Denaturant-Dependent Folding of Bovine Pancreatic Trypsin Inhibitor Mutants with Two Intact Disulfide Bonds†

Mark R. Hurle,‡ Cara B. Marks,§,‖ Phyllis Anne Kosen,‡ Stephen Anderson,*,§,⊥ and Irwin D. Kuntz*,‡

Department of Pharmaceutical Chemistry, University of California at San Francisco, San Francisco, California 94143, and Department of Cardiovascular Research, Genentech, South San Francisco, California 94080

Received September 14, 1989; Revised Manuscript Received December 14, 1989

ABSTRACT: The equilibrium and kinetic behavior of the guanidine hydrochloride (Gdn-HCl) induced unfolding/refolding of four bovine pancreatic trypsin inhibitor (BPTI) mutants was examined by using ultraviolet difference spectroscopy. In three of the mutants, we replaced the buried 30–51 disulfide bond with alanine at position 51 and valine (Val30/Ala51), alanine (Ala30/Ala51), or threonine (Thr30/Ala51) at position 30. For the fourth mutant, the solvent-exposed 14–38 disulfide was substituted by a pair of alanines (Ala14/Ala38). All mutants retained the 5–55 disulfide. Experiments were performed under oxidizing conditions; thus, both the unfolded and folded forms retained two native disulfide bonds. Equilibrium experiments demonstrated that all four mutants were destabilized relative to wild-type BPTI. However, the stability of the 30–51 mutants increased with the hydrophobicity of the residue substituted at position 30. Kinetic experiments showed that all four mutants contained two minor slow refolding phases with characteristics of proline isomerization. The specific behavior of the phases depended on the location of the disulfide bonds. The major unfolding/refolding phase for each of the 30–51 mutants was more than an order of magnitude slower than for Ala14/Ala38 or for BPTI in which the 14–38 disulfide bond was specifically reduced and blocked with iodoacetamide [Jullien, M., & Baldwin, R. L. (1981) *J. Mol. Biol. 145*, 265–280]. Since this effect is independent of the stability of the protein, it is consistent with a model in which the proper docking of the interior residues of the protein is the rate-limiting step in the folding of these mutants.

One generally accepted hypothesis of biochemistry is that the stable, native conformation of a globular protein is governed by its amino acid sequence (Anfinsen, 1973). To investigate the possible interactions that may be responsible for this dependence, some research groups have focused their attention on the consequences of single amino acid substitutions on protein structure and/or protein stability [for recent reviews, see Goldenberg (1988a) and Alber and Matthews (1988)]. The manner in which the sequence *directs* the rapid conformational change from the unfolded to the native state (the *folding mechanism*) can be investigated by using complementary approaches.

For many years, denaturant-induced unfolding and refolding experiments have been used as an in vitro model for investigating the energetics of protein folding (Kim & Baldwin, 1982; Jaenicke, 1980; Tanford, 1968, 1970). Further information on folding mechanisms can be obtained by examining the effect of single amino acid substitutions on protein stability and the kinetics of folding, thereby determining the role of a residue at a specific position. Proteins that have been analyzed in this manner include dihydrofolate reductase (Perry et al., 1987; Garvey & Matthews, 1989), the α subunit of tryptophan synthase (Beasty et al., 1986; Hurle et al., 1986), iso-1-(Ramdas & Nall, 1986) and iso-2-cytochrome *c* (White et al., 1987), and barnase (Matouschek et al., 1989).

Bovine pancreatic trypsin inhibitor (BPTI),[1] which contains three disulfide bonds and only 58 amino acid residues, has been the model system for examination of almost all aspects of protein structure, including neutron diffraction and X-ray crystallography (Wlodawer et al., 1987), two-dimensional NMR (Wüthrich et al., 1982), hydrogen exchange (Tüchsen & Woodward, 1985), energy minimization (Levitt, 1981), and molecular dynamics (Levitt & Sharon, 1987). A thorough investigation of the formation of native BPTI structure from a totally reduced state has been performed by Creighton and co-workers (Creighton, 1978; Goldenberg & Creighton, 1985).

One advantage of using a small protein such as BPTI as a model system for folding studies is that, in principle, experimental results and theoretical predictions can be directly compared. However, the energetics of folding garnered from "disulfide folding" experiments on BPTI are difficult to use in molecular calculations because of the connectivity differences between reduced ("unfolded") and oxidized ("folded") proteins. In contrast, calorimetric and denaturant-induced folding experiments determine energies that can easily be related to theoretical values, and which can be obtained over

† Supported by Genentech, Inc., and by NSF Grant DMB8606901, NIH Grant GM19267 (to I.D.K.), DARPA Grant N00014-86-K-0757, and a grant from Biotechnology Resources Education Program (Molecular Biology Institute, UCLA).

* Address correspondence to these authors.

‡ University of California at San Francisco.

§ Genentech.

‖ Present address: Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, England.

⊥ Present address: Center for Advanced Biotechnology and Medicine, Rutgers University, 675 Hoes Lane, Piscataway, NJ 08854.

[1] Abbreviations: Gdn-HCl, guanidine hydrochloride; BPTI, bovine pancreatic trypsin inhibitor; NMR, nuclear magnetic resonance; RCAM(14–38)BPTI, wild-type BPTI in which cysteines 14 and 38 have been specifically reduced and carboxyamidomethylated; Val30/Ala51, Ala30/Ala51, and Thr30/Ala51, BPTI in which cysteine 51 is replaced by alanine and cysteine 30 is replaced by valine, alanine, or threonine, respectively; Ala14/Ala38, BPTI in which cysteines 14 and 38 are replaced by alanines at positions 14 and 38; TCA, trichloroacetic acid; FPLC, fast protein liquid chromatography; EDTA, ethylenediaminetetraacetic acid; KP_i, potassium phosphate; AU, absorbance units.

0006-2960/90/0429-4410$02.50/0 © 1990 American Chemical Society

a wide range of temperature and pH conditions. Moreover, the phenomena highlighted by refolding of chemically denatured protein may be quite distinct from those exhibited by folding that involves disulfide bond formation; thus, additional aspects of the mechanism may be revealed. Although the folding mechanism of fully oxidized BPTI is assuredly not the folding mechanism of BPTI in vivo, the transition between native and unfolded protein is due to the same physical forces that govern protein folding in vivo (Kim & Baldwin, 1982).

Unfortunately, there are technical problems with the denaturant-induced unfolding of wild-type BPTI. With its three disulfide bonds, BPTI is extremely stable at neutral pH and is resistant to unfolding by common chemical denaturants (urea or guanidine hydrochloride) or temperatures under 90 °C (Kassell, 1970; Moses & Hinz, 1983). Thus, the value of wild-type BPTI as a model system for folding studies of this type is limited. However, the stability of BPTI is decreased substantially upon specific reduction and subsequent carboxymethylation (RCOM) or carboxyamidomethylation (RCAM) of the 14–38 disulfide bond (Vincent et al., 1971).

The calorimetric (Schwarz et al., 1987) and denaturant-induced (Vincent et al., 1971; Jullien & Baldwin, 1981) folding of the RCAM(14–38)BPTI protein has previously been examined. In this paper, we examine the folding and stability of four mutants of BPTI that lack either the buried 30–51 disulfide (Val30/Ala51, Ala30/Ala51, and Thr30/Ala51) or the solvent-exposed 14–38 disulfide (Ala14/Ala38). Results from such experiments can test predictions (Levitt, 1981) made concerning the folding mechanism of BPTI.

Materials and Methods

Materials

Bacterial Strains and Plasmids. The constructions of the plasmids encoding the gene for the BPTI mutants Cys30/Cys51 → Ala30/Ala51 (p481) and Cys14/Cys38 → Ala14/Ala38 (p475) have been described (Marks et al., 1987a,b). These plasmids are pBR322 derivatives containing genes that confer ampicillin and tetracycline resistance and have been transformed into *Escherichia coli* (Marks et al., 1987b). The BPTI gene is preceded by an alkaline phosphatase promoter and STII signal sequence (Chang et al., 1987), resulting in secretion of the mature form of BPTI, with correct processing of the signal peptide. The plasmids for the genes encoding the BPTI mutants Cys30/Cys51 → Val30/Ala51 (p627) and Thr30/Ala51 (p631) were constructed in a similar manner. Amino acid analysis was consistent with the mutations determined by DNA sequencing.

Purification. The purification procedure is a modification of that described elsewhere (Marks et al., 1986). The reduced stability of the BPTI mutants required that the initial trichloroacetic acid (TCA) treatment of the crude extracts be performed at 2.0% (instead of 2.5%). Extracts containing the less stable Thr30/Ala51 mutant (see Results) were treated with 0.5% TCA. The resulting supernatant was then titrated to pH 7.5 with sodium hydroxide and loaded onto a chymotrypsin–Affi-Gel affinity column (Affi-Gel 10 obtained from Bio-Rad, Inc.) equilibrated in 0.1 M triethanolamine, 0.3 M NaCl, and 10 mM $CaCl_2$ at a flow rate not exceeding 2 mL/min. The protein was eluted with 0.25 M KCl/HCl at pH 1.8–2.0. Typically, BPTI coeluted with a light yellow contaminant, which was eliminated by Sephadex G-25 size exclusion chromatography in 50 mM ammonium bicarbonate, pH 7.8.

The resulting protein solution (>95% BPTI) was lyophilized and subsequently dissolved in 50 mM ammonium acetate, pH 4. Filtration through a 0.2-μm filter preceded loading onto a Mono-S cation-exchange FPLC column (Pharmacia). BPTI was eluted by an ammonium acetate gradient from 50 mM to 5 M (in acetate) at pH 4 and subsequently lyophilized. BPTI eluted as a single peak on FPLC, as well as on a reverse-phase HPLC column (M. Randal, personal communication). It was thus highly likely that all mutants contained both remaining disulfides. The lyophilized protein, which contained a small amount of acetate detectable by NMR, was dissolved in 50 mM ammonium bicarbonate, pH 7.8, and desalted by using a Sephadex G-25 or Bio-Rad PD-10 column equilibrated in the same buffer. After lyophilization, BPTI was stored sealed as a powder at 20 °C. All folding experiments were generally performed within 2 weeks after final purification.

Folding experiments were performed in 10 mM potassium phosphate, pH 7.8, and 0.2 mM EDTA, 15 °C (standard buffer). Protein concentration was determined by using the extinction coefficient ϵ_{280} = 5400 M^{-1} cm^{-1} [0.80 $(mg/mL)^{-1}$ cm^{-1}] for wild-type protein (Kassell, 1970). This value was confirmed for the Val30/Ala51 protein by examining the absorbance of a measured amount of lyophilized protein dissolved in standard buffer and for the Ala30/Ala51 and Ala14/Ala38 mutants by amino acid analysis of a protein solution with known absorbance and norleucine added as an internal control. Subsequent concentrations were determined at the absorbance maximum at 277 nm, with the extinction coefficient ϵ_{277} = 5720 M^{-1} cm^{-1} [0.85 $(mg/mL)^{-1}$ cm^{-1}].

Chemicals. Guanidine hydrochloride (Gdn-HCl) (ultrapure grade) was purchased from Schwarz/Mann and used without further purification. Desiccation was performed before weighing. All other chemicals were of reagent grade.

Methods

Equilibrium Experiments. UV difference spectra were acquired by using the tandem cell technique (Herskovits, 1967) with a double-beam Kontron Uvikon 860 spectrophotometer (Research Instruments International, San Diego, CA). Solutions were prepared by adding, by microliter syringe, 100 μL of protein solution in standard buffer (see Materials) to Corex tubes containing various concentrations of Gdn-HCl in 700 μL of standard buffer, resulting in 800-μL solutions of ~1 mg/mL protein in the appropriate Gdn-HCl concentrations. Solutions were incubated >90 min (≫5 half-lives) to ensure that equilibration had occurred. The samples were then transferred to a single half-cell of the sample cuvette, with the second half-cell containing 800 μL of standard buffer. The two half-cells of the reference cuvette contained protein solution in standard buffer and a Gdn-HCl solution of the appropriate concentration with no protein.

The difference spectrum of the two cells was then recorded over the range 350–250 nm. The virtually flat absorbance change between 350 and 320 nm could be extended to intercept with the isosbestic point at ~255 nm. Spectra that did not meet this condition (~5% of all spectra) were eliminated from further consideration. For each spectrum, this base line was determined, and the difference in absorbance between the peak at 287 nm and the base line was recorded as the absorbance of the solution. The absorbance was then converted to an extinction coefficient by using Beer's law, and the dependence of the observed extinction coefficient, ϵ_{OBS}, on Gdn-HCl concentration was determined.

Equilibrium parameters were derived by fitting the data to the equation for a two-state unfolding process N ⇌ U, $F_{app} = K/(1 + K)$. In this equation, F_{app} represents the apparent fraction of unfolded material, defined as $F_{app} = (\epsilon_{OBS} - \epsilon_N)/(\epsilon_U$

$- \epsilon_N$), where ϵ_N and ϵ_U are the extinction coefficients of the native and unfolded forms, respectively. The equilibrium constant, $K = [U]/[N]$, is described by a change in free energy, ΔG, that varies linearly with denaturant concentration (Pace, 1975; Schellman, 1978):

$$\Delta G = \Delta G^\circ - A[\text{Gdn-HCl}]$$

The quantity ΔG° represents the difference in free energy between the native and unfolded forms in the absence of denaturant at the experimental temperature and pH, while the parameter A represents the dependence of the free energy of unfolding on denaturant concentration. The midpoint of the transition, C_M, is calculated by setting $\Delta G = 0$. The displacement of the transition, $\Delta\Delta G$, due to a mutation is determined at the reference midpoint by substituting the value of the reference midpoint along with the ΔG° and A values for the mutant into the above equation (Cupo & Pace, 1983). For example, if the mutant is Ala30/Ala51 and the reference protein is Val30/Ala51, the value of $\Delta\Delta G$ is (using Table I, see Results)

$$\Delta\Delta G = 6.0 - 1.20(5.41) = -0.5 \text{ kcal/mol}$$

This value of $\Delta\Delta G$ demonstrates that the mutant Ala30/Ala51 is 0.5 kcal/mol less stable than the reference protein (Val30/Ala51) at the midpoint of denaturation of the reference protein (5.41 M).

Errors are standard deviations reported by the fitting procedure except where otherwise noted.

Kinetic Experiments. As an initial check, tandem cells were used in kinetic unfolding and refolding experiments on the Val30/Ala51 protein, and difference spectra were taken after completion of the reaction. As the difference spectra recorded in both the unfolding and refolding directions were virtually identical with the corresponding equilibrium spectra, reversibility was established. This was confirmed by the observation that refolded Val30/Ala51 demonstrated >90% of its original binding to trypsin (data not shown). To further increase the observed signal, therefore, all subsequent kinetic experiments used full 1-cm cells.

Unfolding reactions were started by an 8-fold dilution of protein, using microliter syringes, into a cuvette containing various amounts of Gdn-HCl so that the desired final Gdn-HCl concentration in a final volume of 800 μL was achieved. In refolding, the protein was initially dissolved in a 8 M Gdn-HCl solution and incubated 30 min at 25 °C and >60 min at the final temperature (25 or 15 °C). This was done to ensure complete unfolding, as the relaxation time is less than 30 s at this Gdn-HCl concentration (see Results), as well as complete equilibration of any multiple unfolded forms (slowest $\tau \sim 500$ s at 25 °C, 1500 s at 15 °C). Refolding reactions were then started by an 8-fold dilution of this protein/Gdn-HCl solution into the cuvette. Upon dilution, the solutions were manually mixed, and the time dependence of the absorbance change was recorded by an Wyse PC286 computer (Wyse Technology, San Jose, CA) interfaced to the spectrophotometer. The final protein concentration was 0.5–1.5 mg/mL (~150 μM).

The resulting data were transferred to a VAX 8650 and analyzed by using the nonlinear least-squares fitting program from RS/1 (BBN Software Products Corp., 1987). Kinetic data were fitted to the equation $A(t) = \sum_i A_i(i) \exp(-t/\tau_i) + A_\infty$, where $A(t)$ is the absorbance at time t, A_i is the absorbance of phase i at zero time, τ_i is the relaxation time for phase i, and A_∞ is the absorbance at infinite time. The data were initially fitted to one, two, or three exponential phases, depending on the conditions of the experiment. Final determination of the number of phases depended on the behavior of the residuals from various fits (Motulsky & Ransnas, 1987).

Table I: Equilibrium Free Energies of Unfolding for BPTI Mutants[a]

mutant	$\Delta G^{\circ b}$	A^c	$C_M{}^d$	$\Delta\Delta G^{b,e}$
V30A51	7.8 ± 0.7	1.5 ± 0.1	5.41	
A30A51	6.0 ± 0.7	1.20 ± 0.06	4.99	−0.5
T30A51	4.6 ± 0.2	1.21 ± 0.05	3.81	−2.0
A14A38	6.9 ± 0.3	1.28 ± 0.06	5.38	0.0

[a] Parameters represent fits to data of Figure 1 (see Methods). Errors represent standard deviations. [b] kcal (mol of protein)$^{-1}$. [c] kcal (mol of protein)$^{-1}$ (M Gdn-HCl)$^{-1}$. [d] Molar Gdn-HCl. [e] Negative values denote lower stability.

Results

Equilibrium Unfolding

Unfolding was monitored by UV difference spectroscopy in the region between 250 and 350 nm; since BPTI contains four tyrosine residues, changes in the extinction coefficient at 287 nm reflect perturbations of its tertiary structure. Attempts to obtain satisfactory urea-induced equilibrium unfolding data for the Val30/Ala51 BPTI mutant in standard buffer (10 mM KP_i and 0.2 mM EDTA, pH 7.8) were unsuccessful, as the protein did not appear to be denatured significantly by 9 M urea at 25 °C (data not shown).

However, the ϵ_{287} of Val30/Ala51 decreased substantially upon addition of large amounts of Gdn-HCl. Further investigation showed that the sigmoidal transition of $\Delta\epsilon_{287} \sim -950$ M^{-1} cm^{-1} appeared to be complete above 6.5 M Gdn-HCl. The kinetics of unfolding and refolding (see below) were slower and thus more easily observed at 15 °C than at 25 °C; therefore, conditions of standard buffer at 15 °C were used for more complete studies. Reversibility was then established by demonstrating that the transition curves of Val30/Ala51 BPTI unfolded and refolded to the same Gdn-HCl concentration were coincident (Figure 1A).

The Gdn-HCl-induced equilibrium unfolding data for the BPTI mutants Val30/Ala51, Ala30/Ala51, and Thr30/Ala51 at 15 °C are shown in Figure 1A and for the Ala14/Ala38 mutant in Figure 1B. For ease of comparison, the data are represented by the normalized term F_{app} defined as the apparent fraction of unfolded material at various Gdn-HCl concentrations (see Methods). The total amplitude of unfolding for the four mutants differed by <10%, suggesting that the tertiary and unfolded structures of all four mutants had similar tyrosine environments. The BPTI mutants Val30/Ala51 and Ala14/Ala38 unfold at higher denaturant concentrations than the mutants Ala30/Ala51 and Thr30/Ala51.

The unfolding of most single-domain proteins appears to follow a simple two-state reaction, $N \rightleftharpoons U$, where N and U represent native and unfolded forms, respectively (Tanford, 1968). The free energy change of this reaction can be approximated by a linear dependence on denaturant concentration (Pace, 1975; Schellman, 1978). For all four BPTI mutants, the free energy of denaturation is linearly dependent on the Gdn-HCl concentration throughout the transition zone for all four mutants (data not shown). Thus, in the absence of conflicting evidence, the transition can be treated as a two-state reaction. Table I displays the calculated values for the change in free energy in the absence of denaturant, ΔG°, the dependence of this change on [Gdn-HCl], A, and the midpoint of the transition, C_M.

The simplest method of evaluating mutants of a single protein is to compare the midpoints of unfolding (Cupo & Pace, 1983). It can be seen that the Ala14/Ala38 and Val30/Ala51 mutants have virtually identical midpoints. However, the residue at position 30 (Val, Ala, or Thr) has a

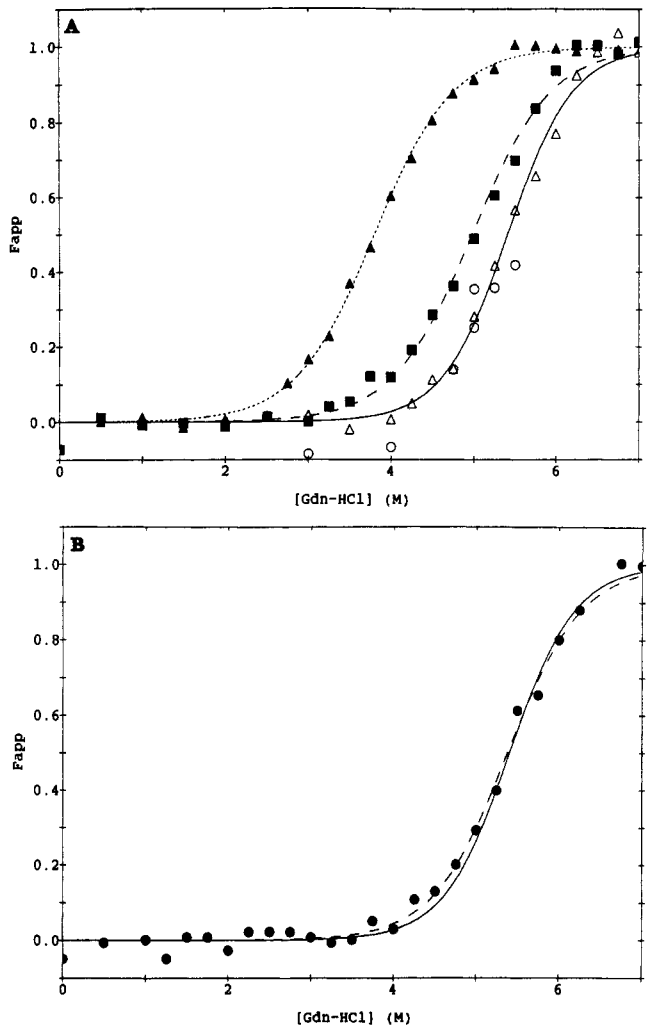


FIGURE 1: Fraction of the total change in extinction coefficient at 287.5 nm, F_{app}, as a function of Gdn-HCl concentration for the BPTI mutant proteins (A) Val30/Ala51 (△, —), Ala30/Ala51 (■, - - -), and Thr30/Ala51 (▲, ···) and (B) Ala14/Ala38 (●, - - -) and Val30/Ala51 (—). Buffer was 10 mM potassium phosphate, pH 7.8, and 0.2 mM EDTA, 15 °C. Lines represent fits to a two-state model as described in the text. Open circles in (A) represent points obtained from difference spectra after kinetic refolding experiments for the Val30/Ala51 protein.

significant effect on the stability of BPTI mutants containing alanine at position 51. Replacement of valine by alanine or threonine decreases the midpoint of unfolding by 0.4 and 1.6 M Gdn-HCl, respectively. The quantity A represents the dependence of the free energy change on Gdn-HCl; thus, the product of A and the change in midpoint gives the free energy change due to mutation, $\Delta\Delta G$. The values of A are substantially the same for all of the mutants except Val30/Ala51 (Table I). The 25% larger value of A for the Val30/Ala51 mutant is probably not especially significant within 2 M of the Val30/Ala51 midpoint. By use of the midpoint of the Val30/Ala51 mutant as reference, it can be calculated that the Ala30 and Thr30 substitutions destabilize the protein by 0.5 and 2.0 kcal/mol, respectively.

Kinetics of Folding

The dependencies of the relaxation times of unfolding and refolding on the final concentration of Gdn-HCl were obtained by using manual mixing methods for each of the four BPTI mutants. Unfolding experiments were performed by raising [Gdn-HCl] from 0 to the final concentration and observing the transient response, while in refolding experiments, [Gdn-HCl] was reduced from 8 M to the final concentration.

30–51 Mutants. The relaxation times of unfolding and refolding of the BPTI mutants lacking the 30–51 disulfide bond, Val30/Ala51, Ala30/Ala51, and Thr30/Ala51, were found to be dependent on the final Gdn-HCl concentration in a manner that was independent of the substitution at position 30 (Figure 2, parts A–C, respectively).

Two slow phases of refolding, τ_1 and τ_2, were observed at low concentrations of Gdn-HCl. Their relaxation times appeared to be independent of Gdn-HCl concentration and were virtually identical for all three mutants, with values of 1500 and 260 s, respectively. The amplitudes of these two phases accounted for 10% and 14%, respectively, of the total absorbance change due to unfolding. These amplitudes were independent of Gdn-HCl concentration at low concentrations corresponding to the native base line (Figure 1); at higher Gdn-HCl concentration, the amplitudes decreased along with decreasing population of the native form at equilibrium (data not shown). To further characterize the τ_1 and τ_2 relaxation times, the activation enthalpies for refolding of the Val30/Ala51 protein to 2 M Gdn-HCl were determined, yielding 27 ± 4 and 27 ± 3 kcal/mol, respectively, for the range 11–25 °C (data not shown). The relaxation times were also found to be independent of protein concentration in the range 0.5–1.5 mg/mL (data not shown).

An additional refolding phase, τ_3, was seen at intermediate Gdn-HCl concentrations. The relaxation time of this phase appeared to reach a maximum near the equilibrium midpoint for each mutant, becoming virtually Gdn-HCl-independent at higher denaturant concentrations. The amplitude of this phase dominated the total observed amplitude at these concentrations, and, within experimental error, accounted for the remainder of the total extinction coefficient change expected from the equilibrium transition. The amplitude increased with denaturant concentration, reaching a maximum near the unfolding midpoint, and decreased at higher concentrations (data not shown). It should be noted that the rapid nature of the τ_3 phase makes the estimation of its amplitudes and relaxation times at concentrations below the midpoint of unfolding extremely questionable.

Unfolding experiments demonstrated two phases that completely accounted for the amplitude expected from the equilibrium studies. The faster of these phases showed an identical relaxation time with the τ_3 phase of refolding, demonstrating that this phase reflects a kinetically reversible reaction (Utiyama & Baldwin, 1981). This phase accounted for 85% of the total amplitude expected from equilibrium results. The slower of the two phases, τ_4, containing the remaining 15% of the expected total amplitude, exhibited a relaxation time that decreased sharply with increasing Gdn-HCl concentrations. The contrasting nature of these two unfolding phases results in similar relaxation times at high Gdn-HCl concentrations (Figure 2). The fitting program (see Methods) discriminated only one phase at these concentrations; however, the fit was relatively poor. Thus, any dependence of amplitude on denaturant concentration is difficult to determine; however, it appeared that the relative amplitudes of the two phases were independent of denaturant concentration (data not shown). It is important to note that for all three mutants lacking the 30–51 disulfide bond the amplitude observed in equilibrium unfolding appeared to be accounted for by the observed kinetic phases.

Ala14/Ala38 Mutant. The Gdn-HCl-induced unfolding and refolding of RCAM(14–38)BPTI, in which the 14–38 disulfide bond has been selectively reduced and carboxyamidomethylated, has been previously examined (Jullien &

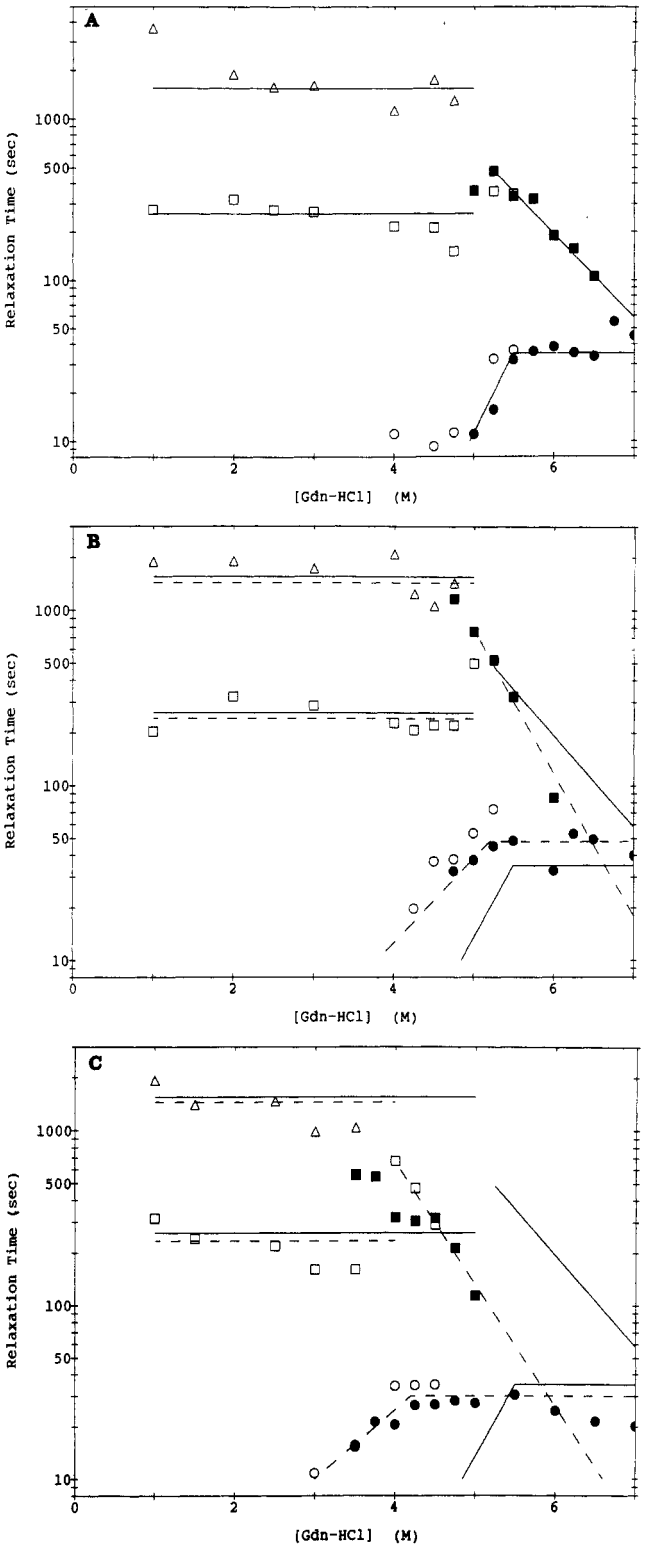


FIGURE 2: Relaxation times at the indicated final Gdn-HCl concentration for the three refolding (○, □, △) and two unfolding (●, ■) phases for (A) Val30/Ala51, (B) Ala30/Ala51, and (C) Thr30/Ala51 BPTI mutant proteins. All experiments were performed in the buffer described in Figure 1. The solid lines represent the values found for the Val30/Ala51 protein. Dashed lines in (B) and (C) represent the values found for the Ala30/Ala51 and Thr30/Ala51 mutants, respectively. The lines represent averages for the denaturant-independent relaxation times of the intermediate and slow refolding phases and are drawn to aid the eye for the slow unfolding phase and the fast unfolding/refolding phase.

Baldwin, 1981). Refolding monitored by absorbance revealed two slow relaxation times, which were proposed to involve proline isomerization (Levitt, 1981; Jullien & Baldwin, 1981).

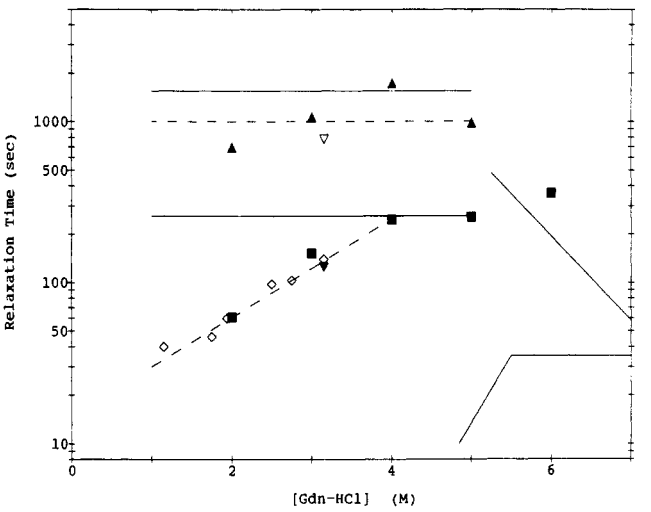


FIGURE 3: Relaxation times at the indicated final Gdn-HCl concentration for the two refolding phases (■, ▲) of the Ala14/Ala38 BPTI mutant. All experiments were performed in the buffer described in Figure 1. The solid lines and dashed lines represent the values found for the Val30/Ala51 (Figure 2A) and Ala14/Ala38 proteins, respectively. Lines are drawn to represent the average denaturant-independent relaxation time of the slow phase and the apparent denaturant-dependent relaxation time of the intermediate phase. Additional points from refolding of RCAM(14–38) monitored by absorbance (▼, ▽) or fluorescence (◇) are taken from Jullien and Baldwin (1981).

Table II: Relaxation Times and Amplitudes of Refolding of BPTI Mutants[a,b]

disulfide eliminated	τ_1[c]	τ_2[c]	τ_3[c]	A_1[d]	A_2[d]	A_3[d]
30–51	1500	260	10–50[e]	10	14	~75
14–38	1150	50–300[e]	n.o.[f]	10	15	n.o.[f]

[a] Buffer used is the same as in Figures 1–3. [b] Relaxation times and amplitudes are independent of [Gdn-HCl] unless otherwise indicated. [c] Relaxation times, τ_i, for phase i, in seconds. [d] Amplitudes, A_i, for phase i, expressed as a percentage of the total expected from equilibrium studies. [e] [Gdn-HCl] dependent. [f] Refolding phase not observed.

Unfolding was too rapid to measure by manual mixing techniques.

The Gdn-HCl-induced refolding relaxation times of the Ala14/Ala38 mutant are shown in Figure 3. Two slow relaxation times, τ_1 and τ_2, were observed in the 2–5 M Gdn-HCl range. The slowest relaxation time, τ_1, appeared to be independent of [Gdn-HCl], with a value of 1150 s. The faster relaxation time, τ_2, appeared to depend on [Gdn-HCl] and ranged from 60 to 360 s. The τ_1 and τ_2 phases accounted for 10% and 15% of the total amplitude expected from equilibrium measurements, respectively. Although the relaxation time of the τ_2 phase is now Gdn-HCl dependent, the amplitudes of both the τ_1 and τ_2 phases remain unchanged from those of the 30–51 mutants (Table II). The Gdn-HCl-dependent τ_3 phase seen for the 30–51 mutants was too rapid to be observed in refolding of the Ala14/Ala38 mutant using manual mixing techniques.

Like RCAM(14–38)BPTI (Jullien & Baldwin, 1981), unfolding of the Ala14/Ala38 BPTI mutant was more than an order of magnitude faster than unfolding of the 30–51 BPTI mutants. Unfolding jumps at 15 °C were virtually complete within 10 s, as <2% of the expected change was observed after this time.

Discussion

The results presented above demonstrate that mutants of BPTI which lack one of the three wild-type disulfide bonds

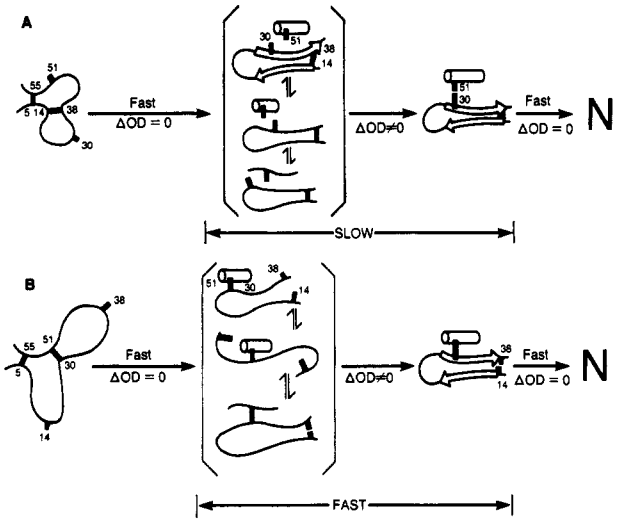


FIGURE 4: Proposed folding model for disulfide mutants of BPTI, as implied by the framework model. Note that this model does not take proline isomerization into account. (A) Molecules lacking the 30–51 disulfide bond fold rapidly to a collapsed state in which the improper arrangement of the interior is most likely. Rearrangement to the proper orientation occurs slowly with a corresponding change in extinction coefficient. Further minor adjustments are then made quickly, with no extinction coefficient change. (B) The 14–38 mutants, which contain the 30–51 disulfide, fold with very little rearrangement of the interior necessary. Note that the absence of the 14–38 disulfide bond implies a reduction in the likelihood of β-sheet formation in the collapsed state.

can be reversibly unfolded by Gdn-HCl. A schematic drawing of BPTI (Figure 4) shows that the molecules lack either the solvent-exposed 14–38 disulfide bond (Ala14/Ala38) or the buried 30–51 disulfide (Val30/Ala51, Ala30/Ala51, Thr30/Ala51). The 14–38 disulfide connects the ends of the two major β strands in BPTI, while the 30–51 connects the major α helix with one of the strands. The folding kinetics of the two sets of molecules are somewhat different, and the stabilities of the 30–51 molecules span a 2 kcal/mol range. The possible reasons for these effects are discussed below.

Effects Due to Mutation at Position 30

The stability and folding kinetics of BPTI mutants are directly affected by single amino acid replacements at position 30. For example, the Val30→Thr substitution results in a 2 kcal/mol destabilization. Crystallographic (Wlodawer et al., 1987) and two-dimensional NMR (Wüthrich et al., 1982) studies have shown that the 30–51 disulfide bond of wild-type BPTI is buried. Preliminary analysis of the Ala30/Ala51 mutant from both X-ray (C. Eigenbrot, M. Randall, and A. Kossiakoff, personal communication) and NMR (C. Eads and I. Kuntz, unpublished results) studies suggests that there are no major structural perturbations upon mutation. It thus appears that this energetically sensitive position is buried in the folded structure of the 30–51 mutants.

Temperature-sensitive mutations of T4 lysozyme, selected by genetic techniques, were generally found to occur in positions of low side-chain mobility and low solvent accessibility (Alber et al., 1987). The characteristics of residue 30 of BPTI are similar. Mobility of the side chains can be estimated by using the isotropic thermal *B* factors from X-ray crystal studies and averaging over the β and γ positions. The values for wild-type BPTI, which contains the 30–51 disulfide, are 12.5 Å^2 for Cys30 and 11 Å^2 for Cys51, significantly less mobile than the identical atoms of the average BPTI residue (16.3 Å) (data not shown). It should be noted that inclusion of all side-chain atoms, as was done for T4 lysozyme (Alber et al., 1987), would increase the average mobility significantly.

Table III: Relationship between Kinetics[a] and Equilibria[b] of 30–51 Mutants

mutant	$\tau_3^{c,d}$(ref)	$\tau_4^{c,e}$(unf)	C_M^f
V30A51	5[g]	200	5.41
A30A51	15	100	4.99
T30A51	30	20	3.81

[a] Gdn-HCl-dependent relaxation times at indicated [Gdn-HCl] from Figure 2. [b] Stability measurements from Table I. [c] Seconds. [d] Refolding relaxation time at 4 M Gdn-HCl. [e] Unfolding relaxation time at 6 M Gdn-HCl. [f] Molar Gdn-HCl. [g] Estimated by extrapolation.

Preliminary calculations on the 30 and 51 β positions in the Ala30/Ala51 mutant show mobilities similar to those of the wild-type structure (C. Eigenbrot and A. Kossiakoff, personal communication). Calculations on the solvent accessibility of the 30–51 disulfide in wild-type BPTI show that only 5.7 Å^2 of the 30–51 disulfide is exposed to solvent (Cys30–C_β methylene) (data not shown). Thus, the sensitivity of the stability of BPTI to substitutions at position 30 is consistent with the effects reported for T4 lysozyme.

The changes in the free energy of denaturation for Val30/Ala51, Ala30/Ala51, and Thr30/Ala51 (Table I) are consistent with free energies of transfer of the amino acids valine, alanine, and threonine, respectively, from the interior of a protein to its surface (Chothia, 1976). Thus, the hydrophobicity of residue 30 appears to have a significant effect on the stability of the protein. Furthermore, the 2 kcal/mol destabilization due to the buried Val→Thr mutation is similar to the effects seen in other proteins with Val→Thr substitutions at buried positions. In the α subunit of tryptophan synthase, with Glu49 as the wild-type amino acid, the Val49→Thr mutation destabilizes the protein by 2.4–3.2 kcal/mol (Yutani et al., 1987). In T4 lysozyme, where Ile3 is the wild-type amino acid, the Val3→Thr substitution destabilizes the protein by 1.1–1.9 kcal/mol (Matsumura et al., 1988). In both of these studies, the stability of mutants at the respective position was correlated with the hydrophobicity of the residue. Further, the wild-type amino acid in both of these studies was larger than Val or Thr.

These studies lead to the suggestions that (a) the hydrophobicity at position 30 of BPTI is important and (b) the stability of the BPTI 30–51 mutants will increase with increasing hydrophobicity until the volume of the disulfide is reached. Other forces at position 30 might also be important, but given the present absence of detailed mutant structures, the more direction-dependent interactions (hydrogen bonds) cannot be evaluated. Further substitution of other residues at this position is necessary to examine the validity of these hypotheses.

The kinetics of folding and unfolding were also altered by mutation at residue 30. While the denaturant-independent relaxation times of refolding were unaffected, the relaxation times of the Gdn-HCl-dependent phases appear to reflect the stability of the mutant (Figure 2 and Table III). The major refolding phase (τ_3) becomes slower in the less stable mutants, while the minor unfolding phase (τ_4) becomes faster. The origin of this minor unfolding phase is unclear, but it may reflect the presence of two slowly interconverting native forms in solution (see below). The contrasting nature of the denaturant dependence of the two unfolding relaxation times may reflect differences in the position of each transition state (Hurle et al., 1987; Tanford, 1970) along separate pathways.

Effects Due to Position of Disulfide Bonds

Stability. Two different effects on stability are important in the evaluation of the disulfides: (a) entropic effects on the

unfolded forms and (b) hydrophobic or electrostatic forces in the native forms. The entropic effect can be estimated by using the matrix-based algorithm of Lin et al. (1984). This algorithm takes into account multiple disulfides, and calculations on the 14–38 and 30–51 molecules suggest that the loss of each disulfide increases the conformational entropy of the unfolded form by 4.15 and 4.24 kcal/mol, respectively, at 25 °C (data not shown). This difference is much less than the variation in stability of the three 30–51 mutants (Table I). Further substitutions at these positions are necessary to completely assess the relative importance of the forces at these two positions. Knowledge of these interactions may assist in the design of disulfides engineered to increase protein stability (Wells & Estell, 1988).

Folding Kinetics. All three 30–51 mutants have unfolding relaxation times longer than 20 s. However, unfolding of both RCAM(14–38)BPTI (Jullien & Baldwin, 1981) and Ala14/Ala38 is complete within 30 s ($\tau \ll 10$ s). The major refolding phase, τ_3, is also too rapid to be observed in the Ala14/Ala38 mutant. Stopped-flow fluorescence studies of RCAM(14–38)BPTI showed that this refolding phase has an extremely rapid relaxation time of 40 ms at 25 °C, 1.9 M Gdn-HCl, pH 6.8 (Jullien & Baldwin, 1981). In summary, techniques which monitor the burial of tyrosines (i.e., tertiary packing) demonstrate that proteins which contain the 5–55 and 14–38 disulfides (30–51 mutants) fold and unfold slowly, while those which contain the 5–55 and 30–51 disulfides (14–38 mutants) fold and unfold rapidly. These characteristics appear to be independent of the stability of a particular 30–51 or 14–38 mutant.

The 5–55 and 30–51 disulfides of wild-type BPTI are buried in the hydrophobic core of the molecule. The 14–38 disulfide is solvent-exposed and links the ends of two β strands. In molecules missing the 30–51 disulfide bond, both the 14–38 and 5–55 disulfide bonds are maintained in the native and unfolded states. In the 14–38 mutants, the 30–51 and 5–55 disulfides are always present. Formation of the α helix and β sheet, as well as the subsequent or concurrent docking of these elements to give the final tertiary structure, is presumed to be involved in the processes of folding in both of these molecules.

Figure 4 illustrates the probable differences in the folding behavior of the two classes. Since the 14–38 disulfide bond is only present in the 30–51 mutants, it seems reasonable to assume that the formation of the β sheet is favored in the 30–51 mutants. Since the 30–51 disulfide bond is only present in the 14–38 mutants, it is expected that formation of the correct core geometry is favored in the 14–38 mutants. The intrinsic rate of helix formation should not be affected by the presence or absence of the 30–51 disulfide; in any case, the rate of helix formation model systems has been shown to occur within 10^{-5}–10^{-7} s (Kim & Baldwin, 1982). The experimental data show that folding of the 14–38 mutants is faster than that of the 30–51 mutants, suggesting that the loss of the 30–51 disulfide may result in a requirement to search for the correct orientation of packed residues. This leads to the hypothesis that the energy of the transition state between the folded and unfolded states of BPTI reflects the need to organize the protein interior.

Folding Models. Two major models for folding mechanisms that are currently under discusion are the *jigsaw-puzzle* model (Harrison & Durbin, 1985) and the *framework* model (Kim & Baldwin, 1982). In the jigsaw-puzzle model, proteins are suggested to fold by multiple pathways, while the framework model suggests that proteins fold in a sequential series of steps. In the framework model, the intermediates are rather well-defined (possibly corresponding to partially populated secondary structures), while in the jigsaw-puzzle model, the intermediates may correspond to a wider range of structures. The distribution and population of early folding intermediates in the folding of ribonuclease A (Udgaonkar & Baldwin, 1988) and cytochrome *c* (Roder et al., 1988) have been determined by 2D NMR, and the authors have suggested that their data support the framework model. We have not yet detected the formation of intermediates in our folding experiments, so we cannot address this point.

A second question concerns the manner in which a mutation affects the kinetics of folding. Harrison and Durbin (1985) suggested that in the framework model a mutation at some point in a sequential pathway might result in perturbation of a particular step. This might result in a large change in the observed folding kinetics. On the other hand, if multiple pathways connect the native and unfolded states (jigsaw-puzzle model), perturbation of one pathway does not imply that other pathways would be affected. Thus, the folding kinetics should be relatively impervious to mutational variation. Our data and those of Matthews and co-workers (Beasty et al., 1986, 1987), in which they also showed that mutations can drastically affect folding rates without affecting stability, are inconsistent with this proposal. However, they do not preclude the applicability of the jigsaw-puzzle model to steps *before* the rate-limiting step of folding.

An open question is whether the protein still follows the same folding pathway upon these mutations (framework model) or a different pathway which is *characteristically* much slower (revised jigsaw-puzzle model). The similar denaturant dependence and amplitudes seen in the mutants of tryptophan synthase (Beasty et al., 1986) suggest that the same folding pathway is followed in these cases. The more drastic effects of removing cross-links in the BPTI mutants, along with the lack of knowledge concerning the denaturant dependence of the folding step of the Ala14/Ala38 mutant, preclude any firm conclusion about the folding characteristics of wild-type BPTI.

These observations are also in conflict with the cardboard-box model of Goldenberg and Creighton (1985), in which they suggested that the energy required to progress over the transition state (to open or to fold the box) is directly related to the stability of the protein (the stiffness of the flaps). These experimental results suggest that there does not need to be an inverse relation between the stability of the native state and that of the transition state.

Our results also could have important consequences for engineering disulfides into proteins to decrease their rates of unfolding (C. Mitchinson and J. Wells, personal communication), as they demonstrate that the presence or absence of disulfides can affect both the unfolding and refolding rates of a protein.

Relationship to Disulfide Folding Kinetics. The effects of substitutions of the 30–51 and 14–38 disulfides on disulfide refolding kinetics have previously been examined (Marks et al., 1987b; P. A. Kosen, unpublished results). The disulfide refolding rates of the Ala14/Ala38, Thr14/Thr38 (Marks et al., 1987b), and Ser14/Ser38 (Goldenberg, 1988b) BPTI mutants, as well as RCAM(14–38)BPTI (Creighton, 1977a), are slower than that of wild type, while the disulfide refolding rate of Ala30/Ala51 is equal to or greater than that of wild type (Marks et al., 1987b; P. A. Kosen, unpublished results). The results are in contrast to those observed for denaturant-dependent refolding. In such refolding experiments, Ala14/Ala38 refolding is faster than that of Ala30/Ala51.

However, denaturant-induced refolding rates reflect folding steps that occur after the correct disulfides have been formed, processes which may occur after the rate-limiting step of disulfide refolding (Creighton, 1988).

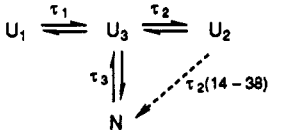
Details of Folding Mechanism

Proline Isomerization and Refolding. Examination of the unfolding and refolding behavior of a protein yields information on the kinetic mechanism of folding (Utiyama & Baldwin, 1986). Multiple refolding phases have often been observed and have usually been shown to be due to various aspects of cis–trans proline isomerization reactions (Nall, 1985). As discussed by Jullien and Baldwin (1981) and by Levitt (1981), these reactions can be classified into three major types, in which the isomerization reaction occurs after (type I), during (II), or before (III) folding. Type II and III prolines, termed "essential", are those that reflect isomerizations that interfere with folding and are most easily detected. Type III reactions have denaturant-independent relaxation times with characteristic activation energies of ~20 kcal/mol. Type II reactions have Gdn-HCl-dependent relaxation times with activation energies similar to those of folding reactions (<10 kcal/mol); thus, only double-jump experiments (Brandts et al., 1975; Jullien & Baldwin, 1981) can distinguish between the two.

As interpreted by Levitt (1981), these three classifications result from the extent of native-form destabilization due to the nonnative isomeric form of the proline. (1) If the stability of the native form is relatively unaffected by the presence of the nonnactive isomer, the protein is able to fold irrespective of the isomeric state of the proline. The proline isomerizes *after* folding and thus is classified as type I. (2) If the destabilizing effect of the nonnative proline is similar to the stability of the folded form, the proline appears as type II. (3) Finally, if the nonnative proline destabilizes the protein by an amount greater than the stability of the protein, the protein cannot fold until the proline is in its native configuration. The proline thus isomerizes *before* folding and is classified as type III.

The two slow phases, τ_1 and τ_2, observed in refolding of the two-disulfide BPTI mutants have the characteristics of proline isomerization reactions. The three 30–51 BPTI mutants have similar Gdn-HCl-independent relaxation times at 15 °C of 1500 and 260 s, respectively (Figure 2). Thus, in the 30–51 mutants, both phases appear to belong to the class of type III isomerizations. Although two slow phases are also seen in refolding of the Ala14/Ala38 mutant, the relaxation time of the faster of these two is now dependent on Gdn-HCl concentration. Studies on the fluorescence-monitored refolding of RCAM(14–38)BPTI (Jullien & Baldwin, 1981) demonstrated a phase with a Gdn-HCl-dependent relaxation time remarkably similar to that of the Ala14/Ala38 mutant (Figure 3) double-jump experiments showed that this slow phase was due to proline isomerization. Thus, in molecules lacking the 14–38 disulfide, the τ_1 phase is due to a type III isomerization, but the faster τ_2 phase appears to result from a type II isomerization. It appears that the τ_2 phase, which may be due to the same proline residue in all the mutants studied, falls into either the type II or type III category depending on the location of the disulfide bonds in the particular mutant.

BPTI contains prolines at positions 2, 8, 9, and 13, and the X-ray structure shows that all four are in the trans conformation in the native form (Wlodawer, 1987). In the 30–51 mutants, we postulate that two proline residues *must* be in the proper isomeric form (trans) before folding may proceed. The remaining two prolines are classified as "nonessential". However, some folding appears to occur before or during isomerization for one of these two prolines in the 14–38 mutants, since the τ_2 phase has characteristics of a type II reaction. A folding mechanism which is consistent with the data is



where in molecules missing the 14–38 disulfide, folding and the τ_2 reaction occur simultaneously (type II).

Since the amplitudes of the various phases are identical in the 14–38 and 30–51 molecules, the equilibria between the various isomeric unfolded forms of the 30–51 and 14–38 molecules may also be similar. However, the energetics of the native structure are altered, such that the cis form of one of the two prolines is no longer sufficiently destabilizing to prohibit some degree of folding. Conversion of refolding relaxation times from denaturant independent to denaturant dependent has also been observed for some mutants of the α subunit of tryptophan synthase (Hurle et al., 1986; Stackhouse et al., 1988). It is possible that the τ_1 and τ_2 phases have different sources in the 14–38 and 30–51 mutants. Although this seems unlikely, identification of the prolines responsible for these phases will be made by using single-site mutagenesis in both classes of two-disulfide BPTI proteins.

Energy minimization experiments based on the wild-type BPTI native structure demonstrated that isomerization about each of the four proline peptide bonds (at position 2, 8, 9, or 13) affected the computed stability of the molecule (Levitt, 1981). In this study, Levitt suggested that the cis form of Pro8 was sufficiently destabilizing that it could prohibit folding (type III). The cis forms of Pro2 and Pro9 were predicted to be somewhat less destabilizing and could inhibit, but not prevent, folding (type II). The cis form of Pro13 did not affect the computed stability of native BPTI and thus was proposed to exist in equilibrium with the trans isomer in the native structure (type I). Site-directed mutagenesis has yielded results consistent with an assignment of τ_1 to Pro8 and τ_2 to Pro9 (Hurle, unpublished data).

Unfolding. Unfolding of the 30–51 mutants of BPTI was biphasic, which is unusual for monomeric proteins (Creighton, 1988). The major phase comprised 85% of the total unfolding amplitude and had a relaxation time virtually independent of Gdn-HCl concentration. The minor phase accounted for the remaining 15% of the amplitude with a relaxation time that decreased with increasing Gdn-HCl concentrations. This major unfolding phase and the τ_3 major refolding phase appear to reflect a reversible reaction, since unfolding and refolding relaxation times are similar at identical final denaturant concentrations. However, the relation between the minor unfolding phase and any refolding phase is unclear. The biphasic unfolding could reflect (1) conversion of a trans proline to the cis form after complete unfolding (type I isomerization), (2) a transiently populated intermediate, or (3) separate unfolding of a minor native form. These possibilities are discussed separately below.

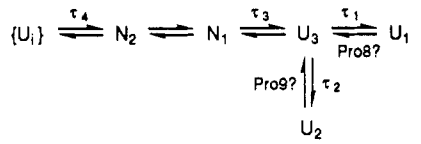
Isomerization in the unfolded form could only be detected by absorbance at 287 nm if one of the four tyrosine residues was perturbed. Tyrosine 10 is located in a hexapeptide of BPTI containing three prolines (PPYTGP) and is the only tyrosine that might be affected by isomerization of a random-coil unfolded form. However, a large difference in solvation of the one tyrosine would be necessary to observe a change which is 15% of that caused by solvation of all four tyrosines upon complete unfolding. Such an effect seems

unlikely to occur in the unfolded form, unless this state deviates significantly from a random coil.

A transiently populated intermediate would be observed if the sequence of unfolding reactions were N–(fast, τ_3)→I–(slow, τ_4)→U, where the fast phase accounted for 85% of the observed amplitude. In this case, a single refolding phase with a relaxation time of τ_4 would likely be observed, as the U → N reaction would be rate-limited by the U → I transition. However, refolding is controlled by τ_3, which is an order of magnitude faster than τ_4. Thus, this possibility seems unlikely.

The biphasic unfolding of the (30–51)BPTI mutants may reflect the presence of two slowly interconverting native forms in solution. As mentioned above, Levitt (1981) showed that cis and trans forms of Pro13 were both energetically possible in the native structure of wild-type BPTI. A similar hypothesis was postulated for dihydrofolate reductase from *E. coli* (Perry et al., 1987), and mutation of proline residues 21 and 25 appeared to affect the populations of the native forms (J. J. Onuffer and C. R. Matthews, personal communication). Interestingly, there is evidence from CD (Kosen et al., 1981) and NMR (G. Wagner, personal communication) for conformational flexibility at the 14–38 disulfide bond of wild-type BPTI.

A second native form results in the model



where $\{U_i\}$ represents a collection of unfolded forms with populations proportionately similar to $\{U_1, U_2, U_3\}$. The source of the minor unfolding phase τ_4 is a subject for further investigation.

Summary

Initial studies on the folding of mutants of BPTI containing only two disulfide bonds have demonstrated that the behavior of slow refolding phases is significantly affected by the positions of the disulfide bonds. It is suggested that the energy of the transition state between the folded and unfolded states reflects the need to organize the interior of BPTI and that the 30–51 disulfide bond limits the conformational search required in this slow step. However, the kinetics do not appear to be dramatically dependent on the particular residue substituted in place of the disulfide. This may be due to the conservative nature of the residues substituted thus far (Ala, Val, Thr); larger residues may have a greater effect (Matsumura et al., 1988; Garvey & Matthews, 1989). The residues substituted in BPTI appear to have effects on stability that correlate with their hydrophobicity.

A folding mechanism for BPTI mutant molecules has been proposed in which the isomerization of two proline residues limits folding of a significant fraction of the BPTI molecules, consistent with earlier suggestions (Jullien & Baldwin, 1981). The rate of one of the two refolding steps assigned to proline isomerization appears to be altered, depending on whether the 14–38 or 30–51 disulfide is present, suggesting that the adaptability of the native structure to the incorrect cis isomer is changed by the cross-link. Future work will initially focus on mutation of the four proline residues to assign these slow refolding phases to various molecular events.

ACKNOWLEDGMENTS

We gratefully acknowledge Mike Brochier for performing fermentations of the various strains discussed in this paper, Mike Randal for assistance with purifications, John Altman and Drs. Bjorn Nilsson, Charles Eads, Bob Matthews, and Robert Baldwin for helpful discussion, and Genentech for facility support. We thank Drs. Charles Eigenbrot and Tony Kossiakoff for access to the wild-type and Ala30/Ala51 BPTI thermal factors and John Altman for assistance in the computation of wild-type BPTI solvent-accessible surface area.

REFERENCES

Alber, T., & Matthews, B. W. (1988) *Methods Enzymol. 154*, 511–533.

Alber, T., Dao-Pin, S., Nye, J. A., Muchmore, D. C., & Matthews, B. W. (1987) *Biochemistry 26*, 3754–3758.

Anfinsen, C. B. (1973) *Science 181*, 223–230.

BBN Software Products Corp. (1987) *RS/1 User's Guide*, VAX VMS/UNIX Version 3.0, BBN Software Products Corp., Cambridge, MA.

Beasty, A. M., Hurle, M. R., Manz, J. T., Stackhouse, T., Onuffer, J. J., & Matthews, C. R. (1986) *Biochemistry 25*, 2965–2974.

Beasty, A. M., Chrunyk, B. A., Gittelman, M. S., Herndon, C. S., Hurle, M. R., Manz, J. T., Onuffer, J. J., Perry, K. M., Stackhouse, T., Touchette, N. A., Tweedy, N., & Matthews, C. R. (1988) *Protein Structure, Folding and Design* (Oxender, D. L., Ed.) Vol. 2, pp 321–333, Liss, New York.

Brandts, J. F., Halvorson, H. R., & Brennan, M. (1975) *Biochemistry 14*, 4953–4963.

Chang, C. H., Rey, M., Bochner, B., Heyneker, H., & Gray, G. (1987) *Gene 55*, 189–196.

Chothia, C. (1976) *J. Mol. Biol. 105*, 1–14.

Creighton, T. E. (1977a) *J. Mol. Biol. 113*, 275–293.

Creighton, T. E. (1977b) *J. Mol. Biol. 113*, 313–328.

Creighton, T. E. (1978) *Prog. Biophys. Mol. Biol. 33*, 231–297.

Creighton, T. E. (1986) *Methods Enzymol. 131*, 83–106.

Creighton, T. E. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 5082–5086.

Cupo, J. F., & Pace, C. N. (1983) *Biochemistry 22*, 2654–2658.

Garvey, E. P., & Matthews, C. R. (1989) *Biochemistry 28*, 2083–2093.

Goldenberg, D. P. (1988a) *Annu. Rev. Biophys. Biophys. Chem. 17*, 481–507.

Goldenberg, D. P. (1988b) *Biochemistry 27*, 2481–2489.

Goldenberg, D. P., & Creighton, T. E. (1985) *Biopolymers 24*, 167–182.

Harrison, S. C., & Durbin, R. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 4028–4030.

Herskovits, T. T. (1967) *Methods Enzymol. 11*, 748–775.

Hurle, M. R., Tweedy, N. B., & Matthews, C. R. (1986) *Biochemistry 25*, 6356–6360.

Hurle, M. R., Michelotti, G. A., Crisanti, M. M., & Matthews, C. R. (1987) *Proteins 2*, 54–63.

Jaenicke, R., Ed. (1980) *Protein Folding, Proceedings of the 28th Conference of the German Biochemical Society*, p 587, Elsevier/North-Holland Biomedical Press, Amsterdam.

Jullien, M., & Baldwin, R. L. (1981) *J. Mol. Biol. 145*, 265–280.

Kassell, B. (1970) *Methods Enzymol. 19*, 844–852.

Kim, P. S., & Baldwin, R. L. (1982) *Annu. Rev. Biochem. 51*, 459–489.

Kosen, P. A., Creighton, T. E., & Blout, E. R. (1981) *Biochemistry 20*, 5744–5754.

Levitt, M. (1981) *J. Mol. Biol. 145*, 251–263.

Levitt, M., & Sharon, R. (1987) *Crystallography in Molecular Biology*, (Moras, D., Drenth, J., Strandberg, B., Suck, D., & Wilson, K., Eds.) pp 197–205, Plenum Press, New York.

Lin, S. H., Konishi, Y., Denton, M. E., & Scheraga, H. A. (1984) *Biochemistry 23*, 5504–5512.
Marks, C. B., Vasser, M., Ng, P., Henzel, W., & Anderson, S. (1986) *J. Biol. Chem. 261*, 7115–7118.
Marks, C. B., Naderi, H., Kosen, P. A., Kuntz, I. D., & Anderson, S. (1987a) *Science 235*, 1370–1373.
Marks, C. B., Naderi, H., Kosen, P. A., Kuntz, I. D., & Anderson, S. (1987b) in *Protein Structure, Folding and Design* (Oxender, D. L., Ed.) Vol. 2, pp 335–340, Liss, New York.
Matouschek, A., Kellis, J. T., Jr., Serrano, L., & Fersht, A. R. (1989) *Nature 340*, 122–126.
Matsumura, M., Becktel, W. J., & Matthews, B. W. (1988) *Nature 334*, 406–410.
Moses, E., & Hinz, H.-J. (1983) *J. Mol. Biol. 170*, 765–776.
Motulsky, H. J., & Ransnas, L. A. (1987) *FASEB J. 1*, 365–374.
Nall, B. T. (1985) *Comments Mol. Cell. Biophys. 3*, 123–143.
Pace, C. N. (1975) *CRC Crit. Rev. Biochem. 3*, 1–43.
Pace, C. N., Grimsley, G. R., Thomson, J. A., & Barnett, B. J. (1988) *J. Biol. Chem. 263*, 11820–11825.
Perry, K. M., Onuffer, J. J., Touchette, N. A., Herndon, C. S., Gittelman, M. S., Matthews, C. R., Chen, J.-T., Mayer, R. J., Taira, K., Benkovic, S. J., Howell, E. E., & Kraut, J. (1987) *Biochemistry 26*, 2674–2682.
Ramdas, L., & Nall, B. T. (1986) *Biochemistry 25*, 6959–6964.
Richardson, J. A. (1981) *Adv. Protein Chem. 34*, 167–339.
Roder, H., Elove, G. A., & Englander, S. W. (1988) *Nature 335*, 700–704.
Schellman, J. A. (1978) *Biopolymers 17*, 1305–1322.
Schwarz, H., Hinz, H.-J., Mehlich, A., Tschesche, H., & Wenzel, H. R. (1987) *Biochemistry 26*, 3544–3551.
Stackhouse, T. M., Onuffer, J. J., Matthews, C. R., Ahmed, S. A., & Miles, E. W. (1988) *Biochemistry 27*, 824–832.
Tanford, C. (1968) *Adv. Protein Chem. 23*, 121–282.
Tanford, C. (1970) *Adv. Protein Chem. 24*, 1–95.
Tüchsen, E., & Woodward, C. (1985) *J. Mol. Biol. 185*, 405–419.
Udgaonkar, J. B., & Baldwin, R. L. (1988) *Nature 335*, 694–699.
Utiyama, H., & Baldwin, R. L. (1986) *Methods Enzymol. 131*, 51–71.
Vincent, J.-P., Chicheportiche, R., & Lazdunski, M. (1971) *Eur. J. Biochem. 23*, 401–411.
Wells, J. A., & Estell, D. A. (1988) *Trends Biochem. Sci. 13*, 291–297.
White, T. B., Berget, P. B., & Nall, B. T. (1987) *Biochemistry 26*, 4358–4366.
Wlodawer, A., Deisenhofer, J., & Huber, R. (1987) *J. Mol. Biol. 193*, 145–156.
Wüthrich, K., Wider, G., Wagner, G., & Braun, W. (1982) *J. Mol.* Biol. *155*, 311–319.
Yutani, K., Ogasahara, K., Tsujita, T., & Sugino, Y. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 4441–4444.

Primary Structure of Non-Histone Chromosomal Protein HMG2 Revealed by the Nucleotide Sequence†

Hitoshi Shirakawa,‡ Koh-ichiroh Tsuda,§ and Michiteru Yoshida*,‡

Department of Biological Science and Technology, Science University of Tokyo, Yamazaki, Noda-shi, Chiba 278, Japan, and Department of Agricultural Chemistry, Tohoku University, 1-1 Tsutsumidori-Amamiyamachi, Sendai 980, Japan

Received June 22, 1989; Revised Manuscript Received September 13, 1989

ABSTRACT: The isolation and sequencing of a cDNA clone for the entire sequence of pig thymus non-histone protein HMG2 are described. cDNA the size of 1153 nucleotides contains an open reading frame of 627 nucleotides. The 5′-untranslated region of 146 nucleotides is extremely rich in GC residues whereas the 3′-untranslated region of 380 nucleotides is rich in AT residues. The open reading frame encodes 209 amino acids, which contain a unique continuous run of 23 acidic amino acids at the C-terminal. The deduced amino acid sequence is 79% homologous to that of HMG1 protein from the same source which we reported [Tsuda, K., Kikuchi, M., Mori, K., Waga, S., & Yoshida, M. (1988) *Biochemistry 27*, 6159–6163]. In addition, the hydropathy index profiles of both proteins are very similar, supporting that they have similar structural features. Northern analysis of poly(A+) RNA reveals that a single-sized mRNA codes for HMG2 protein. Southern analysis suggests that the HMG2 coding gene is homogeneous within the pig thymus genome.

The relatively low molecular weight proteins in eukaryotic non-histone nuclear proteins are called high mobility group (HMG)[1] proteins. HMG1 and HMG2 in four major HMG proteins have remarkable structural and functional similarities to each other. These proteins show a preferential binding to single-stranded DNA (Bidney & Reeck, 1978; Isackson et al., 1979; Yoshida & Shimura, 1984; Hamada & Bustin, 1985), unwind double-stranded DNA structure (Yoshida & Shimura, 1984; Makiguchi et al., 1984; Javaherian et al., 1978, 1979), and remove the transcriptional block caused by left-handed

† The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J02895.

* Correspondence should be addressed to this author.
‡ Science University of Tokyo.
§ Tohoku University.

[1] Abbreviations: HMG, high mobility group; bp, base pair(s); cDNA, complementary DNA; SDS, sodium dodecyl sulfate.

0006-2960/90/0429-4419$02.50/0 © 1990 American Chemical Society