Protein Science (1997), *6*:1563–1576. Cambridge University Press. Printed in the USA.

Mutational analysis of the BPTI folding pathway: II. Effects of aromatic → leucine substitutions on folding kinetics and thermodynamics

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 14, 1997)

Abstract

The rates of the individual steps in the disulfide-coupled folding and unfolding of eight BPTI variants, each containing a single aromatic to leucine amino acid replacement, were measured. From this analysis, the contributions of the four phenylalanine and four tyrosine residues to the stabilities of the native protein and the disulfide-bonded folding intermediates were determined. While the substitutions were found to destabilize the native protein by 2 to 7 kcal/mol, they had significantly smaller effects on the intermediates that represent the earlier stages of folding, even when the site of the substitution was located within the ordered regions of the intermediates. These results suggest that stabilizing interactions contribute less to conformational stability in the context of a partially folded intermediate than in a fully folded native protein, perhaps because of decreased cooperativity among the individual interactions. The kinetic analysis also provides new information about the transition states associated with the slowest steps in folding and unfolding, supporting previous suggestions that these transition states are extensively unfolded. Although the substitutions caused large changes in the distribution of folding intermediates and in the rates of some steps in the folding pathway, the kinetically-preferred pathway for all of the variants involved intramolecular disulfide rearrangements, as observed previously for the wild-type protein. These results suggest that the predominance of the rearrangement mechanism reflects conformational constraints present relatively early in the folding pathway.

Keywords: aromatic residues; bovine pancreatic trypsin inhibitor; disulfide bonds; folding kinetics; hydrophobic residues; protein folding; stability

During the process by which a disordered polypeptide folds into a well-defined three-dimensional structure, numerous stabilizing interactions must form. While each of the individual interactions is intrinsically quite weak, once they are all formed in the native protein they balance almost exactly the large loss in conformational entropy associated with folding (Baldwin, 1986; Creighton, 1990; Dill, 1990; Makhatadze & Privalov, 1995). Partially structured intermediates that form during folding are presumably stabilized by similar interactions but relatively little is known about the roles of individual interactions in such species. Studies of several folding intermediates have revealed the presence of ordered regions with structures quite similar to those found in the corresponding native proteins (Kim & Baldwin, 1990; Matthews, 1993b; Dobson et al., 1994; Fersht, 1995; Ptitsyn, 1995; Creighton et al., 1996) but there is evidence that even the most ordered regions are considerably more flexible than in the native protein (Hughson et al., 1991; Barrick & Baldwin, 1993; Marmorino & Pielak, 1995; Peng et al., 1995; Miranker & Dobson, 1996). As a consequence, the cooperativity among individual interactions in folding intermediates may be less pronounced than in native proteins, perhaps reducing the contribution of each interaction.

Dissecting the contributions of individual interactions to conformational stability is an intrinsically difficult problem. High-resolution structural methods can identify interactions, but cannot assess their energetic contributions. Experiments with covalently-modified proteins, generated by either chemical or genetic meth-

Reprint requests to: David P. Goldenberg, Department of Biology, University of Utah, Salt Lake City, Utah 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; DTT_{S}^{S} and DTT_{SH}^{SH}, the disulfide and dithiol forms of dithiothreitol; GuHCl, guanidinium chloride; Tris-HCl, tris(hydroxymethyl)-aminomethane hydrochloride; EDTA, ethylenediamintetraacetic acid; HPLC, high performance liquid chromatography.

ods, can provide estimates of the contributions of individual residues but must be interpreted with caution because altering one site in a protein may perturb surrounding interactions. Nonetheless, a great deal of progress has been made in understanding the energetics of fully folded proteins, particularly in cases where mutational, thermodynamic, and structural methods have been applied in concert (Alber, 1989; Goldenberg, 1992; Matthews, 1993a).

Mutational methods have also been applied to the analysis of intermediates in the folding of several proteins, including the α-subunit of tryptophan synthetase (Beasty et al., 1986; Tsuji et al., 1993), dihydrofolate reductase (Garvey et al., 1989; Perry et al., 1989), barnase (Matouschek et al., 1990; Serrano et al, 1992; Fersht 1995), and bovine pancreatic trypsin inhibitor (BPTI) (Goldenberg et al., 1989; Zhang & Goldenberg, 1993). These studies have shown that amino acid replacements can have clearly distinguishable effects on different steps in a folding mechanism, and have facilitated the identification of residues that contribute to the formation of particular intermediates.

In the accompanying paper (Zhang & Goldenberg, 1997), we have described the initial characterization of a series of BPTI variants designed to probe the roles of eight aromatic residues in the disulfide-coupled BPTI folding pathway. In each of these variants, one of the four Phe or four Tyr residues of wild-type BPTI has been replaced by Leu, a somewhat smaller hydrophobic residue with different packing properties. These substitutions were each found to cause substantial, but differing effects, on the distributions of disulfide-bonded intermediates that accumulate during folding and unfolding. Some replacements acted specifically to destabilize kinetically-trapped species that accumulate relatively late during the folding of the wild-type protein, leading to a simplification of the intermediate profile and increased rates of forming the native protein with three disulfides. Other, chemically equivalent, substitutions were found to reduce the tendency of native-like structure to form early in folding and led to greatly broadened intermediate distributions. Although the aromatic → Leu substitutions cause pronounced changes in the observed intermediate distributions, important features of the folding pathway appeared to be conserved for the mutant proteins, in particular the role of intramolecular rearrangements.

In the present paper, we describe a detailed kinetic analysis of the folding pathways for the aromatic → Leu mutants. The rates and equilibria of the individual steps making up the pathway have been measured, enabling us to estimate the effects of the replacements on the conformational stabilities of the major intermediates and the fully folded protein. In addition, the kinetic measurements provide information about the relative contributions of alternative folding pathways (involving direct disulfide formation and disulfide rearrangement, respectively) and about the nature of the major transition states in the two mechanisms. Together with recent NMR studies of the major intermediates, the studies described in this and the accompanying paper provide a detailed description of how the individual aromatic residues contribute to a network of cooperative interactions during the folding of a small protein.

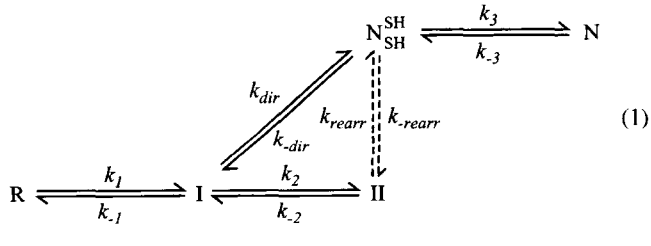
Results

The folding and unfolding of the aromatic → Leu BPTI variants were studied under the same conditions used previously for the wild-type protein and other mutants (Creighton & Goldenberg, 1984; Goldenberg et al., 1989). As in the previous studies, the oxidized and reduced forms of dithiothreitol (DTT^{S}_{S} and DTT^{SH}_{SH}, respectively) were used to promote disulfide formation and reduction. These reagents were used, rather than the more physiological glutathione, because the mixed disulfides formed between DTT^{SH}_{SH} and protein thiols are much less stable than those formed with glutathione. This simplifies considerably the distribution of species that accumulate during folding and, in addition, assures that the observed rates of disulfide formation are proportional to the rate of the intramolecular step in the process (Creighton & Goldenberg, 1984). By adjusting the concentrations of the thiol and disulfide reagents, the relative stabilities of species with zero, one, two, or three disulfides can be manipulated without the addition of denaturants or changes in temperature, thus allowing for the measurement of folding and unfolding rates under nearly identical conditions.

To follow the kinetics of folding or unfolding, the reactions were quenched by the addition of iodoacetate, which reacts irreversibly with the free thiols, and the blocked intermediates were separated by nondenaturing gel electrophoresis. Although electrophoretic separation is not as effective as that provided by HPLC, this method does resolve the fully reduced protein, the major populations with one and two disulfides and the native protein. While some of the amino acid replacements alter significantly the distributions of intermediates, the experiments described in the accompanying paper indicated that the distributions within the major populations did not change significantly with time during folding or unfolding, suggesting that the species within the populations were in rapid equilibrium. The groups of rapidly equilibrated species with one or two disulfides, designated I and II, respectively, were therefore treated as kinetically homogeneous classes in the analyses described here.

Although some intermediates may undergo intramolecular rearrangement during trapping with iodoacetate, it is unlikely that the rearrangements will affect the kinetic analysis. The rearrangements during trapping are most likely to occur when an intermediate with buried thiols is in rapid equilibrium with other species in which thiols are exposed. For example, [30–51,14–38] has been found to rearrange to other two-disulfide species ([30–51,5–14] and [30–51,5–38]) during trapping (Weissman & Kim, 1991). In these cases, however, the interconverting species are components of the same kinetic population, so that measurements of the total concentrations of the populations are unaffected by the rearrangements.

Following gel electrophoresis, the relative concentrations of the resolved species were estimated by video densitometry. The observed time-dependent changes in the concentrations of the various species were then compared to those predicted by numerical simulations based on the following kinetic scheme:



(1)

In the scheme shown above, arrows representing steps that are second order and are, therefore, sensitive to the concentration of DTT^{S}_{S} or DTT^{SH}_{SH}, are drawn with solid lines, while the intramolecular rearrangements among two-disulfide intermediates are represented with dashed lines. The rate constants used in the simulations were adjusted manually to give a good fit between the observed and

calculated data. For each mutant, a single set of rate constants was used to simulate all of the folding and unfolding experiments. For one of the variants, F4L BPTI, the kinetically-trapped species [5–55,14–38] accumulated significantly, and an additional, step I ↔ [5–55,14–38], was included in the simulation to account for the appearance of this species. In the simulations, the changes in concentrations of all species, including DTT^{S}_{S} and DTT^{SH}_{SH}, were incorporated explicitly. The rate constants determined by this procedure are estimated to be accurate within a factor of two.

In the following sections, we describe the specific experiments used to determine the rate constants for the various steps in the folding of the mutant proteins.

Unfolding kinetics

Unfolding experiments were carried out by mixing native protein with various concentrations of DTT^{SH}_{SH}. As demonstrated in the previous paper, the first step in the unfolding of each of the BPTI variants was the reduction of the 14–38 disulfide to convert the native protein to the native-like two-disulfide intermediate, N^{SH}_{SH}. The rate constant for this step, k_{-3}, was determined from the disappearance of the native protein (Fig. 1) and, as shown in Table 1, was found to vary from 25 $s^{-1}M^{-1}$ (Y23L) to 900 $s^{-1}M^{-1}$ (Y35L), versus 30 $s^{-1}M^{-1}$ for the wild-type protein. Three substitutions at sites close to the 14–38 disulfide, Y10L, F33L, and Y35L, increased the rate of reducing this disulfide by least 10-fold. The other five substitutions had relatively small effects on the rate of this step, probably because they are farther from the 14–38 disulfide in the native protein.

The two-disulfide intermediate generated by the reduction of the 14–38 disulfide, N^{SH}_{SH}, can unfold either via direct reduction to generate one-disulfide intermediates (I) or via a rearrangement process that yields other two-disulfide species (II) (Creighton, 1977a). The contributions of the two reactions were dissected by analyzing chemically modified forms of the mutant proteins in which the 14–38 disulfide was reduced and the resulting free thiols were alkylated with iodoacetate. Since free thiols are required for any intramolecular rearrangement, the modified forms, designated N^{SCM}_{SCM}, could only unfold via direct reduction, and the reduction of this species should reflect the process of the direct reduction of N^{SH}_{SH}, with the second-order rate constant k_{-dir}, As illustrated in Figure 2, the rates of reduction varied substantially among the mutants. Even when incubated with a much lower concentration of DTT^{SH}_{SH} (1 mM), the N^{SCM}_{SCM} form of F45L was reduced at a rate much greater than that of F4L, Y10L, and Y21L at higher concentrations of DTT^{SH}_{SH} (20 or 50 mM). The second-order rate constant for the direct reduction of each mutant protein, k_{-dir}, was determined by comparing the observed and simulated kinetics at various concentration of DTT^{SH}_{SH}. Once k_{-dir} was determined, the first-order rate constant for the intramolecular rearrangement (k_{-rearr}) was estimated from unfolding kinetics of the unmodified proteins at various DTT^{SH}_{SH} concentrations (Fig. 1).

The rate constants for the direct reduction for some mutants (the F4L, Y10L, Y21L, F33L, and Y35L variants) could also be determined directly from the unfolding rates of unmodified proteins by determining the dependence of the observed rates on the DTT^{SH}_{SH} concentration, using the procedure described by Mendoza et al.

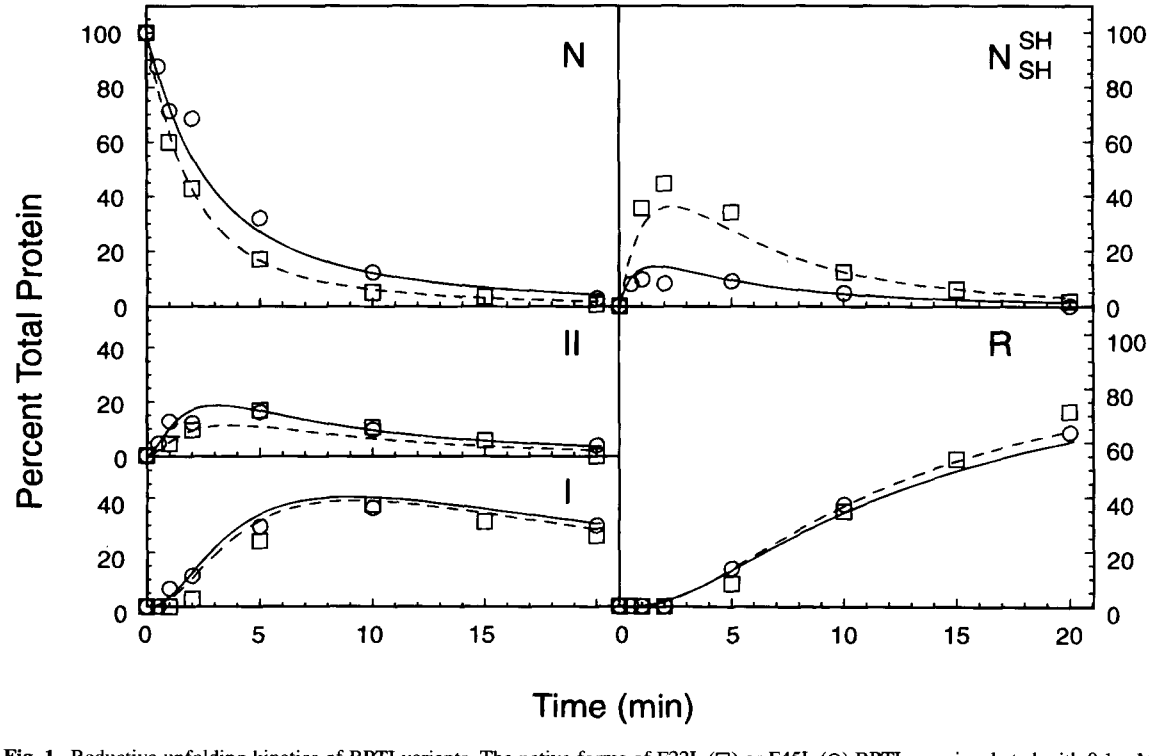


Fig. 1. Reductive unfolding kinetics of BPTI variants. The native forms of F22L (□) or F45L (○) BPTI were incubated with 0.1 mM DTT^{SH}_{SH} at pH 8.7, 25 °C. At the times indicated on the abscissa, samples of the reactions were quenched by the addition of iodoacetate, and the trapped species were fractionated by nondenaturing gel electrophoresis. The relative concentrations of the different forms were quantified by video densitometry of Coomassie-blue stained gels. The curves shown were generated by numerical integration of the rate expressions corresponding to the scheme shown in Equation 1, using the rate constants listed for the corresponding variants in Tables 1 and 2.

Table 1. *Rate constants for individual steps in the unfolding of wild-type BPTI and the aromatic → Leu variants*

Variant	$N \rightarrow N_{SH}^{SH}$ k_{-3} ($s^{-1}M^{-1}$)	$N_{SH}^{SH} \rightarrow II$ k_{-rearr} (s^{-1})	$N_{SH}^{SH} \rightarrow II$[a] k_{-dir} ($s^{-1}M^{-1}$)	$II \rightarrow I$ k_{-2} ($s^{-1}M^{-1}$)	$I \rightarrow R$ k_{-1} ($s^{-1}M^{-1}$)
Wild type[b]	30	1.2×10^{-5}	1.2×10^{-4}	250	20
F4L	50	9×10^{-5}	1.3×10^{-3} (5.0×10^{-3})	150	35
Y10L	300	1.0×10^{-4}	2.5×10^{-3} (4×10^{-3})	250	20
Y21L	45	1.1×10^{-3}	2.0×10^{-2} (8.8×10^{-2})	250	30
F22L	80	9.0×10^{-3}	0.7	250	45
Y23L[c]	25	1.0×10^{-1}	11	250	95
F33L	300	5.0×10^{-4}	3.5×10^{-2} (5×10^{-2})	250	30
Y35L[d]	900	3.5×10^{-4}	4×10^{-2} (3.4×10^{-2})	250	25
F45L	60	3.0×10^{-2}	8	150	35

[a]From the modified proteins with Cys 14 and 38 blocked. The values in parentheses are from measurements with the unmodified proteins at various concentrations of DTT_{SH}^{SH}, as described by Mendoza et al. (1994).
[b]From Creighton and Goldenberg (1984) and Goldenberg (1988).
[c]From Goldenberg et al. (1989), with the exception of k_{-dir}.
[d]From Zhang and Goldenberg (1993).

(1994). The values of k_{-dir} determined by this method were very similar to those determined with the modified forms of the mutant proteins (Table 1), consistent with other results indicating that modification of the thiols of Cys 14 and 38 has relatively little effect on the kinetic properties of the resulting protein (Creighton, 1977a; Marks et al., 1987; Goldenberg, 1988).

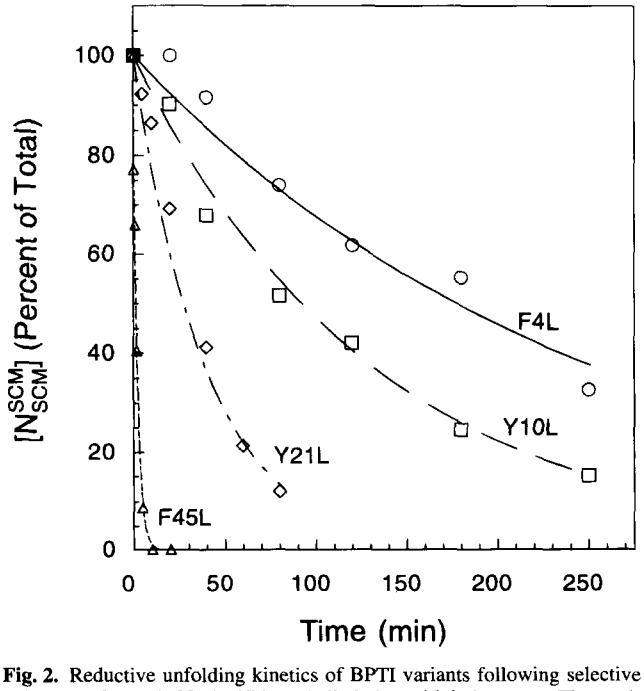


Fig. 2. Reductive unfolding kinetics of BPTI variants following selective reduction of the 14–38 disulfide and alkylation with iodoacetate. The modified forms (N_{SCM}^{SCM}) of F4L, Y10L, Y21L, and F45L BPTI were incubated with 50, 50, 20, and 1 mM DTT_{SH}^{SH}, respectively. The disappearance of the N_{SCM}^{SCM} form was monitored by gel electrophoresis as described in the legend to Figure 1. The curves shown were generated by numerical simulations, using the rate constants in Tables 1 and 2.

Each of the eight aromatic → Leu substitutions increased the rates of both direct reduction of N_{SH}^{SH} and intramolecular rearrangement, but to different degrees. The second-order rate constant for direct reduction (k_{-dir}) varied from 1.2×10^{-4} $s^{-1}M^{-1}$, for the wild-type protein, to 11 $s^{-1}M^{-1}$, for the Y23L variant, and the first-order rate constants for the rearrangement (k_{-rearr}) ranged from 1.2×10^{-5} s^{-1} to 0.1 s^{-1} (Table 1). The increases in the rearrangement rate constants displayed a striking correlation with those of the direct reduction reaction. When plotted on logarithmic axes, the rate constants show a clear linear relationship, as observed previously for a larger collection of BPTI variants at pH 8.0 (Mendoza et al., 1994). Since the logarithms of the rate constants are proportional to the free energy differences between the ground state and the transition state, the observed correlation suggests that the transition states for rearrangement and direct reduction respond similarly to the amino acid replacements.

For most of the mutants, the rates of the late steps in unfolding (II → I and I → R) were not readily determined from unfolding experiments because there was no significant accumulation of one- and two-disulfide species other than N_{SH}^{SH}. For these mutants, the rate constants in the late steps of unfolding (k_{-1} and k_{-2}) were determined from the kinetics of folding in the presence of added DTT_{SH}^{SH}, as described in a subsequent section. For some mutants, such as F22L and F45L, however, the unfolding of N_{SH}^{SH} was sufficiently fast that both one-and two-disulfide species accumulated to measurable levels (Fig. 1). In these cases, a single set of rate constants could account for the unfolding kinetics as well as the kinetics of folding in the presence of added DTT_{SH}^{SH}.

The estimated k_{-1} and k_{-2} values for all of the mutants are listed in Table 1. In contrast to the large effects seen on the early steps in unfolding, the aromatic → Leu substitutions had only small effects

on the last two steps. For the step II → I, the largest change in the second-order rate constant was less than two-fold, within the range of experimental error. The second-order rate constants that were determined for the last step were also close to that measured for the wild-type protein, with the exceptions of the F22L and Y23L variants. For these two mutants, the rate was increased approximately two- or four-fold, respectively.

Unfolding reactions were also carried out in the presence of added DTT^{S}_{S}, where the reverse reactions of some of the unfolding steps become significant. As shown in Figure 3 for the Y10L and F22L variants, the native protein was rapidly and essentially irreversibly converted to N^{SH}_{SH} in the presence of DTT^{SH}_{SH} alone but, when DTT^{S}_{S} was added, the reverse reaction became significant, leading to biphasic kinetics for the disappearance of the native protein. The rapid phase corresponds to the establishment of a steady-state equilibrium between the native protein and N^{SH}_{SH}, while the slower phase represents the disappearance of N^{SH}_{SH}. With k_{-3}, k_{-rearr}, and k_{-dir} determined from the experiments described above, the second-order rate constant (k_3) for the last folding step (N^{SH}_{SH} → N) could be estimated from the effect of added DTT^{S}_{S} on the unfolding kinetics. The rate constants estimated in this way were also consistent with the results of the folding experiments described in the following section.

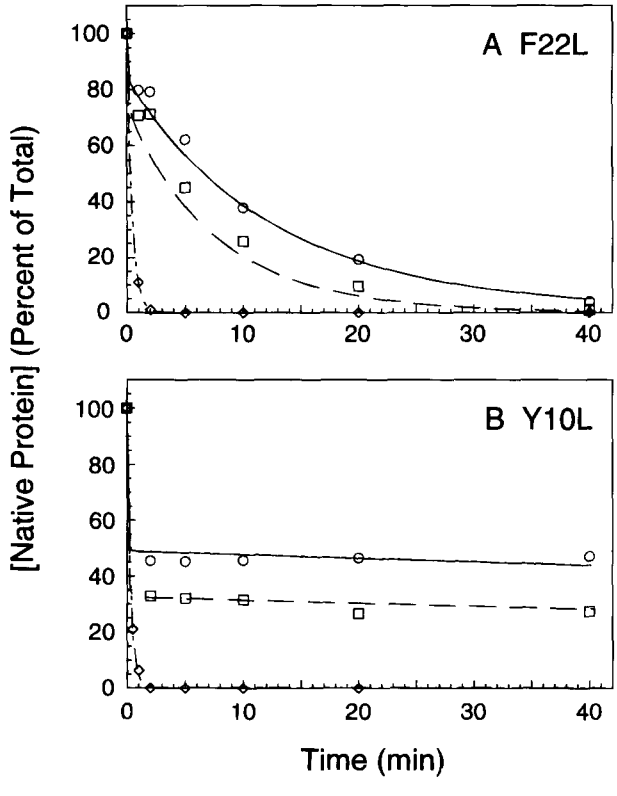


Fig. 3. Effect of added oxidized dithiothreitol, DTT^{S}_{S}, on the reductive unfolding kinetics of BPTI variants F22L (**A**) and Y10L (**B**). The F22L variant was unfolded in the presence of 0.5 mM DTT^{SH}_{SH} and 0 (◇), 40 (□), or 80 (○) mM DTT^{S}_{S}, while the experiments with the Y10L variant utilized 0.5 mM DTT^{SH}_{SH} and 0 (◇), 20 (□), or 40 (○) mM DTT^{S}_{S}. The disappearance of the native protein was monitored electrophoretically, as in Figures 1 and 2. The curves shown were generated by numerical simulations, using the rate constants in Tables 1 and 2.

Folding kinetics

The fully reduced forms of the mutant proteins were refolded in the presence of DTT^{S}_{S}, with and without added DTT^{SH}_{SH}. Examples of the folding experiments are shown in Figure 4 for the Y10L and F45L mutants refolded in the presence of 40 mM DTT^{S}_{S}. Unlike the unfolding reactions, in which the early steps were essentially unidirectional, the folding reaction was not unidirectional under the conditions used for these experiments. Because DTT^{SH}_{SH} is a potent reducing reagent, even the low concentrations produced by oxidizing the protein led to significant back reactions. To better quantify this effect, refolding was carried out in the presence of added DTT^{SH}_{SH}, as illustrated in Figure 5 for the Y10L variant. As shown, the addition of DTT^{SH}_{SH} led to biphasic kinetics for the disappearance of R, reflecting a steady-state equilibration between the fully reduced protein and the one-disulfide intermediates and a decreased overall rate of forming the native protein. By comparing the observed kinetics with those predicted by simulations based on Equation 1, the forward rate constants k_1, k_2, and k_{rearr}, as well as the reverse rate constants k_{-1} and k_{-2}, were estimated. For F4L BPTI there was also significant accumulation of [5–55,14–38], and an additional reaction, I ↔ [5–55,14–38] with forward and reverse rate constants was incorporated into Equation 2 for this mutant. Although this step was able to account for the observed appearance of [5–55,14–38], it is likely that this species is also formed via intramolecular rearrangement of other two-disulfide species, as has been found for the wild-type protein (Creighton & Goldenberg, 1984; Weissman & Kim, 1992).

The rate constant (k_{dir}) for the step I → N^{SH}_{SH} was estimated from the constraints imposed by the other rate constants and the principle of microscopic reversibility. For a reaction cycle such as that linking I, II, and N^{SH}_{SH} (Equation 1), the product of rate constants representing a clockwise path around the cycle must equal the product of rate constants for the counter-clockwise path. The value of k_{dir} was, therefore, calculated from the other five rate constants according to

$$k_{dir} = \frac{k_2 \cdot k_{rearr} \cdot k_{-dir}}{k_{-2} \cdot k_{-rearr}} \qquad (2)$$

For each of the mutants, the value of k_{dir} was increased by about one order of magnitude but was still much smaller than k_2, the rate of the competing reaction in which other two-disulfide forms are generated. Thus, the one-disulfide intermediates (I) are most likely to be converted to the two-disulfide intermediates in population II, which must then undergo intramolecular rearrangements to generate N^{SH}_{SH}. The preference for the rearrangement mechanism was also demonstrated by measuring the folding rates of the modified proteins with Cys 14 and 38 blocked, N^{SCM}_{SCM}. For each mutant, the rate of folding was much lower for the modified protein, as demonstrated previously for the wild-type protein and the Y35L variant (Creighton, 1977a; Goldenberg, 1988; Zhang & Goldenberg, 1993), thus demonstrating that the direct pathway is too slow to account for the folding rates of the unmodified proteins.

In striking contrast with the large effects seen on the rates of unfolding, the aromatic → Leu replacements caused only modest changes in any of the forward rate constants making up the BPTI folding pathway (Table 2). As discussed further below, the observed pattern of kinetic effects suggests that the early stages of folding are relatively tolerant of amino acid replacements that may destabilize any partially folded structure, while the native protein

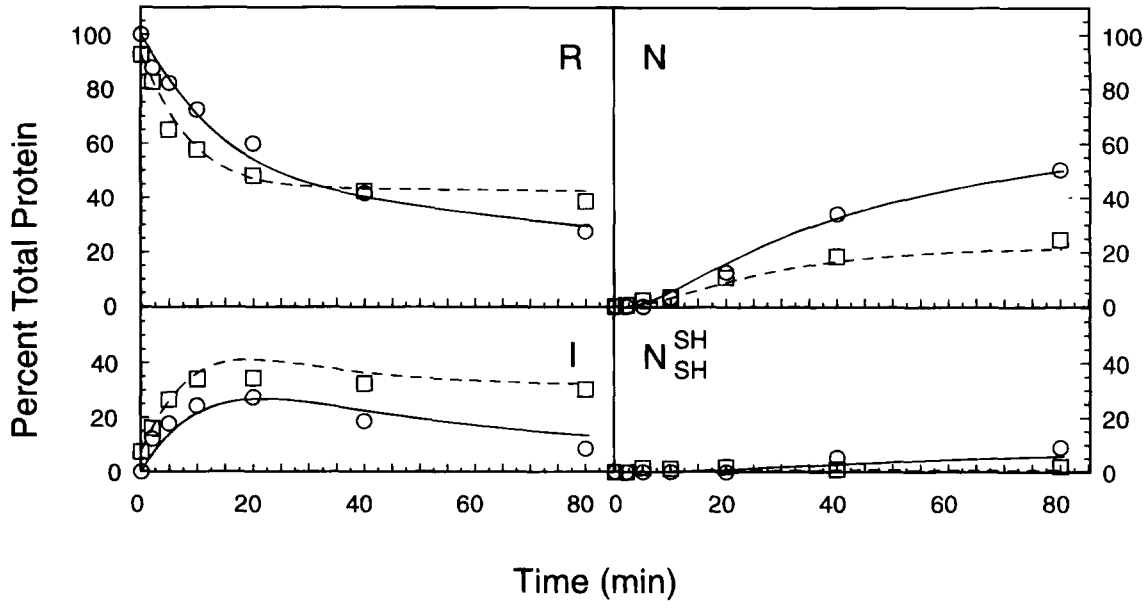


Fig. 4. Disulfide-coupled refolding of BPTI variants. The fully reduced and unfolded forms of Y10L (○) and F45L (□) BPTI were incubated in the presence of 40 mM DTT^{S}_{S} at pH 8.7, 25 °C. At the times indicated on the abscissa, samples of the reactions were quenched with iodoacetate and analyzed electrophoretically. The curves shown were generated by numerical simulations, using the rate constants in Tables 1 and 2.

and intermediates representing the late stages of folding are more easily perturbed. Even when the distribution of folding intermediates is greatly broadened, as shown in the accompanying paper for the Y23L and F45L BPTI variants, the rates of subsequent steps are only slightly affected, perhaps because additional avenues of disulfide formation become accessible.

Discussion

The detailed kinetic analysis described here provides quantitative information about the effects of the aromatic → Leu replacements on the individual steps in the BPTI folding pathway. The resulting rate constants can be used to calculate the effects of the substitu-

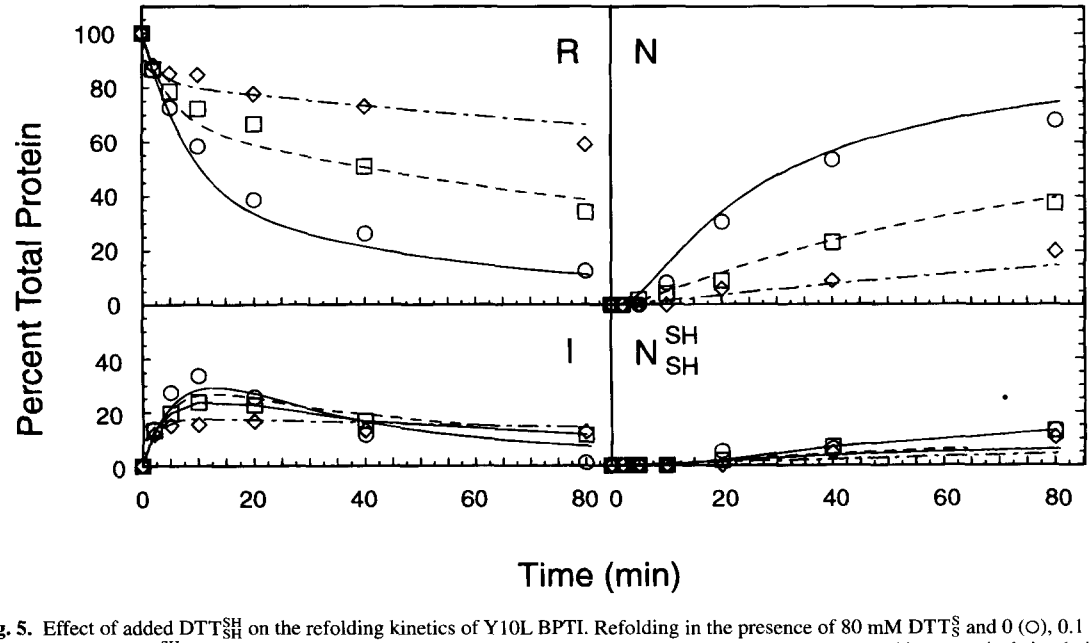


Fig. 5. Effect of added DTT^{SH}_{SH} on the refolding kinetics of Y10L BPTI. Refolding in the presence of 80 mM DTT^{S}_{S} and 0 (○), 0.1 (□), or 0.25 mM (◇) DTT^{SH}_{SH} was monitored electrophoretically, as in Figure 4. The curves shown were generated by numerical simulations, using the rate constants in Tables 1 and 2.

Table 2. *Rate constants for the folding of wild-type BPTI and the aromatic → Leu variants*

Variant	R → I k_1 ($s^{-1}M^{-1}$)	I → II k_2 ($s^{-1}M^{-1}$)	I → N_{SH}^{SH} k_{dir} ($s^{-1}M^{-1}$)	II → N_{SH}^{SH} k_{rearr} (s^{-1})	N_{SH}^{SH} → N k_3 ($s^{-1}M^{-1}$)
Wild type[a]	0.022	0.030	5.3×10^{-6}	5.0×10^{-3}	5.7
F4L	0.019	0.025	3.5×10^{-5}	5.0×10^{-3}	4.0
Y10L	0.015	0.030	3.8×10^{-5}	8.0×10^{-3}	3.5
Y21L	0.015	0.035	3.1×10^{-5}	5.0×10^{-3}	5.0
F22L	0.018	0.015	4.7×10^{-5}	1.0×10^{-2}	2.5
Y23L[b]	0.020	0.012	6.4×10^{-5}	2.0×10^{-2}	0.8
F33L	0.022	0.028	7.8×10^{-5}	1.0×10^{-2}	2.0
Y35L[c]	0.020	0.013	6.4×10^{-5}	1.0×10^{-2}	0.13
F45L	0.022	0.010	1.8×10^{-4}	1.0×10^{-2}	1.0

[a]From Creighton and Goldenberg (1984) and Goldenberg (1988).
[b]From Goldenberg et al. (1989), with exception of k_{dir}.
[c]From Zhang and Goldenberg (1993).

tions on the conformational stabilities of the various intermediates making up the pathway and provide new information about the factors that determine the folding mechanism for the wild-type protein.

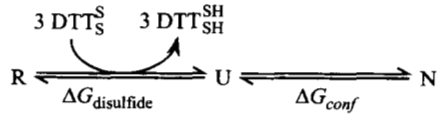
Roles of the aromatic residues in stabilizing native BPTI and N_{SH}^{SH}

From the measured forward and reverse rate constants, the effects of substitutions on the stability of the native protein can be calculated according to

$$\Delta\Delta G = RT \ln(K_{wt}/K_{mutant}) \quad (3)$$

where R and T are the gas constant and temperature, respectively, and K_{wt} and K_{mutant} are the overall equilibrium constants for folding the wild-type and mutant proteins. The equilibrium constants are calculated as the product of the forward rate constants divided by the product of the reverse rate constants. As defined above, the value of $\Delta\Delta G$ is positive for a substitution that destabilizes the native protein. The effects of the substitutions on the intermediates making up the pathway can be calculated in the same way and are listed in Table 3 for N, N_{SH}^{SH}, and the populations I and II.

Although this method of measuring the energetic effects of an amino acid replacement is different from that more commonly used, in which changes in conformational stability are measured by thermal or denaturant-induced unfolding, the mutational effects on the disulfide-coupled equilibrium for forming the native protein are expected to be nearly equivalent to those measured for unfolding in the absence of disulfide reduction. This point is illustrated by considering the following hypothetical mechanism for the folding of a protein with three disulfides:



(4)

In this scheme, which is mechanistically unrealistic but thermodynamically useful, the first step is the formation of the three disulfides without concomitant folding. In the second step, the unfolded, but disulfide bonded, protein (U) folds to its native conformation. It is the free energy change for this second step that would be measured in, for instance, thermal unfolding experiments. Since the unfolded forms with and without the disulfides (U and R, respectively) are presumed to be equally devoid of stabilizing interactions, amino acid replacements are expected to have minimal effects on $\Delta G_{disulfide}$, and a mutational destabilization determined from Equation 4 should be roughly equivalent to the change in conformational stability. Although the same arguments can be applied to a unique intermediate with a stable well-defined structure, such as N_{SH}^{SH}, the situation for intermediate populations, such as I and II, is somewhat more complex and is discussed in a later section.

The kinetic data indicate that the native protein was substantially destabilized by all eight substitutions, with the effects ranging from 2.2 kcal/mol (F4L) to 7.1 kcal/mol (Y23L) (Table 3). As illustrated in Figure 6A, where the effects of the replacements at the various sites in the native protein are represented by a color code, there is no clear correlation between the degree of destabilization and the position of the altered site. Strikingly, however, the three sites at which replacement with Leu had the largest effects,

Table 3. *Destabilization of the folding intermediates and native BPTI caused by the aromatic → Leu substitutions*

Variant	I (kcal/mol)	[30–51][a] (kcal/mol)	II (kcal/mol)	N_{SH}^{SH} (kcal/mol)	N (kcal/mol)
F4L	0.4	<0.2	0.2	1.7	2.2
Y10L	0.2	<0.2	0.2	1.2	2.9
Y21L	0.5	1.5	0.4	3.1	3.4
F22L	0.6	0.8	1.0	4.5	5.6
Y23L	1.0	>1.5	1.5	6.1	7.1
F33L	0.2	0.8	0.3	2.1	4.1
Y35L	0.2	<0.2	0.9	2.3	6.5
F45L	0.3	1.5	0.7	4.9	6.3

[a]Calculated from the fraction of [30–51] in the one-disulfide intermediates, listed in Table 1 of the accompanying paper (Zhang & Goldenberg, 1997) using the graph shown in Figure A.1 and assuming that $f_{[30-51]}^{wt} = 0.6$ and $K_f^{wt} = 20$.

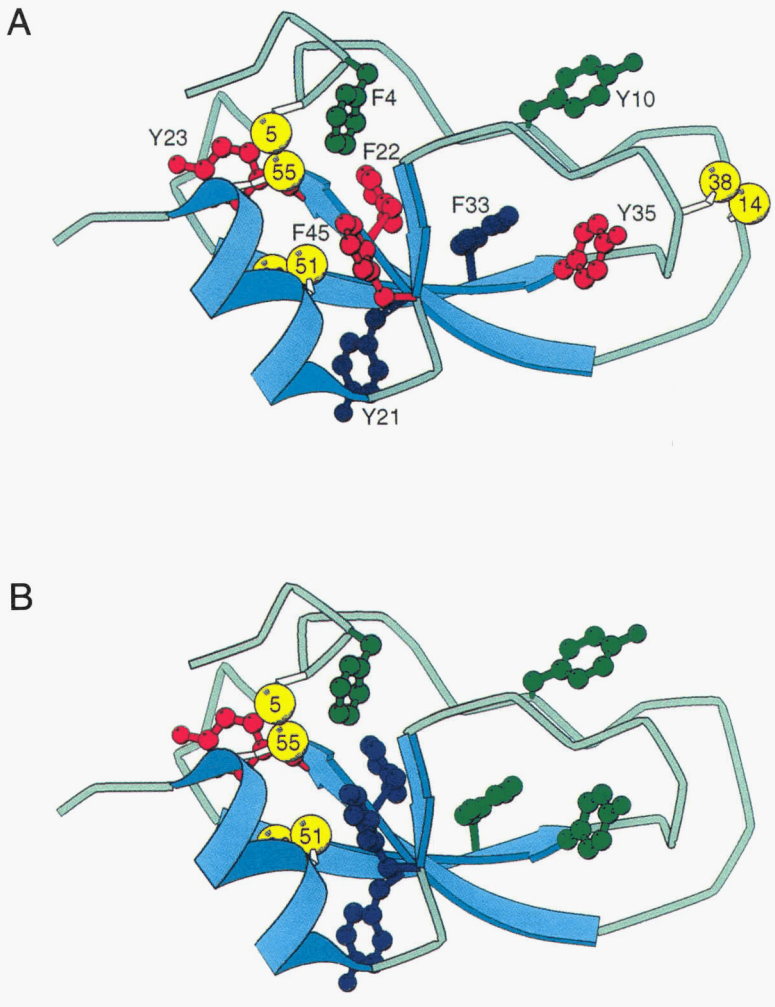


Fig. 6. Schematic representations of **(A)** native BPTI and **(B)** the native-like intermediate N_{SH}^{SH}, summarizing the effects of the aromatic → Leu amino acid replacements on the stabilities of these species with respect to the fully reduced protein. Both of the ribbon diagrams were drawn using 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 sites at which replacing an aromatic residue with Leu causes a destabilization of 5–7 kcal/mol are colored red. The sites of substitutions that lead to a destabilization of 1–3 kcal/mol are colored green, while intermediate effects are indicated by purple coloring.

Tyr 23, Tyr 35, and Phe 45, are also the residues that have been shown by Wagner et al. to have the most rigidly positioned side chains, as measured by ring-flip rates (Wagner et al., 1987). These three sites are also among the most highly buried and evolutionarily-conserved residues in the protein. These patterns are consistent with other evidence indicating that buried and rigid residues generally contribute more to protein stability than do more exposed and mobile sites (Creighton, 1983; Goldenberg, 1985; Alber et al., 1987; Alber, 1989; Goldenberg, 1992).

The immediate precursor of the native protein, N_{SH}^{SH}, was also significantly destabilized by all eight mutations, with the effect ranging from 1.2 kcal/mol (Y10L) to 6.1 kcal/mol (Y23L). The destabilizations of N_{SH}^{SH} were, however, all significantly smaller than those of the native protein, even though high-resolution NMR studies indicate that the intermediate has a folded conformation very similar to that of the form with all three disulfides (Naderi et al., 1991). In addition, the pattern of effects was quite different from that seen in the native protein and showed a clear correlation with the positions of the altered residues in the native protein (Fig. 6B). For substitutions F4L, Y21L, F22L, Y23L, and F45L, the destabilization of N_{SH}^{SH} corresponded to 80% or more of the effect of the same substitution on the native protein. In contrast, the destabilization of N_{SH}^{SH} caused by the other substations, which are all located near the 14–38 disulfide in the native protein, was less than 50% of the destabilization of the native protein, indicating that the side-chains of Tyr 10, Phe 33, and Tyr 35 make the majority of their stabilizing contributions only when the final disulfide of the native protein is formed.

The pattern of effects on N and N_{SH}^{SH} is consistent with the previous suggestion that the folded conformation of BPTI is composed of two substructures that can be partially decoupled energetically (Oas & Kim, 1988; Goldenberg et al., 1992). One region includes the α-helix and β-sheet of the native protein, as well as the 30–51 and 5–55 disulfides, while the other region consists of the two loop segments, linked by the 14–38 disulfide, that form the trypsin binding site. In the absence of the 14–38 disulfide, the stability of the folded structure is much less sensitive to amino acid replacements in the loop segments than it is after the final disulfide

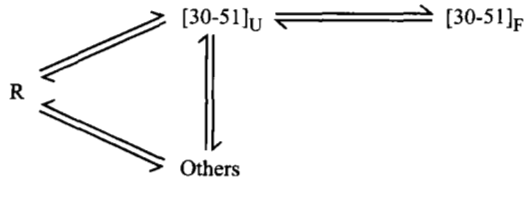
is formed, perhaps because this region is more flexible in N_{SH}^{SH} than in the native protein. Aromatic → Leu replacements in the α/β region, on the other hand, destabilize N_{SH}^{SH} to nearly the same extent as they do N, suggesting that this region is almost as rigid in the intermediate as it is in the native protein.

The interface between the α-helix and β-sheet also appears to be energetically most important for stabilizing the major one-disulfide intermediate, [30–51], as discussed below and in the accompanying paper. Thus, the region of the protein implicated in stabilizing the earlier intermediate retains its relative importance in the last intermediate, N_{SH}^{SH}. Once the final disulfide is formed, however, this distinction is lost, as residues distributed throughout the native protein make contributions that appear uncorrelated with their position in the overall structure.

Roles of aromatic residues in stabilizing early intermediates

In contrast to the large effects on N and N_{SH}^{SH}, the aromatic → Leu replacements caused only small decreases in the stability of the population of one-disulfide intermediates (Table 3). The largest effect was caused by the Y23L substitution, which destabilized population I by only 1.0 kcal/mol. The interpretation of these effects is somewhat more complex than for the case of the native protein or N_{SH}^{SH}, however, because the conformational stability of an individual intermediate in the population is not so tightly linked to the overall stability of the population. Even if a substitution greatly destabilizes the folded conformation of some of the molecules in a population, the major effect may be a redistribution of the various species, rather than a net destabilization of the population as a whole.

As a first approximation, the relationship between conformational stability of the major one disulfide intermediate, [30–51], and the measured stability of population I can be analyzed using the following scheme:



(5)

Among the one-disulfide intermediates, [30 51], represents approximately 60% of the steady-state population during the folding of the wild-type protein (Creighton, 1974; Weissman & Kim, 1991). Under the conditions of the experiments described here (25 °C, pH 8.7), the rest of the population is a mixture representing most or all of the 14 other possible one-disulfide species. In the scheme above, the folded form of [30–51] is in rapid equilibrium with the unfolded conformation and the other one-disulfide intermediates, which are assumed not to possess stable structure. The stability of I, as measured in the experiments described here, is defined by the equilibrium between R and the entire population of one-disulfide species. The conformational stability of [30–51], on the other hand, is defined by the equilibrium between $[30–51]_U$ and $[30–51]_F$, The more stable the folded form, the larger the fraction of [30–51] in the population of one-disulfide intermediates and the greater the stability of I.

Using the model described by Equation 5, the relationship between the destabilization of the population ($\Delta\Delta G_I$) and the change in conformational stability of [30–51] ($\Delta\Delta G_{[30-51]}$), can be derived, as outlined in the appendix. This relationship predicts that even large destabilizations of the folded conformation of [30–51] will lead to only small changes in the overall stability of the population of one-disulfide intermediates: If the equilibrium constant for the folding of [30–51] lies in the range from 10 to 100 (Creighton, 1988; Staley, 1993), the maximum expected destabilization of I is only about 0.5 kcal/mol. That the observed destabilization of I was somewhat larger than this for the Y23L mutant suggests that some other members of the population also contain structure that is destabilized by the substitution.

While the stability of the population is relatively insensitive to the conformational stability of [30–51], the relative concentration of this species in the population is expected to decrease substantially as the folded conformation of [30–51] is destabilized. Using the simplified model of Equation 5, the relationship between the concentration of [30–51], expressed as the fraction of the population, and its conformational stability is derived in the appendix. This relationship has been used to estimate the conformational destabilization of [30–51] caused by the aromatic → Leu substitutions, using the data presented in the accompanying paper, and these results are included in Table 3.

The estimated changes in the conformational stability of [30-51] ranged from essentially zero, for the F4L, Y10L, and Y35L substitutions, to more than 1.5 kcal/mol, in the case of Y23L. Because the latter substitution decreased the level of [30-51] to less than detectable levels in the population, this estimate represents a lower bound on the destabilization. The remaining substitutions all caused smaller, but measurable, decreases in the relative level of [30–51], corresponding to destabilizations of 0.8 to 1.5 kcal/mol. Although there is significant uncertainty in these calculations, they indicate that the stability of [30–51] is much less sensitive to amino acid replacements than is the native protein, even when the site of the substitution lies within the most ordered region of the intermediate.

These observations indicate that the folded portions of the intermediate are energetically quite distinct from the fully folded protein. The greater tolerance toward the mutational perturbations most likely reflects a lower degree of cooperativity among the individual stabilizing interactions within the partially folded molecules.

The overall stability of the population of two-disulfide intermediates, II, was also relatively insensitive to the amino acid replacements. As described in the accompanying paper, however, the accumulation of the major component of II, [30–51,14–38] was decreased significantly by the F4L substitution and was essentially eliminated by the others. Arguments similar to those applied to the one-disulfide intermediates lead to estimates of at least 1 to 3 kcal/mol for the destabilization of [30–51,14–38]. That all of the substitutions destabilized this species significantly is consistent with NMR studies indicating that it has a very native-like conformation (van Mierlo et al., 1991).

The other major two-disulfide intermediate, [5–55,14–38], is kinetically distinct from both the members of population II and N_{SH}^{SH}. As described in the accompanying paper, the accumulation of this kinetically-trapped species was eliminated at pH 8.7 by all of the amino acid replacements examined here, except the F4L substitution. For this mutant, the destabilization of [5–55,14–38] was estimated to be about 2 kcal/mol. This value serves as a minimum estimate for the destabilization of this species by the other aromatic → Leu replacements, indicating that all of the aromatic residues contribute significantly to the stability of this native-like species.

Characterization of the major transition states in the BPTI folding pathway

During the folding of wild-type BPTI, the slowest intramolecular steps are associated with the formation of N_{SH}^{SH}, from either the one-disulfide intermediates (I) via direct disulfide formation or from the two-disulfide intermediates (II) via intramolecular rearrangement (Creighton, 1977a; Creighton & Goldenberg, 1984; Goldenberg, 1988; Weissman & Kim, 1992, 1995). The reverse processes are also the slowest steps in unfolding under the experimental conditions typically used (0.1–10 mM DTT_{SH}^{SH}). In an earlier mutational analysis of the transition states, Mendoza et al. (1994) found that 18 amino acid replacements, distributed throughout the structure of folded structure of BPTI, each caused nearly parallel increases in the rates for the rearrangement and direct reduction of N_{SH}^{SH}. This pattern indicated that the transition states for the two processes were structurally similar, and it was suggested that the results were most easily explained by a model in which both transition states were extensively unfolded. The data presented here, which include both folding and unfolding rate constants, provide additional support for this model.

In Figure 7, the rate constants for direct reduction and rearrangement of N_{SH}^{SH} (k_{-dir} and k_{-rearr}, expressed as $RT\ln(k)$) are plotted versus the destabilization of N_{SH}^{SH}, relative to the fully reduced and unfolded protein. For both processes, there is a clear linear correlation between the logarithm of the rate constant and the destabilization of N_{SH}^{SH}. Further, the slopes of both correlation lines are close to unity (0.88 for rearrangement and 1.08 for direct reduction). Since the quantity $RT\ln(k)$ is related to the transition-state free energy change by a constant of addition, the correlations indicate that each of the amino acid replacements destabilizes N_{SH}^{SH} with respect to the two transition states by nearly the same amount as N_{SH}^{SH} is destabilized with respect to the unfolded protein.

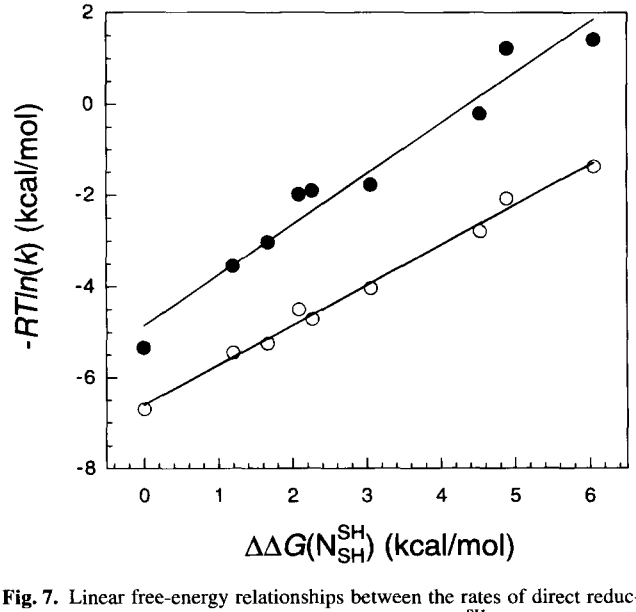


Fig. 7. Linear free-energy relationships between the rates of direct reduction (○) or intramolecular rearrangement (●) of N_{SH}^{SH} and the thermodynamic destabilization of this species. The rate constants are expressed in the form $RT\ln(k)$, which is related by a constant of addition to the corresponding difference in the free energies of N_{SH}^{SH} and the transition states. The least-squares line for the direct reduction rate constant has a slope of 1.1, while that for the rearrangement rate constant has a slope of 0.88. The correlation coefficients (R^2) for the two lines are 0.96 and 0.99, respectively.

It thus appears that the increase in the unfolding rates of N_{SH}^{SH} via either process can be accounted for almost entirely from the destabilization of N_{SH}^{SH}, rather than any stabilization of the transition states relative to the fully reduced state. Given the dispersed locations of the substituted residues in the native structure and the significant effects of the substitutions on the intermediates and the native protein, this pattern suggests that many, if not all, of the stabilizing interactions present in N_{SH}^{SH} are disrupted in the transition states, consistent with other evidence that the transition states are extensively unfolded (Weissman & Kim, 1992; Mendoza et al., 1994).

As noted previously (Mendoza et al., 1994), the major transition states for the BPTI folding pathway appear to be more extensively unfolded than those characterized for other protein folding reactions that do not involve disulfide formation and rearrangement (e.g., Serrano et al., 1992; Itzhaki et al., 1995; Daggett et al., 1996). This difference may reflect the additional steric constraints imposed by the thiol-disulfide chemistry. During unfolding, breaking either of the buried disulfides remaining in N_{SH}^{SH} appears to require disruption of all of the native-like structure in this species. Conversely, during folding, forming the buried 5–55 disulfide of the native protein, once the 30–51 is already present, may require extensive unfolding of any stable conformation in order to accommodate the chemical transition state, as discussed further in the following section.

Contributions of the rearrangement and direct disulfide-formation pathways to the folding of BPTI

One of the most striking features of the BPTI folding pathway is the role of intramolecular rearrangements in the formation of the three disulfides found in the native protein. The kinetically preferred folding mechanism involves at least one intramolecular disulfide rearrangement step, in which the immediate precursor of the native protein (N_{SH}^{SH}) is formed from the rearrangement of other two-disulfide intermediates (Creighton, 1977a; Goldenberg, 1988; Weissman & Kim, 1992). Recent studies, however, have generated some controversy about the significance and origins of the rearrangement process. When Weissman and Kim re-examined the distribution of BPTI folding intermediates, they found that the levels of intermediates with non-native disulfides were significantly lower than had been estimated previously (Weissman & Kim, 1991). In particular, [30–51,14–38] was found to be much more highly populated than the two non-native components of II, [30–51,5–14] and [30–51,5–38]. Since [30–51,14–38] and [5–55,14–38] are the only two-disulfide intermediates that accumulated significantly during the folding of the wild-type protein at neutral pH, it might appear that the rearrangements in BPTI folding arise only because of the stability of these species, neither of which can readily form a third disulfide (Weissman & Kim, 1992, 1995).

Our previous report (Zhang & Goldenberg, 1993) showed, however, that the rearrangement pathway is preferred even when the two kinetically trapped species are substantially destabilized by the Y35L substitution, indicating that the preference for the rearrangement mechanism is not solely the consequence of the stabilities of these intermediates. The experiments described here extend these observations to the seven other aromatic → Leu mutant proteins. For each of the BPTI variants examined, the rate of forming N_{SH}^{SH} from the one-disulfide species was at least 100-fold slower than that for the formation of other two disulfide species. The vast

majority of molecules thus form two-disulfide intermediates other than N_{SH}^{SH} and then must undergo intramolecular rearrangements to form this species, which is the only two-disulfide intermediate that readily forms a third disulfide.

The preference for the rearrangement pathway reflects not so much a strong tendency to form the non-native two-disulfide intermediates, which are formed at rates comparable to those observed for other disulfide-formation steps involving non-folded polypeptide segments, but rather the unusually low rate of directly forming N_{SH}^{SH} (Creighton, 1977a; Goldenberg, 1988; Darby et al., 1995). One factor that seems likely to contribute to the low rate of this reaction is the structure present in the major one-disulfide intermediate, [30–51] (van Mierlo et al., 1993; Staley & Kim, 1994). Although this structure appears to be very similar to that present in the native protein and is presumably compatible with the structure of the 5–55 disulfide, the transition-state for forming this disulfide may require significant perturbation of the structure of [30–51]. The transition-state for disulfide formation, and reduction, requires the placement of three sulfur atoms in the vicinity of the space occupied by two sulfur atoms in the di-thiol and disulfide ground states. In addition, the bond angles found in the transition-state are significantly different from those of the disulfide (Singh & Whitesides, 1993). Just as the structures of enzymes discriminate between the geometries of ground and transition states, preferentially stabilizing the latter, the structure present in [30–51] may tend to preferentially stabilize the ground state, leading to slow formation of the 5–55 disulfide.

The suggestion that the partially folded structure of [30–51] may inhibit direct formation of N_{SH}^{SH} is supported by our observation that the aromatic → Leu substitutions generally increased the rate of the step I → N_{SH}^{SH}. In contrast, all of the other disulfide-formation steps were either retarded or unaffected by the destabilizing substitutions. The increase in rate observed for the direct formation of N_{SH}^{SH} was, however, relatively small, even for the mutant proteins for which [30–51] was substantially destabilized. The small size of this effect might be explained by a lack of cooperativity in the unfolding of [30–51], leading to residual structure in even the most destabilized mutants. Also, the observed rate constant for the direct formation of N_{SH}^{SH} depends not only on the rate of the microscopic step, [30–51] → N_{SH}^{SH}, but also on the relative concentration of [30–51] in population I. The mutations that destabilize [30–51] also decrease the concentration of this species, thus tending to compensate for any increase in the microscopic rate constant.

Summary

The results presented in this and the accompanying paper provide an unusually detailed description of the energetic effects of amino acid replacements on a protein folding pathway, particularly when interpreted in light of the recent NMR studies of the disulfide-bonded folding intermediates. The eight aromatic → Leu substitutions examined destabilize to varying extents the native-like and partially-folded intermediates and generally lead to a broadening of the distribution of folding intermediates, as illustrated by the conformational funnels shown in the accompanying paper. The magnitude of the destabilization of a particular intermediate depends upon both the site of the substitution and the extent of folded structure in the intermediate, with the more native-like intermediates displaying the greatest sensitivity.

Although the substitutions cause substantial changes in the stabilities and distributions of the intermediates, some features of the pathway are relatively insensitive to the perturbations. For all of the variants, the predominant pathway involves intramolecular rearrangements of two-disulfide intermediates, even when the native-like intermediates are severely destabilized. Also, the eight variants, along with the wild-type protein, display a striking linear free-energy relationship between the stability of N_{SH}^{SH} and the rates of this species' direct reduction or intramolecular rearrangement. These patterns suggest that the preference for the rearrangement mechanism in the folding of the wild-type protein arises from conformational constraints imposed relatively early in the folding pathway and that the rearrangements involve extensive disruption of structure present in the early intermediates.

While some features of the BPTI folding pathway may reflect the special steric constraints imposed by the requirement to form disulfide bonds that are buried in the final folded structure, others may be more general. In particular, the increase in the energetic effects of the amino acid replacements as more of the polypeptide becomes ordered suggests that the cooperativity among individual interactions increases as the protein folds, a phenomenon that may be important in defining the folding mechanisms of many proteins.

Materials and methods

Preparation of protein variants

The mutant proteins used in these experiments were produced in recombinant *Escherichia coli* and purified as described in the accompanying paper.

Preparation of selectively reduced and alkylated proteins

For each of the BPTI variants that unfolded relatively slowly (F4L, Y10L, F22L, Y21L, F22L, F33L, and Y35L), the 14–38 disulfide of the native protein was selectively reduced by incubating 250 μM protein with 4 mM DTT_{SH}^{SH} for 1 min at 25 °C in the presence of 0.1 M Tris-HCl pH 8.7, 0.2 M KCl, and 1 mM EDTA. For the Y23L and F45L variants, which unfolded much more rapidly, 30 μM native protein was incubated with 20 mM DTT_{SH}^{SH} for 30 s in the presence of 0.1 M Tris-HCl pH 7.3, 0.2 M KCl, and 1 mM EDTA. The resulting thiols were blocked by adding Na-iodoacetate to a final concentration of 0.2 M and incubating for an additional 2 min. The modified proteins were then purified by reversed phase HPLC, and the identities of the modified Cys thiols were confirmed by peptide mapping, as described in the accompanying paper.

Measurement of refolding and unfolding kinetics

The conditions and procedures used to study the kinetics of unfolding and refolding of the BPTI variants were essentially identical to those described previously (Creighton & Goldenberg, 1984). Folding or unfolding reactions contained 0.1 M Tris-HCl pH 8.7, 0.2 M KCl, 1 mM EDTA, 30 μM protein and various concentrations of reduced and oxidized dithiothreitol. Oxidized dithiothreitol was purified as described by Creighton (1977b) to remove less stable disulfides that can increase the apparent rates of disulfide formation. Samples from refolding reactions were quenched by mixing 50 μL of the reaction mixture with 12.5 μL of a quenching solution containing 0.5 M Na-iodoacetate and 0.25 M Tris-HCl

pH 6.8. After a 2 min incubation at room temperature, half of the sample was put on ice, and half was mixed with 25 mg solid urea and allowed to react for an additional 20 min at room temperature. Samples from unfolding reactions were treated in the same way, except that the urea treatment was omitted. The trapped intermediates were separated by non-denaturing gel electrophoresis using a low-pH discontinuous buffer system (Reisfeld et al., 1962; Goldenberg, 1989). The relative concentrations of the native, reduced and partially disulfide-bonded forms were determined by video densitometry of the Coomassie-blue stained gels. Rate constants for the individual steps in folding and unfolding were estimated by comparing the observed time-dependent changes in the concentrations of the various species with those predicted by numerical integration of the rate expressions making up the folding model.

Appendix

Experiments of the type described in this paper provide a means of measuring the free energy differences among the kinetically-distinct populations making up a disulfide-coupled folding pathway. These populations are frequently made up of multiple species with different disulfide bonds, and even those molecules with the same disulfides may exist in multiple conformational states. The apparent stability of a population depends upon the free energies of the individual components and on the linked equilibria among these species (Creighton, 1988). These relationships are outlined here for the case of the one-disulfide intermediates in the BPTI folding pathway, providing equations that can be used to estimate the effects of amino acid replacements on the conformational stability of the major intermediate.

During the refolding of wild-type BPTI under the conditions used for the kinetic measurements (pH 8.7, 25 °C), [30–51] is the most abundant one-disulfide intermediate, representing approximately 60% of the population (Creighton, 1974; Weissman & Kim, 1991). No other single one-disulfide intermediate represents more than about 10% of this population. NMR studies of analogs of [30–51] have demonstrated that this intermediate has a partially folded conformation in which approximately 50% of the polypeptide chain has a conformation very similar to that of the native protein (van Mierlo et al., 1993; Staley & Kim, 1994). If this structure is assumed to undergo a two-state transition, an admittedly first-order approximation, the conformational stability of the intermediate can be expressed as

$$\Delta G_{[30-51]} = -RT \ln \frac{[[30-51]_U]}{[[30-51]_F]} = RT \ln(K_f^{wt}) \quad (A.1)$$

where $[30–51]_U$ and $[30–51]_F$ refer to the unfolded and folded forms of the intermediate, and K_f^{wt} is the equilibrium constant for folding [30–51]:

$$K_f^{wt} = \frac{[[30-51]_F]}{[[30-51]_U]}. \quad (A.2)$$

The equilibria linking these two forms with the other one-disulfide intermediates, collectively designated "Others," and the fully reduced protein are shown schematically in Equation 5 of the text.

The destabilization of the folded conformation of the intermediate caused by a mutation is given by

$$\Delta\Delta G_{[30-51]} = RT \ln\left(\frac{K_f^{wt}}{K_f^{mut}}\right), \quad (A.3)$$

where K_f^{mut} is the equilibrium constant for the folding of [30–51] for the mutant protein. This destabilization can be detected as both a decrease in the fraction of [30–51] in population I and a decrease in the overall stability of the population.

The fraction of [30–51] in the population of one-disulfide intermediates for the wild-type protein is given by

$$f_{[30-51]}^{wt} = \frac{[[30-51]_F] + [[30-51]_U]}{[\text{others}] + [[30-51]_F] + [[30-51]_U]}. \quad (A.4)$$

If the equilibrium constant between the unfolded form of [30–51] and the other one-disulfide intermediates is defined as

$$K_{other} = \frac{[[30-51]_U]}{[\text{others}]}, \quad (A.5)$$

then the fraction of [30-51] can be written as

$$f_{[30-51]}^{wt} = \frac{1 + K_f^{wt}}{1 + K_f^{wt} + K_{other}}. \quad (A.6)$$

An analogous expression can be written for the fraction of [30–51] for the mutant protein, $f_{[30-51]}^{mut}$. While K_f^{mut} is expected to be different from K_f^{wt}, K_{other} can be assumed, to a first approximation, to be unaffected by the amino acid replacement, since this term represents an equilibrium among predominantly unfolded forms.

By rearranging Equation A.3 and substituting in Equation A.6, the fraction of [30–51] in the one-disulfide intermediates can be expressed in terms of the destabilization of $[30–51]_F$ and the parameters for the wild-type protein:

$$f_{[30-51]}^{mut} = \frac{f_{[30-51]}^{wt}\left(1 + K_f^{wt}\cdot\exp\left(\frac{-\Delta\Delta G_{[30-51]}}{RT}\right)\right)}{1 + K_f^{wt} + K_f^{wt}\cdot f_{[30-51]}^{wt}\left(\exp\left(\frac{-\Delta\Delta G_{[30-51]}}{RT}\right) - 1\right)}. \quad (A.7)$$

In this expression, K_{other} does not appear because it is not independent of the other two terms in Equation A.6, K_f^{wt} and $f_{[30-51]}^{wt}$.

In order to apply Equation A.7, estimates of $f_{[30-51]}^{wt}$ and K_f^{wt} are required. While the fraction of [30–51] has been well-established to be about 0.6, there is more uncertainty with respect to the conformational equilibrium constant, K_f^{wt}. Using linkage arguments similar to those presented here, Creighton (1988) has estimated that K_f^{wt} lies in the range of 20–100. A similar range is predicted from amide hydrogen exchange rates measured in an analog of [30–51] (Staley, 1993).

In Figure A.1, Equation A.7 is used to calculate the fraction of [30–51] expected when the folded conformation of this species is destabilized by various amounts, assuming that $f_{[30-51]}^{wt} = 0.6$ and $K_f^{wt} = 10$, 20, or 100. As shown by the curves, destabilizing $[30–51]_N$ by 1 to 2 kcal/mol is expected to reduce the total steady-state concentration of [30–51] from 60% to about 10%, but further destabilization causes relatively small changes in the level of this species, since nearly all of the intermediate is already in the unfolded state. The curves in Figure A.1 were used to estimate the decrease in conformational stability of [30–51] caused by the aromatic → Leu substitutions, as listed in Table 3. Since [30–51] was detectable at a level of 15% or greater for all of the substitutions, with the exception of Y23L, the value of $\Delta\Delta G_{[30-51]}$ is estimated for these mutants to be less than 2 kcal/mol.

The same linkage relationships can also be used to estimate the change in the overall stability of the population of one disulfide intermediates associated with a substitution that destabilizes $[30–51]_N$, The equilibrium constant for forming the population of one-disulfide intermediates from the fully reduced protein is given by

$$K_I = \frac{[[30-51]_F] + [[30-51]_U] + [\text{others}]}{[R]}, \quad (A.8)$$

which for the wild-type protein can be written as

$$K_I^{wt} = \frac{[[30-51]_U]}{[R]}\cdot(1 + K_{other} + K_f^{wt}). \quad (A.9)$$

The observed destabilization of the population, relative the fully reduced protein, is

$$\Delta\Delta G_I = RT \ln\left(\frac{K_I^{wt}}{K_I^{mut}}\right) = RT \ln\left(\frac{1 + K_{other} + K_f^{wt}}{1 + K_{other} + K_f^{mut}}\right). \quad (A.10)$$

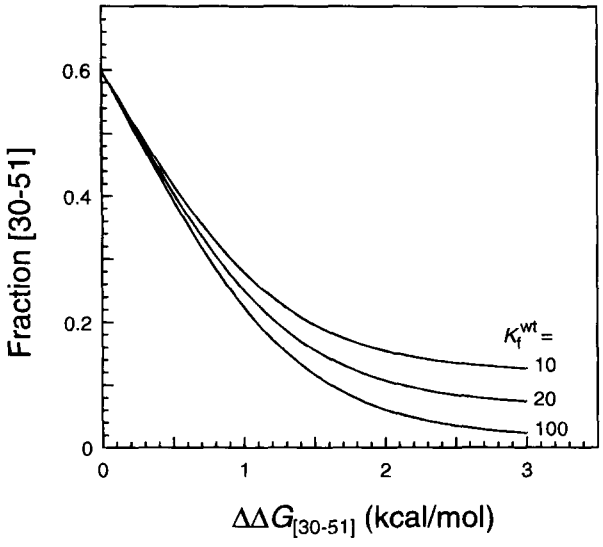


Fig. A.1. Predicted relationship between the destabilization of the folded conformation of [30–51], $\Delta\Delta G_{[30-51]}$, and the fraction of this intermediate present at steady state in the population of one-disulfide intermediates. The curves were calculated using Equation A.7, assuming that the fraction of [30–51] for the wild-type protein, $f_{[30-51]}^{wt}$, is 0.6 and that the equilibrium constant for the folding of the wild-type intermediate, K_f^{wt}, is 10, 20, or 100, as indicated.

By substituting Equations A.5 and A.6 into Equation A.10, the destabilization of population I can be expressed in terms of the destabilization of $[30-51]_N$:

$$\Delta\Delta G_I = -RT \ln\left(1 + \frac{f_{[30-51]}^{wt} \cdot K_f^{wt} \cdot \left(\exp\left(\frac{-\Delta\Delta G_{[30-51]}}{RT}\right) - 1\right)}{1 + K_f^{wt}}\right). \quad \text{(A.11)}$$

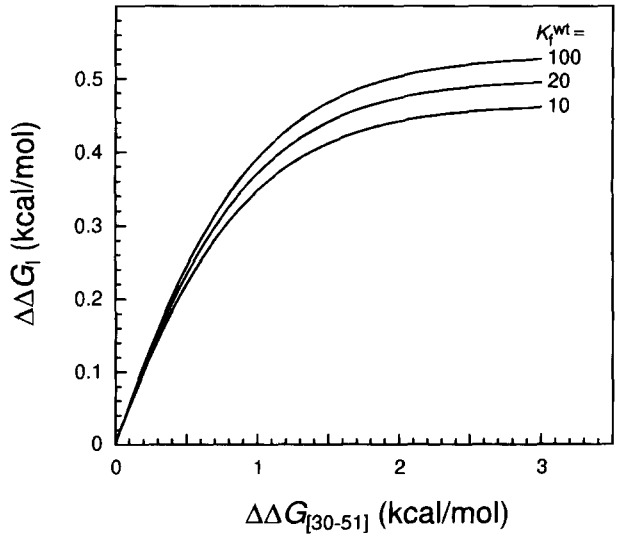


Fig. A.2. Predicted relationship between the destabilization of the folded conformation of [30–51], $\Delta\Delta G_{[30-51]}$, and the overall destabilization of the population of one-disulfide intermediates, $\Delta\Delta G_I$, The curves were calculated using Equation A.11, assuming that the fraction of [30–51] for the wild-type protein, $f_{[30-51]}^{wt}$, is 0.6 and that the equilibrium constant for the folding of the wild-type intermediate, K_f^{wt}, is 10, 20, or 100, as indicated.

Equation A.13 was used to generate the curves shown in Figure A.2, in which the expected destabilization of population I is plotted as a function of the destabilization of $[30-51]_N$, assuming that $f_{[30-51]}^{wt} = 0.6$ and $K_f^{wt} =$ 10, 20, or 100. As shown in the figure, the maximum expected destabilization of the population is only about 0.5 kcal/mol, even when a relatively high value of K_f^{wt} is assumed. The observed destabilization for the Y23L mutant was somewhat greater than this (1 kcal/mol), which may indicate that some of the intermediates other than [30–51] have folded conformations that are destabilized by the substitution.

Curves such as those shown in Figure A.2 could, in principle, be used to estimate the destabilization of the folded conformation of [30–51]. However, the measurements of $\Delta\Delta G_I$ are relatively imprecise, and it was felt that $\Delta\Delta G_{[30-51]}$ could be more reliably estimated from the effects of the mutation on the distribution of one-disulfide intermediates.

The model used to derive Equations A.9 and A.13 is based upon simplifying assumptions that introduce substantial uncertainty into any quantitative application of these relationships. In particular, the unfolding of [30–51] may be less cooperative than implied by Equation A.1, and species other than $[30-51]_N$ may have structure that is destabilized by amino acid replacements. Nonetheless, this analysis highlights the relatively weak nature of the thermodynamic linkage between the overall stability of a population of folding intermediates and the conformational stabilities of individual members of the population, a phenomenon that may influence the interpretation of many folding experiments.

In any experimental study of protein folding, intermediates are defined by their observable properties, and a folding intermediate should probably be thought of as a population of molecules that have in common some experimentally defined feature and are kinetically indistinguishable on the time scale of the experiments. In the case of the BPTI pathway, "I" represents a population of species that all have a single disulfide, while [30–51] represents a sub-population of molecules with a specific disulfide. In the case of a folding reaction that does not involve disulfide formation, an intermediate population might be defined, for instance, by the fluorescence of a Trp side chain or the protection of particular amide hydrogens from exchange with solvent. If the conformational stability of some of the molecules within a population is perturbed, as by an amino acid replacement or a change in solution conditions, the distribution of molecules within the population is likely to be altered. This change, in turn, is likely to be linked to a change in the overall stability of the population with respect to other experimentally defined states, such as the fully unfolded or native states. However, the overall stability of the population is likely to be less sensitive to the perturbation than is the stability of an individual component of the population, as shown in Figure A.2. For this reason, the effects of amino acid replacements on folding intermediates may tend to be masked as the members of a population are redistributed.

Acknowledgments

We thank 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.

Alber T, Sun D-P, Nye JA, Muchmore DC, Matthews BW. 1987. Temperature-sensitive mutations of bacteriophage T4 lysozyme occur at sites with low mobility and low solvent accessibility in the folded protein. *Biochemistry 26*:3754–3758.

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

Barrick D, Baldwin RL. 1993. The molten globule intermediate of apomyoglobin and the process of protein folding. *Protein Sci 2*:869–876.

Beasty AM, Hurle MR, Manz JT, Stackhouse T, Onuffer JJ, Matthews CR. 1986. Effects of the phenylalanine-22 → leucine, glutamic acid-49 → methionine,

glycine-234 → aspartic acid, and glycine-234 → lysine mutations on the folding and stability of the alpha subunit of tryptophan synthase from *Escherichia coli*. *Biochemistry 25*:2965–2974.
Creighton TE. 1974. The single-disulphide intermediates in the refolding of reduced pancreatic trypsin inhibitor. *J Mol Biol 87*:603–624.
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. Energetics of folding and unfolding of pancreatic trypsin inhibitor. *J Mol Biol 113*:295–312.
Creighton TE. 1983. An empirical approach to protein conformation, stability and flexibility. *Biopolymers 22*:49–58.
Creighton TE. 1988. On the relevance of non-random polypeptide conformations for protein folding. *Biophys Chem 31*:155–162.
Creighton TE. 1990. Protein folding. *Biochem J 270*:1–16.
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.
Daggett V, Li A, Itzhaki LS, Otzen DE, Fersht AR. 1996. Structure of the transition state for folding of a protein derived from experiment and simulation. *J Mol Biol 257*:430–440.
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.
Dill KA. 1990. Dominant forces in protein folding. *Biochemistry 29*:7133–7155.
Dobson CM, Evans PA, Radford SE. 1994. Understanding how proteins fold: The lysozyme story so far. *Trends Biochem Sci 19*:31–37.
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.
Garvey EP, Swank J, Matthews CR. 1989. A hydrophobic cluster forms early in the folding of dihydrofolate reductase. *Proteins 6*:259–266.
Goldenberg DP. 1985. Dissecting the roles of individual interactions in protein stability: Lessons from a circularized protein. *J Cell Biochem 29*:321–335.
Goldenberg DP. 1988. 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. 1992. Mutational analysis of protein folding and stability. In: Creighton TE, ed. *Protein folding*. New York: W. H. Freeman. pp 353–403.
Goldenberg DP, Berger JM, Laheru DA, Wooden S, Zhang JX. 1992. Genetic dissection of pancreatic trypsin inhibitor. *Proc Natl Acad Sci USA 89*:5083–5087.
Goldenberg DP, Frieden RW, Haack JA, Morrison TB. 1989. Mutational analysis of a protein-folding pathway. *Nature 338*:127–132.
Hughson FM, Barrick D, Baldwin RL. 1991. Probing the stability of a partly folded apomyoglobin intermediate by site-directed mutagenesis. *Biochemistry 30*:4113–4118.
Itzhaki LS, Otzen DE, Fersht AR. 1995. The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: Evidence for a nucleation-condensation mechanism for protein folding. *J Mol Biol 254*:260–288.
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.
Makhatadze GI, Privalov PL. 1995. Energetics of protein structure. *Adv Protein Chem 47*:307–425.
Marks CB, Naderi H, Kosen PA, Kuntz ID, Anderson S. 1987. Mutants of bovine pancreatic trypsin inhibitor lacking cysteines 14 and 38 can fold properly. *Science 235*:1370–1373.
Marmorino JL, Pielak GJ. 1995. A native tertiary interaction stabilizes the A state of cytochrome c. *Biochemistry 34*:3140–3143.
Matouschek A, Kellis JJ, Serrano L, Bycroft M, Fersht AR. 1990. Transient folding intermediates characterized by protein engineering. *Nature 346*:440–445.
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.
Miranker AD, Dobson CM. 1996. Collapse and cooperativity in protein folding. *Curr Opin Struct Biol 6*:31–42.
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.
Oas TG, Kim PS. 1988. A peptide model of a protein folding intermediate. *Nature 336*:42–48.
Peng ZY, Wu LC, Schulman BA, Kim PS. 1995. Does the molten globule have a native-like tertiary fold? *Philos Trans R Soc Lond B Biol Sci 348*:43–47.
Perry KM, Onuffer JJ, Gittelman MS, Barmat L, Matthews CR. 1989. Long-range electrostatic interactions can influence the folding, stability, and cooperativity of dihydrofolate reductase. *Biochemistry 28*:7961–7968.
Ptitsyn OB. 1995. Structures of folding intermediates. *Curr Opin Struct Biol 5*:74–78.
Reisfeld RA, Lewis UJ, Williams DE. 1962. Disk electrophoresis of basic proteins and peptides on polyacrylamide gels. *Nature 195*:281–283.
Serrano L, Matouschek A, Fersht AR. 1992. The folding of an enzyme. III. Structure of the transition state for unfolding of barnase analysed by a protein engineering procedure. *J Mol Biol 224*:805–818.
Singh R, Whitesides GM. 1993. Thiol-disulfide interchange. In: Patai S, Rappoport Z, eds. *The chemistry of sulphur-containing functional groups*. Chichester (England): John Wiley & Sons. pp 633–658.
Staley JP. 1993. Structural studies of early intermediates in the folding pathway of bovine pancreatic trypsin inhibitor. [Ph.D thesis]. Cambridge, Massachusetts: Massachusetts Institute of Technology.
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.
Tsuji T, Chrunyk BA, Chen X, Matthews CR. 1993. Mutagenic analysis of the interior packing of an alpha/beta barrel protein. Effects on the stabilities and rates of interconversion of the native and partially folded forms of the alpha subunit of tryptophan synthase. *Biochemistry 32*:5566–5575.
van Mierlo, CPM, Darby NJ, Keeler J, Neuhaus D, Creighton TE. 1993. Partially folded conformation of the (30–51) intermediate in the disulphide folding pathway of bovine pancreatic trypsin inhibitor. ^{1}H and ^{15}N resonance assignments and determination of backbone dynamics from 15N relaxation measurements. *J Mol Biol 229*:1125–1146.
van Mierlo, CPM, Darby NJ, Neuhaus D, Creighton TE. 1991. (14–38, 30–51) double-disulphide intermediate in folding of bovine pancreatic trypsin inhibitor: A two-dimensional ^{1}H nuclear magnetic resonance study. *J Mol Biol 222*:353–371.
Wagner G, Bruhwiler D, Wüthrich K. 1987. 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.
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: I. Effects of aromatic → leucine substitutions on the distribution of folding intermediates. *Protein Sci 6*:1549–1562.