Characterization of a Folding Intermediate of Apoplastocyanin Trapped by Proline Isomerization†

Shohei Koide, H. Jane Dyson, and Peter E. Wright*

Department of Molecular Biology, The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, California 92037

Received August 2, 1993®

ABSTRACT: The unfolding and refolding transitions of French bean apoplastocyanin (apo-Pc), a β-sandwich protein, have been characterized. The apoprotein is stabilized by sodium sulfate and can be reversibly unfolded by guanidine hydrochloride (GuHCl). However, in contrast to holo-Pc, apo-Pc is unstable at low ionic strength, suggesting that the copper ion stabilizes the holoprotein. The equilibrium unfolding transition monitored by peptide circular dichroism (CD) and tyrosine fluorescence is described by a two-state model. The kinetics of the unfolding transition were monitored using a manual mixing technique and are consistent with a single two-state transition. In contrast, the kinetics of the refolding reaction measured by fluorescence and CD show two transitions with different rates. The relaxation time of the slower phase (800–1000 s) is almost independent of GuHCl concentration. The faster phase was observed only under strongly native conditions, and its relaxation time is GuHCl-dependent. Double-jump experiments and acceleration by cyclophilin demonstrate that both phases involve *cis–trans* isomerization of proline residues. The changes in fluorescence associated with the two phases are more than 150% of the total change expected from equilibrium experiments, indicating the presence of intermediate(s) with fluorescence greater than the unfolded state. Amide hydrogen-exchange experiments coupled with two-dimensional NMR spectroscopy demonstrate the formation of an intermediate in the very slow refolding reaction in which amide protons in the β-sheets are weakly protected from exchange. No CD evidence for nativelike β-sheet formation was found for this intermediate. The NMR experiments suggest that the intermediate is compact with flexible β-sheets and altered packing of the hydrophobic core. It has many of the characteristics of a molten globule. However, the ^{1}H NMR spectrum of the intermediate exhibits a small number of shifted resonances that indicate the presence of specific tertiary interactions in a localized region. A mechanism for refolding of apoplastocyanin is proposed that includes two slow steps corresponding to *trans* → *cis* isomerization of two prolines.

The mechanism by which an unfolded polypeptide chain spontaneously folds into a unique three-dimensional structure remains one of the fundamental problems in molecular biology. Elucidation of the structures of intermediates formed on folding pathways is the goal of many experiments; however, the highly cooperative nature of the protein-folding process makes it difficult to characterize folding intermediates in detail. Transient intermediates of small single-domain proteins have been characterized by low-resolution techniques such as disulfide bond formation (Creighton, 1990) and circular dichroism (CD)[1] kinetics (Kuwajima et al., 1987; Sugawara et al., 1991). More detailed structural characterization of folding intermediates has been achieved for several proteins by the combined use of hydrogen-exchange techniques and multidimensional NMR [see, for example, Roder et al. (1988), Udgaonkar and Baldwin (1988), Bycroft et al. (1990), and Miranker et al. (1991)]. Protein engineering has also been used to dissect contributions of individual interactions to the stability of intermediates and folding transition states (Matthews, 1987; Matouschek et al., 1989). Together, these experiments provide support for a framework model for protein folding, in which elements of secondary structure are formed rapidly and tertiary interactions are developed at a later stage (Kim & Baldwin, 1982). Further support for the framework model comes from experiments which show that many short linear peptide fragments of proteins have a propensity to adopt secondary structure in aqueous solution (Wright et al., 1988; Dyson & Wright, 1993). Helices, nascent helix, reverse turns, and hydrophobic clusters have all been observed (Dyson & Wright, 1991). These results suggest the significance of local secondary structural propensities for the initiation of protein folding.

Information on the structures of equilibrium unfolding intermediates has come from the characterization of states which are found in several proteins under specific denaturing conditions (Ptitsyn, 1987; Kuwajima, 1989). These intermediates, the so-called molten globules, in general show high secondary structure content as judged by CD but lack fixed, nativelike tertiary interactions. Molten globules have been postulated to be universal folding intermediates, but only in two cases, α-lactalbumin and apomyoglobin, is there direct evidence that they are actually formed on the kinetic folding pathway (Kuwajima, 1989; Jennings & Wright, 1993).

At the present time, the most detailed information on protein-folding pathways and intermediates comes from studies of all-α or α/β proteins. Support for the framework model derives principally from studies of helical proteins. Folding studies of only relatively few proteins consisting entirely of

† This research was supported by Grant GM38794 from the National Institutes of Health. S.K. was supported by a fellowship from the Human Frontier Science Program Organization.

* Author to whom correspondence should be addressed.

® Abstract published in *Advance ACS Abstracts,* October 15, 1993.

[1] Abbreviations: Pc, plastocyanin; ANS, 8-anilino-1-naphthalenesulfonate; EDTA, ethylenediaminetetraacetic acid; GuHCl, guanidine hydrochloride; SDS, sodium dodecyl sulfate; CD, circular dichroism; 1D, one dimensional; 2D, two dimensional; NMR, nuclear magnetic resonance; TOCSY, total correlation spectroscopy.

0006-2960/93/0432-12299$04.00/0

β-structure have been reported so far (Goto & Hamaguchi, 1982; Ropson et al., 1990; Rudolph et al., 1990; Bychkova et al., 1992; Varley et al., 1993). It is possible that the mechanism of folding of β proteins might differ from that of α or α/β proteins, since a β-sheet is formed by long-range interactions rather than local interactions as in the case of a helix. A theoretical study suggests that the rate of β-structure formation is slow unless it is particularly stable (Finkelstein, 1991). In the case of one all-β protein, β-lactoglobulin, rapid formation of nonnative secondary structure was observed (Kuwajima et al., 1987). In addition, β-sheet formation has not generally been observed in peptides, except when stabilized by disulfide bonds or by metal complexation (Oas & Kim, 1988; Lee et al., 1989). An NMR study (Dyson et al., 1992a,b) of peptide fragments derived from all-α and all-β proteins suggested that the folding of proteins that consist largely of β-structure may be fundamentally different from that of other structural motifs. Peptide fragments derived from the sequence of an all-helical protein, myohemerythrin, were found to possess significant propensities for secondary structure formation (Dyson et al., 1992a). In contrast, the fragments from a Greek key β-sandwich protein, plastocyanin (Pc), showed surprisingly little propensity to adopt folded conformations in aqueous solution (Dyson et al., 1992b).

In view of the relative lack of detailed experimental data on folding pathways of β-sheet proteins, we have undertaken folding experiments on apoplastocyanin. Pc is a small (99 residue) blue copper protein which participates in electron transfer in photosynthesis (Sykes, 1985). X-ray crystal structures and NMR solution structures of Pc from several sources have been determined (Guss & Freeman, 1983; Moore et al., 1988, 1991; Collyer et al., 1990). The complete assignment of the ^{1}H NMR spectrum of French bean Pc and determination of its solution structure have been accomplished in our laboratory (Chazin & Wright, 1988; Moore et al., 1991). The structures of the proteins derived from poplar (Guss & Freeman, 1983) and French bean (Moore et al., 1991) are almost indistinguishable, consistent with the similarity in their amino acid sequences. The global fold of Pc is best described as a β-sandwich formed from two β-sheets. The protein contains only a single turn of α-helix. A cysteine, a methionine, and two histidine side chains are involved in copper ligation in the holoprotein. In order to simplify folding studies, we removed the bound copper from the protein. The crystal structure of poplar apo-Pc closely resembles that of the holo-Pc (Garrett et al., 1984), and ^{1}H NMR spectra of the holo- and apoproteins suggest that the same is true for the French bean Pc used in the folding studies described in this paper.

EXPERIMENTAL PROCEDURES

Materials. Ultrapure GuHCl was purchased from United States Biochemicals and 8-anilino-1-naphthalenesulfonate (ANS) from Molecular Probes (Eugene, OR); other chemicals were of analytical grade. Recombinant mouse cyclophilin was kindly provided by Dr. J. G. Sutcliffe (Hasel et al., 1991).

Preparation of Apo-Pc. Holo-Pc was purified from French bean (*Phaseolus vulgaris*) leaves as described previously (Ramshaw et al., 1973). The oxidized Pc had an absorbance ratio, A_{278}/A_{597}, of 1.1–1.2 and ran as a single band on SDS–polyacrylamide gel electrophoresis. The Pc was stored under liquid nitrogen in Tris-HCl buffer (10 mM) containing EDTA (0.1 mM). Apo-Pc was prepared by thawing the holoprotein and exchanging the buffer to potassium phosphate (50 mM, pH 7.0) containing NaCl (1 M) and EDTA (0.1 mM) using an Amicon ultrafiltration unit with a YM-1 membrane. Solid sodium dithionite was added to a final concentration of 5 mM; excess dithionite was removed by repetitive concentration and dilution with buffer. Solid KCN was added to a final concentration of 50 mM, and the solution was kept under argon at 4 °C for 12 h. KCN was then removed by repetitive ultrafiltration with the phosphate buffer supplemented with 2 mM 2-mercaptoethanol. Apo-Pc, which contained less than 3% of the original copper as judged by A_{597} after oxidation with potassium ferricyanide, was stored in liquid nitrogen. The protein concentration was calculated using $A_{278} = 4.2 \times 10^3\ M^{-1}\ cm^{-1}$ on the basis of the aromatic amino acid content (Gill & von Hippel, 1989).

Equilibrium Unfolding. GuHCl solutions were prepared gravimetrically. All solutions contained potassium phosphate (50 mM, pH 7.0), Na_2SO_4 (0.5 M), 2-mercaptoethanol (2 mM), and EDTA (0.1 mM). Buffer and GuHCl solutions were treated with Chelex resin (Bio-Rad) and filtered through a 0.22 μM membrane filter before use. Circular dichroism (CD) spectra in the 205–260-nm range were recorded on an Aviv 61DS spectropolarimeter using a 2-mm cell. Fluorescence intensity at 303 nm was measured with excitation at 268 nm on an Aminco-Bowman Model 20 spectrofluorometer using a 1 × 1 cm cell. The fluorescence amplitude was normalized in terms of the ratio to that of the sample in 5 M GuHCl. Protein concentrations were 10–50 μM for CD spectroscopy and 2–25 μM for fluorescence. All samples were equilibrated at the appropriate GuHCl concentration at 25 °C for 3 h or more before measurements were made.

Kinetics of Folding. Kinetics of unfolding and refolding were followed by CD and fluorescence. Reactions were initiated by a concentration jump of GuHCl using manual mixing; 100–200 μL of protein solution in the native (0 M GuHCl) or unfolding (4 or 5 M GuHCl) buffers was diluted into the appropriate buffer in a 1-cm cuvette equipped with a small stirring bar. The final volume was 2.5 mL, and the final protein concentration was 1–10 μM for fluorescence and 10 μM for CD. The dead time of the mixing was 10–15 s for fluorescence and 30–50 s for CD experiments.

Fitting of Equilibrium and Kinetic Data. The equilibrium unfolding of apo-Pc was characterized by the linear extrapolation method (Schellman, 1978; Santoro & Bolen, 1988). All parameters were determined by nonlinear least-squares fitting of the following equation using the program Igor (WaveMetrics) on a Macintosh computer:

$$X_{(D)} = \{(X^0_N + m_N[D]) + (X^0_U + m_U[D])\exp[-(\Delta G^0_{N-U} + m_G[D])/RT]\}/\{1 + \exp[-(\Delta G^0_{N-U} + m_G[D])/RT]\}$$

where $X_{(D)}$ is the value of the specified optical parameter (θ_{226} or F_{303}) at denaturant concentration [D]; X^0_N and X^0_U represent the calculated values of X linearly extrapolated to [D] = 0 for the native and unfolded protein, respectively; m_N and m_U are the slopes of the native and unfolded baselines, respectively; ΔG^0_{N-U} is the free energy change for unfolding in the absence of denaturant; m_G represents the slope of the free energy change versus denaturant concentration, assuming a linear relationship between the free energy of unfolding, ΔG_{obs}, and [D]; R is the gas constant; and T is the absolute temperature. The midpoint of the transition was determined by dividing ΔG^0_{N-U} by $-m_G$.

The kinetic data were resolved into phases by nonlinear least-squares fitting of the equation

$$X(t) = \sum_i X_i \exp(-k_i t) + X_\infty$$

where $X(t)$ is the amplitude of fluorescence or ellipticity at time t, X_i is the total amplitude of the change associated with

phase i, k_i is the apparent rate constant of phase i, and X_∞ is the amplitude at equilibrium. The number of phases was increased until no significant improvement of the fit was observed as judged from the residual errors. Simulation of kinetic mechanisms was performed using HopKINSIM on a Macintosh computer (Barshop et al., 1983).

Double-Jump Experiments. The native protein was unfolded by transferring it into unfolding buffer containing 5 M (final) GuHCl at time 0. Under these conditions, the unfolding reaction is estimated to be completed in less that 1 s. After incubation for a variable time (between 20 s and 2 h) at 25 °C, the unfolded protein was transferred into buffer without GuHCl in a cuvette, and the refolding reaction was followed by fluorescence. The final protein concentration was 5 μM, and the residual GuHCl concentration was 0.20 M. The experiments were also performed using CD. In this case, the protein was unfolded in 4 M GuHCl, and the final concentrations of the protein and GuHCl in the refolding buffer were 10 μM and 0.32 M, respectively.

ANS Binding. ANS binding was monitored by refolding apo-Pc in a refolding buffer containing 200 μM ANS. The final concentrations of protein and GuHCl were 5 μM and 0.5 M, respectively. Two minutes after initiation of refolding, a fluorescence spectrum from 450 to 600 nm was recorded in 40 s with excitation at 370 nm. A spectrum was also recorded after refolding was completed.

Hydrogen-Exchange Experiments. Apo-Pc was first exchanged by repetitive ultrafiltration into D_2O buffer containing potassium phosphate (50 mM), NaCl (1 M), 2-mercaptoethanol (10 mM), and EDTA (0.1 mM) (pH* 7.0, direct meter reading uncorrected for isotope effect). The protein was unfolded in the D_2O buffer containing 2.2 M GuHCl and kept at 25 °C for 1 h. The refolding reaction was initiated by 1:21 dilution in D_2O buffer without GuHCl. After a delay of 90 s to allow completion of the faster refolding reactions, the competition experiment was initiated by 1:9 dilution with either potassium phosphate (50 mM, pH 5.9) containing NaCl (1 M), 2-mercaptoethanol (10 mM), and EDTA (0.1 mM) in H_2O or sodium acetate (50 mM, pH 5.0) containing NaCl (1.5 M), 2-mercaptoethanol (10 mM), and EDTA (0.1 mM) in H_2O. The final pH values were 6.0 and 5.2, respectively. The solution was kept at 25 °C for 1 h and then cooled in an ice/water bath. Further procedures were performed at 4 °C. The pH of the solution was adjusted to 6.0 by adding 1 M K_2HPO_4 solution. Solid glutathione (reduced) was added to a final concentration of 10 mM, followed by addition of 200 mM $CuSO_4$ to a final concentration of 2 mM. After 1 h, solid ammonium sulfate was added to 45% saturation (262 g added per liter). No precipitation was observed. The solution was then applied to a 25-mL phenyl-Sepharose fast-flow (Pharmacia) column equilibrated in 50 mM potassium phosphate buffer (pH 6.0) containing 50% saturated ammonium sulfate (295 g added per liter). The reconstituted holo-Pc ran as a blue band after oxidation with buffer containing 1 mM ferricyanide and elution with a linear gradient of ammonium sulfate (50–0% saturation) on an FPLC system. Incorporation of copper was confirmed by the absorbance ratio A_{278}/A_{597}. The yield of reconstituted protein was about 30% in 1 h, though nearly complete reconstitution could be achieved in a longer period. Holo-Pc was concentrated to 0.5 mL in an ultrafiltration unit under argon and reduced with solid sodium dithionite. Buffer was exchanged using a 3-mL Sephadex G-25SF (Pharmacia) spin column equilibrated in potassium phosphate (100 mM, pH* 6.0) containing EDTA (50 μM) in D_2O. The final concentration of holo-Pc was 1–2 mM. The sample was transferred to an NMR tube and sealed under argon with a rubber septum after addition of a small amount of sodium dithionite. It was then frozen on dry ice and stored at −70 °C. Prior to acquisition of the NMR spectrum, each sample was equilibrated for 3 h at 30 °C. Samples in which the protein was completely refolded in D_2O prior to the exchange step at pH 6.0 and 5.2 were used as "deuterated controls". Holo-Pc was used without refolding or reconstitution as a "protonated control".

Intrinsic exchange rates for individual amide protons were calculated using the second-order specific base rate constant (k_B) of 1.24×10^8 M^{-1} s^{-1} for the exchange from D to H of an alanine amide proton (Jeng, 1991), an activation energy of 17.5 kcal (Englander & Mayne, 1992), and the revised "Molday factors" (Molday et al., 1972; Bai et al., 1993).

NMR Spectroscopy. All spectra were recorded on a Bruker AMX500 spectrometer. Two-dimensional (2D) ^1H TOCSY spectra (Braunschweiler & Ernst, 1983) with a mixing period of 40 ms were recorded at 30 °C. A total of 512 t_1 increments of 96 transients was collected, with a total acquisition time of 24 h. FIDs consisted of 2048 complex data points and spectral widths of 10 and 7 kHz were used in the ω_2 and ω_1 dimensions, respectively. Data were processed using a modified version of the FTNMR program (Hare Research). A Lorentz-to-Gaussian window function was used in the ω_2 dimension and a shifted sine-bell in ω_1. Acquisition and processing of the data from each experiment were identical.

Slowly exchanging amide protons were identified using published assignments (Chazin & Wright, 1988). Peak heights of NH–C$^\alpha$H cross peaks in the TOCSY spectra were measured using the FTNMR program and normalized against phenylalanine side-chain cross peaks. For Val28, the NH–C$^\beta$H cross peak was used due to overlap of the NH–C$^\alpha$H cross peak with another resonance. Amide protons for which peak heights in the deuterated control were less than 40% of those in the protonated control were used as probes. The normalized proton occupancy was calculated as

$$I_{rel} = (I_C - I_D)/(I_H - I_D)$$

where I_C, I_D, and I_H denote cross-peak heights in the competition, deuterated control, and protonated control spectra, respectively. Errors were estimated by measuring the standard deviation of noise in "empty" portions of the NH–C$^\alpha$H region of the spectra.

Chemical shifts in the 1D NMR spectra used in the kinetic refolding experiments are referenced to the Val39 C$^\gamma$H resonance at −0.32 ppm (Chazin & Wright, 1988). The spectra in GuHCl were referenced to the sharp high-field singlet resonance at 0.18 ppm which is probably due to contamination with silicone grease from the lyophilizer.

RESULTS

Properties of Apo-Pc. Apo-Pc, prepared by removing copper with cyanide, was unstable even at neutral pH. In 50 mM potassium phosphate buffer at pH 7.0, the CD spectrum of the protein changed gradually, indicating slow unfolding. Apo-Pc is stabilized in its folded conformation by high salt (1 M NaCl or 0.5 M Na_2SO_4) and thiol-containing reducing agents. Under these conditions, the CD spectrum of apo-Pc was almost identical to that of holo-Pc, and no change in the far-UV CD spectrum of apo-Pc was detected after long incubation at 25 °C. The NMR spectrum of apo-Pc is also similar to that of holo-Pc, and the fine structure in the UV absorption spectrum of apo-Pc resembles that of the holoprotein. Together, these results suggest that the native apoprotein has secondary and tertiary structures similar to

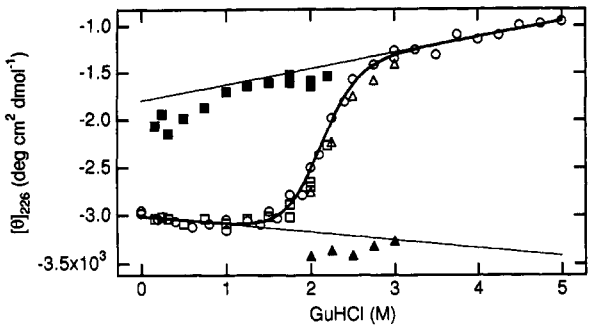


FIGURE 1: Equilibrium and kinetic data for the unfolding and refolding of apo-Pc by GuHCl. The measurements were carried out by measuring changes in ellipticity at 226 nm. The unfolding equilibrium (○) was measured at a protein concentration of 50 μM. The initial points of the unfolding (▲) and refolding (■) kinetics were obtained by extrapolation to time 0. The final points of the unfolding (△) and refolding (□) kinetics were obtained by extrapolation to time infinity. The curve indicates the best fit of the two-state transition curve to the equilibrium data by the linear extrapolation method. The straight lines show the native and unfolded baselines.

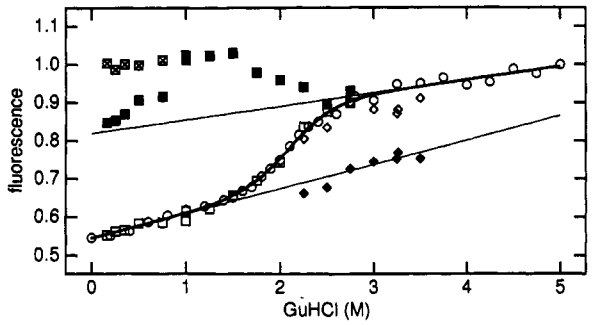


FIGURE 2: Equilibrium and kinetic data for unfolding and refolding monitored by Tyr fluorescence. The fluorescence amplitude at 303 nm after excitation at 268 nm was monitored and was normalized in terms of the ratio to that of the sample in 5 M GuHCl. All measurements were made at a protein concentration of 5 μM. (○) shows the unfolding equilibrium, and the thick curve indicates the best fit of the two-state transition equation to the equilibrium data. The straight lines show the native and unfolded baselines. The initial and final points of the unfolding kinetics obtained by extrapolation to $t = 0$ or $t = \infty$ are shown by (◆) and (◇), respectively. The initial points of the very slow and slow phases of the refolding kinetics are shown by (■) and (crosshatched □), respectively. They were obtained by extrapolating the respective phases to time 0. (□) indicates the final point of the refolding kinetics.

those of the holoprotein. This is not surprising since the three-dimensional structures of apo- and holo-Pc from poplar are virtually identical in the crystalline state (Garrett et al., 1984).

Equilibrium Unfolding. The GuHCl-induced unfolding transition of apo-Pc was determined in the presence of 0.5 M sodium sulfate at 25 °C and pH 7.0. Figure 1 shows the unfolding transition monitored by change in ellipticity at 226 nm. Figure 2 shows the transition monitored by change in intrinsic fluorescence intensity at 303 nm (excitation 268 nm). Since the protein lacks Trp, the fluorescence should originate from the three Tyr residues. No shift of the fluorescence emission maximum was observed upon unfolding. The curves shown are the fits to a two-state model assuming a linear dependence of the free energy of unfolding on GuHCl concentration (Schellman, 1978). The free energy of unfolding in the absence of GuHCl, ΔG^0_{N-U}, was calculated to be 5.8 ± 0.6 and 5.3 ± 0.8 kcal mol^{-1} from changes in ellipticity and fluorescence, respectively (Table I). The corresponding slopes of the free energy change with denaturant concentration (m_G) are −2.7 ± 0.4 and −2.5 ± 0.4 kcal mol^{-1} M^{-1}. These data indicate that the unfolding curves obtained by CD and fluorescence spectroscopy are almost identical within experimental error. No significant difference was observed in the equilibrium unfolding over protein concentrations ranging from 2.5 to 50 μM, monitored by fluorescence. The transition induced by GuHCl was completely reversible at low protein concentration (≤10 μM). However, aggregation was observed in the refolding reaction at a higher protein concentration (≥50 μM). The transitions are fit well by a simple two-state model, and the transition curves obtained by CD and fluorescence spectroscopy are almost identical, suggesting that only native and unfolded forms of apo-Pc are present in detectable population at equilibrium.

Folding Kinetics. The kinetics of unfolding and refolding of apo-Pc were followed by changes in fluorescence intensity and ellipticity. The unfolding and refolding reactions were initiated by a concentration jump of the denaturant using manual mixing techniques.

The unfolding reaction is well described by a single exponential decay. Figure 3C shows the relaxation time ($1/k_{app}$) of unfolding as a function of final concentration of GuHCl. The logarithm of the relaxation time is linearly dependent on the GuHCl concentration, and both spectroscopic methods give similar relaxation times as a function of GuHCl concentration. The unfolding reaction becomes too fast to be monitored by manual mixing techniques at high GuHCl concentration. Thus, the kinetics could be characterized only in a limited concentration range.

The initial points (amplitude at $t = 0$) in fluorescence (Figure 2) and in ellipticity (Figure 1) are, within experimental error, on the extended native baseline. Therefore, the amplitudes observed for the unfolding reaction are in good agreement with those determined by equilibrium measurements, suggesting that no intermediates are accumulated during unfolding.

In the refolding reaction, two phases were detected by fluorescence, while a single phase was observed by CD spectroscopy (Figure 3). The difference may be attributed to the low sensitivity and the long dead time of mixing in the CD experiment. The change in fluorescence intensity is well fitted by two exponential decays under strongly native conditions ([GuHCl] ≤ 0.75 M), while a single exponential was sufficient to fit the data under marginally native conditions. The relaxation times of the refolding as a function of denaturant concentration are shown in Figure 3. The faster phase resolved under strongly native conditions will be referred

Table I: Parameters Characterizing the Fits of the Equilibrium Unfolding Transition of Apo-Pc[a]

detection	X^0_N	m_N (M^{-1})	X^0_U	m_U (M^{-1})	ΔG°_{N-U} (kcal mol^{-1})	m_G (kcal mol^{-1} M^{-1})	midpoint (M)
θ_{226}	−3010 (30)	−83 (41)	−1800 (110)	170 (30)	5.8 (0.6)	−2.7 (0.3)	2.1
F_{303}	0.545 (0.007)	0.064 (0.009)	0.82 (0.02)	0.035 (0.005)	5.3 (0.8)	−2.5 (0.4)	2.1

[a] The equilibrium unfolding of apo-Pc was characterized, and the parameters were determined as described in the text. X^0_N and X^0_U represent the evaluated values of the specified optical parameter linearly extrapolated to [D] = 0 for the native and unfolded protein, respectively, m_N and m_U are the slopes of the native and unfolded baselines, respectively, ΔG°_{N-U} is the free energy change for unfolding in the absence of denaturant, m_G represents the slope of the free energy in terms of denaturant concentration, and midpoint is the denaturant concentration at the midpoint of the transition. Standard deviations are shown in parentheses.

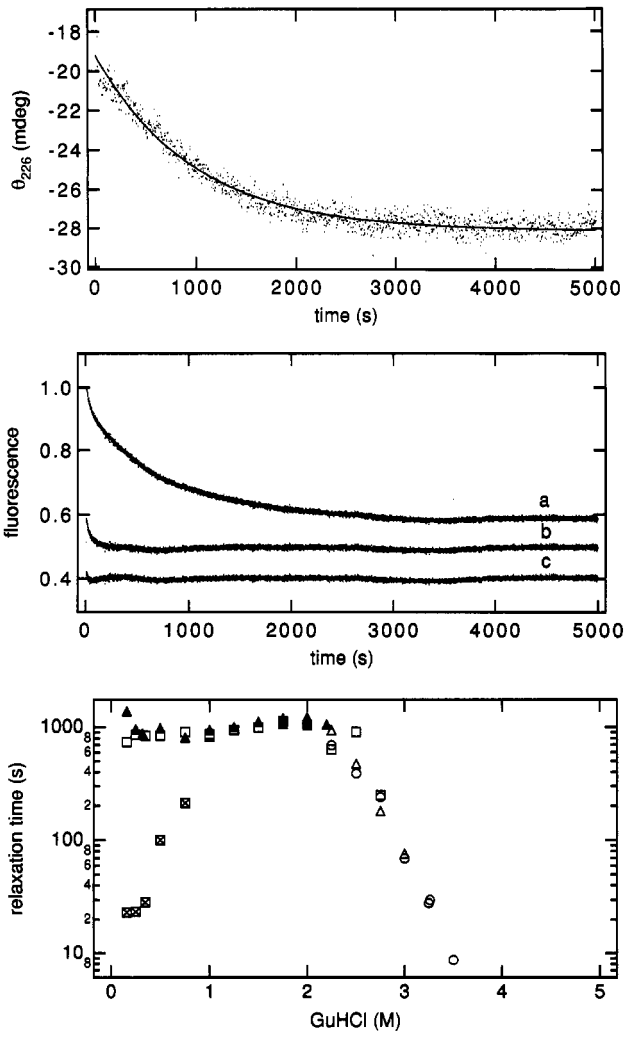


FIGURE 3: Kinetics of refolding at 0.5 M GuHCl monitored by CD (A, top) and by fluorescence (B, middle). In (A), the fit of a single exponential is also shown. In (B), the observed decay (a) and the residuals of the fit to a single exponential (b) and double exponentials (c) are shown. The residuals are offset for comparison. (C, bottom) Dependence of the relaxation time on the final GuHCl concentration. (○) and (△) correspond to the unfolding reactions observed by fluorescence and CD, respectively. (□) and (▲) correspond to the very slow phase detected by fluorescence and the single phase observed by CD, respectively, in the refolding kinetics. (Crosshatched □) corresponds to the slow phase of refolding detected by fluorescence.

to as the "slow" phase, the other being referred to as the "very slow" phase; the "fast" phase refers to reactions which are too fast to be detected by manual mixing. The relaxation time of the very slow phase is about 800–1000 s and exhibits marginal dependence on GuHCl concentration. The relaxation times of the very slow phase detected by fluorescence and CD are in excellent agreement, indicating that they correspond to the same reaction. The relaxation time of the slow phase depends on the GuHCl concentration, although it appears to reach a constant value at low GuHCl concentration (Figure 3).

Under marginally native conditions, the initial points of the refolding reaction detected by CD are close to the unfolded baseline (Figure 1). Thus the amplitude of the very slow phase accounts for almost the entire change expected from the equilibrium study. As the residual GuHCl concentration decreases, the amplitude of the very slow phase gradually decreases.

Since the refolding reaction is very slow, we recorded the CD spectrum immediately after the initiation of refolding; acquisition started 60 s after initiation, and the spectrum was

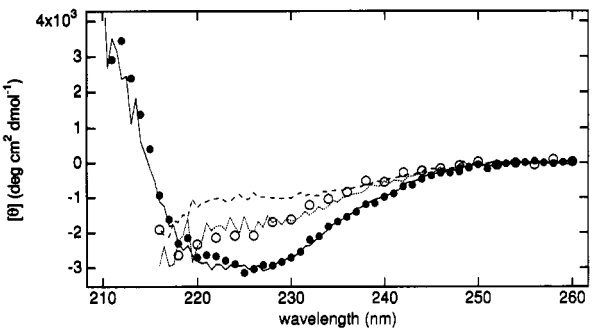


FIGURE 4: CD spectrum taken 60 s after the initiation of refolding in 0.32 M GuHCl (○). The spectrum of the same sample taken after 2 h (●), that of the native protein (continuous line), that of the unfolded protein in 5 M GuHCl (broken line), and a calculated spectrum of the unfolded protein in 0.32 M GuHCl (dashed line) are also shown. The protein concentration was 10 μM in the refolding reaction and 50 μM for the native protein. The spectrum of the unfolded protein was calculated by linear extrapolation of the unfolded baseline (Figure 1) to 0.32 M at each wavelength.

recorded in 60 s (Figure 4). It is expected that only 10% of the very slow refolding species would be converted to the native form within the acquisition period. The spectrum is similar to that calculated for the unfolded protein and differs significantly from that of the native protein. This confirms that the very slow folding phase accounts for the major change in ellipticity at wavelengths longer than 220 nm. However, it should be noted that there is a significant difference between the extrapolated spectrum in 0.32 M GuHCl and the experimental spectrum in 5 M GuHCl of the unfolded state because of the slope of the unfolded baseline (Figure 1).

In contrast to the CD measurements, the initial fluorescence intensity following initiation of refolding is much higher than expected for the unfolded protein from the equilibrium experiments (Figure 2). As the residual GuHCl concentration decreases, the amplitude of the very slow phase decreases with concomitant increase of the amplitude of the slow phase. The relaxation times and the relative amplitude of the two phases are independent of protein concentration from 1 to 10 μM (data not shown), showing that the observed kinetics are not due to aggregation. The decrease of the very slow phase amplitude is parallel to that observed in the kinetic CD experiments (Figure 1). The initial points seem to be linearly related to the GuHCl concentration under native conditions; in the transition region (1.5–2.5 M) the initial point approaches the unfolded baseline at increasing GuHCl concentration. These results suggest that one or more intermediates with a higher fluorescence intensity are formed within the dead time of mixing.

Proline Isomerization. Since the refolding reaction observed by manual mixing is slow, even under strongly native conditions, and the relaxation time of the slow phase is independent of final GuHCl concentration, the folding may well be limited by the *cis–trans* isomerization of prolines (Brandts et al., 1975; Kim & Baldwin, 1982). We performed several tests to examine this possibility.

First, the slow isomerization in the unfolded protein was examined by double-jump experiments. Apo-Pc was unfolded for a short time and transferred to refolding conditions. Figure 5 shows the observed amplitude as a function of the duration of the unfolding time; the amplitude of the slow and very slow phases increases as the duration of unfolding increases. The same experiment was also performed by CD, and a similar slow buildup of the amplitude was found (data not shown). These results clearly establish that a slow isomerization process in the unfolded protein produces the two phases. They also

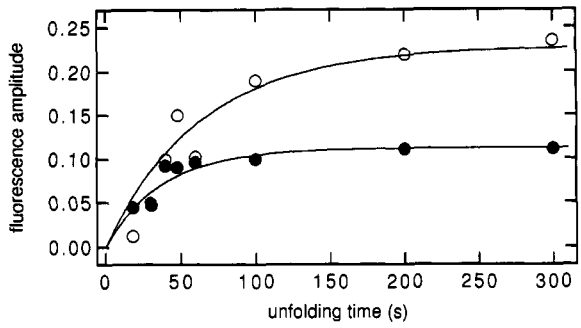


FIGURE 5: Double-jump experiments on apo-Pc. The amplitudes of the slow phase (●) and the very slow phase (○) detected by fluorescence are normalized as described in Figure 2. The curves are the theoretical time course according to the unfolding mechanism described under Discussion, where the rate constants of *trans* to *cis* and *cis* to *trans* isomerization were set at 1.55×10^{-2} and 1.55×10^{-4} s^{-1} for Pro16 and 2.1×10^{-2} and 5.25×10^{-3} s^{-1} for Pro36, respectively, and the unfolding rate was set at 4 s^{-1}.

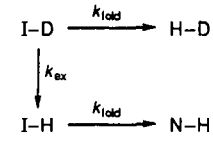
indicate that the fast refolding species, which is predominant when folding is initiated immediately after denaturation, folds so rapidly that its refolding reaction cannot be detected by the manual mixing method. Second, the temperature dependence of the relaxation time of the very slow refolding phase at 0.32 M GuHCl was monitored by CD, and its activation enthalpy was determined to be 23.9 ± 0.5 kcal mol^{-1} (data not shown), which is within the range expected for proline isomerization (Brandts et al., 1975). Finally, the refolding reaction was followed by fluorescence in the presence of cyclophilin, which has peptidyl–prolyl isomerase activity (Hasel et al., 1991). In the presence of 5 μM cyclophilin, the slow and very slow phases were accelerated by factors of 2.1 and 3.9, respectively. The relative fluorescence intensity of the slow phase was about 33% of the total intensity, independent of the cyclophilin concentration (data not shown). Because both phases are affected by cyclophilin, we can conclude that the *cis–trans* isomerization of at least two prolines is responsible for the appearance of the two phases and that proline isomerism is the rate-determining step under strongly native conditions.

Experiments To Probe the Nature of the Kinetic Intermediates. An intermediate is trapped before the very slow refolding step. Several experiments were undertaken to probe the nature of this intermediate. A diagnostic test for a molten globule is a strong affinity for the hydrophobic fluorescent probe ANS (Semisotnov et al., 1991), which detects exposed hydrophobic surfaces. We carried out a refolding experiment in the presence of ANS. Fluorescence spectra of ANS in the presence and absence of the native apoprotein were almost identical, indicating no strong binding of ANS to native apo-Pc. Little fluorescence enhancement (≤20% relative to native apo-Pc) was observed in the spectrum recorded 2 min after initiation of refolding, in contrast to the substantial enhancement caused by typical molten globule intermediates (Semisotnov et al., 1991).

The refolding reaction was followed by size exclusion chromatography, to detect formation of compact intermediate species. In a refolding experiment, in which unfolded protein was injected into a column equilibrated under conditions where intermediate species should be present at high concentrations, a peak was observed at a retention time similar to that of the native apoprotein, consistent with the formation of a compact species. However, since control experiments using a different chromatography column showed a peak with a longer retention time than that of the native protein, we could not discount the possibility that the intermediate(s) was (were) binding to both columns, and we therefore cannot draw any conclusions from this experiment about the compactness of the intermediate.

Hydrogen-Exchange Experiments. The intermediate which folds to the native apoprotein in the very slow refolding reaction was characterized by measuring the protection of backbone amide protons from solvent exchange. Because problems with precipitation were encountered in refolding buffer containing 0.5 M Na_2SO_4, even at low protein concentrations, the refolding buffer was replaced with one containing 1 M NaCl. In this buffer, the protein does not precipitate during refolding at concentrations up to *ca.* 100 μM. Control experiments were performed to show that the folding properties in the NaCl buffer are similar to those in the Na_2SO_4 buffer. Under the new conditions, equilibrium unfolding apparently involves a two-state transition with a midpoint of 1.2 M GuHCl. Two kinetic phases, with relaxation times similar to those measured using the Na_2SO_4 buffer, were observed in the refolding reaction at 0.1 M GuHCl, both in H_2O and D_2O solutions. In the D_2O buffer, the relaxation times are 32 s and 1040 s for the slow and very slow phases, respectively. The fluorescence amplitude associated with the very slow phase is approximately 60% of the total amplitude observed in refolding. The relaxation time of the very slow phase does not change significantly at the lower pH used in the competition step (750 and 830 s at pH 6.0 and 5.2, respectively).

Hydrogen-exchange experiments were performed in the competition mode, in which amide exchange competes with the refolding reaction (Schmid & Baldwin, 1979; Kim, 1986; Miranker et al., 1991). The competition was initiated 90 s after the start of refolding, which allows the faster reactions to go to completion (*ca.* 94%) before the competition step and populates only the intermediate of interest and the native species. A similar delay was used in the characterization of a folding intermediate of ribonuclease A (Udgaonkar & Baldwin, 1988). The contribution of the native species to the hydrogen-exchange behavior is measurable; the means whereby the measurements were corrected for this contribution are discussed below. The competition method applied after an interval of 90 s enables identification of amide protons that are protected due to the formation of hydrogen-bonded intermediates in the very slow refolding reaction. A simplified scheme for the competition experiment is (Schmid & Baldwin, 1979; Kim, 1986; Miranker et al., 1991)



where I and N denote the intermediate and native states, respectively; D and H indicate deuteron and proton at a given amide, respectively; k_{ex} and k_{fold} are rate constants for hydrogen-exchange and refolding reactions, respectively. Assuming that the unfolding rate, back-exchange rate (from H to D in this case), and exchange rate in the native state are negligible, the proton occupancy for a given amide is approximated by

$$\text{proton occupancy} = \frac{k_{ex}}{k_{ex} + k_{fold}} = \frac{k_{int}/k_{fold}}{k_{int}/k_{fold} + P} \quad (1)$$

where k_{int} is the intrinsic hydrogen-exchange rate constant in an unstructured peptide and P is the protection factor (k_{int}/k_{ex}). Under the conditions of the competition experiment, the ratios of the intrinsic exchange rates to the rate of the very slow refolding reaction, k_{int}/k_{fold}, are 22–480 at pH 5.2, depending on the local amino acid sequence (Bai et al., 1993). At pH 6.0, k_{int}/k_{fold} should be 5.7 times faster than at pH 5.2, since exchange is catalyzed by hydroxide ion at neutral pH

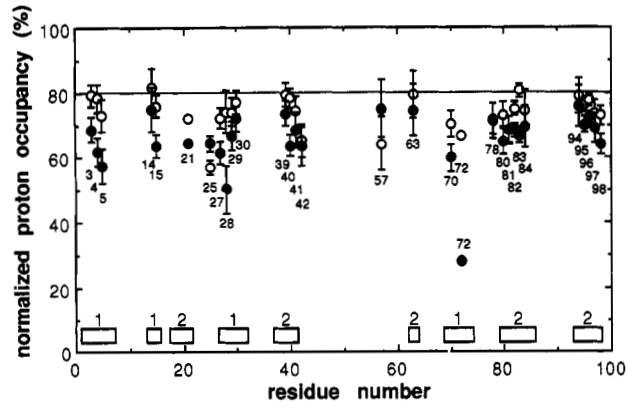


FIGURE 6: Normalized proton occupancy of an amide proton plotted versus residue number. Open circles show occupancies after the competition at pH 6.0 and closed circles show those at pH 5.2. Proton occupancy is normalized by $(I_C - I_D)/(I_H - I_D) \times 100$, where I_C, I_D, and I_H denote peak heights of an amide proton in the competition, deuterated control, and protonated control spectra, respectively. Errors are estimated from the standard deviation of noise in "empty" portions in the NH–C$^\alpha$H region of the TOCSY spectra; where no error bars appear, the estimated error would be contained within the symbols. The positions of the β-strands are indicated (1 and 2 indicate sheet I and sheet II, respectively), and the amide protons used as probes are labeled with their residue number in the sequence of Pc. The line at 80% occupancy represents an estimate of the baseline for complete exchange, taking into account the composition of the buffer (90% H_2O/10% D_2O) and the estimated contribution of already folded native apo-Pc (*ca.* 10%).

(Englander & Mayne, 1992). Therefore, if an amide proton is not protected, exchange occurs before significant formation of the native state, and nearly full proton occupancy should be observed; if it is protected, so that k_{ex} is comparable to or smaller than k_{fold}, lower proton occupancy should be observed. The upper limit of proton occupancy for any amide proton probe should be 90% in the case of complete exchange, because the competition was carried out in 90% H_2O/10% D_2O solution; the lower limit is 0% for complete protection.

The competition reaction was performed at two pHs, with deuterated and protonated control experiments for each. After the competition reaction, the refolded apo-Pc was reconstituted with copper, and a 2D TOCSY spectrum was recorded. Proton occupancies for all amide protons which are highly protected in the reconstituted holo-Pc were measured from cross-peak intensities in the TOCSY spectra. Under the experimental conditions used, a total of 46 amide protons were observed to be highly protected in the protonated control holo-Pc; the remaining protons exchange too rapidly for observation under the conditions used. A total of 29 amide protons which, in the deuterated control, had low proton occupancies (less than 40% of that in the protonated control) were used as probes for the presence of secondary structure in the intermediate. From the proton occupancies in the deuterated controls, minimal protection factors for the probe amides in the native apoprotein are estimated as 7000. The relative proton occupancy at each site was determined as described under Experimental Procedures and is plotted for each residue in Figure 6.

At pH 6.0, the majority of the probes have proton occupancies close to 70–80%, which is lower than expected for complete exchange. Occupancies lower than for complete exchange (90%) can be interpreted in two ways: (i) an intermediate is formed in which amide protons are protected from exchange; (ii) a small amount of the native molecule is formed in addition to an intermediate, and amide protons are highly protected in the native molecule but not significantly protected in the intermediate. These two possibilities can be distinguished by the pH dependence of proton occupancies, since k_{ex} is pH-dependent (faster at higher pH), while k_{fold} is essentially pH-independent for apo-Pc. The occupancies for the probes near the C-terminus are very similar at pH 6.0 and 5.2, which suggests the presence of a species with highly protected probes, presumably the native protein, at the start of the competition reaction. This is not unexpected, since the competition was started a significant time (90 s) after initiation of refolding. The small amount of native protein already formed at the beginning of the competition period is excluded from the competition reaction. Because they are highly protected, probes in the native protein remain deuterated throughout the competition period. The consequent decrease in the population of the intermediate should uniformly lower the upper limit of proton occupancy for all probes. For apo-Pc we observe a lowering of occupancy for all probes of about 10%, so that, combined with the effect of 10% D_2O in the final buffer, the baseline value for complete exchange is about 80%.

Except for protection due to the small population of the native species, occupancy for most probes at pH 6.0 is generally close to the baseline value (80%). However, for several residues, in particular, Val 72, proton occupancies at pH 5.2 are significantly lower when compared both to the baseline value and to the corresponding occupancy at pH 6.0, clearly indicating that these amide protons are protected at the start of the competition period by formation of the intermediate species. According to eq 1, the occupancy should decrease at lower pH, where k_{int} is smaller and the protection factor P is comparable in size to k_{int}/k_{fold}. Comparison of protection between residues with different k_{int}'s is difficult without data from competition experiments performed over a wide range of pH. However, for residues with similar k_{int} the proton occupancy directly relates to their protection factors. The calculated value of k_{int} for Val72 is one of the lowest, and there are several probes which have similar k_{int}'s (Val3, Leu5, Val28, Val40, and Leu63), all of which have higher occupancy than Val72 (Figure 6). Assuming that 80% occupancy corresponds to complete exchange in the intermediate, and using the calculated k_{int} values, the protection factor P for each probe was calculated (Table II). Val72 is the most protected amide proton; Tyr70, which has higher k_{int}, is protected to a similar extent.

The possibility that the observed protection is determined by equilibrium unfolding of an intermediate can be eliminated, since then all amides should have identical protection factors. The intermediate is not significantly destabilized at the lower pH, since the occupancy for many probes decreases as the pH is lowered, indicating that the amide proton probes remain well protected. One can be confident that probes which have significantly lower occupancy at pH 5.2 than that at pH 6.0 (Figure 6) are protected in the intermediate (Table II). Other amides with higher k_{int} could also be protected to a similar extent, but because of the high intrinsic exchange rate, the protection is not detectable at these pHs. Experiments at much lower pH, where apo-Pc is not stable, would be required to allow unambiguous evaluation. The significantly protected amide protons are distributed throughout the two β-sheets of holo-Pc (Figure 7). The competition experiments clearly show that protection of amide protons occurs faster than the rate-determining step of the slow refolding process; however, the degree of protection in the intermediate (*ca.* 50) is much smaller than in the native apoprotein ($\geq$7000).

Direct Measurement of the Slow Refolding Process by 1D NMR. The very slow refolding reaction was characterized by 1D ^{1}H NMR spectroscopy at low temperature. The equilibrium unfolding transition monitored by CD does not

Table II: Protected Amides in the Very Slow Folding Process of Apo-Pc As Identified from the Hydrogen-Exchange Competition Experiments and Their Protection Factors

amide	proton acceptor[a]	protection factor[b]	amide	proton acceptor[a]	protection factor[b]
sheet I			sheet II		
Val3	Val28O	4 ± 2	Val21	Thr97O	17 ± 1
Leu4	Val15O	10 ± 3	Val40	Tyr83O	7 ± 2
Leu5	Lys30O	12 ± 4	Tyr80	Val96O	36 ± 12
Val15	Leu4O	11 ± 3	Phe82	Gly94O	35 ± 7
Ile27	Val72O	14 ± 3	Tyr83	Val40O	22 ± 5
Val28	Leu1O	13 ± 5	Lys95	Ser17O	34 ± 10
Tyr70	Phe29O	52 ± 15	Val96	Tyr80O	5 ± 1
Val72	Ile27O	49 ± 2	Val98	Gly78O	14 ± 4

[a] Acceptors of the amide protons in the solution structure of holo-Pc (Moore et al., 1991). [b] The protection factor is calculated on the assumption that the proton occupancy corresponding to complete exchange in the intermediate is 80% (see text). Errors do not include possible error in this assumption.

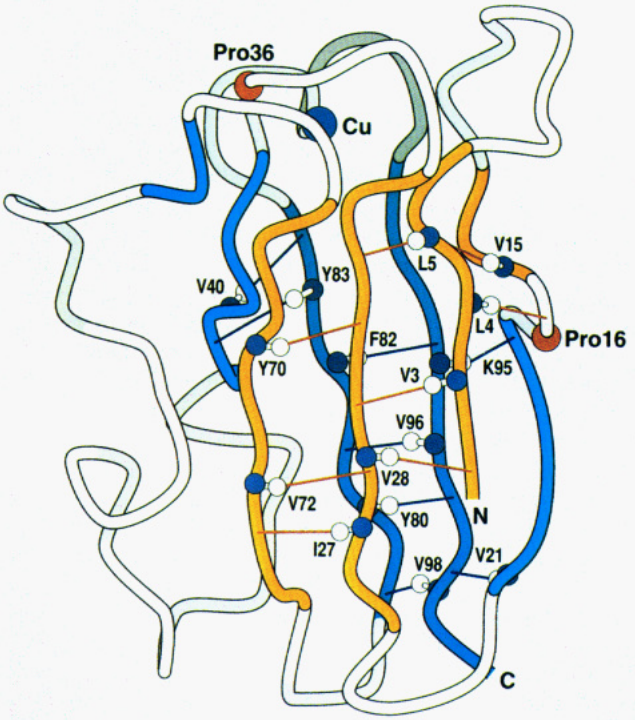


FIGURE 7: Schematic drawing of the solution structure of French bean holo-Pc (Moore et al., 1991). Strands in sheet I (front) are drawn in yellow and those in sheet II (back) in blue. The amide protons protected in the intermediate are shown as white spheres attached to their amide nitrogens (dark spheres) with residue numbers. Hydrogen bonds involving the protected amides in the holoprotein are drawn in red in sheet I and in blue in sheet II. The two *cis*-proline residues and a copper atom are shown as spheres with labels. N and C indicate the N- and C-termini of the peptide chain.

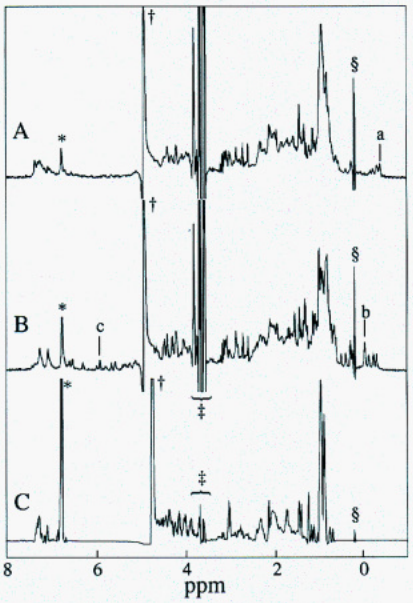


FIGURE 8: ^{1}H NMR spectra of apo-Pc recorded 22 min (A) and 15 h (B) after the initiation of refolding. Apo-Pc was unfolded in D_2O containing 50 mM potassium phosphate, 1 M NaCl, 50 μM EDTA, 1 mM dithiothreitol-d_{11}, and 2.2 M GuHCl (pH* 7.0), and refolding reaction was started by a 1:22 dilution with the buffer without GuHCl. NMR spectra were recorded at 5 °C, and 512 transients were accumulated in 11 min. A Lorentz-to-Gaussian window function was used prior to Fourier transformation. Spectra A and B are from an identical sample, processed in an identical manner and plotted on a same scale. (C) Spectrum of unfolded apo-Pc in buffer containing 2.5 M GuHCl. The protein concentration was 70 μM in (A) and (B) and 300 μM in (C). Peaks used in kinetic analyses (Figure 9) are indicated by a, b, and c. According to the assignment of holo-Pc (Chazin & Wright, 1988), peak b (–0.05 ppm) may correspond to $C^{\gamma}H_3$ resonances of Ile27 and Ile55; peak c (5.92 ppm) may correspond to either C^{α} protons or phenylalanine ring protons. *, †, ‡, and § denote peaks from GuHCl, residual water, EDTA and 2-mercaptoethanol, and impurity, respectively.

differ significantly between 5 and 25 °C (data not shown). To the best of our knowledge, no systems have been reported to have different folding mechanisms over this small range of temperature, and we therefore expect the results of the 1D NMR experiment to be applicable to refolding at 25 °C. The refolding reaction was initiated at 0 °C on ice, and a series of 1D NMR spectra were recorded at 5 °C. Figure 8A shows the spectrum recorded 22 min after initiation; under these conditions, the refolding reaction is slowed considerably, so that the very slow folding reaction has not commenced to any significant extent. A spectrum recorded 15 h after initiation of refolding is almost identical to that of the native apoprotein under the same conditions (Figure 8B). The spectrum recorded after 22 min differs from that of the unfolded protein in 2.5 M GuHCl (Figure 8C) in that several resonances are shifted from random-coil positions. In particular, some aliphatic resonances are observed to have surprisingly large upfield shifts, into the region between 0.5 and –0.5 ppm (Figure 8A), suggesting the presence of specific tertiary interactions in the intermediate state. We note that these signals do not arise from the presence of a small amount of fully folded native protein, since many resolved and prominent resonances in the native apo-Pc spectrum (Figure 8B) are absent from the spectrum of the intermediate. Also, the resonance at –0.4 ppm is shifted to higher field than any resonances in the native state.

The time course of the very slow refolding reaction can be followed by 1D NMR. Figure 9 shows the time-dependent changes in the intensities of the three resonances identified in Figure 8A,B. The time courses are fitted well by single exponentials with relaxation times of 7.5–10.1 h, in good agreement with the value (5.2 h) estimated for the very slow refolding reaction at 5 °C from the measured relaxation time

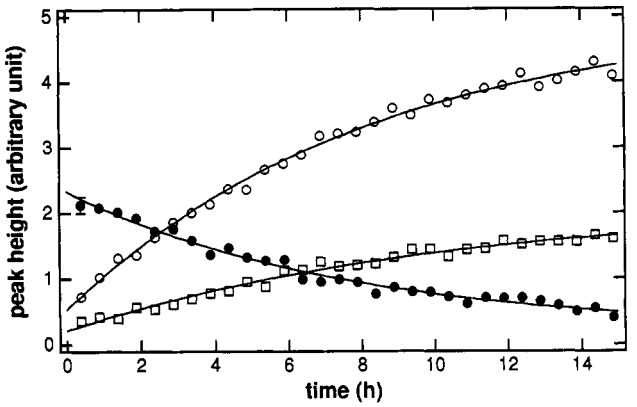


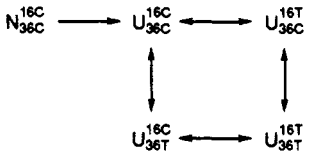
FIGURE 9: Kinetics of refolding monitored by 1D NMR. Peak heights of peaks a (●), b (○), and c (□) shown in Figure 8 are plotted as a function of refolding time. The experiment was performed as described in Figure 8, and the 1D NMR spectrum was recorded every 30 min. Curves show the best fits of a single exponential. Relaxation times are determined as 7.5 (0.8), 8.2 (0.6), and 10.1 (1.5) h for a, b, and c, respectively, the numbers in parentheses indicating standard deviations. The error shown on the first point of peak a is estimated from the standard deviation of noise in the spectrum, and errors are essentially identical in all the spectra.

at 25 °C and the activation energy. The time course for peak b indicates that formation of over 90% of the native molecule is associated with the observed reaction.

DISCUSSION

Kinetic Model for Unfolding. In the unfolding reaction, no intermediates were observed. However, the unfolded protein consists of a mixture of slow and very slow refolding species. The double-jump experiments indicate the presence in the equilibrium unfolded state of a small amount of a fast folding species, which refolds within the dead time of manual mixing. The slow and very slow refolding phases, which are accelerated by cyclophilin, are attributed to isomerization of *trans*-prolines in the unfolded protein to the *cis* isomer in the native state. French bean and poplar holo-Pc contain two *cis*-proline residues, Pro16 and Pro36 (Moore et al., 1991; Guss & Freeman, 1983). NMR studies of peptide fragments of Pc indicate that most of the peptides derived from Pc have low propensity to form secondary structures and adopt predominantly extended conformations in aqueous solution (Dyson et al., 1992b). In the peptide fragments, Pro16 exists predominantly in the *trans* isomer with no detectable population of *cis*, while approximately 20% of Pro36 is in the *cis* state. It is reasonable to expect that both prolines will adopt similar *cis*/*trans* ratios at equilibrium in the unfolded protein. The remaining prolines in the native protein are *trans*, and they are uniformly *trans* in the peptide fragments as well. Thus, it is reasonable to assign the two prolines responsible for the slow refolding species to Pro16 and Pro36. Scheme I shows a kinetic model that explains the isomerization in unfolded apo-Pc. N and U denote the native and unfolded states, respectively, and the *cis*/*trans* isomerization states are indicated for Pro16 and Pro36.

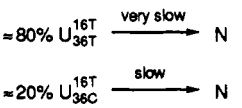
Scheme I



Kinetic Model for Refolding. In the refolding reaction under strongly native conditions, two kinetic phases are observed. The kinetic results for the refolding reaction can be interpreted in terms of a sequential model, in which unfolded protein refolds through two reactions that follow each other, or a parallel model in which two (or more) species in the unfolded state refold on two separate pathways. The model for unfolding (Scheme I) indicates that four species can be present in the unfolded protein. The double-jump experiment suggests that U_{36C}^{16C}, with the correct proline isomers, refolds in the dead time of the manual mixing kinetic experiments. The two slower refolding phases could then, in principle, involve three species with incorrect prolines. On the reasonable assumption that *cis*/*trans* ratios of the prolines are similar in the unfolded protein and in the peptide fragments, Pro16 would be predominantly *trans* and Pro36 would be ~80% *trans* and 20% *cis* in the unfolded protein (Dyson et al., 1992b). Accordingly, only two of the four unfolded species, U_{36C}^{16T} and U_{36T}^{16T}, would be well populated at equilibrium; U_{36T}^{16C} would not be significantly populated at equilibrium and can therefore be neglected.

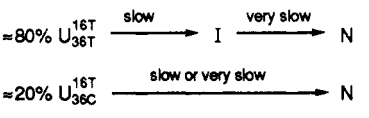
The very slow phase contributes the major change in the fluorescence and CD kinetic experiments and appears to be responsible for the majority of the native protein formed during the hydrogen-exchange experiments. In a parallel model, the refolding reaction of U_{36T}^{16T} would correspond to the very slow phase and that of U_{36C}^{16T} to the slow phase. Although not explicitly indicated in Scheme II, intermediates may well be formed on the parallel folding pathways.

Scheme II



An alternative, sequential, model could also be invoked to explain the data. In this model, the slow phase must precede the very slow phase to account for the fact that it is observable experimentally. The model also requires at least one folding intermediate on the pathway. The refolding of the minor unfolded species, U_{36C}^{16T}, may have a slow or very slow rate. I represents a folding intermediate. Although any model must have a parallel character since there are two unfolded species, we term Scheme III a sequential model because the folding of the major unfolded species is sequential and includes both of the experimentally observed rates.

Scheme III

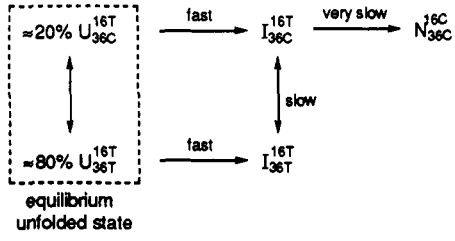


The hydrogen-exchange experiments provide the key to allow discrimination between the two models. After 90 s of refolding at 25 °C, the starting point for hydrogen-exchange competition, the population of the native state is estimated to be about 10% (as discussed previously). On the basis of the fluorescence and CD refolding kinetics, the slow phase is almost complete in 90 s and the very slow phase is approximately 8% completed in this time. According to Scheme II, 26% of the native protein should be formed 90 s after the initiation of refolding. In contrast, 8% of the native protein would be formed after 90 s according to Scheme III, if the refolding of U_{36C}^{16T} is associated with the very slow phase. [If the refolding of U_{36C}^{16T} is associated with the slow phase, the sequential mechanism is essentially the same as the parallel one, and 26% of the native protein should be formed after 90 s.] The absence of resonances characteristic of the native

apoprotein in the first 1D NMR spectrum recorded after initiation of refolding together with the kinetics observed by 1D NMR clearly indicates that over 90% of the native apoprotein is formed only on the very slow time scale and that little, if any, native protein refolds on a faster time scale (Figure 9). We conclude that the sequential model with U_{36C}^{16T} refolding on the very slow time scale (Scheme III) fits the experimental data better than the parallel model (Scheme II).

The fluorescence-refolding experiments, in which the initial points are much higher than the unfolded baseline (Figure 2), indicate formation of at least one intermediate species with distinctive fluorescence properties prior to the proline isomerization steps. It is most likely that the intermediate formed first from U_{36C}^{16T} would be I_{36C}^{16T}, and similarly, I_{36T}^{16T} would be the first intermediate formed from U_{36T}^{16T}. Since both the major and minor species should fold with the very slow rate, as discussed above, we can connect their folding pathways by assuming that the final intermediate before the formation of the native form is I_{36C}^{16T}. Thus we assign the slow refolding step to the conversion of $I_{36T}^{16T} \rightarrow I_{36C}^{16T}$ by isomerization of Pro36 and the very slow step to conversion of I_{36C}^{16T} to the fully folded native form by isomerization of Pro16. The simplest scheme that includes these steps and accounts for all of the experimental data is shown in Scheme IV.

Scheme IV



Other features of the refolding of apo-Pc can be explained on the basis of this model. To explain the fluorescence profile of the refolding kinetics (Figure 2), the highly fluorescent intermediate should be I_{36T}^{16T}, and I_{36C}^{16T} should have lower fluorescence according to Scheme IV. The dependence of the relaxation time of the slow phase on residual GuHCl concentration (Figure 3) can be explained by a kinetic coupling between isomerization and folding (Kiefhaber et al., 1992); under strongly native conditions, the isomerization of Pro36 is the rate-determining step of the reaction from I_{36T}^{16T} to I_{36C}^{16T}, while under marginally native conditions, some folding process not involving proline isomerism becomes the rate-determining step. In the latter case, I_{36C}^{16T} may be destabilized, to the point where the highly fluorescent intermediate I_{36T}^{16T} is accumulated. This would be consistent with the observation that the larger fluorescence amplitude is apparently associated with the very slow phase under marginally native conditions (Figure 2).

Although we omit U_{36T}^{16C} in the refolding model shown above, U_{36T}^{16C} should be accumulated transiently during the unfolding reaction (Scheme I) and might be detected by double-jump experiments. Computer simulations were carried out to fit the data obtained from the double-jump experiment. We assumed that the two prolines isomerize independently and used the same *cis/trans* population ratios for the two prolines. A refolding model that includes U_{36T}^{16C}, which is assumed to refold on the slow time scale, fits the data (Figure 5). The proposed mechanism fits all of the presently available experimental data. We plan to perform mutagenesis experiments to directly identify the prolines involved in the isomerization processes and to probe further the apoplastocyanin folding pathway.

The Trapped Intermediate. Since the refolding reaction is very slow, we were able to apply a variety of techniques to investigate the major, very slow refolding phase of apo-Pc. According to the proposed model (Scheme IV), the reaction I_{36C}^{16T} to N_{36C}^{16C} should be the rate-determining step for refolding of U_{36T}^{16T} and U_{36T}^{16C}. Therefore, the folding intermediate characterized by hydrogen-exchange and 1D NMR experiments should correspond to I_{36C}^{16T}.

The hydrogen-exchange experiment demonstrates that several amide protons are partially protected from exchange in the intermediate (Table II, Figure 6), suggesting that a number of hydrogen bonds are present. These protected amides are distributed throughout the two β-sheets found in the native protein. It is probable that nativelike β-sheet structure is formed in the intermediate, since amide protons are protected for pairs of residues (4–15, 27–72, 40–83, and 80–96) which have mutual hydrogen bonds in the native holoprotein (Guss & Freeman, 1983; Moore et al., 1991). Although we cannot entirely eliminate the possibility of nonnative hydrogen bond formation, the observation of protected amide pairs provides strong evidence that the secondary structure is similar to that in the native protein. However, it should be noted that the amide protons are only weakly protected in the intermediate, certainly to a much lesser extent than in the native apoprotein, indicating that the β-sheet structure is much less stable in the intermediate.

The 1D NMR spectrum of the intermediate recorded immediately after initiation of refolding (Figure 8A) reveals several resonances shifted from random-coil positions. While the spectrum shows more dispersion than that of the unfolded state (Figure 8C), the dispersion is considerably less than for the native state (Figure 8B). The absence of resonances characteristic of the native apoprotein (Figure 8A), as well as the changes in the spectrum with time (Figure 9), indicates that the initial spectrum corresponds to that of the intermediate, with only a minor contribution from the native state. From the 1D NMR spectrum it is clear that many of the defined structural features of the native state are absent from that of the intermediate. Few resonances are observed in the NMR spectrum of the intermediate in the region between 5 and 6 ppm, where $C^{\alpha}H$ resonances of residues in β-sheet are frequently observed (Wishart et al., 1991); most of the 23 residues with $C^{\alpha}H$ chemical shifts downfield of 5 ppm in holo-Pc are located in the β-sheets (Chazin & Wright, 1988; Moore et al., 1991), and if the sheets were stably and rigidly folded in the intermediate, the $C^{\alpha}H$ resonances of 8 of the residues in Table II (Leu4, Tyr70, Phe82, Tyr83, Lys95, Val96, and Val98) should be observable between 4.9 and 5.4 ppm. The high-field aromatic resonances of Phe29 and Phe82 (5.7–6.3 ppm), two residues which are packed together in the core of the native protein, are absent from the spectrum of the intermediate. Several upfield-shifted resonances are observed in the spectrum of the intermediate, some of which are further upfield than the highest field resonances in the native apoprotein (–0.32 ppm); the presence of these high-field resonances indicates the presence of a local structured region(s) in which aliphatic side chains are packed specifically against aromatic residues. Taken together, these results suggest that the β-sheets in the intermediate remain flexible and are not packed as specifically or tightly as in the native apoprotein. Further, the intermediate has a significantly altered packing of the hydrophobic core. The formation of the tightly packed sheets in the native protein from the loosely arranged structurein the intermediate is thus blocked by the

presence of an incorrect proline isomer.

Slow refolding processes limited by proline isomerization have been studied by means of hydrogen exchange for two ribonucleases (Udgaonkar & Baldwin, 1988, 1990; Mullins et al., 1993), which are both α/β proteins. In both proteins, rapid protection of probes in the β-sheets is observed. In the case of ribonuclease A, protection factors of the probes exceed 1000 at 0.4 s from the initiation of folding. In both proteins, rapid buildup of a nativelike peptide CD spectrum is also detected, and the intermediates with the incorrect proline isomers are found to be enzymatically active (Kim & Baldwin, 1982; Kiefhaber et al., 1990). These observations indicate that an incorrect proline isomer does not necessarily prevent formation of highly organized secondary structure. In the refolding of apo-Pc, however, both the lack of highly protected amide protons in the intermediate and the limited dispersion of its NMR spectrum suggest that the intermediate has an incorrect proline at a crucial position for the formation of the native structure. According to the model (Scheme IV), this proline would be Pro16. It is not difficult to imagine how the *trans* isomer would prevent proper packing of the two sheets, because *cis*-Pro16 occurs in a tight turn (type VI) (Guss & Freeman, 1983; Moore et al., 1991) that directly connects the two β-sheets (Figure 7).

The region of the protein with the most protected amide protons corresponds to a part of the sequence where initiation of secondary structure formation is expected. NMR studies of peptide fragments of Pc (Dyson et al., 1992b) showed hydrophobic clustering involving residues Ile27–Phe29. In the native protein, the amide protons and carbonyl oxygens of Val72 and Ile27 are mutually hydrogen bonded (Guss & Freeman, 1983; Moore et al., 1991). The present competition experiments show that Tyr70, Val72, Ile27, and Val28 are protected from hydrogen exchange (Table II), suggesting stabilization of secondary structure in this region. The density of hydrophobic side chains in this region is very high (Ile27, Val28, Phe29, Tyr70, Val71, and Val72), and hydrophobic clustering involving these residues may be important in the initiation of refolding, as suggested previously on the basis of studies of peptide fragments (Dyson et al., 1992b).

The CD spectrum taken immediately following initiation of refolding (Figure 4) is significantly different from that of the native apoprotein, suggesting that the secondary structure of the intermediate is different. The CD spectrum of native Pc is in good agreement with a theoretical spectrum calculated from the three-dimensional structure (Manning & Woody, 1987). The minimum of the spectrum at 226 nm also agrees with that of the pure component for antiparallel β-sheet which was recently determined from a data set containing proteins with high β-sheet content (Perczel et al., 1992). The CD spectrum of apo-Pc can therefore be explained largely by its secondary structure. However, a theoretical study predicts that the CD spectrum of two closely packed β-sheets is strongly dependent on the intersheet separation and their relative orientation (Manning & Woody, 1987). Accordingly, the difference in ellipticity between the intermediate and the native apoprotein could well be attributable to flexibility and altered packing of the β-sheets, consistent with the NMR data.

Is the Intermediate a Molten Globule? There is increasing evidence for the participation of molten globule states in protein-folding pathways (Kuwajima, 1989; Ptitsyn et al., 1990; Ptitsyn, 1991; Jennings & Wright, 1993). Several of the properties of the intermediate are consistent with its being a molten globule. Secondary structure is present, with rather small protection factors of the amide protons. Similar results have been obtained for the molten globule intermediate of apomyoglobin, for which protection factors are in the same range as those of the apo-Pc intermediate (Hughson et al., 1990). The NMR spectrum of the apo-Pc intermediate is poorly dispersed, with the exception of a few resonances in the upfield region. All of these properties are consistent with a molten globule, where secondary structure is present but specific tertiary structure is absent. However, there is evidence of specific hydrophobic interactions in the apo-Pc intermediate, and the large exposed hydrophobic surface typical of a molten globule is absent, as shown by the ANS-binding experiment. We therefore conclude that the intermediate species cannot be strictly described as a molten globule according to the currently accepted definitions, although it shows some of the properties of a molten globule. We also note that most examples of molten globule intermediates are observed early in the folding pathway, as transient species (Ptitsyn et al., 1990), whereas the apo-Pc intermediate is long-lived and occurs relatively late in folding.

Implications for Biosynthesis and in Vivo Folding of Plastocyanin. The slow folding of apoplastocyanin and its instability relative to the holoprotein are of considerable interest in view of what is known about the biosynthesis and folding of plastocyanin *in vivo*. Apoplastocyanin is synthesized in the cytoplasm as a higher molecular weight precursor that is imported into the chloroplast posttranslationally (Merchant & Bogorad, 1986b; Smeekens et al., 1986; Li et al., 1990). Following import the protein is processed to its mature size, and copper is incorporated. An essential requirement for transport across a membrane is that a protein be maintained in an unfolded or incompletely folded conformation while awaiting translocation (Randall & Hardy, 1989; Hartl & Neupert, 1990). While chaperones undoubtedly participate in the folding and transport processes *in vivo*, the intrinsically slow rate of folding of apoplastocyanin might well be advantageous for efficient translocation to the thylakoid lumen by extending the time during which the polypeptide would be competent for entry into the export pathway.

Plastocyanin biosynthesis in certain green algae and cyanobacteria is regulated by the availability of copper (Wood, 1978; Sandmann & Boger, 1980). Studies of regulation in *Chlamydomonas* show that biosynthesis, transport, and processing of pre-apoplastocyanin are unaffected in cells grown in copper-deficient media; regulation occurs by rapid degradation of the nature apo-Pc within the chloroplasts (Merchant & Bogorad, 1986a,b). Regulation by rapid proteolytic degradation *in vivo* is understandable in view of our observation that the folded native conformation of apo-Pc is only marginally stable toward denaturation. Although apo-Pc adopts a folded structure that is very similar to that of the holoprotein at high salt concentrations (Garrett et al., 1984), the much lower stability of the folded apoprotein would render it highly susceptible to proteolytic degradation under the conditions that are expected to prevail within the chloroplast.

CONCLUSIONS

Apo-Pc provides a rare opportunity to characterize the structure of intermediates formed on the folding pathway of a predominantly β protein. We have shown that fluorescence, CD, and NMR spectroscopic information is consistent with a folding pathway for apo-Pc that includes two slow steps associated with proline isomerization from *trans* to *cis*. A well-populated intermediate state trapped before the final, very slow refolding step has been characterized by hydrogen-exchange competition experiments and 1D NMR spectroscopy. Further detailed characterization of the intermediate as well

as of earlier events in the formation of the intermediate could provide novel insights into the initiation of β-sheet formation and into the folding pathways of proteins with the Greek key β-barrel topology. Such experiments are in progress.

ACKNOWLEDGMENT

We thank Dr. J. G. Sutcliffe for providing mouse cyclophilin, Dr. R. M. Ghadiri for access to the fluorometer, Drs. P. A. Jennings, M.-F. Jeng, and D. Morikis for helpful discussions, and L. Tennant for technical assistance. We also thank Dr. S. W. Englander for making available the revised Molday factors before publication.

REFERENCES

Bai, Y., Milne, J. S., Mayne, L., & Englander, S. W. (1993) *Proteins 17*, 75–86.

Barshop, B. A., Wrenn, R. F., & Frieden, C. (1983) *Anal. Biochem. 130*, 134–145.

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

Braunschweiler, L., & Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521–528.

Bychkova, V. E., Berni, R., Rossi, G. L., Kutyshenko, V. P., & Ptitsyn, O. B. (1992) *Biochemistry 31*, 7566–7571.

Bycroft, M., Matouschek, A., Kellis, J. T., Jr., Serrano, L., & Fersht, A. R. (1990) *Nature 346*, 488–490.

Chazin, W. J., & Wright, P. E. (1988) *J. Mol. Biol. 202*, 623–636.

Collyer, C. A., Guss, J. M., Sugimura, Y., Yoshizaki, F., & Freeman, H. C. (1990) *J. Mol. Biol. 211*, 617–632.

Creighton, T. E. (1990) *Biochem. J. 270*, 1–16.

Dyson, H. J., & Wright, P. E. (1991) *Annu. Rev. Biophys. Biophys. Chem. 20*, 519–538.

Dyson, H. J., & Wright, P. E. (1993) *Curr. Opin. Struct. Biol. 3*, 60–65.

Dyson, H. J., Merutka, G., Waltho, J. P., Lerner, R. A., & Wright, P. E. (1992a) *J. Mol. Biol. 226*, 795–817.

Dyson, H. J., Sayre, J. R., Merutka, G., Shin, H.-C., Lerner, R. A., & Wright, P. E. (1992b) *J. Mol. Biol. 226*, 819–835.

Englander, S. W., & Mayne, L. (1992) *Annu. Rev. Biophys. Biomol. Struct. 21*, 243–265.

Finkelstein, A. V. (1991) *Proteins: Struct., Funct., Genet. 9*, 23–27.

Garrett, T. P. J., Clingeleffer, D. J., Guss, J. M., Rogers, S. J., & Freeman, H. C. (1984) *J. Biol. Chem. 259*, 2822–2825.

Gill, S. C., & von Hippel, P. H. (1989) *Anal. Biochem. 182*, 319–326.

Goto, Y., & Hamaguchi, K. (1982) *J. Mol. Biol. 156*, 891–910.

Guss, J. M., & Freeman, H. C. (1983) *J. Mol. Biol. 169*, 521–563.

Hartl, F.-U., & Neupert, W. (1990) *Science 247*, 930–938.

Hasel, K. W., Glass, J. R., Godbout, M., & Sutcliffe, J. G. (1991) *Mol. Cell. Biol. 11*, 3484–3491.

Hughson, F. M., Wright, P. E., & Baldwin, R. L. (1990) *Science 249*, 1544–1548.

Jeng, M.-F. (1991) Ph.D. Thesis, University of Pennsylvania, Philadelphia, PA.

Jennings, P. A., & Wright, P. E. (1993) *Science* (in press).

Kiefhaber, T., Quaas, R., Hahn, U., & Schmid, F. X. (1990) *Biochemistry 29*, 3061–3070.

Kiefhaber, T., Kohler, H.-H., & Schmid, F. X. (1992) *J. Mol. Biol. 224*, 217–229.

Kim, P. S. (1986) *Methods Enzymol. 131*, 136–156.

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

Kim, P. S., & Baldwin, R. L. (1990) *Annu. Rev. Biochem. 59*, 631–660.

Kuwajima, K. (1989) *Proteins 6*, 87–103.

Kuwajima, K., Yamaya, H., Miwa, S., Sugai, S., & Nagamura, T. (1987) *FEBS Lett. 221*, 115–118.

Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A., & Wright, P. E. (1989) *Science 245*, 635–637.

Li, H.-M., Theg, S. M., Bauerle, C. M., & Keegstra, K. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 6748–6752.

Manning, M. C., & Woody, R. W. (1987) *Biopolymers 26*, 1731–1752.

Matouschek, A., Kellis, J. T., Jr., Serrano, L., & Fersht, A. R. (1989) *Nature 340*, 122–126.

Matthews, C. R. (1987) *Methods Enzymol. 154*, 498–511.

Merchant, S., & Bogorad, L. (1986a) *Mol. Cell. Biol. 6*, 462–469.

Merchant, S., & Bogorad, L. (1986b) *J. Biol. Chem. 261*, 15850–15853.

Miranker, A., Radford, S. E., Karplus, M., & Dobson, C. M. (1991) *Nature 349*, 633–636.

Molday, R. S., Englander, S. W., & Kallen, R. G. (1972) *Biochemistry 11*, 150–158.

Moore, J. M., Case, D. A., Chazin, W. J., Gippert, G. P., Havel, T. F., Powls, R., & Wright, P. E. (1988) *Science 240*, 314–317.

Moore, J. M., Lepre, C. A., Gippert, G. P., Chazin, W. J., Case, D. A., & Wright, P. E. (1991) *J. Mol. Biol. 221*, 533–555.

Mullins, L. S., Pace, N. P., & Raushel, F. M. (1993) *Biochemistry 32*, 6152–6156.

Oas, T. G., & Kim, P. S. (1988) *Nature 336*, 42–48.

Perczel, A., Park, K., & Fasman, G. D. (1992) *Proteins 13*, 57–69.

Ptitsyn, O. B. (1987) *J. Protein Chem. 6*, 273–293.

Ptitsyn, O. B. (1991) *FEBS Lett. 285*, 176–181.

Ptitsyn, O. B., Pain, R. H., Semisotnov, G. V., Zerovnik, E., & Razgulyaev, O. I. (1990) *FEBS Lett. 262*, 20–24.

Ramshaw, J. A. M., Brown, R. H., Scawen, M. D., & Boulter, D. (1973) *Biochim. Biophys. Acta 303*, 269–273.

Randall, L. L., & Hardy, S. J. S. (1989) *Science 243*, 1156–1159.

Roder, H., Elöve, G. A., & Englander, S. W. (1988) *Nature 335*, 700–704.

Ropson, I. J., Gordon, J. I., & Frieden, C. (1990) *Biochemistry 29*, 9591–9599.

Rudolph, R., Siebendritt, R., Nesslaüer, G., Sharma, A. K., & Jaenicke, R. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 4625–4629.

Sandmann, G., & Boger, P. (1980) *Planta 147*, 330–334.

Santoro, M. M., & Bolen, D. W. (1988) *Biochemistry 27*, 8063–8068.

Schellman, J. (1978) *Biopolymers 17*, 1305–1322.

Schmid, F. X., & Baldwin, R. L. (1979) *J. Mol. Biol. 135*, 199–215.

Semisotnov, G. V., Rodionova, N. A., Razgulyaev, O. I., Uversky, V. N., Gripas, A. F., & Gilmanshin, R. I. (1991) *Biopolymers 31*, 119–128.

Smeekens, S., Bauerle, C., Hageman, J., Keegstra, K., & Weisbeek, P. (1986) *Cell 46*, 365–375.

Sugawara, T., Kuwajima, K., & Sugai, S. (1991) *Biochemistry 30*, 2698–2706.

Sykes, A. G. (1985) *Chem. Soc. Rev. 14*, 283–315.

Udgaonkar, J. B., & Baldwin, R. L. (1988) *Nature 335*, 694–699.

Udgaonkar, J. B., & Baldwin, R. L. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8197–8201.

Varley, P., Gronenborn, A. M., Christensen, H., Wingfield, P. T., Pain, R. H., & Clore, G. M. (1993) *Science 260*, 1110–1113.

Wishart, D. S., Sykes, B. D., & Richards, F. M. (1991) *FEBS Lett. 293*, 72–80.

Wood, P. M. (1978) *Eur. J. Biochem. 87*, 9–19.

Wright, P. E., Dyson, H. J., & Lerner, R. A. (1988) *Biochemistry 27*, 7167–7175.