Protein Science (1997), 6:1549–1562. Cambridge University Press. Printed in the USA.

Mutational analysis of the BPTI folding pathway: I. Effects of aromatic → leucine substitutions on the distribution of folding intermediates

JIAN-XIN ZHANG[1] AND DAVID P. GOLDENBERG
Department of Biology, University of Utah, Salt Lake City, Utah 84112

(RECEIVED February 24, 1997; ACCEPTED April 17, 1997)

Abstract

The roles of aromatic residues in determining the folding pathway of bovine pancreatic trypsin inhibitor (BPTI) were analyzed mutationally by examining the distribution of disulfide-bonded intermediates that accumulated during the refolding of protein variants in which tyrosine or phenylalanine residues were individually replaced with leucine. The eight substitutions examined all caused significant changes in the intermediate distribution. In some cases, the major effect was to decrease the accumulation of intermediates containing two of the three disulfides found in the native protein, without affecting the distribution of earlier intermediates. Other substitutions, however, led to much more random distributions of the intermediates containing only one disulfide. These results indicate that the individual residues making up the hydrophobic core of the native protein make clearly distinguishable contributions to conformation and stability early in folding: The early distribution of intermediates does not appear to be determined by a general hydrophobic collapse. The effects of the substitutions were generally consistent with the structures of the major intermediates determined by NMR studies of analogs, confirming that the distribution of disulfide-bonded species is determined by stabilizing interactions within the ordered regions of the intermediates. The plasticity of the BPTI folding pathway implied by these results can be described using conformational funnels to illustrate the degree to which conformational entropy is lost at different stages in the folding of the wild-type and mutant proteins.

Keywords: aromatic residues; bovine pancreatic trypsin inhibitor; conformational funnels; disulfide bonds; hydrophobic residues; protein folding

The identification and characterization of partially folded intermediate states have historically been major goals in the study of protein folding mechanisms. Early experiments demonstrating the transient accumulation of such species provided important evidence supporting the view that proteins fold via non-random pathways (Baldwin, 1975; Creighton, 1978). Intermediates have now been identified in the folding of many different proteins, and the structures of several intermediates have been characterized by high-resolution techniques (Kim & Baldwin, 1990; Matthews, 1993b; Dobson et al., 1994; Fersht, 1995; Creighton et al., 1996). Nonetheless, many questions remain about the nature of these species and their significance. Relatively little is known, for instance, about the factors that determine the conformations and stabilities of partially folded molecules, thereby causing some regions of a polypeptide to become ordered before others do. More fundamentally, it is not clear that the intermediates that are most easily detected experimentally actually promote formation of the fully folded molecule. Recent studies suggest, in fact, that folding is often fastest and most efficient when the accumulation of intermediates is minimized (Zhang & Goldenberg, 1993; Sosnick et al., 1994; Huang & Oas, 1995; Schindler et al., 1995).

Mutational analysis can provide important information about both the factors that stabilize particular folding intermediates and the roles of these intermediates. Amino acid replacements that destabilize folding intermediates can implicate specific residues and interactions in favoring formation of the intermediates (Goldenberg, 1988a, 1992a; Goldenberg et al., 1989; Fersht et al., 1992;

Reprint requests to: David P. Goldenberg, Department of Biology, University of Utah, Salt Lake City, UT 84112; e-mail: goldenberg@bioscience.utah.edu.

[1] Current address: Department of Pathology, Harvard University Medical School, Boston, MA 02115.

Abbreviations: BPTI, bovine pancreatic trypsin inhibitor. Amino acid replacements are indicated by the wild-type residue type (using the one-letter code for the 20 standard amino acids), followed by the residue number and the mutant residue type. The disulfides of native BPTI and folding intermediates are indicated by the residue numbers of the disulfide-bonded cysteine residues; N^{SH}_{SH}, native-like two-disulfide intermediate containing the 30–51 and 5–55 disulfides. Other intermediates are indicated by square brackets enclosing the disulfide bonds they contain; GSSG, oxidized glutathione; DTT^{SH}_{SH}, the dithiol form of dithiothreitol; GuHCl, guanidinium chloride; IAEDANS, *N*-iodoacetyl-*N'*-(5-sulfo-1-naphthyl)ethylenediamine; Tris-HCl, tris(hydroxymethyl)-aminomethane hydrochloride; EDTA, ethylenediamintetraacetic acid; HPLC, high performance liquid chromatography.

Matthews, 1993b; Fersht, 1995). By altering the distribution of intermediates during folding, mutations can also help resolve questions about whether a specific species promotes or hinders formation of the native protein (Kiefhaber et al., 1990; Zhang & Goldenberg, 1993).

One of the most detailed descriptions of a folding reaction presently available comes from the study of bovine pancreatic trypsin inhibitor (BPTI) (Creighton, 1978, 1990, 1992; Creighton & Goldenberg, 1984; Weissman & Kim, 1991; Goldenberg, 1992b). BPTI is a globular protein of 58 amino acids, stabilized by three disulfide bonds (Fig. 1A). Its folding is thermodynamically coupled with disulfide formation, making it relatively easy to trap, identify, and characterize disulfide-bonded intermediates in the refolding of the reduced and unfolded protein. A folding pathway for BPTI, in which intermediates are defined by the disulfides they contain, has been elucidated and is illustrated schematically in Figure 2. The kinetic roles of the various intermediates and the energetics of the pathway have been determined from extensive kinetic experiments, in which the rates of interconversion among the various species have been measured directly (Creighton, 1978; Creighton & Goldenberg, 1984; Goldenberg, 1988b; Darby et al., 1992, 1995; Weissman & Kim, 1992, 1995; Darby & Creighton, 1993; Dadlez & Kim, 1996). In addition, the conformations of the major intermediates have been characterized by high-resolution NMR spectroscopy or X-ray crystallography (Eigenbrot et al., 1990; Naderi et al., 1991; van Mierlo et al., 1991a, 1991b, 1993, 1994; Staley & Kim, 1992, 1994). Relatively little is known about the roles of individual residues other than the cysteines in defining the BPTI folding pathway, however, and there has been considerable controversy about the relative importance of some intermediates (Weissman & Kim, 1991, 1992; Creighton, 1992; Goldenberg, 1992b).

In this and the accompanying paper, we describe a mutational analysis designed to assess the roles in the folding mechanism of the residues that make up the hydrophobic core of native BPTI. The core is composed almost entirely of eight aromatic side chains: four phenylalanines (residues 4, 22, 33, and 45) and four tyrosines (residues 10, 21, 23, and 35). These residues are distributed throughout the sequence and the three-dimensional structure of the native protein (Fig. 1A). For each of the aromatic residues, a large fraction of the side chain is buried in the native protein, and these side chains make the largest individual contributions to the hydrophobic core (Fig. 1B). Collectively, they make up about 36% of the total buried non-polar surface area. In order to assess the contributions of the aromatic residues to folding and stability, eight BPTI variants, each with one of the residues replaced with Leu, were constructed, and the folding mechanisms of these variants were compared with that of the wild-type protein. Initial studies with two of these variants have been described in previous papers (Goldenberg et al., 1989; Zhang & Goldenberg, 1993).

This paper describes the effects of the aromatic → Leu substitutions on the distributions of disulfide-bonded intermediates detected during both folding and unfolding. Some of the substitutions were found to eliminate nearly all specificity in the formation of the disulfide-bonded intermediates, while other substitutions selectively decreased the accumulation of particular species, leading to a simplified distribution. Strikingly, however, all of the mutant proteins were able to fold efficiently, and some formed their three disulfides more rapidly than did the wild-type protein. As observed previously for the wild-type protein, the kinetically preferred folding pathway for each of the mutant proteins involves intramolecular rearrangements of intermediates containing two-

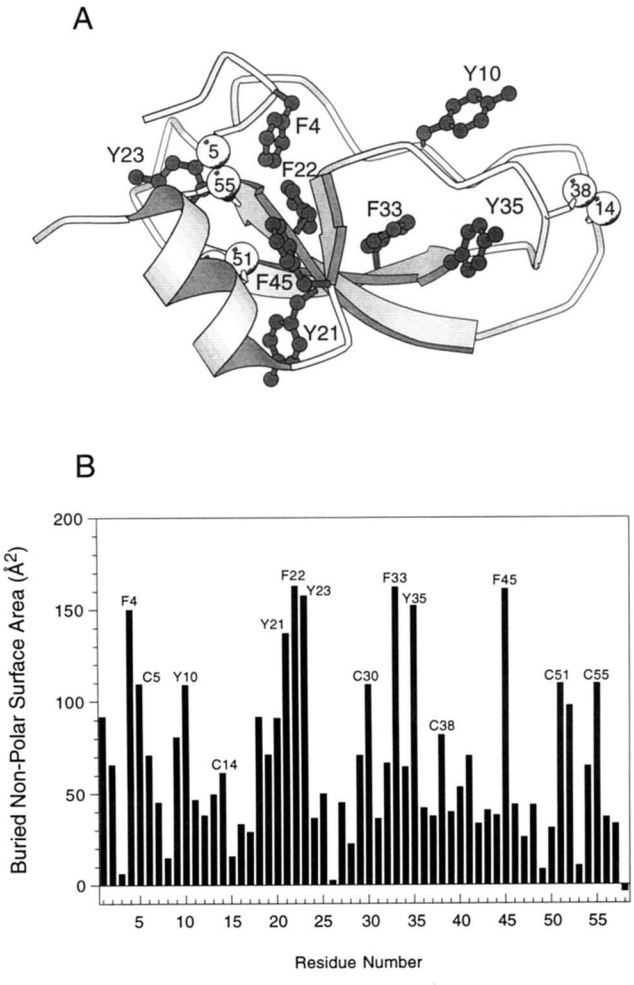

Fig. 1. Locations of the eight aromatic residues in native wild-type BPTI. **A:** Three-dimensional structure of wild-type BPTI, drawn from the atomic coordinates of the Form II crystals (entry 4pti in the Brookhaven Protein Data Bank; Wlodawer et al., 1987) using the program MOLSCRIPT (Kraulis, 1991). The side chains of the aromatic residues and the disulfide-bonded Cys residues are shown as balls and sticks. **B:** Buried non-polar surface area of residues in native BPTI. Buried surface area was calculated as the difference in accessible surface area in the native and unfolded protein, using the algorithm of Lee and Richards (1971), implemented in the program ACCESS by T.J. Richmond, and the group radii used by Chothia (1975). To approximate the solvent exposure of each residue in the unfolded polypeptide, accessibilities were calculated for Arg-Gly, Gly-X-Gly, and Gly-Ala peptides, corresponding to the N-terminal, internal, and C-terminal residues of BPTI, respectively. The peptide model for each residue was constructed with an extended backbone conformation (ϕ = −120°, ψ = 140°) and side-chain dihedral angles corresponding to the predominant rotomer conformation identified by Ponder and Richards (1987). All nitrogen and oxygen atoms were classified as polar, while carbon and sulfur atoms were classified as non-polar.

disulfide bonds. The results indicate that the accumulation of specific intermediates is quite sensitive to amino acid replacements, and that the various aromatic residues play clearly distinguishable roles in determining specificity in the formation of the disulfide-bonded intermediates. The ability of the protein to fold and some general features of the pathway are, however, quite tolerant of even those substitutions that have the most drastic effects on the intermediate distribution.

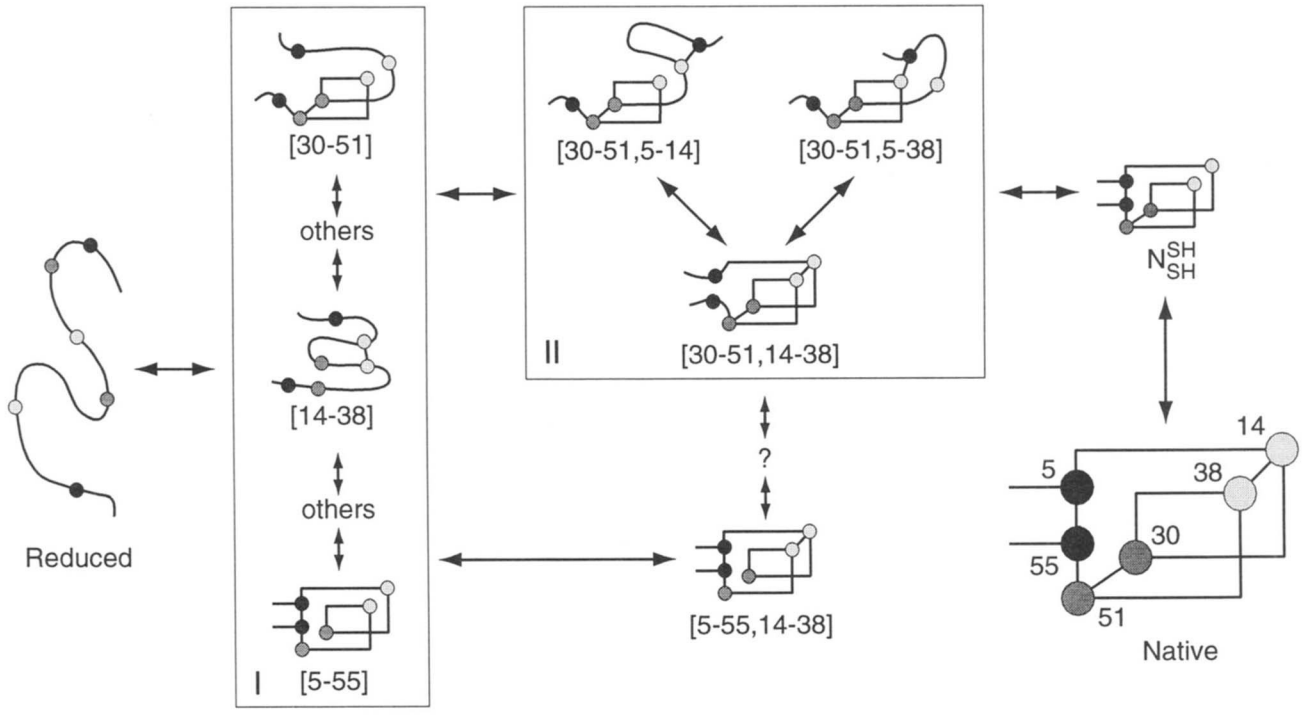

Fig. 2. The BPTI folding pathway. The fold of the native protein is indicated schematically, and the intermediates are drawn to show qualitatively the degree to which they contain native-like structure, as determined by NMR spectroscopy of analogs. Intermediates grouped together in the boxes labeled I and II interconvert rapidly on the time scale of the experiments described here. The major exception is [5–55], which interconverts more slowly than the other one-disulfide species. Adapted, with permission, from Figure 3 of Goldenberg (1992b).

Results

Design and initial characterization of aromatic → Leu BPTI variants

Leucine was chosen as a replacement for Phe and Tyr residues because this is a relatively conservative substitution but is likely to cause measurable stability differences. Although substitutions of aromatic residues with Leu are usually tolerated during evolution (Bordo & Argos, 1991), single replacements of this type have been found to destabilize folded proteins significantly (Kellis et al., 1988; Goldenberg et al., 1989; Coplen et al., 1990; Eriksson et al., 1993). The replacement decreases the size of the side chain and eliminates any interaction specific for aromatic rings. Although both Leu and the aromatic residues have branched γ-carbons, the geometry of the γ-carbon is different, leading to differences in the preferred side-chain dihedral angles observed in folded proteins (Ponder & Richards, 1987). As a consequence, the replacements with leucine may disrupt the close packing of the atoms around the substitution site.

Initial experiments with the mutant proteins indicated that each had an overall conformation very similar to that of the wild-type protein. All had electrophoretic mobilities on non-denaturing polyacrylamide gels that were indistinguishable from that of the wild-type protein, and they all inhibited trypsin stoichiometrically at μM concentrations.

To further characterize the structures of the native proteins, one-dimensional ^{1}H NMR spectra of the mutant and wild-type proteins were compared and found to share the wide chemical shift dispersion characteristic of the wild-type protein (Fig. 3). Similarities among the spectra were most apparent in the regions corresponding to backbone amide and aromatic side-chain protons. Resonances assigned previously to the Tyr 23 side chain, for example, were observed for all of the mutants except Y23L. Another distinguishing feature of the spectrum for the wild-type protein is the unusually large down-field chemical shift displayed by the amide proton of Tyr 23. This peak was clearly detected for all of the mutants except F22L and F45L, in which the altered side chains are positioned in close proximity to the Tyr 23 amide group. Since chemical shifts are highly dependent on protein conformation, the spectra suggest that the substitutions do not greatly alter the folded conformation.

Although more detailed analysis of the mutant proteins might well reveal local structural differences among the proteins, the many structural studies of mutant proteins suggest that a single replacement of a hydrophobic residue by another hydrophobic residue of only slightly smaller size is unlikely to cause atomic shifts of more than a few angstroms (Matthews, 1993a). Structures of five BPTI variants have been solved by X-ray crystallography, and are generally very similar to that of wild-type BPTI (Housset et al., 1991; Danishefsky et al., 1993). In some of these mutants, aromatic residues are replaced by Ala, resulting in crevices open to the exterior of the molecule and small rearrangements among the atoms lining the crevices (Danishefsky et al., 1993; Kim et al., 1993). One of the largest observed structural changes arising from an amino acid replacement in any protein is seen in the crystal structure of Y35G BPTI (Housset et al., 1991), in which backbone atoms near the altered residue are shifted by as much as 6 Å. Even for this mutant, however, the overall fold is similar to that of the

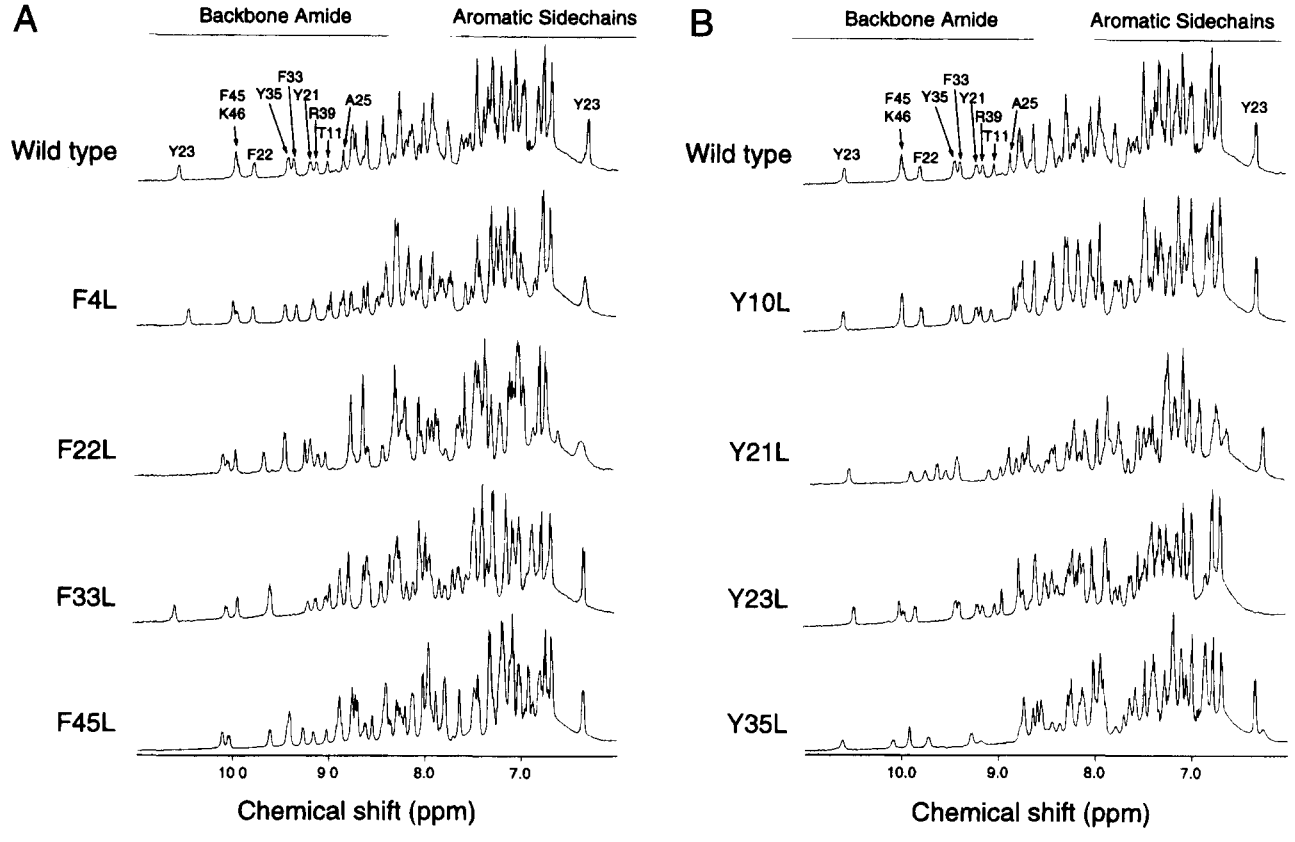

Fig. 3. Proton NMR spectra of the native forms of wild-type BPTI and the aromatic → Leu variants. Spectra were collected at pH 4.5 and room temperature as described in Materials and methods. The well-resolved amide proton resonances and that of the Tyr 23 side chain were identified from the published assignments for wild-type BPTI (Wagner et al., 1987a, 1987b).

wild-type protein. We thus believe that it is unlikely that the overall conformation of the protein is altered by the much more conservative aromatic → Leu substitutions, though details of packing at the atomic level may well be perturbed.

General effects of the aromatic → Leu substitutions on the distribution of folding intermediates

During the refolding of reduced wild-type BPTI, the distribution of disulfide-bonded intermediates is dominated by approximately six species, depending somewhat on pH (Creighton, 1974, 1975; Creighton & Goldenberg, 1984; Weissman & Kim, 1991) (Fig. 2). The major one-disulfide intermediate at pH 8.7 is [30–51], which represents approximately 60% of the one-disulfide species at steady state. At pH 7.3, [5–55] also accumulates significantly, with it and [30–51] each representing about 40% of the one-disulfide population. The major two-disulfide intermediates include each of the three possible species containing two of the three disulfides found in the native protein. In addition, two species each containing a non-native disulfide, [30–51,5–14] and [30–51,5–38], accumulate to detectable levels at pH 8.7, but not at pH 7.3. Recently, an additional one-disulfide intermediate, containing the native 14–38 disulfide, was identified in the refolding of BPTI (Dadlez & Kim, 1995, 1996; Ferrer et al., 1995). Although this species is formed quite rapidly, it also rearranges rapidly and is present at very low concentrations at steady-state (Dadlez & Kim, 1995, 1996). This intermediate was not detected in the experiments described here, in which only the steady-state distribution of one-disulfide intermediates was examined.

To examine the effects of the aromatic → Leu substitutions on the distributions of folding intermediates, the native mutant proteins were reduced and unfolded, and then refolded in the presence of oxidized glutathione (GSSG). The resulting disulfide-bonded intermediates were trapped by acidification and fractionated by reversed-phase HLPC, as shown in Figure 4. The numbers of disulfides in the isolated intermediates were determined by gel electrophoresis following reaction with iodoacetate, as described in Materials and methods, and the identities of the disulfides in the major intermediates (i.e., those representing more than about 5% of the total protein) were established by peptide mapping using the procedure of Weissman and Kim (1991).

Substitutions Y23L and F45L were found to broaden dramatically the distribution of folding intermediates. Although there were a few prominent peaks in the chromatograms, further analysis indicated that each of these peaks contained more than one species. These substitutions, therefore, caused a nearly uniform distribution of folding intermediates. Unlike the wild-type protein, the intermediate distributions for the Y23L and F45L variants were not affected detectably by the change in pH from 8.7 to 7.3.

In contrast to the broad distributions seen for the Y23L and F45L variants, a simplified intermediate distribution was observed for the folding of Y35L BPTI at both pH 8.7 and 7.3 (Fig. 4) (Zhang & Goldenberg, 1993). Similar, though less extreme, effects were seen for the remaining five mutants, F4L, Y10L, Y21L, F22L,

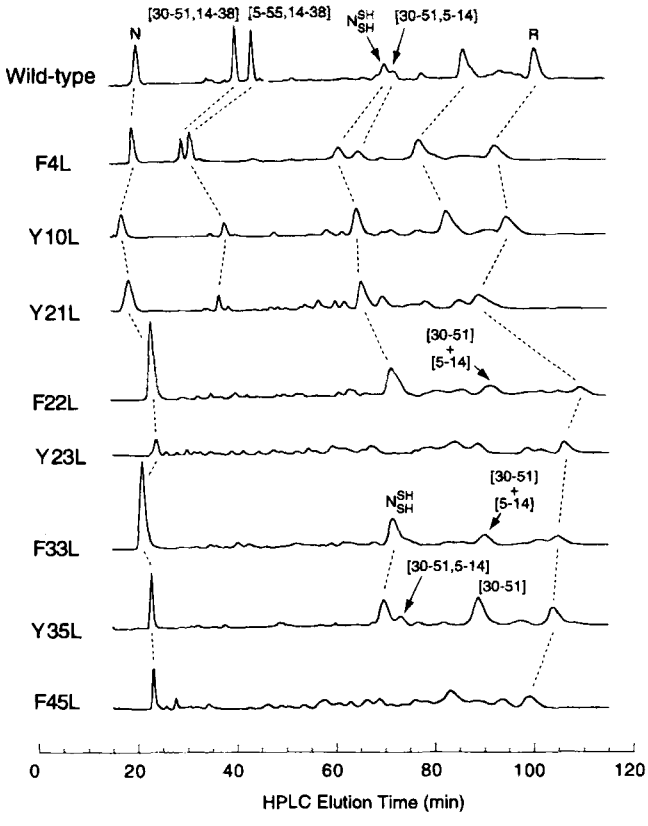

Fig. 4. Reversed-phase HPLC separation of intermediates trapped in the refolding of wild-type BPTI and the eight aromatic → Leu variants. Each protein was fully reduced and unfolded and then refolded in the presence of 0.1 mM GSSG for 5 min at pH 8.7, 25 °C. The reactions were quenched by the addition of formic acid, and the intermediates were fractionated by HPLC. The major intermediates were identified by peptide mapping, as described in the text.

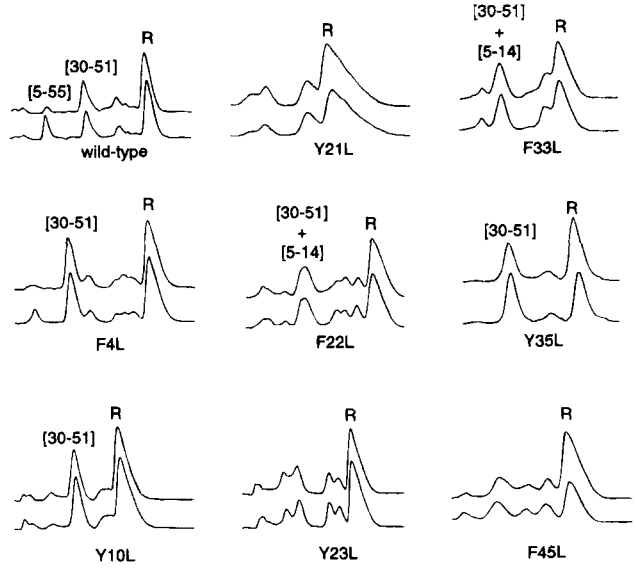

Fig. 5. Comparison of the one-disulfide intermediates detected at pH 8.7 (upper traces) and pH 7.3 (lower traces) during the refolding of wild-type BPTI and the aromatic → Leu variants. Folding experiments were carried out as described in the legend to Figure 4 and in Materials and methods. For the upper traces, refolding reactions were allowed to proceed at pH 8.7 for 2 min before trapping with formic acid. For the lower traces, the reactions were quenched after 20 min of folding at pH 7.3. Because the rates of the thiol-disulfide exchange reactions decrease at lower pH, the two sets of chromatograms represent approximately the same extent of reaction progress. Note the increased accumulation of [5–55] during the folding of the wild-type protein at the lower pH. In contrast, the intermediate profiles for the mutants other than F4L are largely unaffected by pH.

and F33L. For each of these mutants, there was a significant decrease in the accumulation of one or both of the native-like two-disulfide intermediates, [5–55,14–38] and [30–51,14–38]. For some of the mutants, especially Y21L and F22L, the decreased accumulation of these species was accompanied by the appearance of a large number of other species that presumably contain non-native disulfides.

Effects on the one-disulfide intermediates

The effects of the aromatic → Leu substitutions on the distributions of the one-disulfide intermediates are highlighted in Figure 5, where the regions of the chromatograms containing these early intermediates in the folding of the mutant and wild-type proteins are shown. For each variant, the distributions seen at pH 8.7 (upper trace) and 7.3 (lower trace) are compared.

On the basis of their effects on the one-disulfide intermediates, the substitutions can be divided into three categories (Table 1). The first category includes F4L, Y10L, and Y35L, none of which detectably altered the distribution at pH 8.7. During the folding of these three variants, [30–51] accounted for approximately 60% of the one-disulfide intermediates, as for the wild-type protein. The second category consists of variants F22L and F33L, for which the accumulation of [30–51] decreased to about 30%. The reduced accumulation of this intermediate with a native disulfide was accompanied by a substantial accumulation of [5–14], which contains a non-native disulfide. The third category includes substitutions Y21L, Y23L, and F45L, which gave rise to the largest effects on [30–51]. During the folding of the Y21L and F45L variants, [30–51] was detectable, but accounted for only about 15% of the total population. For Y23L, no peak containing [30–51] could be identified. Substantial levels of other, non-native, intermediates, were detected in the folding of these proteins.

Decreasing the pH from 8.7 to 7.3 causes a pronounced change in the distribution of one-disulfide intermediates in the folding of

Table 1. *Steady-state levels of the major one-disulfide intermediates as a percentage of the total population of one-disulfide species at pH 8.7, 25 °C*

Variant	[30–51] (%)	[5–14] (%)	[5–30] (%)	[38–51] (%)
Wild type	60	—	—	—
F4L	60	—	—	—
Y10L	60	—	—	—
Y21L	15	25	15	15
F22L	30	20	—	—
Y23L	—	25	15	—
F33L	30	20	—	—
Y35L	60	—	—	—
F45L	15	—	—	—

the wild-type protein, with the level of [5–55] increasing to approximately 40% (Weissman & Kim, 1991). In contrast, the distributions of one-disulfide intermediates in the folding of the aromatic → Leu mutants were nearly identical at the two pH values (Fig. 5). For none of the mutants was a major peak corresponding to [5–55] detectable at either pH. Although a small peak with the expected elution time and pH dependence of [5–55] was seen in the case of the F4L mutant, there was not sufficient material to identify this species by peptide mapping. The greatly reduced level of accumulation of [5–55] suggests that all eight of the aromatic residues contribute to the stability of this intermediate.

The spectrum of one-disulfide intermediates for each BPTI variant appeared to remain constant during the time course of the folding reactions, suggesting that the intermediates were in rapid equilibrium with one another.

Effects on the two-disulfide intermediates

The major effect of each of the amino-acid replacements on the two-disulfide intermediates was to decrease the accumulation of one or both of two native-like intermediates, [30–51,14–38] and [5–55,14–38]. During the folding of the wild-type protein, [5–55, 14–38] acts as a kinetic trap under both neutral and slightly alkaline conditions (pH 7.3 and 8.7, respectively) (Creighton & Goldenberg, 1984; Weissman & Kim, 1991). This species neither forms a third native disulfide nor rearranges to other two-disulfide intermediates on the time scale of most folding experiments. At pH 8.7, the second species, [30–51,14–38], does not readily form a third disulfide, but does rearrange, via non-native intermediates, to form [30–51,5–55], which rapidly forms the [14–38] disulfide (Creighton, 1977; Weissman & Kim, 1992, 1995). At neutral pH, however, rearrangements of [30–51,14–38] are extremely slow, causing this species to act as a kinetic trap also (Weissman & Kim, 1991). As a consequence, the decreased accumulation of these two-disulfide intermediates in the folding of the mutant proteins was accompanied by an increased yield of the native protein at late times, as shown in Figure 6. This figure shows the portions of the chromatograms containing N, [30–51,14–38] and [5–55,14–38] trapped after three hours of folding at pH 7.3.

The accumulation of [30–51,14–38] at both pH 8.7 (Fig. 4) and pH 7.3 (Fig. 6) was nearly eliminated by all of the aromatic substitutions except F4L, which reduced the level of this species to about 50% of that seen with the wild-type protein. A greater range of effects was seen on [5–55,14–38]. This species was detectable at high levels in the folding of the F4L, Y10L and Y21L variants. Lower, but detectable amounts were also seen for the F22L and F33L variants. For the remaining three mutants, Y23L, Y35L, and F45L, [5–55,14–38] was not detectable at either of the pH values examined.

During the folding of the wild-type protein, the third native-like two-disulfide intermediate, [5-55,30-51] (also designated N^{SH}_{SH}) accumulates transiently at pH 8.7. At pH 7.3, formation of this species is so slow that it does not accumulate significantly, though other evidence indicates that it is the immediate precursor to the native protein at both pH 8.7 and 7.3 (Creighton, 1977a; Weissman & Kim, 1992, 1995). N^{SH}_{SH} was readily detected during the folding of the F4L, Y10L, Y21L, F22L, F33L, and Y35L variants, but not with Y23L or F45L BPTI (Fig. 4).

Of the two non-native two-disulfide intermediates that accumulate during the folding of the wild-type protein, only one, [30–51, 5–14] was detected at significant levels in the refolding of any of the mutant proteins. For Y35L, [30–51,5–14] was observed at both pH 8.7 and 7.3. In contrast, this species was detected only at the higher pH in the folding of F4L BPTI, as is the case for the wild-type protein.

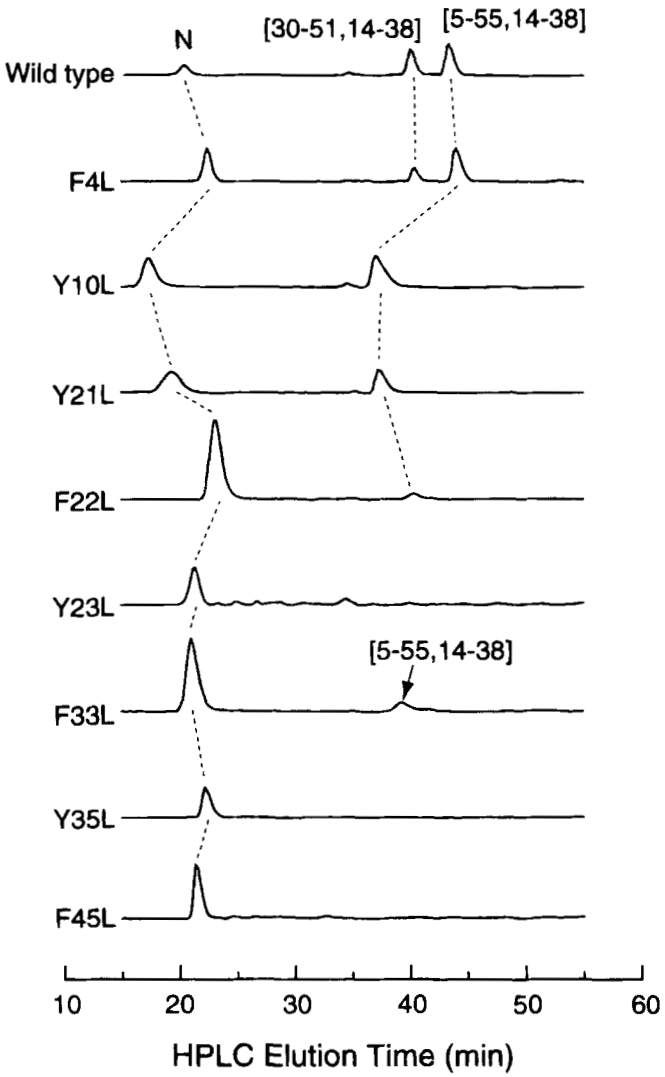

Fig. 6. Native and native-like two-disulfide forms present after folding for 3 h at pH 7.3. Folding experiments were carried out as described in the legend to Figure 4 and in Materials and methods. All of the aromatic → Leu replacements decrease the accumulation of the kinetically trapped two-disulfide intermediates, leading to the increased yield of the native protein with three-disulfides.

Effects of the aromatic → Leu substitutions on the distribution of unfolding intermediates

When native wild-type BPTI is incubated with reduced dithiothreitol (DTT^{SH}_{SH}), the 14–38 disulfide is first reduced to produce N^{SH}_{SH}. Further unfolding of N^{SH}_{SH} in the presence of mM DTT^{SH}_{SH} concentrations (via either direct reduction or rearrangement) is extremely slow, with a half time of many hours (Creighton, 1977a). The resulting intermediates with one or two disulfides are much more rapidly reduced and are not usually detectable. Because the interconversion of N^{SH}_{SH} with the other intermediates is very slow,

this species is treated as a distinct kinetic class in the analysis of the folding pathway. Of the three species containing two naive disulfides, N_{SH}^{SH} is the only one that forms a third native disulfide at a rate consistent with observed rates of complete refolding. Thus, N_{SH}^{SH} is both the immediate precursor of the native protein and the first intermediate in reductive unfolding (Creighton, 1977a; Weissman & Kim, 1992, 1995).

To determine whether N_{SH}^{SH} plays the same roles in the folding and unfolding of the mutant proteins, the distributions of unfolding intermediates were analyzed at pH 8.7. For each of the eight mutants, N_{SH}^{SH} was the first and major intermediate detected during unfolding. Figure 7 illustrates the distributions of intermediates during the unfolding of Y23L and F45L BPTI. In contrast to the distributions seen during folding of these variants, where no single intermediate dominated (Fig. 4), the unfolding of these two mutants was clearly characterized by the predominance of N_{SH}^{SH}. The distinct behavior of this species was even more apparent for the other six mutants. These results indicate that N_{SH}^{SH} should be regarded as a kinetic class distinct from the other two-disulfide intermediates.

Discussion

The predominance of a relatively small number of intermediates is a hallmark of the BPTI folding pathway. In the 1970s, Creighton found that only a few intermediates accumulate significantly during the folding of this protein, an observation that provided some of the first experimental evidence for a specific folding mechanism (Creighton, 1974, 1975, 1978). Using newer methods, Weissman and Kim (1991) recently re-examined the distribution of disulfide-bonded folding intermediates. Although these authors observed a distribution somewhat different from that reported in the earlier studies, they also concluded that the distribution is dominated by a few of the possible disulfide-bonded species.

The results presented here demonstrate that the non-random distribution is very sensitive to amino acid replacements. This sensitivity of the folding intermediates to the aromatic → Leu substitutions contrasts with the generally observed tolerance of proteins to amino acid replacements. Many studies of mutant protein structures indicate that the overall fold of a protein is remarkably tolerant to amino acid replacements (Alber, 1989; Pakula & Sauer, 1989; Goldenberg, 1992a; Matthews, 1993a). Even when a replacement leads to a large destabilization, the result is generally a protein that fails to fold at all, rather than one that folds to a radically different conformation (Lattman & Rose, 1993). In contrast, some, but not all, aspects of the folding process appear to be quite sensitive to mutations.

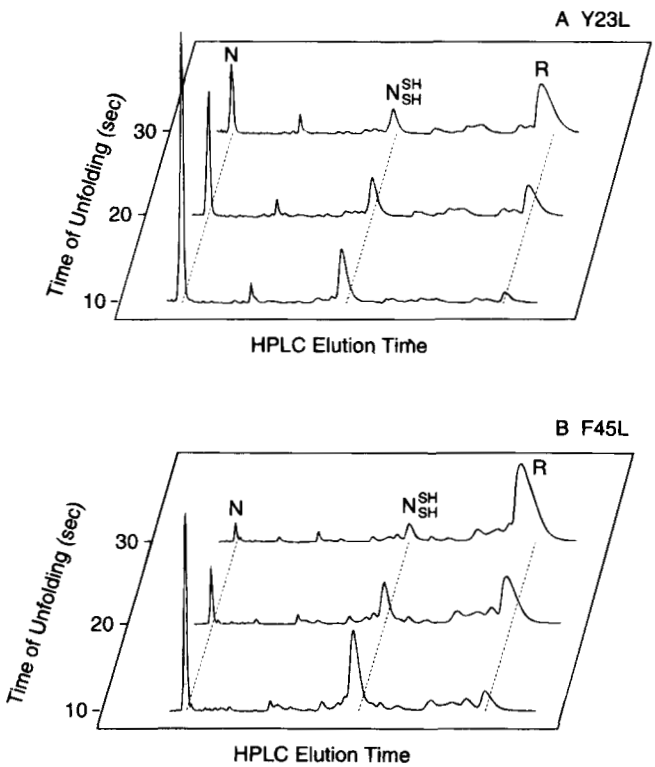

Fig. 7. Intermediates trapped in the reductive unfolding of **(A)** Y23L and **(B)** F45L BPTI. Native proteins were incubated for the indicated times in the presence of 1 mM DTT_{SH}^{SH} at pH 8.7, 25 °C. Following trapping with formic acid, the reaction mixtures were fractionated by reversed-phase HPLC, as in Figures 4, 5, and 6. The major two-disulfide intermediate in the unfolding of both proteins was identified as [5-55,30-51] (N_{SH}^{SH}) by peptide mapping.

Folding mechanisms of the aromatic → Leu variants

Important constraints on possible folding mechanisms can be deduced from consideration of which intermediates interconvert rapidly and which appear to be kinetically distinct. For all of the mutants examined, the distribution of one-disulfide intermediates appeared constant during the course of folding at both pH 8.7 and 7.3, suggesting that the major one-disulfide intermediates equilibrated rapidly with one another on the time scale of these experiments and should be considered as a single kinetic class. Apart from [5–55,14–38] and N_{SH}^{SH}, the relative levels of the two-disulfide intermediates also appeared to be constant, suggesting that these intermediates can be treated as a single kinetic class. On the other hand, N_{SH}^{SH} appeared for all of the mutants to be kinetically distinct from the other two disulfide intermediates. This is seen most clearly by comparing the distribution observed during unfolding, where N_{SH}^{SH} is the major intermediate at early times (Fig. 7) and refolding, where the other two-disulfide intermediates are present at higher relative levels (Fig. 4).

Since N_{SH}^{SH} does not appear to be in rapid equilibrium with other two-disulfide intermediates that accumulate during folding, a minimal mechanism consistent with both folding and unfolding data includes the following steps:

$$R \rightarrow I \rightarrow II \rightarrow N_{SH}^{SH} \rightarrow N \qquad (1)$$

where I and II are populations of one- and two-disulfide intermediates, respectively. As described in the accompanying paper, this scheme can account quantitatively for both folding and unfolding kinetics for each of the mutants at pH 8.7. The one exception is the F4L mutant, for which the formation of [5–55,14–38] must also be explicitly included. While it is difficult to rule out the possibility that some of the native protein is formed via direct disulfide formation in [5–55,14–38] or [30–51,14–38], the data are collectively most consistent with a mechanism in which the majority of molecules fold via N_{SH}^{SH}. In this respect the wild-type folding mechanism appears to be retained by the aromatic → Leu mutants.

Roles of the aromatic residues in determining specificity in the one-disulfide folding intermediates

Each of the aromatic → Leu substitutions was found to decrease the relative levels of some of the intermediates that predominate during the refolding of the wild-type protein. Because the levels of the species making up populations I and II are determined by their relative thermodynamic stabilities, the mutationally induced changes can be used to identify residues that play important roles in stabilizing the intermediates and determining specificity in the early stages of folding.

The predominance of [30–51] in population I is thought to play a particularly important role in defining the productive folding pathway of wild-type BPTI. Since all of the major intermediates in population II contain the 30–51 disulfide, this one-disulfide species is likely to be their major precursor. This interpretation is supported by recent NMR analyses of [30–51] analogs, in which the free Cys residues 5, 14, 38, and 55 were replaced by either Ser (van Mierlo et al., 1993) or Ala (Staley & Kim, 1994). The analogs have a partially folded conformation, represented schematically in Figure 8, that contains the α-helix and most of the β-sheet of the native protein. Some regions of the analogs are disordered, however, including the N-terminal 15 residues and residues 37 to 41. It is likely that the flexible regions of [30–51] allow the 5–14, 5–38, and 14–38 disulfides of population II to form readily.

Early studies by Creighton showed that the predominance of [30–51] could be abolished by the addition of urea or GuHCl, so that the 15 possible one-disulfide intermediates are populated nearly evenly (Creighton, 1977b). This observation indicated that the hydrophobic effect is a major force in determining specificity in the early stages of folding. The effects of the aromatic substitutions indicate, however, that the various hydrophobic residues are not equally involved in stabilizing [30–51], even though most are located within ordered regions of the intermediate (Fig. 8).

The mutational results indicate that residues Tyr 21, Tyr 23, and Phe 45 (colored red in Fig. 8) play the largest role in stabilizing the intermediate. In the native protein and the [30–51] analogs, each of these side chains is located near the 30–51 disulfide and the interface between the α-helix and β-sheet. Smaller effects on [30–51] were observed when Leu replaced Phe 22 or Phe 33 (colored purple in Fig. 8). These side chains are also located within the β-sheet, but on the side opposite from Tyr 21, Tyr 23, and Phe 45. Another aromatic residue, Tyr 35, is located at the end of the sheet farthest from the 30–51 disulfide, and the Leu replacement at this site caused no significant change in the level of the [30–51] intermediate. Substitutions of two other residues, Phe 4 and Tyr 10 (colored green in Fig. 8, along with Tyr 35), were also tolerated with no measurable effect on [30–51]. Consistent with the small effects on the intermediate, these residues are located in the N-terminal segment that is disordered in the [30–51] analogs.

Although many of the aromatic residues are located in the ordered regions of [30-51], not all of their side chains are likely to be fully buried in the intermediate. To assess the environment of the side chains in [30–51], solvent accessible surface areas were calculated using a model of the native protein in which the N-terminal 15 residues were simply deleted. In this hypothetical fragment, the side chains of Tyr 21, Tyr 23, and Phe 45 have nearly as much of their surface buried as in the intact native protein, consistent with the observation that replacement of these residues caused the largest effects on [30–51]. In contrast, Phe 22 and Phe 33, at which intermediate mutational effects were observed, are considerably more exposed in the fragment than in the native protein. Phe 4 and Tyr 10, located in the disordered N-terminal segment of [30–51], are presumably as exposed in the intermediate as they are in the

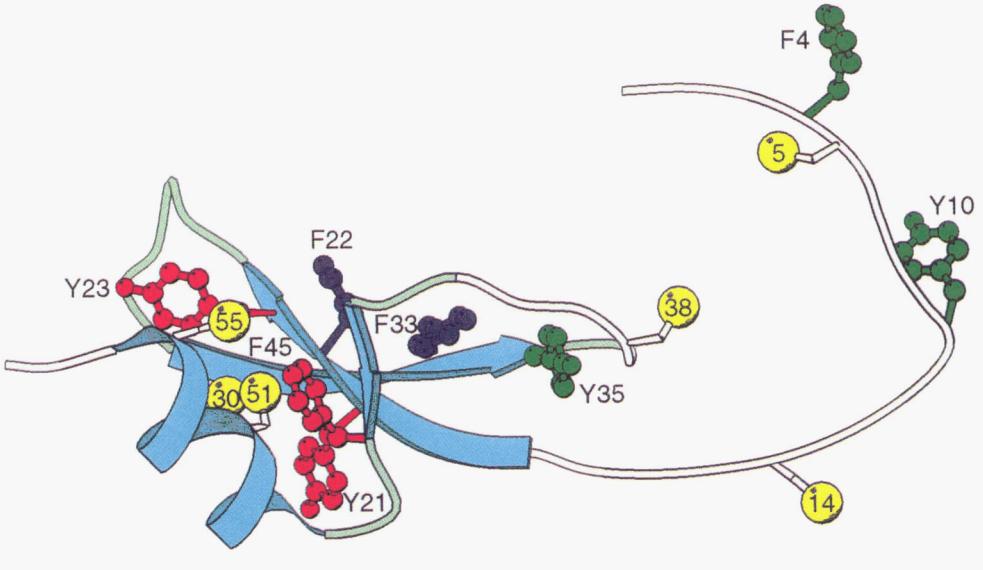

Fig. 8. Schematic representation of the partially folded one-disulfide intermediate [30–51], summarizing the effects of the aromatic → Leu substitutions on the accumulation of this species during folding. The regions of the intermediate shown by NMR studies (van Mierlo et al., 1993; Staley & Kim, 1994) to have well-ordered conformations are shown in blue or light green, to indicate regions of regular secondary structure or irregular, but ordered, structure, respectively. The polypeptide segments that do not appear to be ordered are shown as white ribbons in an arbitrary extended conformation. The side chains of those aromatic residues at which replacement by Leu caused little or no change in the apparent stability of [30–51] are colored dark green. The sites at which substitutions reduced the accumulation of this species to 15% or less of population I are indicated by red side chains. Sites at which intermediate mutational effects were observed are colored purple. Drawn with the program MOLSCRIPT (Kraulis, 1991).

fully unfolded protein. The environment of Tyr 35 is more difficult to assess, as it is located close to the border between ordered and disordered fragments, but the mutational results suggest that this residue does not participate in stabilizing interactions.

The predominance of [5–55] at neutral pH is important for the formation of [5–55,14–38], a kinetically trapped species (Creighton & Goldenberg, 1984; Weissman & Kim, 1991). The results shown here indicate that all of the aromatic residues contribute significantly to the stability of this intermediate, in contrast to the more selective effects of the same substitutions on [30–51]. Alanine-scanning mutagenesis also demonstrates that each of the eight aromatic residues contributes to the stability of a [5–55] analog (Yu et al., 1995). These results are consistent with NMR studies indicating that analogs of [5–55] take on a fully folded conformation, with stabilities highly dependent on the residues used to replace the cysteines normally buried in the intermediate (Darby et al., 1991; van Mierlo et al., 1991b; Staley & Kim, 1992; Schulman & Kim, 1994). The structure of this intermediate thus appears to be determined by the same residues that stabilize the native protein but, because of its marginal stability, is much more sensitive to perturbations than is the protein with all three disulfides.

For four of the mutants (Y21L, F22L, Y23L, and F33L), the decreased accumulation of [30–51] and [5–55] was accompanied by an increase in the relative level of a non-native intermediate, [5–14]. Similar effects have been observed for other types of replacements (G. Bulaj and D.P. Goldenberg, unpubl.). Although the structural basis for the accumulation of [5–14] remains to be elucidated, the appearance of this species during the folding of the mutant proteins suggests that it has conformational properties quite different from the native-like conformations found in the other intermediates.

Roles of the aromatic residues in determining specificity in the two-disulfide folding intermediates

Population II in the refolding of wild-type BPTI includes [30–51, 14–38], [30–51,5–14], and [30–51,5–38], with the native-like species making up at least 90% of the population (Creighton, 1975; Weissman & Kim, 1991). The relative importance of these three intermediates in defining the folding pathway has been a subject of considerable debate (Weissman & Kim, 1991, 1992; Creighton, 1992; Darby et al., 1995). Creighton and colleagues have shown that proteins in which the thiol of either Cys 14 or 38 is modified individually can readily form the native-like species equivalent to N_{SH}^{SH}, even though these substitutions eliminate the possibility of [30–51,14–38] forming (Creighton, 1977a; Darby et al., 1995). The variants with only one Cys residue replaced fold via non-native intermediates at rates that account quantitatively for the folding rate of the unmodified protein. These authors thus argue that the native-like intermediate does not play a central role in folding and emphasize the role of the non-native species. Weissman and Kim (1991, 1992), on the other hand, emphasize the importance of the native-like intermediate because of its greater stability and the high level to which it accumulates during folding.

Our results show that each of the eight aromatic residues contributes to defining the composition of population II. All of the aromatic → Leu replacements decreased the accumulation of [30–51,14–38], but two of the substitutions, Y35L and F4L, did not appear to affect the non-native intermediate [30–51,5–14]. An analog of [30–51,14–38] has a conformation very similar to the native protein, consistent with the observation that all of the aromatic residues contribute to its stability (van Mierlo et al., 1991a). The non-native intermediate [30–51,5–14], in contrast, has a partially folded conformation quite similar to that of [30–51] (van Mierlo et al., 1994). The observation that the non-native intermediate accumulates to at least wild-type levels during the folding of the F4L and Y35L variants is consistent with the NMR evidence suggesting that formation of the non-native 5–14 disulfide is accompanied by the formation of little or no additional structure.

Although all of the substitutions examined here altered the distribution of intermediates making up population II, none of them adversely affected the ability of the population to undergo the subsequent step in the folding pathway, the rearrangement yielding N_{SH}^{SH}. Indeed, kinetic studies described in the accompanying paper (Zhang & Goldenberg, 1997) revealed that the rate of this step at pH 8.7 was actually enhanced slightly by several of the substitutions. This effect most likely arises from the tendency of the replacements to disfavor [30–51,14–38] relative to the non-native intermediates, which are more prone to undergo intramolecular rearrangements. At pH 7.3, where [30–51,14–38] is extremely stable and very slow to rearrange, the destabilizing substitutions promote folding much more dramatically. These results argue against an essential role for the native-like intermediate.

Another two-disulfide intermediate, [5–55,14–38], contains two of the disulfides and essentially all of the folded structure of the native protein. This species, however, is best thought of as a kinetic trap, since it is very slow either to form a third disulfide or undergo intramolecular rearrangements (Creighton & Goldenberg, 1984). The effects of the substitutions on the accumulation of [5–55, 14–38] appear to be correlated with their effects on the stability of the fully folded protein, as measured in the accompanying paper (Zhang & Goldenberg, 1997). The three mutants for which the accumulation of this species was greatest, F4L, Y10L, and Y21L, were also those which displayed the smallest destabilization of the native protein, ranging from 2.2 to 3.4 kcal/mol (relative to the fully reduced state). Somewhat greater destabilizations, 4.1 to 5.6 kcal/mol, were measured for the F22L and F33L substitutions, for which the levels of [5–55,14–38] were detectable but greatly reduced. The three mutants for which this species was not detectable, Y23L, Y35L, and F45L, were destabilized by 6.3 to 7.1 kcal/mol. The decreased accumulation of [5–55,14–38] most likely reflects thermodynamic destabilization of this species, causing it to be either formed more slowly or to rearrange more rapidly. As with the case of [30–51,14–38], the reduced levels of [5–55,14–38] were accompanied by increased folding rates (Fig. 6).

Some of the aromatic → Leu substitutions were also found to alter the levels of the third native-like two-disulfide species, N_{SH}^{SH}. As discussed earlier, however, this species does not, in general, interconvert rapidly with the intermediates making up population II, and the relative level to which it accumulates is determined primarily by kinetic factors. As a consequence, interpreting the effects of the mutations on this species requires a more complete kinetic analysis, as described in the accompanying paper.

As discussed in detail above, the results of this mutational analysis are remarkably consistent with the NMR studies of intermediate analogs. Those intermediates with fully folded conformations, especially [5–55] and [30–51,14–38], were affected by all of the aromatic → Leu substitutions. In contrast, more selective effects were seen on the partially folded intermediates [30–51] and [30–51,5–14], and the patterns of mutational effects are easily rationalized in terms of the structures of these intermediates. Quantitative measurements of the mutational effects on the intermedi-

ates, presented in the accompanying paper, demonstrate, however, that the effects of a substitution depend greatly upon dynamic aspects of the intermediates, as well as their average conformations. In general, the effects of the substitutions are greatest in the most fully folded intermediates, even when the altered sites lie in ordered regions of the earlier intermediates. Mutational analysis, particularly when combined with structural studies, thus provides a means of examining the increased cooperativity of a protein as the native structure is formed.

The BPTI folding pathway as a conformational funnel

The ability of single amino acid replacements to broaden substantially the distribution of folding intermediates, without greatly affecting the competence of the protein to fold, implies a great deal of plasticity in the BPTI folding pathway. This aspect of the pathway can be described using the notion of a "conformational funnel" (Fig. 9), a graphical representation that has been employed recently to summarize the results of theoretical studies of protein folding (Dill et al., 1995; Wolynes et al., 1995; Dill & Chan, 1997). In this representation, the conformational entropy of the polypeptide chain is indicated by the cross-sectional area of a funnel. The top of the funnel, with the largest cross-section, represents the fully unfolded protein, and the bottom of the funnel represents the native protein, with a greatly reduced entropy. The length of the funnel represents progress of the folding reaction, usually quantified in terms of the number of stabilizing interactions or the total stabilizing energy. For the funnels representing the BPTI folding pathway, distance along the length of the funnels was used to represent the steps in the reaction scheme shown in Equation 1. For simplicity, the native-like kinetic trap, [5–55, 14–38], is not incorporated in this representation.

In drawing the conformational funnels, the total loss of conformational entropy was estimated assuming that each residue can take on eight conformations of equal energy in the reduced and unfolded state but has a unique conformation in the native state (Baldwin, 1986; Privalov, 1979). The total decrease in conformational entropy for folding calculated in this way is 240 cal/deg·mol. The loss in conformational entropy due to forming a disulfide bond in a disordered chain, without concomitant structure formation, was estimated from polymer statistics to be about 20 cal/deg·mol (Schellman, 1955), a relatively small fraction of the total entropy loss associated with folding. Although the entropy loss attributed to disulfide formation is expected to differ for the various possible Cys pairs, with various numbers of intervening residues, these differences are predicted to be comparable to the other uncertainties in these calculations (i.e., less than 10 cal/deg·mol). Based on the NMR studies indicating that approximately one-half of the polypeptide chain is folded in the [30–51] intermediate, forming this species from the fully reduced state was assumed to result in the loss of one-half of the conformational entropy of the unfolded protein. The native-like intermediates [5–55], [30–51, 14–38], [5–55, 14–38], and N_{SH}^{SH} were assumed to have lost 95% of the conformational entropy of the unfolded protein. The entropies of intermediate populations I and II were calculated as weighted averages of the entropies of their constituents. Although these estimates are extremely crude, they make it possible to describe, at least qualitatively, an important aspect of folding.

As illustrated in Figure 9, the initial decrease in entropy can be altered greatly by changes in pH or by amino acid replacements. The sharpest narrowing of the conformational funnel is seen for the wild-type protein at neutral pH. Under these conditions, about 40% of the molecules have lost nearly all of their conformational entropy, and another 40% have lost about half upon forming the first disulfide. At pH 8.7, the most native-like one-disulfide intermediate, [5–55], is greatly destabilized, almost certainly because of the tendency of the Cys 30 and Cys 51 thiols to ionize. As a consequence, there is significantly less loss of conformational entropy. For the Y35L variant, the distribution of one-disulfide intermediates at both pH 7.3 and 8.7 is similar to that of the wild-type protein at the higher pH. In contrast, the Y23L substitution leads to a more random distribution of one-disulfide intermediates that are likely to contain little or no specific structure. In this case, there is only a relatively small decrease in conformational entropy corresponding to the formation of a disulfide in a disordered polypeptide.

During the folding of the wild-type protein, at either pH 8.7 or 7.3, nearly all of the conformational entropy of the reduced protein is lost with the formation of population II, which is dominated by

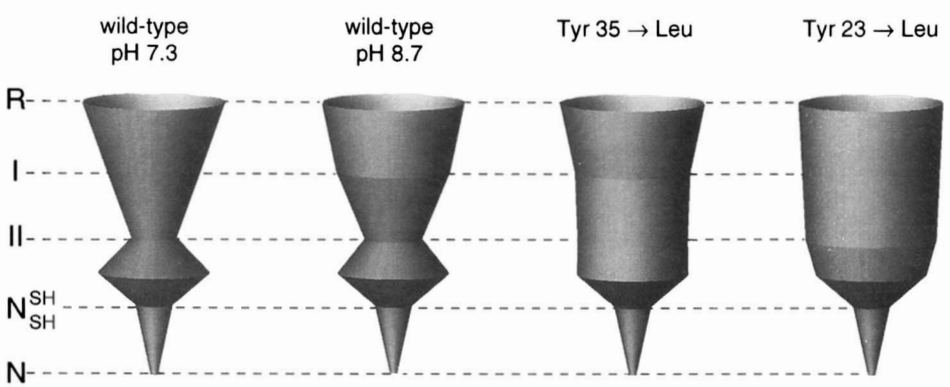

Fig. 9. Conformational funnels illustrating the effects of pH and amino-acid replacements on the BPTI folding pathway. The loss of conformational entropy upon folding is represented as the decreasing cross-sectional area of the funnels. The length of the funnels represents progress along the folding mechanism, according to the mechanism shown in Equation 1. The transition-state for the slowest intramolecular step, the rearrangement of population II to form N_{SH}^{SH} is shown as a bulge in the funnels representing folding of the wild-type protein, indicating that significant unfolding of the major component of II, [30–51, 14–38], is required for this process. The aromatic → Leu replacements lead to substantial broadening of the intermediate populations and make the loss of conformational entropy more progressive than for the wild-type protein.

the native-like [30–51,14–38]. For both Y35L and Y23L BPTI, however, there is only a slight narrowing of the conformational funnel between populations I and II. For these, and the other aromatic → Leu mutants, there appears to be little additional ordered structure formed, and the only decrease in conformational entropy is due to the formation of the additional disulfide.

For the wild-type protein, the immediate precursor to the native protein, N_{SH}^{SH}, is known to have nearly all of the structure of the native state, as does the major component of population II, [30–51, 14–38]. As a consequence, there is relatively little net loss in conformational entropy during either the rearrangement of II to N_{SH}^{SH} or the formation of the final disulfide. The rearrangements must, however, involve at least one non-native intermediate (most likely the minor components of II, [30–51,5–14] and [30–51, 5–38]), and there is evidence from mutational and other studies that the transition state for the rearrangements are extensively unfolded (Weissman & Kim, 1992; Mendoza et al., 1994; Zhang & Goldenberg, 1997). The loss of structure in the transition state is represented as a bulge in the conformational funnels.

Because population II is less ordered for the Y35L mutant, the expansion of the conformational funnel associated with the rearrangement step is less dramatic than for the wild-type protein. For the Y23L mutant, this portion of the funnel is shown as being even smoother, since population II appears to be a nearly random distribution of two-disulfide intermediates. Since only some of the possible two-disulfide intermediates can form the correct transition state, the conformational funnel for this mutant is drawn to indicate a modest loss of entropy in the transition state. In this case, nearly all of the conformational entropy of the reduced protein is lost between II and N_{SH}^{SH}.

Although drawn in Figure 9 as smooth surfaces, the interiors of conformational funnels are likely to contain barriers representing slow interconversions among some intermediate species. In the case of the BPTI folding pathway, the largest barriers arise when the conformational entropy decreases sharply before N_{SH}^{SH} has formed, leading to the formation of [30–51,14–38], which acts as a kinetic trap at neutral pH, along with [5–55,14–38]. By broadening the distribution of molecules making up population II, the aromatic → Leu substitutions increase the likelihood that a molecule will undergo the intramolecular rearrangements required to form N_{SH}^{SH}. The apparent folding rate is thereby increased for all of the mutants examined here.

When the top portion of the conformational funnel becomes too broad, however, the rate of folding may decrease because a smaller fraction of molecules are competent to undergo the subsequent steps. The Y23L and F45L mutants, for instance, display overall folding rates that are significantly lower than that of Y35L (Fig. 4), even though the kinetically trapped species are eliminated for all three of these mutants. There thus appears to be an optimum shape for a conformational funnel: The top of the funnel should show an initial narrowing that helps limit the conformational search, but this narrowing should not be so severe as to trap the protein in a native-like conformation that is unable to form the final disulfide.

While the analysis presented in this paper provides a qualitative picture of the effects of the aromatic → Leu substitutions on the BPTI folding pathway, summarized in the conformational funnels of Figure 9, a more complete description of the energetic effects of these substitutions can be obtained by measuring the rates and equilibria of the interconversions among the various species. These measurements are described in the accompanying paper (Zhang & Goldenberg, 1997).

Materials and methods

Preparation of BPTI variants

Mutant forms of BPTI were produced in *Escherichia coli* using the expression system described previously (Goldenberg, 1988b). The proteins were produced as fusions to the leader peptide of the OmpA protein, an *E. coli* outer membrane protein, and exported to the periplasmic space, where the leader peptide was removed and the polypeptides folded to yield native protein. Mutations in the BPTI coding sequence were constructed using the method of Kunkel (1985).

For preparation of protein for folding experiments, *E. coli* HB101 bacteria carrying the appropriate plasmid were grown at 30 °C with 12–14 L fresh air/min passing through the culture in supplemented M9 medium, which contained 6 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 0.5 g/L NH_4Cl, 0.2 g/L yeast extract, 6 g/L casein, 4 g/L glucose, 40 mg/L L-tryptophan, 10 mg/L L-proline, 36 mg/L L-cysteine, 2 mg/L thiamine-HCl, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, and 50 mg/L ampicillin. Four hours after inoculating 10 L of this medium with 200 mL of a saturated culture grown in 2×TY medium (containing 16 g/L tryptone, 10 g/L yeast extract, and 5 g/L NaCl), isopropyl β-D-thiogalactoside (IPTG) was added to a final concentration of 0.2 mM to induce expression of the mutant BPTI gene. After an additional 18 h of growth, the medium was cooled to 10–15 °C, and then concentrated to a volume of 2 L using a tangential-flow hollow-fiber filter (Kontes). The concentrated bacteria were then centrifuged to yield 100–120 g of wet cell paste, which was stored at −70 °C.

The frozen bacterial pellet from a 10 L culture was resuspended in 400 mL of extract buffer, containing 0.05 M Tris-HCl (pH 8.0), 0.05 M NaCl, 5 mM EDTA, and 0.5 mM oxidized glutathione. After resuspension, the bacteria were lysed in a French pressure cell at 16,000 psi, followed by the addition of 100 mg phenylmethylsulfonyl fluoride (dissolved in 10 mL ethanol) and 6 mL of 0.5 M $CaCl_2$, 0.5 M $MgCl_2$ containing 4 mg deoxyribonuclease I. The lysate was stirred at 4 °C for 1 h and then diluted to a total volume of 800 mL with 0.05 M Tris-HCl, pH 8.0, 0.05 M NaCl, 5 mM EDTA. Fifty mL of 0.1 M EDTA were added to the diluted lysate. The lysate was then centrifuged for 30 min at 12,000 RPM in a Beckman JA-14 rotor. The supernatant was applied to a ZetaPrep SP-100 sulfopropionyl ion exchange cartridge (CUNO Inc.) equilibrated in 0.05 M Tris-HCl, pH 8.0, 0.05 M NaCl, 5 mM EDTA. After washing with 1 L of the equilibration buffer, the BPTI was eluted with a 0.05 M to 1 M NaCl linear gradient with a total volume of 1 L. The BPTI-containing fractions were pooled and concentrated to about 30 mL by ultrafiltration using a Millipore PCAC membrane with a nominal molecular weight cutoff of 1,000. The concentrated BPTI was applied to a 3 cm diameter × 100 cm long column of Sephadex G-50 (fine) equilibrated and eluted with 0.1 M NH_4HCO_3. The BPTI-containing fractions were pooled and lyophilized. Final purification was obtained by reversed-phase HPLC using a Vydac C_{18} semi-preparative column (10 mm inside diameter × 25 cm long), with a gradient of 12 to 60% acetonitrile in 0.1% trifluoroacetic acid. The quantity of BPTI at each purification step was estimated using a spectrophotometric assay for trypsin inhibitor activity (Kassell, 1970). The concentrations of purified samples were also determined by their absorbance at 280 nm, assuming an extinction coefficient of 5,400 M^{-1} (for the wild type and phenylalanine mutants) or 4,050 M^{-1} (for tyrosine mutants). The concentrations determined by the two methods agreed within 5% for each purified mutant protein.

NMR spectroscopy

One-dimensional ^1H NMR spectra of the wild-type and mutant proteins were recorded by Scott Beeser and Terrence Oas using a Varian Unity 500 spectrometer at Duke university. NMR samples contained about 1 mg of protein dissolved in 650 μL 20 mM deuterated acetic acid, 17 μg/mL 2,3-D_4-3-trimethylsilypropionate (TMSP), and 10% D_2O, and were adjusted to pH 4.5 with KOH. Solvent suppression was achieved by selective pre-saturation during the relaxation delay (1.5 s). The spectral width was 8,000 Hz, and 2,048 complex data points were acquired for each transient. For each protein, 1,024 transients were collected and averaged. The acquired data were processed and Fourier transformed using the FELIX program from Biosym. The lowest frequency contribution to the free induction decays was removed using a sine-bell deconvolution, so as to further reduce the solvent signal. The data sets were zero-filled to a total of 4,096 complex data points and apodized with a sine-bell function shifted by 90°. The transformed spectra were referenced to the TMSP chemical shift of −0.0045648 parts per million.

Preparation of reduced BPTI

Reduced BPTI was prepared by incubating the native protein with 50 mM reduced dithiothreitol (DTT^{SH}_{SH}), 6 M GuHCl, 0.1 M Tris-HCl pH 8.7, 0.2 M KCl, and 1 mM EDTA for 1 h at room temperature. The reduced protein was purified on a reversed-phase HPLC column (Vydac C_{18}, 10 mm inside diameter × 25 cm long) eluted with a gradient of acetonitrile in 0.1% trifluoroacetic acid at a flow rate of 5 mL/min. The gradients were prepared by mixing two buffer solutions: A (0.1% trifluoroacetic acid) and B (0.1% trifluoroacetic acid and 60% acetonitrile), beginning with 20% B for 10 min, followed by a 20 min convex increase in composition of buffer B to 60%, and finally a 20 min concave increase in composition of buffer B to 100%. The elution of reduced protein was detected by the absorbance at 280 nm. The reduced protein was dried by lyophilization and redissolved in 0.01 M HCl.

Separation of folding intermediates by reversed-phase HPLC

Refolding reactions were carried out under a nitrogen atmosphere and contained 30 μM reduced protein, 0.1 M Tris-HCl pH 8.7, 0.2 M KCl, 1 mM EDTA, and 0.1 mM oxidized glutathione. The reactions were stopped by adding formic acid to a final concentration of 5% and the products were fractionated on a reversed-phase HPLC column (Vydac C_{18}, 4.6 mm inside diameter × 25 cm long) eluted with a gradient of acetonitrile in 0.1% trifluoroacetic acid at a flow rate of 1 mL/min. The elution profiles were monitored by absorbance at 229 nm. For both the wild-type and mutant proteins, the gradients were prepared by mixing two buffer solutions: A (0.1% trifluoroacetic acid) and B (0.1% trifluoroacetic acid and 90% acetonitrile). For each variant, the gradient conditions were empirically adjusted to obtain optimum separation of the disulfide-bonded intermediates. The gradients used for each of the proteins are listed below, where the final concentration of buffer B and the final time of each segment is indicated:

wild type: 25% B (15 min), 28% B (35 min), 29% B (40 min), and 32% B (105 min)

F4L: 28% B (15 min), 31% B (50 min), and 33% B (110 min)

Y10L: 22% B (5 min), 28% B (15 min), 32% B (50 min), and 35% B (110 min)

Y21L: 25% B (5 min), 28% B (15 min), 33% B (50 min), and 34% B (110 min)

F22L: 20% B (5 min), 27% B (15 min), 31% B (50 min), and 33% B (110 min)

Y23L: 25% B (5 min), 28% B (15 min), 30% B (50 min), and 32% B (110 min)

F33L: as for F22L

Y35L: 27% B (15 min), 29% B (20 min), 31% B (110 min), and 35% B (115 min)

F45L: as for Y35L

For the wild-type protein and the Y21L and Y23L variants, the intermediates were separated at 37 °C, while the separations for the other mutants were carried out at room temperature.

Identification of disulfide bonds in folding intermediates

After HPLC separation in the protonated state, the intermediates were first characterized by gel electrophoresis to determine the number of disulfides each contained. The isolated intermediates were treated with 0.5 M Na-iodoacetate to carboxymethylate Cys residues that were not disulfide-bonded. The modified proteins were re-purified by HPLC and subjected to non-denaturing gel electrophoresis (Reisfeld et al., 1962; Goldenberg, 1989). Since proteins with two or four carboxymethylated Cys residues have very different net charges, the intermediates with one or two disulfides have clearly distinguishable electrophoretic mobilities.

For those species that accumulated to a level corresponding to 5% or more of the total protein, the disulfides were identified by the peptide mapping procedure described by Weissman and Kim (1991). The HPLC isolated intermediates were dried by lyophilization and then redissolved in 0.5 M iodoacetate and 0.1 M Tris-HCl pH 8.7 for 1 min to block any free thiols. The blocked intermediates were then re-purified by reversed-phase HPLC. In cases where more than one species were detected after treatment with iodoacetate, the reaction was repeated with various reagent concentrations to determine whether multiple species arose from intramolecular rearrangement during alkylation. Different concentrations of iodoacetate are expected to lead to different rates of carboxymethylation, while the rate of the competing intramolecular rearrangement remains constant. The relative levels of the species are thus expected to change with iodoacetate concentration if they arise from rearrangement during the reaction. Only those species that appeared to be direct alkylation products were characterized further.

After lyophilization, the blocked intermediates were dissolved in a solution containing 10 mM DTT^{SH}_{SH}, 6 M GuHCl, 0.1 M Tris-HCl pH 8.7, and 1 mM EDTA and incubated for 1 h to reduce the remaining disulfides. After this incubation, *N*-iodoacetyl-*N'*-(5-sulfo-1-naphthyl)ethylenediamine (IAEDANS), a fluorescent thiol-labeling reagent, was added to a final concentration of 50 mM and allowed to react with the resulting thiols for 5 min. Following desalting on a Sephadex G-25 column (Pharmacia, prepacked PD-10 column), the labeled protein was mixed with a digestion solution to give final concentrations of 75 μg/mL thermolysin, 2 mM $CaCl_2$, and 0.1 M Tris-HCl (pH 8.7). This mixture

was incubated for 75 min at 65 °C, and the digestion was stopped by adding EDTA and acetic acid to final concentrations of 5 mM and 5%, respectively. The resulting peptides were fractionated by reversed-phase HPLC (Vydac C_{18}), using a very shallow linear gradient of buffers A (0.1% trifluoroacetic acid) and B (0.1% trifluoroacetic acid and 90% acetonitrile). The total time of the gradient was 150 min, during which the acetonitrile concentration was increased linearly from 3.6% to 13.5%. Peptides containing Cys residues that were disulfide-bonded in the trapped intermediates were detected specifically by monitoring the absorbance of the column eluent at 340 nm, where the IAEDANS label absorbs strongly. These peptides were then identified by amino acid analysis and, in some cases, peptide sequencing.

From the identification of the disulfide-bonded Cys residues, the identities of the one-disulfide intermediates were immediately established since there were only two IAEDANS-labeled Cys residues for each one-disulfide intermediate. For the two-disulfide intermediates, however, the pairings were not directly established. The connectivities between the disulfide-bonded Cys residues were therefore assigned by comparing the chromatographic, electrophoretic, and kinetic properties of the two-disulfide intermediates with those of the well-characterized wild-type intermediates.

The method described above worked well for all of the BPTI variants except F4L and F33L. For these variants, peptides containing Cys 5 (for F4L) or 38 (for F33L) were not detectable, presumably because of changes in the thermolysin digestion pattern at the sites of the substitution. In these cases, the peptide mapping using thermolysin was supplemented by analysis of peptides generated by trypsin digestion.

Acknowledgments

We thank Dr. Robert Schackmann for the synthesis of oligonucleotides and peptide sequence analysis, Dr. Dennis Winge for amino acid analyses, and Dr. Pamela Crain for mass spectrometry analyses. These services were provided by the Protein/DNA core facility of the Utah Regional Cancer Center (supported by National Cancer Institute Grant No. 5 P30 CA42014). We also thank Mr. Scott Fergusson for assistance in constructing some of the mutant genes used for this project, Mr. Scott Beeser and Dr. Terrence G. Oas for performing the NMR experiments, and Drs. Grzegorz Bulaj and Marian Price-Carter for helpful discussions and comments on earlier versions of the manuscript. This work was supported by grant no. GM42494 from the U.S. National Institutes of Health and an NIH Biological Chemistry Training Grant (no. GM08537) award to J.-X. Zhang.

References

Alber T. 1989. Mutational effects on protein stability. *Annu Rev Biochem 58*:765–798.

Baldwin RL. 1975. Intermediates in protein folding reactions and the mechanism of protein folding. *Annu Rev Biochem 44*:453–475.

Baldwin RL. 1986. Temperature dependence of the hydrophobic interaction in protein folding. *Proc Natl Acad Sci USA 83*:8069–8072.

Bordo D, Argos P. 1991. Suggestions for "safe" residue substitutions in site-directed mutagenesis. *J Mol Biol 217*:721–729.

Chothia C. 1975. Structural invariants in protein folding. *Nature 254*:304–308.

Coplen LJ, Frieden RW, Goldenberg DP. 1990. A genetic screen to identify variants of bovine pancreatic trypsin inhibitor with altered folding energetics. *Proteins 7*:16–31.

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

Creighton TE. 1975. The two-disulphide intermediates and the folding pathway of reduced pancreatic trypsin inhibitor. *J Mol Biol 95*:167–199.

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

Creighton TE. 1977b. Effects of urea and guanidine-HCl on the folding and unfolding of pancreatic trypsin inhibitor. *J Mol Biol 113*:313–328.

Creighton TE. 1978. Experimental studies of protein folding and unfolding. *Prog Biophys Mol Biol 33*:231–297.

Creighton TE. 1990. Protein folding. *Biochem J 270*:1–16.

Creighton TE. 1992. Protein folding pathways determined using disulphide bonds. *Bioessays 14*:195–199.

Creighton TE, Darby NJ, Kemmink J. 1996. The roles of partly folded intermediates in protein folding. *Faseb J 10*:110–118.

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.

Dadlez M, Kim PS. 1995. A third native one-disulphide intermediate in the folding of bovine pancreatic trypsin inhibitor. *Nature Struct Biol 2*:674–679.

Dadlez M, Kim PS. 1996. Rapid formation of the native 14–38 disulfide bond in the early stages of BPTI folding. *Biochemistry 35*:16153–16164.

Danishefsky AT, Housset D, Kim KS, Tao F, Fuchs J, Woodward C, Wlodawer A. 1993. Crevice-forming mutants in the rigid core of bovine pancreatic trypsin inhibitor: Crystal structures of F22A, Y23A, N43G, and F45A. *Protein Sci 2*:577–587.

Darby NJ, Creighton TE. 1993. Dissecting the disulphide-coupled folding pathway of bovine pancreatic trypsin inhibitor. Forming the first disulphide bonds in analogues of the reduced protein. *J Mol Biol 232*:873–896.

Darby NJ, Morin PE, Talbo G, Creighton TE. 1995. Refolding of bovine pancreatic trypsin inhibitor via non-native disulphide intermediates. *J Mol Biol 249*:463–477.

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

Darby NJ, van Mierlo CMP, Scott GH, Neuhaus D, Creighton TE. 1992. Kinetic roles and conformational properties of the non-native two-disulphide intermediates in the refolding of bovine pancreatic trypsin inhibitor. *J Mol Biol 224*:905–911.

Dill KA, Bromberg S, Yue K, Fiebig KM, Yee DP, Thomas PD, Chan HS. 1995. Principles of protein folding—A perspective from simple exact models. *Protein Sci 4*:561–602.

Dill KA, Chan HS. 1997. From Levinthal to pathways to funnels. *Nature Struct Biology 4*:10–19.

Dobson CM, Evans PA, Radford SE. 1994. Understanding how proteins fold: The lysozyme story so far. *Trends Biochem Sci 19*:31–37.

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.

Eriksson AE, Baase WA, Matthews BW. 1993. Similar hydrophobic replacements of Leu99 and Phe153 within the core of T4 lysozyme have different structural and thermodynamic consequences. *J Mol Biol 229*:747–769.

Fersht AR. 1995. Mapping the structures of transition states and intermediates in folding: Delineation of pathways at high resolution. *Phil Trans Royal Soc Series B 348*:11–15.

Fersht AR, Matouschek A, Serrano L. 1992. The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J Mol Biol 224*:771–782.

Ferrer M, Barany G, Woodward C. 1995. Partially folded, molten globule and molten coil states of bovine pancreatic trypsin inhibitor. *Nature Struct Biol 2*:211–217.

Goldenberg DP. 1988a. Genetic studies of protein stability and mechanisms of folding. *Annu Rev Biophys Biophys Chem 17*:481–507.

Goldenberg DP. 1988b. Kinetic analysis of the folding and unfolding of a mutant form of bovine pancreatic trypsin inhibitor lacking the cysteine-14 and -38 thiols. *Biochemistry 27*:2481–2489.

Goldenberg DP. 1989. Analysis of protein conformation by gel electrophoresis. In: Creighton TE, ed. *Protein structure: A practical approach*. Oxford: IRL Press. pp 225–250.

Goldenberg DP. 1992a. Mutational analysis of protein folding and stability. In: Creighton TE, ed. *Protein folding*. New York: W.H. Freeman. pp 353–403.

Goldenberg DP. 1992b. Native and non-native intermediates in the BPTI folding pathway. *Trends Biochem Sci 17*:257–261.

Goldenberg DP, Frieden RW, Haack JA, Morrison TB. 1989. Mutational analysis of a protein-folding pathway. *Nature 338*:127–132.

Housset D, Kim KS, Fuchs J, Woodward C, Wlodawer A. 1991. Crystal structure of a Y35G mutant of bovine pancreatic trypsin inhibitor. *J Mol Biol 220*:757–770.

Huang GS, Oas TG. 1995. Submillisecond folding of monomeric lambda repressor. *Proc Natl Acad Sci USA 92*:6878–6882.

Kassell B. 1970. Bovine trypsin-kallikrein inhibitor (Kunitz inhibitor, basic pancreatic trypsin inhibitor, polyvalent inhibitor from bovine organs). *Methods Enzymol 19*:844–852.

Kellis JJ, Nyberg K, Sali D, Fersht AR. 1988. Contribution of hydrophobic interactions to protein stability. *Nature 333*:784–786.

Kiefhaber T, Grunert HP, Hahn U, Schmid FX. 1990. Replacement of a cis proline simplifies the mechanism of ribonuclease T1 folding. *Biochemistry 29*:6475–6480.

Kim KS, Tao F, Fuchs J, Danishefsky AT, Housset D, Wlodawer A, Woodward C. 1993. Crevice-forming mutants of bovine pancreatic trypsin inhibitor: Stability changes and new hydrophobic surface. *Protein Sci 2*:588–596.

Kim PS, Baldwin RL. 1990. Intermediates in the folding reactions of small proteins. *Annu Rev Biochem 59*:631–660.

Kraulis PJ. 1991. MOLSCRIPT: A program to produce both detailed and schematic plot of protein structures. *J Appl Crystallogr 24*:946–950.

Kunkel TA. 1985. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc Natl Acad Sci USA 82*:488–492.

Lattman EE, Rose GD. 1993. Protein folding—What's the question? *Proc Natl Acad Sci USA 90*:439–441.

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

Matthews BW. 1993a. Structural and genetic analysis of protein stability. *Annu Rev Biochem 62*:139–160.

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

Mendoza JA, Jarstfer MB, Goldenberg DP. 1994. Effects of amino acid replacements on the reductive unfolding kinetics of pancreatic trypsin inhibitor. *Biochemistry 33*:1143–1148.

Naderi HM, Thomason JF, Borgias BA, Anderson S, James TL, Kuntz ID. 1991. ^{1}H NMR assignments and three-dimensional structure of Ala14/Ala38 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.

Pakula AA, Sauer RT. 1989. Genetic analysis of protein stability and function. *Annu Rev Genet 23*:289–310.

Ponder JW, Richards FM. 1987. Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J Mol Biol 193*:775–791.

Privalov PL. 1979. Stability of proteins: Small globular proteins. *Adv Protein Chem 33*:167–241.

Reisfeld RA, Lewis UJ, Williams DE. 1962. Disk electrophoresis of basic proteins and peptides on polyacrylamide gels. *Nature 195*:281–283.

Schellman JA. 1955. Stability of hydrogen bonded peptide structures in aqueous solution. *Compt Rend Lab Carlsberg, ser chim 29*:230–259.

Schindler T, Herrler M, Marahiel MA, Schmid FX. 1995. Extremely rapid protein folding in the absence of intermediates. *Nature Struct Biol 2*:663–673.

Schulman BA, Kim PS. 1994. Hydrogen exchange in BPTI variants that do not share a common disulfide bond. *Protein Sci 3*:2226–2232.

Sosnick TR, Mayne L, Hiller R, Englander SW. 1994. The barriers in protein folding. *Nature Struct Biol 1*:149–156.

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.

Staley JP, Kim PS. 1994. Formation of a native-like subdomain in a partially folded intermediate of bovine pancreatic trypsin inhibitor. *Protein Sci 3*:1822–1832.

van Mierlo CMP, 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 CMP, Darby NJ, Neuhaus D, Creighton TE. 1991a. (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.

van Mierlo CMP, Darby NJ, Neuhaus D, Creighton TE. 1991b. 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 CMP, Kemmink J, Neuhaus D, Darby NJ, Creighton TE. 1994. ^{1}H NMR analysis of the partly-folded non-native two-disulphide intermediates (30–51,5–14) and (30–51,5–38) in the folding pathway of bovine pancreatic trypsin inhibitor. *J Mol Biol 235*:1044–1061.

Wagner G, Braun W, Havel TF, Schaumann T, Go N, Wüthrich K. 1987a. 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, Bruhwiler D, Wuthrich K. 1987b. Reinvestigation of the aromatic side-chains in the basic pancreatic trypsin inhibitor by heteronuclear two-dimensional nuclear magnetic resonance. *J Mol Biol 196*:227–231.

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

Weissman JS, Kim PS. 1992. 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. 1995. A kinetic explanation for the rearrangement pathway of BPTI folding. *Nature Struct Biol 2*:1123–1130.

Wlodawer A, Deisenhofer J, Huber R. 1987. Comparison of two highly refined structures of bovine pancreatic trypsin inhibitor. *J Mol Biol 193*:145–156.

Wolynes PG, Onuchic JN, Thirumalai D. 1995. Navigating the folding routes. *Science 267*:1619–1620.

Yu MH, Weissman JS, Kim PS. 1995. Contribution of individual side chains to the stability of BPTI examined by alanine-scanning mutagenesis. *J Mol Biol 249*:388–397.

Zhang J-X, Goldenberg DP. 1993. Amino acid replacement that eliminates kinetic traps in the folding pathway of pancreatic trypsin inhibitor. *Biochemistry 32*:14075–14081.

Zhang J-X, Goldenberg DP. 1997. Mutational analysis of the BPTI folding pathway: II. Effects of aromatic → leucine substitutions on folding kinetics and thermodynamics. *Protein Sci 6*:1563–1576.