Characterization of Collapsed States in the Early Stages of the Refolding of Hen Lysozyme†

Charles J. Morgan, Andrew Miranker,‡ and Christopher M. Dobson*

Oxford Centre for Molecular Sciences, New Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QT, U.K.

Received December 23, 1997; Revised Manuscript Received March 27, 1998

ABSTRACT: Early conformational states in the refolding of hen lysozyme from guanidine hydrochloride have been characterized by measuring both the fluorescence and the solvent exchange properties of tryptophan side chains. The indole proton occupancies indicate that at pH 5.5, 25 °C, half the protection against pulse labeling occurs in the dead time (4 ms) of the experiment, with the remaining protection developing with a time constant of 55 ms. Comparison of these data with the protection kinetics of backbone amides and with the fluorescence data provides evidence for hydrophobic collapse involving incorporation of tryptophan residues in a solvent-excluded state in advance of stable secondary structure formation. Analysis of the pH dependence of the indole hydrogen exchange protection is consistent with two or more structurally distinct collapsed states, and indicates that the generation of a correctly folded compact hydrophobic core is a key precursor to the formation of persistent native-like structure during refolding.

An extreme starting point for protein folding is the random coil state, an ensemble of conformations involving only local interactions between residues of the polypeptide chain (*1*, *2*). Proteins in highly denaturing conditions, such as high concentrations of urea or guanidine hydrochloride, can resemble such a state closely, possessing a wide distribution of both individual bond torsion angles and overall dimensions. The process of folding when such an ensemble is placed under conditions where the native state is thermodynamically stable involves the formation of nonlocal interactions within the polypeptide chain. During folding, the formation of native-like contacts between residues directs the protein conformation toward the native state, a process that can be represented as the movement of molecules over an energy surface or landscape (*3*−*5*).

The early stages in the refolding reaction of a protein are particularly important in determining its overall character (*6*). Characterization of the properties and conformations of the earliest states sampled during folding is therefore paramount to our understanding of the folding process and the rules determining folding. A particularly important early event in the folding of a least some proteins is the formation of collapsed structures. For example, in the case of the 10-kDa protein barstar, a collapsed state with solvent exposed side chains has been reported to form with a time constant of 400 μs (*7*), and fluorescence measurements from an ultrarapid mixing technique monitoring the refolding of cytochrome *c* at 0.2 M guanidine hydrochloride have reported that collapse occurs in less than 50 μs (*8*). In the refolding of hen lysozyme, spectroscopic measurements demonstrate that, as well as collapse, structure far in excess of that characteristic of a highly unfolded state forms within the dead time (ca. 4 ms) of the observation (*9*). Although such structure is not sufficiently persistent to protect amide hydrogens from rapid exchange with solvent (*10*), it appears that it is sufficient to commit different populations of molecules to distinct folding routes, along which different intermediates are formed (*9, 11, 12*).

In the work described in this paper, we seek to characterize the collapsed state formed early in the folding of hen lysozyme from its chemically denatured state in guanidine hydrochloride. The four disulfide bonds are maintained throughout the process, and the folding is monitored by probing the properties of the tryptophan side chains. Tryptophan indoles are an excellent reporting group for the study of collapsed states as they are hydrophobic and can be observed by complementary methods, namely, fluorescence and solvent exchange. In the case of lysozyme, three tryptophans are located in the hydrophobic core of the molecular structure (*13*). Moreover, mutation studies have shown that they are important in determining the kinetics of folding (*14*), clustering of hydrophobic groups around tryptophan residues in denatured lysozyme has been reported (*15*), and studies of stable, partially folded states have suggested their key role in stabilizing structure (*16*). In this work we suggest that at least partial and heterogeneous collapse occurs prior to the formation of stable secondary structure.

† C.J.M. was supported in part by a Harris Scholar award, St. Hugh's College, Oxford. A.M. was supported by a Junior Research Fellowship, Christ Church, Oxford. The research of C.M.D. is supported in part by an International Research Scholars award from the Howard Hughes Medical Institute. This is a contribution from the Oxford Centre for Molecular Sciences, which is supported by BBSRC, EPSRC, and MRC.

* Corresponding author. Telephone: +44-(0)-1865-275916. Fax: +44-(0)-1865-275921. E-mail: chris.dobson@chem.ox.ac.uk-

‡ Present address: Department of Molecular Biophysics and Biochemistry, Yale University, 260 Whitney Ave., P.O. Box 208114, New Haven, CT 06520-8114.

S0006-2960(97)03150-4 CCC: $15.00
Published on Web 05/23/1998

Table 1: Resonance Assignment and Solvent Exchange Rates in D_2O Solution of Lysozyme Tryptophan ϵNH Hydrogens

residue	ppm	Hx (s^{-1})[b]
Trp28	9.36	1.5×10^{-5}
Trp62	10.25	$>1.0 \times 10^{-2}$
Trp63	10.36	1.1×10^{-4}
Trp108	10.06	3.6×10^{-5}
Trp111	10.49	$<3.3 \times 10^{-6}$
Trp123	10.88	$<1.8 \times 10^{-6}$

[a] Chemical shift at 12 °C, for protein in 90% H_2O/10% D_2O solution, pH 3.8. [b] Exchange rate at 5 °C in 99.9% D_2O solution, pH 3.9; the rate of exchange for Trp62 is at least 10^{-2} s^{-1} (*42*).

MATERIALS AND METHODS

Hen egg white lysozyme (3× recrystallized) was purchased from the Sigma Chemical Company and dialyzed extensively against water at pH 3 to remove residual sodium acetate. For pulse labeling experiments, lyophilized protein was dissolved in 99.9% D_2O, heated to 80 °C for 10 min, and lyophilized again. This procedure was repeated twice to ensure that deuterons were present at all the labile sites.

The ϵNH resonances of the six tryptophan residues of hen lysozyme are well resolved and have previously been assigned at pH 3.8, 35 °C (*17*). One-dimensional (1D)[1] NMR spectra were collected at 12 °C where the resonances are resolved and the ϵNH protons of Trp28, Trp108, Trp111, and Trp123 have hydrogen deuterium exchange rates (Table 1) in the native state that are sufficiently long ($\tau > 4$ h) to allow for sample manipulation and data collection without significant loss of label.

Refolding experiments monitored by intrinsic fluorescence using an Applied Photophysics SX17-MV stopped-flow fluorimeter were initiated from a 20 mg/mL solution of lysozyme in 6 M guanidine hydrochloride by an 11-fold dilution of the denatured protein solution with 20 mM sodium acetate buffer, pH 5.5. Time-resolved spectra were obtained by performing a series of dilutions and monitoring the intrinsic fluorescence at successive wavelengths, 2 nm apart, using a monochromator with a slit width setting of 9 nm. The dead time, determined from the hydrolysis of 2,4-dinitrophenyl acetate (*18*), was <2 ms. The spectrum of a solution of 1.8 mg/mL lysozyme in 6 M guanidine hydrochloride solution was acquired on the same instrument. An excitation wavelength of 280 nm was used in all intrinsic fluorescence measurements. All spectra are baseline corrected but are uncorrected for a wavelength dependence of photomultiplier response. The intensity weighted average emission wavelength (*19*) is calculated for fluorescence spectra between 300 and 420 nm.

Pulse labeling experiments were carried out using a Biologic QFM-5 rapid mixing device equilibrated at 25 °C using lysozyme in which all labile sites were deuterated. The protein solution contained 20 mg/mL lysozyme in 6 M guanidine deuterochloride in D_2O, pH 6.0. The denatured protein was refolded by dilution from denaturant and pulse labeled at various time points with high pH buffered H_2O according to established methods (*9*, *20*). 10 vol of 20 mM sodium acetate refolding buffer, pH 5.5,was mixed with 1 vol of protein in 6 M guanidine deuterochloride, resulting in a solution at pH 5.2. After the refolding period, a pulse of high pH was achieved by mixing with 5 vol of 200 mM boric acid/sodium borate buffer solution at pH values sufficient to achieve a pH of 8.8, 9.2, 9.7, or 10.1 in the resultant solution. After 8.4 ms of base-catalyzed exchange, the reaction was quenched by the addition of 5 vol of 0.5 M acetic acid, to give a resultant pH close to 4. At this pH, lysozyme is in its native state, and the exchange with solvent of the main chain and side chain probes is sufficiently slow that the distribution of isotope labels can be maintained prior to acquisition of an NMR spectrum. Time zero samples were prepared by heating native lysozyme in a solution containing an isotopic distribution of 94% protons and 6% deuterons, which corresponds to the isotopic content of the labeling pulse. Ultrafiltration (Centricon C10 Amicon) at 4 °C was used to manipulate all solutions before 1D ^{1}H NMR studies. The concentration of protein after this step was ca. 0.3 mM, and the time taken between preparing the sample and acquisition of the spectrum was 4−5 h.

A control experiment was used to assess the extent of labeling of fully folded protein molecules exposed to the pulse labeling procedure. This was conducted by mixing 1 vol of solution containing prepared lysozyme dissolved in 0.54 M guanidine deuterochloride with 10 vol of 0.54 M guanidine hydrochloride in sodium acetate buffer. After 70 ms, the solution pH was raised to values between pH 8.8 and pH 10.1 for 8.4 ms, and then the solution pH was lowered to ca. pH 4 using 0.5 M acetic acid. All pH readings were made using a glass electrode and are uncorrected for solvent isotope effects (*21*). Spectra were obtained using a home-built spectrometer operating at 500 MHz using 580 acquisitions of 8K complex points and a spectral width of 7042 Hz. The sample temperature for acquisition of the spectra was 12 °C. All spectra were apodized with an identical Gaussian window function and zero filled to yield a final digital resolution of 0.4 Hz per point. The proton occupancies were determined as follows. The intensity values of each spectrum were divided by the intensity value of the well-resolved resonance of Leu17 δ-CH_3 at −0.7 ppm. These indole and amide intensity values were divided by the corresponding values from the time zero samples, which were defined to have unit proton occupancy at zero refolding time.

The locations of hydrogen bonds in the 1.33-Å resolution crystal structure of lysozyme (*22*) obtained from the Brookhaven Protein Database (*23*) (accession code 193l) were determined using the program HBPLUS v.3.15 (*24*).

RESULTS

The change in the local environments of tryptophan residues as a function of refolding time was monitored by measurement of the tryptophan fluorescence intensity at discrete emission wavelengths. This approach permits emission spectra of the protein to be determined as a function of refolding time (Figure 1A). Comparison of the fluorescence spectrum of HEWL in 6 M guanidine hydochloride with the spectrum obtained within the dead time of the measurement (<2 ms) reveals a substantial reduction both in the total fluorescence intensity (data not shown) and in

[1] Abbreviations: HEWL, hen egg white lysozyme; 1D, one dimensional; CIDNP, chemically induced dynamic nuclear polarization.

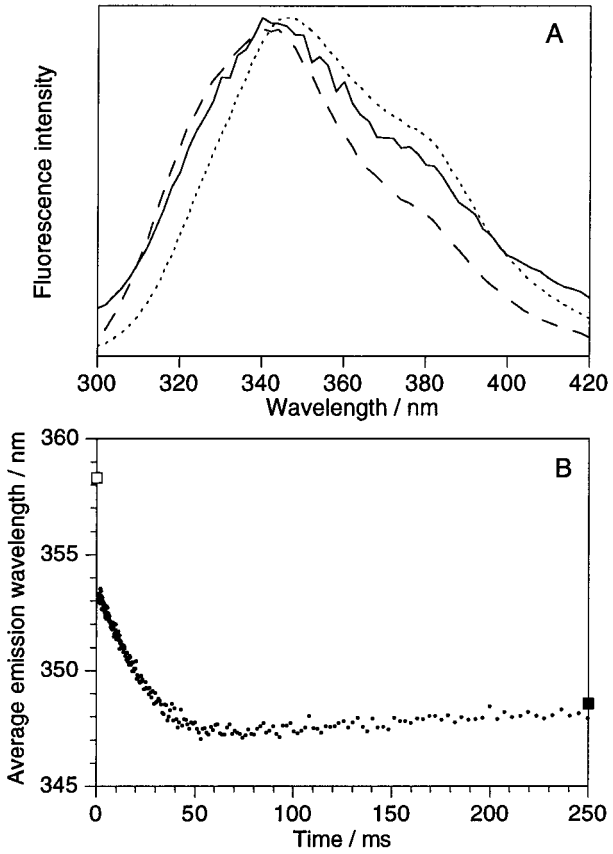

FIGURE 1: (A) Baseline-corrected, intrinsic fluorescence spectra of HEWL, 3 ms (—) and 10 s (− −) after initiation of refolding by 11-fold dilution of protein denatured in 6 M guanidine hydrochloride solution, pH 6.0, with 20 mM sodium acetate buffer, pH 5.5. The spectral intensity was normalized to the total intensity of the final spectrum (− −). For comparison, a baseline-corrected spectrum of HEWL in 6 M guanidine hydrochloride (- - -) is shown and is normalized to the total intensity. (B). Time course of the average emission wavelength during refolding by 11-fold dilution of denatured protein solution with 20 mM sodium acetate buffer, pH 5.5. The average emission wavelength value obtained 10 s after the initiation of refolding is marked as ■ and of HEWL in 6 M guanidine hydrochloride as □.

the calculated intensity weighted average emission wavelength (*19*). The reduction in intensity is consistent with previous reports (*25*) and is attributable, in part, to the lower concentration of guanidine hydrochloride in the bulk solvent (*25*). The average emission wavelength, however, is sensitive only to the degree of burial of tryptophan residues from aqueous solvent (*19*). Approximately half of the total change occurs within the dead time of the experiment, suggesting that a significant number of the indole side chains are included in a hydrophobic environment (Figure 1B). The acquisition of fluorescence spectra as described in this work has significant advantages over methods in which the total intrinsic fluorescence is measured since both the fluorescence quantum yield and the local tryptophan environment can be monitored simultaneously. The contribution of all six tryptophan indoles to the fluorescence spectra, however, precludes a more quantitative interpretation on a residue-specific basis.

The environment of individual indoles was probed by hydrogen exchange pulse labeling monitored by 1D ^{1}H NMR. Figure 2 shows that, for the tryptophan ϵNH proton resonances of Trp28, Trp108, Trp111, and Trp123, the peak

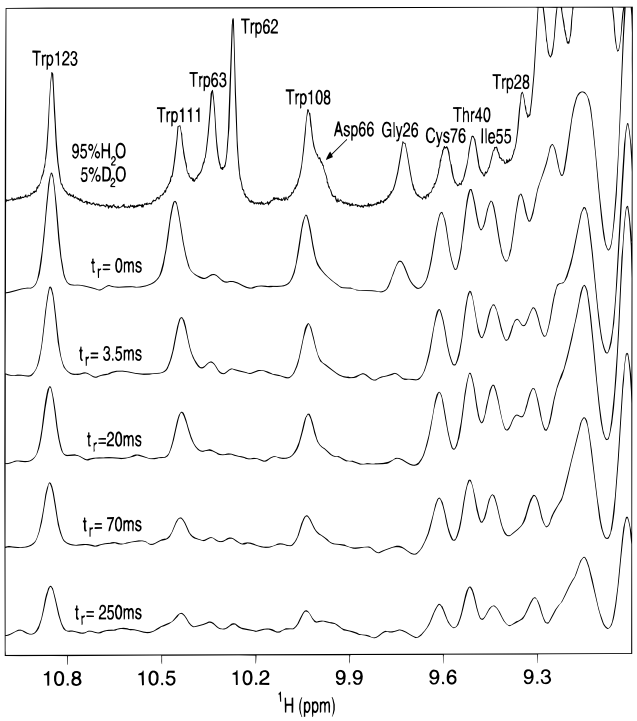

FIGURE 2: Proton occupancies determined from analysis of the 1D ^{1}H NMR spectra of HEWL. The protein was unfolded in 6 M guanidine deuterochloride, refolded for a variable time period, t_r, and then exposed to an 8.4 ms pulse at pH 10.1 in 94% H_2O/6% D_2O. Samples were subsequently transfered to ca. 99% D_2O, pH 3.8, and concentrated to ca. 0.3 mM prior to recording the NMR spectrum. The top trace shows an assigned 1D ^{1}H NMR spectrum of HEWL in 95% H_2O/5% D_2O, pH 3.8, 12 °C.

intensity decreases as the time delay between the dilution of deuterated denaturant and the pulse of high pH increases. This indicates the retention of deuterium at these sites and is consistent with a folding process in which these probes become incorporated in structures that afford protection against solvent exchange. Although the tryptophan ϵNH protons of Trp62 and Trp63 and the amide protons of Gly26 and Asp66 are resolved in the NMR spectrum in 95% H_2O, fractional proton occupancies could not be monitored reliably since in the native state exchange with solvent at these sites is rapid. For the four remaining tryptophan residues, the ϵNH fractional proton occupancy (Ω) data are well represented by a single exponential decay of the form $A \exp(-kt_r) + C$ (Figure 3). The non-zero tryptophan ϵNH fractional proton occupancy values of samples in which deuterated native lysozyme was exposed to a high pH labeling pulse indicate that native protein molecules are subject to labeling (*26*), and in practice this limits the range pH available for the labeling pulse. Since a reasonable agreement was found between the constant offset, *C*, and the fractional proton occupancy of labeled native protein, we attribute the constant offset to labeling of native protein and renormalize the proton occupancy as $(\Omega - C)/(1 - C)$. For example, at pH 10.1, the constant offset was found to vary from 0.14 (Trp28) to 0.38 (Trp123).

Several amide proton resonances are resolvable under the NMR acquisition conditions used here, allowing their protection kinetics to be monitored. The kinetics of protection for the amides contained within the α and the β folding domain are consistent with previously published results (*9*, *27*). In the case of the well-resolved amide resonances in

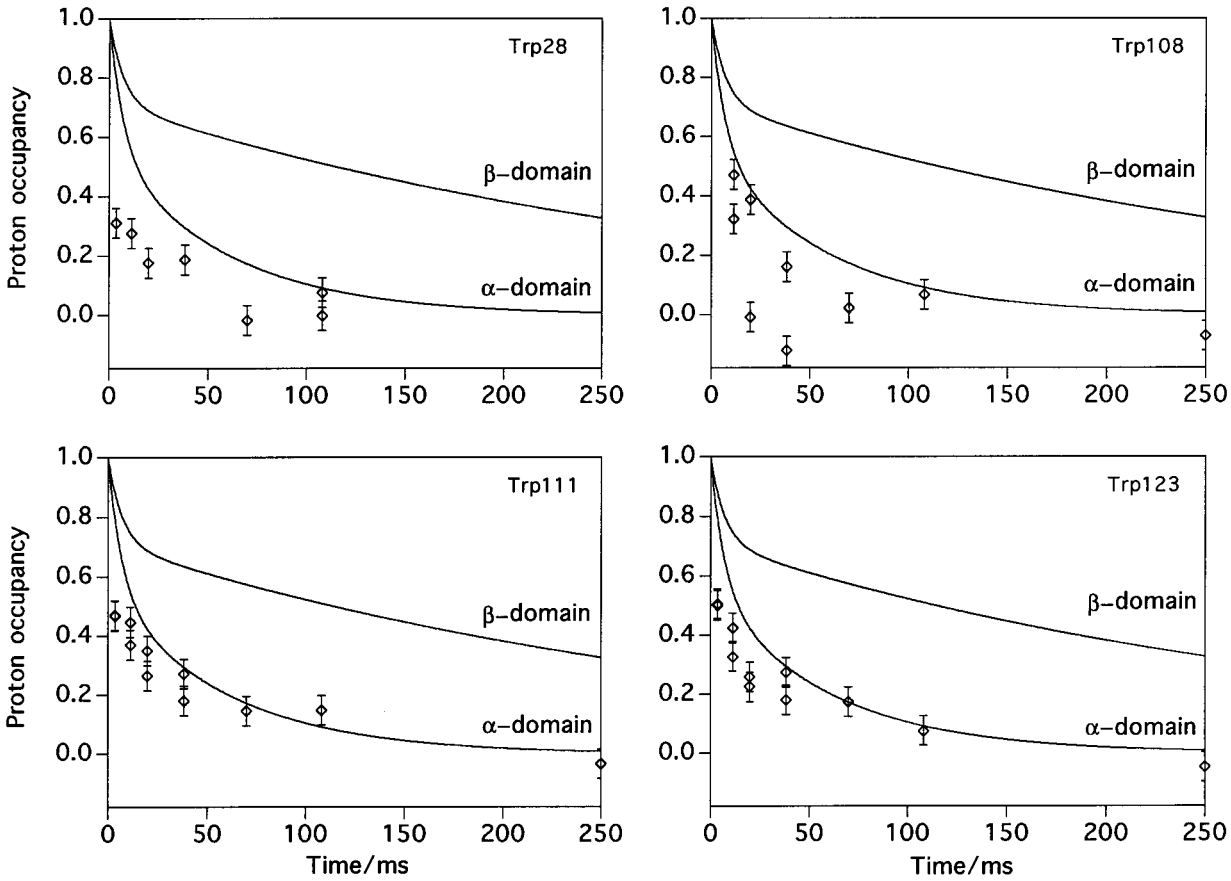

FIGURE 3: Comparison of the protection of main chain and side chain hydrogens against solvent exchange during the refolding of chemically denatured HEWL. The proton occupancies (◇, labeling pulse pH 9.7) of the four tryptophan ϵNH probes, determined in this work by 1D ^{1}H NMR, are compared to the reported behavior (*9*) of amide probes (—). For the α-domain, the proton occupancy is expressed numerically by the equation $0.45 \exp(-t/7) + 0.55 \exp(-t/60)$ and for the β-domain as $0.29 \exp(-t/7) + 0.71 \exp(-t/320)$, where t is the refolding time in milliseconds.

the β folding domain, the mean fractional proton occupancy values of residues Thr40 and Cys76 indicate that 16 ± 6% of molecules acquire protection at 8.2 ± 6.4 ms and 58 ± 4% at 278 ± 63 ms as compared to the reported values of 24% at 7.8 ms and 57% at 333 ms. The kinetics of protection of the amide proton of Ile55, which forms part of the β domain, could not be determined using previous 2D ^{1}H NMR methods. In this work, this residue is shown to have very similar behavior (21 ± 6% at 3.6 ± 2.3 ms, 57 ± 4% at 271 ± 67 ms) to that of other residues located in the β domain and is consistent with observations made in competition experiments (*27*). None of the resonances of amides from the α domain is well resolved in the 1D spectrum, but protection of several amide hydrogens having partially resolved resonances, including that of Leu8 at 8.8 ppm, was clearly faster than for the β domain and broadly consistent with the reported values of 47% at 3.9 ms and 38% at 70 ms. The protection of the indoles against hydrogen exchange, however, differs markedly from that of the amides most notably by the presence of a dead time event, which results in significant hydrogen exchange protection at the earliest measured time point (3.5 ms) and which is not found for the amides (Figure 2). Furthermore, subsequent protection of the indoles from exchange is adequately described using only a single rather than a double exponential decay.

The observation of biphasic kinetics for hydrogen exchange protection can be the result of either parallel or sequential refolding behavior. In the case of parallel behavior, where different populations of molecules fold with different kinetics, the biphasic protection arises directly as a result of the heterogeneity of the kinetic scheme. In the case of a sequential mechanism, however, all molecules follow the same kinetic route. If this route involves the formation of an intermediate that has weaker protection than that of the native state, then the biphasic kinetics can result from the differing degrees of protection. One approach to distinguish between these is to examine the sensitivity of the observed exchange behavior to the pH of the labeling pulse (*28*, *29*). In the experiments presented here, the labeling pulse was varied from pH 8.8 to pH 10.1, while the duration of the pulse was maintained at 8.4 ms. While the reported kinetics of amide protection show little variation in the fractional proton values (*9*), the kinetics of indole hydrogen protection however were found to be highly dependent on the pH of the labeling pulse (Figure 4). Notably, it is the magnitude of the dead time event that is substantially affected, varying from an average of 0.76 ± 0.05 at pH 8.8 to 0.45 ± 0.04 at pH 10.1. The subsequent kinetic phase is virtually unchanged with an average time constant of 55 ± 21 ms and 67 ± 24 ms measured with a labeling pulse of pH 8.8 and pH 10.1, respectively. The average time constant observed for the four indoles over the entire pH range was 50 ± 23 ms, consistent with the observed slow phase of protection 63 ± 13 ms previously reported for the amides of the largely α-helical folding domain of lysozyme.

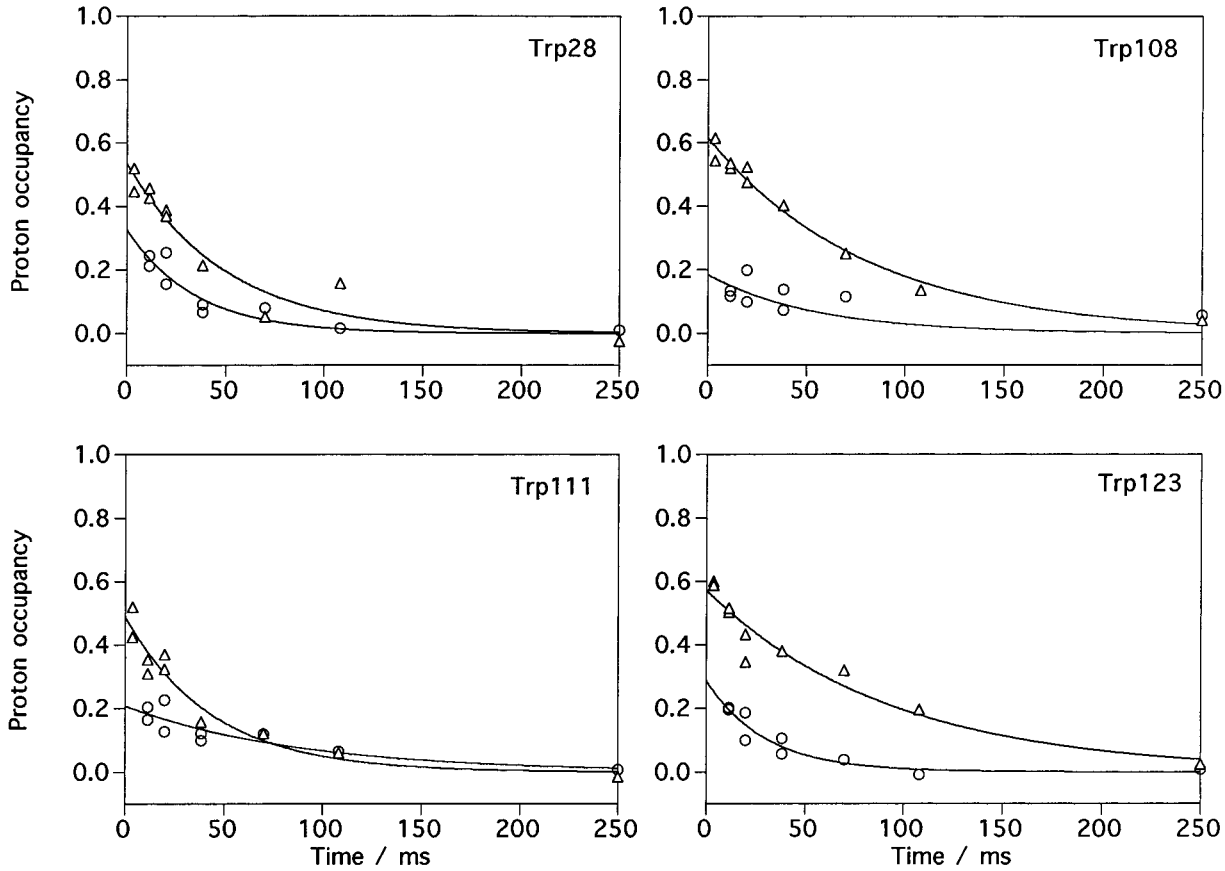

FIGURE 4: Proton occupancies of four tryptophan ϵNH probes as a function of the labeling pH (○, pH = 8.8; Δ, pH = 10.1). The solid lines represent single exponential fits to the data points. Only the extremes of labeling pulse pH used in this study are shown for clarity.

DISCUSSION

The locations of the six tryptophan indole side chains in the structure of hen lysozyme are shown in Figure 5. Trp62 and Trp63 form part of a loop located near to the active-site cleft between the α and β domains, and their ϵNH protons are solvent exposed. The side chain of Trp62 is known to be highly mobile since the side chain atoms have a poorly defined density in the crystal structure, the side chain conformation is completely disordered in the NMR solution structures (*13*), and a low order parameter is reported in ^{15}N NMR relaxation measurements (*30*). Trp28, Trp108, and Trp111 form part of the hydrophobic core of the α domain (*31*). The ϵNH protons of these residues make contact with the carbonyl oxygen of Tyr23, the carbonyl oxygen of Leu56, and a carbonyl oxygen of the side chain of Asn27, respectively. The ϵNH of Trp123 has an interaction with a crystal water and is sufficiently exposed to exhibit polarization in photo-CIDNP experiments (*32*). Nonetheless, the exchange of the ϵNH proton with solvent is sufficiently protected by the native state to allow determination of the fractional proton occupancies of Trp123 in this work.

Even at the earliest measurable stages of refolding, the average emission wavelength of the fluorescence spectrum is significantly lower than that of the spectrum of HEWL in 6 M guanidine hydrochloride. Since this parameter is a measure of the hydrophobicity of the local environment of the tryptophan fluorophores, we conclude that the indole side chains are sequestered from the aqueous, bulk solvent. Further burial is implied by the decrease in the average emission wavelength with a time constant of 33 ms. This observation is in excellent agreement with reports from experiments measuring the accessibility of indole side chains to iodide ions (40 ms) (*25*) and the conversion of collapsed states to a more stable intermediate at 0.54 M guanidine hydrochloride (31 ms) (*33*). In the hydrogen exchange experiments, the proton occupancies for the four observable tryptophan ϵNHs are broadly similar at each refolding time point and include a dead time hydrogen exchange protection event within 3.5 ms. A difference in the intrinsic exchange rates of the tryptophan ϵNHs and the amide NHs in the range of pH used here could in principle provide a trivial explanation for the differences in their hydrogen exchange behavior. Using published data (*34*, *35*), we calculate that the intrinsic exchange rate of an tryptophan ϵNH is ca. 3 times slower than that of an amide NH in an unstructured poly-DL-alanine polymer at pH 10. The slower rate is not of sufficient magnitude to account for the difference in proton occupancies presented here. It appears, therefore, that at the earliest detectable stages of refolding, the indole side chains are included within structure sufficient to protect against exchange.

Protection of the tryptophan ϵNH protons against exchange could arise either from the exclusion of bulk solvent or their involvement in hydrogen bonds to side chain or main chain acceptors. Although neither mechanism can be rejected on the basis of hydrogen exchange data alone, given the blue shift in the average emission wavelength, it seems most likely that the protection of indoles within the collapsed conformations arises largely from bulk solvent exclusion. The average emission wavelength values reported here are wholly con-

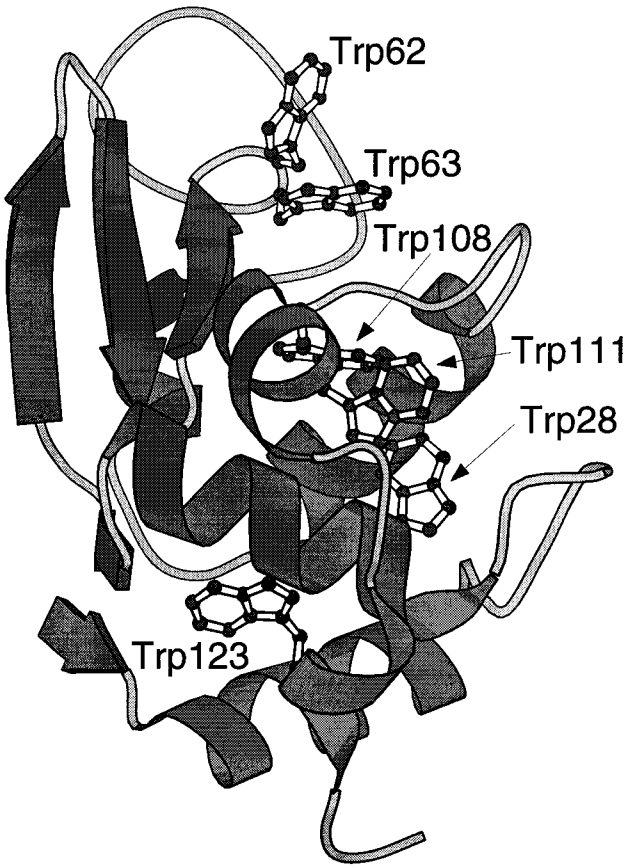

FIGURE 5: Schematic view of the native structure of hen lysozyme including the distribution of the six tryptophan residues. The drawing was generated using the program MOLSCRIPT (*43*) and the 1.33-Å crystal structure of lysozyme (*22*).

sistent with stopped-flow CIDNP measurements of the refolding of HEWL at 20 °C, in 1.4 M urea, pH 5.2, which suggest that 30 ms after the initiation of refolding, no tryptophan residue is significantly accessible to flavin dye (*36*). This, in conjunction with our present observations, suggests that the early events in the refolding of hen lysozyme are well described as a collapse of the polypeptide chain to generate compact structures in which bulk solvent is largely excluded. The ensemble of collapsed conformations that are transiently populated in the refolding of HEWL may in fact resemble the compact states formed at low pH and previously characterized for the structurally related proteins, α-lactalbumin and equine lysozyme (*16*, *37*).

One model that could explain the observation of a hydrogen exchange dead time event followed by a slower kinetic phase is that there is a single, dominant folding pathway in which the collapsed state is homogeneous and partially protects against solvent exchange. The corrected proton occupancy values were fitted to this model using an equation of the form $\exp(-t_r/\tau_r)(1 - \exp(-k_{int}t_p/P))$ where t_r is the refolding time period, τ_r is the refolding time constant, k_{int} is the base-catalyzed intrinsic exchange rate ($4.5 \times 10^{pH-7}$ s^{-1} at 25 °C (*35*)), t_p is the pulse length (8.4×10^{-3} s), and P is the protection factor of an indole in the collapsed state. The first part of this equation, $\exp(-t_r/\tau_r)$, represents the fraction of molecules in the collapsed state prior to the labeling pulse; the second part, $1 - \exp(-k_{int}t_p/P)$, expresses the proportion of these molecules that become labeled with a proton during the pulse. The five-parameter global fit results in a refolding time constant, τ_r, of 53 ms and protection factors of 55, 41, 46, and 32 for Trp28, Trp108, Trp111, and Trp123, respectively. These factors are small as compared to the native state, although they are large as compared to amide hydrogens at this stage of folding (*10*).

An alternative model accounts for the observed hydrogen exchange behavior by suggesting that two distinct populations are formed during a heterogeneous collapse. A proportion of the protein molecules have high protection factors, while the remainder are initially in a partially protected state that subsequently folds to one in which the protection factors are at least 500. The proton occupancy values were fitted in this model to an equation of the form $\delta \exp(-t_r/\tau_r)(1 - \exp(-k_{int}t_p/P))$ where δ is the fraction of partially protected molecules present immediately after mixing. The initial fraction of partially protected molecules, δ, determined by this six-parameter global fit is 0.54; the refolding time constant, τ_r, is 54 ms; and the protection factors, P, for Trp28, Trp108, Trp111, and Trp123 in the partially protected state are 11, 12, 7, and 5, values only slightly greater than those observed in substantially unfolded states (*38*).

The experimental fractional proton occupancy values and their pH dependences are successfully predicted to within our estimate of the experimental errors by both models, although the discrepancy between the calculated and experimental proton occupancy values is slightly smaller if the model of heterogeneous collapse is adopted. A comparison of the experimental, kinetic phase amplitudes for a labeling pulse of pH 9.7 with those calculated for the homogeneous model of collapse indicates that the individual tryptophan protection factors, P, lie in an extremely narrow range (Figure 6A). This clustering of protection factors implies that the local environment of the tryptophan residues in the collapsed state is very similar. Within native proteins, the protection factors of can vary by over 6 orders of magnitude (*39*), and for the previously characterized collapsed states, e.g., the acid states of α-lactalbumin and equine lysozyme, a variation of 3 orders of magnitude has been observed (*16*, *37*). It would be fortuitous for a variation in solvent accessibility to give rise to the nearly identical protection factors observed in this study. They are more likely, therefore, to arise from discrete states, protecting all of the indoles observed, interconverting with states in which the indoles are only very weakly protected. This situation is represented to some degree by the model of heterogeneous collapse in which the incorporation of a label depends on the relative proportions of the collapsed states and their rate of interconversion.

The model of heterogeneous collapse requires the existence of a population of protein molecules with strong protection and another with only weak protection at the early stages of refolding. Although the hydrogen exchange data reported here cannot reliably distinguish this from the sequential model, additional evidence for heterogeneity has been provided by experiments that measure the accessibility of tryptophan fluorophores to quenching by iodide ions as a function of the refolding time (*25*). It was shown that half the total change in quenching occurs within the dead time of the experiment and a further loss of quenching occurs with a time constant of ca. 40 ms at 20 °C. Both the time constant and fraction are consistent with a model of a

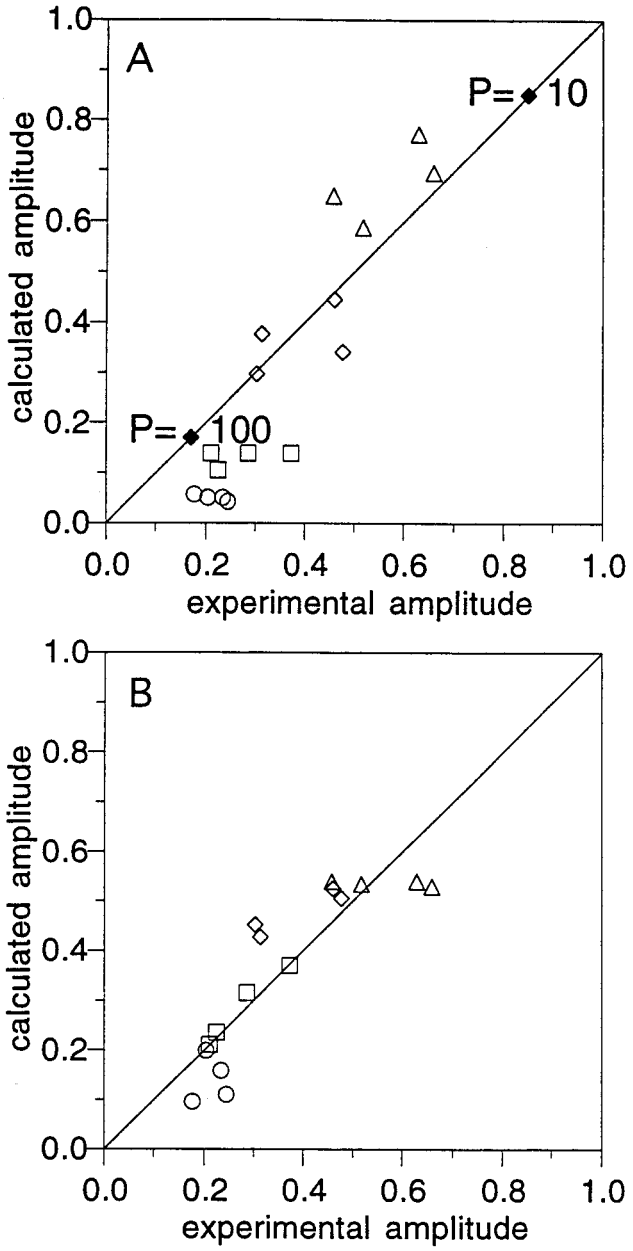

FIGURE 6: Comparison of the experimental kinetic amplitudes, *A*, derived from free fitting of the individual corrected proton occupancies to an equation of the form $A \exp(-kt)$ (a 32-parameter fit) and the amplitudes generated from a global fit to a five-parameter model of homogeneous collapse (A) or to a six-parameter model of heterogeneous collapse (B). As a guide, filled diamonds (♦) in (panel A) represent the homogeneous model amplitudes with a fixed labeling pulse of pH 9.7 for protection factors of 10 (0.85) and 100 (0.17). Experimental data with a variable labeling pulse pH is marked as follows: ○, pH 8.8; □, pH 9.2; ◇, pH 9.7; Δ, pH 10.1.

heterogeneous collapse giving rise to two states with different degrees of solvent exposure.

The four tryptophan residues studied in this work are located in the α folding domain of lysozyme. Amide hydrogen exchange protection data show that ca. 45% of the molecules rapidly (t < 10 ms) protect amides in this region (*9*). This is in good agreement with the proportion of molecules (54%) with highly protected tryptophan indole hydrogens in the heterogeneous model. We suggest that these events reflect the folding of molecules where the core packing of hydrophobic groups, including the tryptophan residues studied here, is sufficiently native-like in the early compact states to protect the indole hydrogens and to form rapidly secondary structure able to protect the amide hydrogens in the α domain. We suggest further that for the remaining molecules both the indole and amide hydrogen protection against exchange takes ca. 60 ms to develop because the initially formed compact state has to reorganize prior to forming native-like structure. This results in the loss of any transiently formed protection of the indole hydrogens as well as the slower formation of persistent secondary structure and amide hydrogen exchange protection.

An important observation in this study is that the kinetics of the development of hydrogen exchange protection for the tryptophan ϵNH of Trp108 are similar to those of the other three tryptophan residues that are located in the α domain. Although Trp108 forms part of the α domain (Figure 5), in the native structure the ϵNH is involved in a hydrogen bond to the carbonyl group of Leu56, located in the β domain. Its protection by native-like interactions is therefore expected to occur only when persistent structure in the β domain develops. The proton occupancy data for this residue are not of sufficient quality to model reliably the pH dependence of labeling independently from that of the other three tryptophans. It is clear, however, that the kinetics of protection are much faster than those for the protection of amide probes within the β domain (Figures 3 and 4). Furthermore, the similarity of the kinetics of protection of Trp108 to the kinetics of protection of amides located within the α domain suggest that, for a proportion of molecules, the protection of Trp108 arises as a consequence of the formation of the partly folded α domain. The exact nature of this protection has not been determined, but the development of non-native-state fluorescence (*25*) suggests that it may arise from the burial of the side chain of Trp108 in the compact, partly folded protein. Moreover, the structural reorganization resulting from the existence of such an interaction is likely to act to slow the folding by contributing to the generation of local minima in the energy surface. This is consistent with the observation that the change of a tryptophan residue for a tyrosine residue at position 108 results in a nearly 10-fold increase in the overall refolding rate (*14*), a result attributable to the destabilization of incorrectly collapsed structures or a reshaping of the energy landscape such that barriers are reduced (*40*).

In conclusion, collapse of the polypeptide chain of hen lysozyme is rapid, and the formation of protective structure occurs readily under refolding conditions involving low temperature and moderate pH. The hydrogen exchange data described here for the tryptophan residues, taken together with previous measurements, suggest that for hen lysozyme the collapse process is heterogeneous. In the model put forward here, some molecules collapse to conformations that closely resemble the native conformation and can fold rapidly; others collapse to 'misfolded' conformations for which rearrangement is necessary before the native state can be formed. For small proteins exhibiting single phase kinetics, collapse appears coincident with the formation of the native state (*41*) as subsequent steps are fast. For larger protein molecules, however, collapse does not necessarily lead to the rapid formation of native structure since the probability of forming non-native interactions during collapse is much greater. For these proteins, folding can be slow

and may involve substantial reorganisation steps prior to the acquisition of the native structure (*6*). The folding of hen lysozyme, under the refolding conditions utilized here, appears to represent a transitional situation in which two distinct types of folding behavior are seen for an ensemble of protein molecules having identical amino acid sequences.

ACKNOWLEDGMENT

We thank Drs. S. E. Radford and P. A. Evans for helpful discussions.

REFERENCES

1. Smith, L. J., Fiebig, K. M., Schwalbe, H., and Dobson, C. M. (1996) *Folding Des. 1*, R95−R106.
2. Shortle, D. R. (1996) *Curr. Opin. Struct. Biol. 6*, 24−30.
3. Bryngelson, J. D., Onuchic, J. N., Socci, N. D., and Wolynes, P. G. (1995) *Proteins: Struct., Funct., Genet. 21*, 167−195.
4. Dobson, C. M., Sali, A., and Karplus, M. (1998) *Angew. Chem., Int. Ed. Engl. 37*, 868−893.
5. Dill, K. A., and Chan, H. S. (1997) *Nat. Struct. Biol. 4*, 10−19.
6. Miranker, A. D., and Dobson, C. M. (1996) *Curr. Opin. Struct. Biol. 6*, 31−42.
7. Nolting, B., Golbik, R., Neira, J. L., Soler Gonzalez, A. S., Schreiber, G., and Fersht, A. R. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 826−830.
8. Chan, C. K., Hu, Y., Takahashi, S., Rousseau, D. L., Eaton, W. A., and Hofrichter, J. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 1779−1784.
9. Radford, S. E., Dobson, C. M., and Evans, P. A. (1992) *Nature 358*, 302−307.
10. Gladwin, S. T., and Evans, P. A. (1996) *Folding Des. 1*, 407−417.
11. Miranker, A. D., Robinson, C. V., Radford, S. E., Aplin, R. T., and Dobson, C. M. (1993) *Science 262*, 896−900.
12. Wildegger, G., and Kiefhaber, T. (1997) *J. Mol. Biol. 270*, 294−304.
13. Smith, L. J., Sutcliffe, M. J., Redfield, C., and Dobson, C. M. (1993) *J. Mol. Biol. 229*, 930−944.
14. Rothwarf, D. M., and Scheraga, H. A. (1996) *Biochemistry 35*, 13797−13807.
15. Schwalbe, H., Fiebig, K. M., Buck, M., Jones, J. A., Grimshaw, S. B., Spencer, A., Glaser, S. J., Smith, L. J., and Dobson, C. M. (1997) *Biochemistry 36*, 8977−8991.
16. Morozova Roche, L. A., Arico Muendel, C. C., Haynie, D. T., Emelyanenko, V. I., VanDael, H., and Dobson, C. M. (1997) *J. Mol. Biol. 268*, 903−921.
17. Redfield, C., and Dobson, C. (1988) *Biochemistry 27*, 122−136.
18. Gutfreund, H. (1969) *Methods Enzymol. 16*, 229−249.
19. Royer, C. A., Mann, C. J., and Matthews, C. R. (1993) *Protein Sci. 2*, 1844−1852.
20. Baldwin, R. L. (1993) *Curr. Opin. Struct. Biol. 3*, 84−91.
21. Glasoe, P. F., and Long, F. A. (1960) *J. Phys. Chem. 64*, 188−193.
22. Vaney, M. C., Maignan, S., Rieskautt, M., and Ducruix, A. (1996) *Acta Crystallogr. Sect. D: Biol. Crystallogr. 52*, 505−517.
23. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) *J. Mol. Biol. 112*, 535−542.
24. McDonald, I. K., and Thornton, J. M. (1994) *J. Mol. Biol. 238*, 777−793.
25. Itzhaki, L. S., Evans, P. A., Dobson, C. M., and Radford, S. E. (1994) *Biochemistry 33*, 5212−5220.
26. Sosnick, T. R., Mayne, L., Hiller, R., and Englander, S. W. (1994) *Nat. Struct. Biol. 1*, 149−156.
27. Miranker, A., Radford, S., Karplus, M., and Dobson, C. (1991) *Nature 349*, 633−636.
28. Roder, H., Elove, G., and Englander, S. (1988) *Nature 335*, 700−704.
29. Udgaonkar, J., and Baldwin, R. (1988) *Nature 335*, 694−699.
30. Buck, M., Boyd, J., Redfield, C., Mackenzie, D. A., Jeenes, D. J., Archer, D. B., and Dobson, C. M. (1995) *Biochemistry 34*, 4041−4055.
31. Blake, C. C. F., Koenig, D. F., Mair, G. A., North, A. C. T., Phillips, D. C., and Sarma, V. R. (1965) *Nature 206*, 757.
32. Broadhurst, R. W., Dobson, C. M., Hore, P. J., Radford, S. E., and Rees, M. L. (1991) *Biochemistry 30*, 405−412.
33. Kiefhaber, T., Bachmann, A., Wildegger, G., and Wagner, C. (1997) *Biochemistry 36*, 5108−5112.
34. Waelder, S. F., and Redfield, A. G. (1977) *Biopolymers 16*, 623−629.
35. Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1993) *Proteins: Struct., Funct., Genet. 17*, 75−86.
36. Hore, P. J., Winder, S. L., Roberts, C. H., and Dobson, C. M. (1997) *J. Am. Chem. Soc. 119*, 5049−5050.
37. Chyan, C. L., Wormald, C., Dobson, C. M., Evans, P. A., and Baum, J. (1993) *Biochemistry 32*, 5681−5691.
38. Buck, M., Radford, S. E., and Dobson, C. M. (1994) *J. Mol. Biol. 237*, 247−254.
39. Gallagher, W., Tao, F., and Woodward, C. (1992) *Biochemistry 31*, 4673−4680.
40. Houry, W. A., Rothwarf, D. M., and Scheraga, H. A. (1995) *Nat. Struct. Biol. 2*, 495−503.
41. Otzen, D. E., Itzhaki, L. S., Elmasry, N. F., Jackson, S. E., and Fersht, A. R. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 10422−10425.
42. Wedin, R. E., Delepierre, M., Dobson, C. M., and Poulsen, F. M. (1982) *Biochemistry 21*, 1098−1103.
43. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.

BI9731504