine residues, maintains a native-like global fold (Wu & Kim, 1997). In another example, a homeobox domain, in which almost half of the residues were simultaneously changed to alanine residues, was demonstrated to be a molten globule

Table 2. Representative non-sequential NOEs

Structure	Proton pair	Distance in WT BPTI (Å)
3_{10} Helix	3Hα-5NH	3.72
	3Hα-6NH	3.73
α-Helix	49Hα-52NH	3.65
	50Hα-53NH	3.71
	52Hα-55NH	3.59
	52Hα-56NH	2.76
β-Sheet	18NH-35NH	3.03
	20NH-33NH	2.59
	20NH-35NH	2.67
	22NH-31NH	3.83
	21Hδ-30Hα	4.16
	21Hδ-31Hα	3.33
	21Hδ-32Hα	4.16
	21Hδ-33Hα	3.75
Other long-range	10Hδ-40Hα	3.17
	11Hα-36NH	3.00
	21Hδ-47Hα	3.00
	21Hδ-48Hα	3.98
	22NH-31Hα	3.75
	22Hα-45NH	3.97
	45NH-21NH	2.82

The underscored pairs are represented in Figure 3(c). Distances were calculated from the crystal structure of the wild-type BPTI (Wlodawer *et al.*, 1984). The shortest distances are listed for prochiral and methyl group protons.

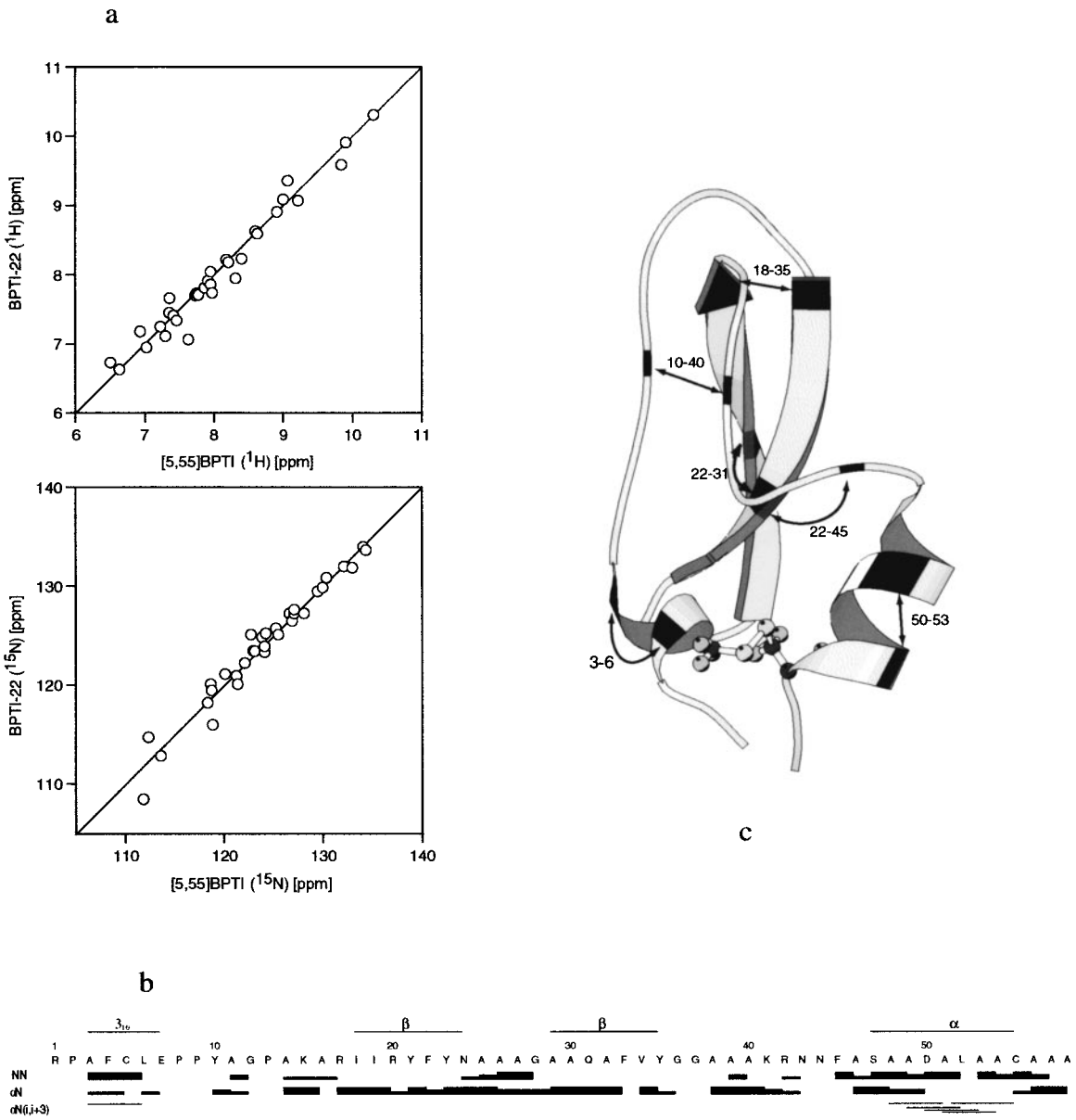

Figure 3. (a) Top; Correlation plot between the ^{1}H chemical shifts of $[5\text{-}55]_{BPTI}$ and BPTI-22. Bottom; ^{15}N chemical shifts of $[5\text{-}55]_{BPTI}$ and BPTI-22 at pH 4.7 and 4 °C. L52 and residues mutated to alanine are not included. For comparison, a line with a slope of one is shown on each panel. (b) Summary of the sequential and medium range NOE connectivities. The sequential NOEs are classified as strong, medium, and weak, and are represented by thick, medium, and thin lines, respectively. A30 and N43 NH-α were not observed because of an overlap with the water signal that is suppressed by presaturation. For reference, the secondary structures observed in the wild-type BPTI (Kabsch & Sander, 1983; Wlodawer *et al.*, 1984) are indicated. (c) Graphical representation (Kraulis, 1991) of the wild-type BPTI backbone with some representative NOEs observed in BPTI-22 indicated by arrows.

that retains residual DNA binding activity (Shang *et al.*, 1994).

Conclusions

Our results indicate that a native BPTI structure can be specified by a simplified sequence in which approximately half of the residues are alanine. The progressively reduced stability observed with multiple alanine mutations indicates a threshold for the transition between a native state to a molten globule state. This result should help in understanding which residues are critical in determining a native protein structure. Furthermore, the additivity observed for the destabilizing effect over a large number of alanine residue substitutions should be useful in the design of mutants with multiple surface residue substitutions.

Materials and Methods

Plasmid construction

A synthetic DNA oligomer coding for BPTI-22 was ligated into the *Hin*dIII/*Bam*HI site of pMMHa to create the pMMHaBPTI-22 plasmid. The pMMHa plasmid, kindly provided by M. Milhollen, was derived from the one used to produce $[5\text{-}55]_{Ala}\Delta M0/M52L$ (Yu *et al.*, 1995; referred to as $[5\text{-}55]_{BPTI}$ here) by adding a His-tag sequence at the N-terminal end of the trpΔLE (Staley & Kim, 1994) leader sequence. Other mutations were introduced by single-stranded mutagenesis (Kunkel, 1985) of the pMMHaBPTI-22 plasmid and verified by DNA sequencing.

Protein expression and purification

BPTI variants were expressed as described by Yu *et al.* (1995), except that carbenicillin was used in place of ampicillin. Inclusion bodies of the expressed proteins were solubilized in 6 M GuHCl and purified with a nickel column. The trpΔLE leader sequence and the His-tag sequence were cleaved with cyanogen bromide and removed by a second passage through the nickel column. The proteins were further purified by reverse phase HPLC, and their identities were confirmed by MALDI-TOF mass spectroscopy (Perceptive Biosystem; Voyager Elite). The calculated masses for $[5\text{-}55]_{BPTI}$, and BPTI-21 through to BPTI-29 are, respectively: 6372, 5967, 5880, 5728, 5682, 5602 and 5544 Da. The measured and calculated molecular masses agreed within 0.1 %.

Equilibrium sedimentation

Measurements were performed in a Beckman XL-A analytical centrifuge. Data were collected at 4 °C using two rotor speeds (35 and 40 krpm), three wavelengths (222, 243 and 282 nm), and three protein concentrations (20, 50 and 100 μM in 20 mM acetate buffer, pH 4.7). Values for the density of solutions and the partial specific volumes of protein species were calculated according to the solvent and amino acid compositions (Laue *et al.*, 1992). For all variants, the data fit an ideal monomeric model of ln (absorbance) *versus* radial distance squared, with no systematic deviation in the residuals.

Circular dichroism spectroscopy

Measurements were made using a 1 cm optical path length cuvette with an Aviv 62DS CD spectrometer equipped with a temperature-control unit. Proteins were dissolved in 20 mM acetate buffer (pH 4.7) to a concentration of 5 μM, as determined by absorbance at 275 nm in 6 M GuHCl, 20 mM sodium phosphate (pH 6.5), using an extinction coefficient based on tryptophan, tyrosine, and disulfide content (Edelhoch, 1967). For the thermal denaturation experiments, the temperature was raised at a rate of 1.5 deg. C min^{-1}. Thermal melts were reversible with superimposable folding and unfolding curves (95 %).

NMR spectroscopy

All samples, devoid of both buffer and salt, were prepared by dissolving lyophilized proteins in H_2O or ^2H_2O, and pH adjusted to 4.8 by titrating with small aliquots of 10 mM NaOH and/or HCl. Protein concentrations varied between 300 μM and 600 μM. Data were collected on a Bruker AMX 500-MHz spectrometer, and processed with FELIX95 software (Biosym). Prior to Fourier transform, 90° shifted sinebell windows were applied in both the t_1 and t_2 directions for the HMQC spectra. The 3D-NOESY-HSQC and 3D-TOCSY-HSQC (Fesik & Zuiderweg, 1988; Marion *et al.*, 1989a,b) were measured with 200 × 512 × 32 points in the t_1, t_2 and t_3 directions, respectively. For the 3D experiments, the t_1 direction was zero-filled to 256 points, and linear prediction was used to increase the data points in the t_3 direction to 64 points.

Resonance assignments for BPTI-22 were performed using DQF-COSY, 2D-NOESY (100 and 150 ms), 2D-TOCSY (65 and 75 ms, at 4 °C and 17 °C), 3D-NOESY-HSQC (150 ms) and 3D-TOCSY-HSQC (75 ms) at 4 °C, without reference to the NMR data of $[5\text{-}55]_{Ala}$ or wt-BPTI.

Trypsin inhibition

Inhibition of trypsin activity was assayed at 15 °C by monitoring the hydrolysis of N^{α}-benzoyl-D,L-arginine *p*-nitroanilide (BAPA) in 0.2 M phosphate buffer (pH 7.0) with 0.1 M KCl, at a trypsin concentration of 280 nM (Kassell, 1970). Inhibitor and trypsin were incubated in buffer for five minutes before the addition of BAPA, and the hydrolysis of BAPA was monitored by measuring absorbance changes over ten minutes at 405 nm.

Acknowledgments

We thank Michael A. Milhollen for providing the pMMHa plasmid, Peter Schmidt, Hiroshi Matsuo (Harvard Medical School) and James C. McKnight (Boston University) for help with the NMR measurements, Brenda A. Shulman for communicating the chemical shift values of $[5\text{-}55]_{Ala}$, James X. Pang for help with mass spectroscopy, Alexandra Evindar for editorial assistance, Jodi M. Harris for graphics assistance, and Lawren C. Wu and other members of the Kim group for helpful discussions and comments on the manuscript. Y.K. was supported by a Human Frontier Science Program Fellowship. This research was supported by the Howard Hughes Medical Institute.

References

Blaber, M., Baase, W. A., Gassner, N. & Matthews, B. W. (1995). Alanine scanning mutagenesis of the alpha-helix 115-123 of phage T4 lysozyme: effects on structure, stability and the binding of solvent. *J. Mol. Biol.* **246**, 317-330.

Bowie, J. U., Reidhaar-Olson, J. F., Lim, W. A. & Sauer, R. T. (1990). Deciphering the message in protein sequences: tolerance to amino acid substitutions. *Science,* **247**, 1306-1310.

Brown, B. M. & Sauer, R. T. (1999). Tolerance of Arc repressor to multiple-alanine substitutions. *Proc. Natl Acad. Sci. USA,* **96**, 1983-1988.

Cordes, M. H., Davidson, A. R. & Sauer, R. T. (1996). Sequence space, folding and protein design. *Curr. Opin. Struct. Biol.* **6**, 3-10.

Cunningham, B. C. & Wells, J. A. (1989). High-resolution epitope mapping of hGH-receptor interactions by alanine- scanning mutagenesis. *Science,* **244**, 1081-1085.

Davidson, A. R. & Sauer, R. T. (1994). Folded proteins occur frequently in libraries of random amino acid sequences. *Proc. Natl Acad. Sci. USA,* **9**, 2146-2150.

Davidson, A. R., Lumb, K. J. & Sauer, R. T. (1995). Cooperatively folded proteins in random sequence libraries. *Nature Struct. Biol.* **2**, 856-864.

Edelhoch, H. (1967). Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry,* **6**, 1948-1954.

Fesik, S. W. & Zuiderweg, E. R. P. (1988). Heteronuclear three-dimensional NMR spectroscopy: a strategy for the simplification of homonuclear two-dimensional NMR spectra. *J. Magn. Reson.* **78**, 588-593.

Gassner, N. C., Baase, W. A. & Matthews, B. W. (1996). A test of the ''jigsaw puzzle'' model for protein folding by multiple methionine substitutions within the core of T4 lysozyme. *Proc. Natl Acad. Sci. USA,* **93**, 12155-12158.

Heinz, D. W., Baase, W. A. & Matthews, B. W. (1992). Folding and function of a T4 lysozyme containing 10 consecutive alanines illustrate the redundancy of information in an amino acid sequence. *Proc. Natl Acad. Sci. USA,* **89**, 3751-3755.

Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers,* **22**, 2577-2637.

Kamtekar, S., Schiffer, J. M., Xiong, H., Babik, J. M. & Hecht, M. H. (1993). Protein design by binary patterning of polar and nonpolar amino acids. *Science,* **262**, 1680-1685.

Kassell, B. (1970). Bovine trypsin-kallikrein inhibitor. *Methods Enzymol.* **19**, 844-852.

Kimura, S., Nakamura, H., Hashimoto, T., Oobatake, M. & Kanaya, S. (1992). Stabilization of *Escherichia coli* ribonuclease HI by strategic replacement of amino acid residues with those from the thermophilic counterpart. *J. Biol. Chem.* **267**, 21535-21542.

Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946-950.

Kunkel, T. A. (1985). Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl Acad. Sci. USA,* **82**, 488-492.

Kuwajima, K. (1989). The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. *Proteins: Struct. Funct. Genet.* **6**, 87-103.

Laue, T. M., Shah, B. D., Ridgeway, T. M. & Pelletier, S. L. (1992). *Computer-aided Interpretation and Analytical Sedimentation Data for Proteins* (Hardings, S. E., Rowe, A. J. & Horton, J. C., eds), The Royal Society of Chemistry, Cambridge, UK.

Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M. & Clore, G. M. (1989a). Overcoming the overlap problem in the assignment of 1H NMR spectra of larger proteins by use of three-dimensional heteronuclear 1H-15N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta. *Biochemistry,* **28**, 6150-6156.

Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A. & Bax, A. (1989b). Three-dimensional heteronuclear NMR of 15N labeled proteins. *J. Am. Chem. Soc.* **111**, 1515-1517.

Matthews, B. W. (1995). Studies on protein stability with T4 lysozyme. *Advan. Protein Chem.* **46**, 249-278.

Ohgushi, M. & Wada, A. (1983). ''Molten-globule state'': a compact form of globular proteins with mobile side-chains. *FEBS Letters,* **164**, 21-24.

Ooi, T., Oobatake, M., Nemethy, G. & Scheraga, H. A. (1987). Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides. *Proc. Natl Acad. Sci. USA,* **84**, 3086-3090.

Padmanabhan, S., Marqusee, S., Ridgeway, T., Laue, T. M. & Baldwin, R. L. (1990). Relative helix-forming tendencies of nonpolar amino acids. *Nature,* **344**, 268-270.

Peng, Z. Y., Wu, L. C., Schulman, B. A. & Kim, P. S. (1995). Does the molten globule have a native-like tertiary fold? *Phil. Trans. Roy. Soc. Lond. B. Biol. Sci.* **348**, 43-47.

Perona, J. J., Tsu, C. A., Craik, C. S. & Fletterick, R. J. (1993). Crystal structures of rat anionic trypsin complexed with the protein inhibitors APPI and BPTI. *J. Mol. Biol.* **230**, 919-933.

Ptitsyn, O. B. (1992). The molten globule state. In *Protein Folding* (Creighton, T. E., ed.), Freeman and Co. New York, USA.

Riddle, D. S., Santiago, J. V., Bray-Hall, S. T., Doshi, N., Grantcharova, V. P., Yi, Q. & Baker, D. (1997). Functional rapidly folding proteins from simplified amino acid sequences. *Nature Struct. Biol.* **4**, 805-809.

Rühlmann, A., Kukla, D., Schwager, P., Bartels, K. & Huber, R. (1973). Structure of the complex formed by bovine trypsin and bovine pancreatic trypsin inhibitor. Crystal structure determination and stereochemistry of the contact region. *J. Mol. Biol.* **77**, 417-436.

Schafmeister, C. E., LaPorte, S. L., Miercke, L. J. & Stroud, R. M. (1997). A designed four-helix bundle protein with native-like structure. *Nature Struct. Biol.* **4**, 1039-1046.

Shang, Z., Isaac, V. E., Li, H., Patel, L., Catron, K. M., Curran, T., Montelione, G. T. & Abate, C. (1994). Design of a ''minimAl'' homeodomain: the N-terminal arm modulates DNA binding affinity and stabilizes homeodomain structure. *Proc. Natl Acad. Sci. USA,* **91**, 8373-8377.

Staley, J. P. & Kim, P. S. (1992). Complete folding of bovine pancreatic trypsin inhibitor with only a single disulfide bond. *Proc. Natl Acad. Sci. USA,* **89**, 1519-1523.

Staley, J. P. & Kim, P. S. (1994). Formation of a native-like subdomain in a partially folded intermediate of bovine pancreatic trypsin inhibitor. *Protein Sci.* **3**, 1822-1832.

Wlodawer, A., Walter, J., Huber, R. & Sjolin, 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.

Wu, L. C. & Kim, P. S. (1997). Hydrophobic sequence minimization of the alpha-lactalbumin molten globule. *Proc. Natl Acad. Sci. USA,* **94**, 14314-14319.
Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids*, Wiley Interscience, New York, USA.
Yu, M. H., Weissman, J. S. & Kim, P. S. (1995). Contribution of individual side-chains to the stability of BPTI examined by alanine-scanning mutagenesis. *J. Mol. Biol.* **249**, 388-397.
Zhang, X. J., Baase, W. A., Shoichet, B. K., Wilson, K. P. & Matthews, B. W. (1995). Enhancement of protein stability by the combination of point mutations in T4 lysozyme is additive. *Protein Eng.* **8**, 1017-1022.

Edited by C. R. Matthews

(Received 29 November 1999; received in revised form 11 February 2000; accepted 11 February 2000)