Kinetics of Compaction during Lysozyme Refolding Studied by Continuous-Flow Quasielastic Light Scattering†

H.-P. Feng and J. Widom*

Department of Biochemistry, Molecular Biology, and Cell Biology, and Department of Chemistry, Northwestern University, Evanston, Illinois 60208-3500

Received June 7, 1994; Revised Manuscript Received August 23, 1994®

ABSTRACT: We recently developed an experiment, termed continuous-flow quasielastic light scattering (QLS), that is capable of monitoring the time evolution of the hydrodynamic diameter of macromolecules or macromolecular assemblies in solution. Here we report the use of this method to directly monitor the kinetics of compaction of the polypeptide chain of hen egg white lysozyme (HEWL) when protein refolding is initiated by 10-fold dilution from 5 M guanidine hydrochloride (GuHCl) at pH 1.5, 23 °C. Previously, such information could only be obtained indirectly, by analysis of the kinetics of binding and release of a fluorescent probe dye. Refolding was also monitored by UV difference absorption spectroscopy to characterize the time scale of the formation of the native environment around the aromatic side chains under the same conditions used in the continuous-flow QLS experiments. We find that HEWL becomes compact within 1 s after the initiation of refolding, the shortest time that is accessible with our first-generation instrument. This time scale is shorter than that for the recovery of the native absorbances in the aromatic region. These results provide direct evidence that the intermediate on the folding pathway of lysozyme is compact. The implications of these results for models of protein folding are discussed.

The amino acid sequence of a protein encodes its native structure and probably also determines the folding pathway [Anfinsen & Scheraga, 1975; for recent reviews, see Kuwajima (1989), Kim and Baldwin (1990), Baldwin (1991), Dill and Shortle (1991), and Matthews (1993)]. Kinetic intermediates exist on the folding pathway when folding *in vitro* is initiated by transferring chemically denatured proteins into native conditions (Kim & Baldwin, 1982). The structure of the kinetic intermediates provides important information for an understanding of the nature of the interactions that direct the folding of polypeptide chains to their final conformations and is therefore the focus of many studies. For some proteins, equilibrium intermediates also exist, when these proteins are exposed to mildly denaturing conditions. The molten globule state is one such intermediate that has been extensively studied and characterized [for reviews, see Ptitsyn (1987), Kuwajima (1989), Baldwin (1991), and Christensen and Pain (1991)]. The conformation of the molten globule state is considered to be intermediate between the conformations of the native and the denatured protein: near native secondary structure content with a rearranging environment around the aromatic residues. The size of the molten globule intermediate is measured to be slightly larger than the native protein. The hydrophobic surface in the molten globule is considered to be partially exposed and is thought to be responsible for the tendency of such equilibrium intermediates to aggregate (Kuwajima, 1989; Christensen & Pain, 1991). Since these structural features are consistent with a partially folded polypeptide chain, the molten globule state was proposed to be an intermediate on the protein folding pathway (Kuwajima et al., 1985; Ptitsyn, 1987; Ptitsyn et al., 1990). For the particular case of α-lactalbumin, the spectral properties of the kinetic folding intermediate were shown to be very similar to those of the equilibrium intermediate, and the structural interpretation of the kinetic intermediate was based on the characteristics of the equilibrium intermediate (Ikeguchi et al., 1986).

When, in the folding process, does a polypeptide chain become compact? This simple question is central to the protein folding problem, yet, surprisingly, at this time no experimental data exist that address it directly.

Ptitsyn and colleagues have developed a method for inferring the compactness of folding proteins indirectly, from measurement of the fluorescence of a hydrophobic probe dye, ANS[1] (Ptitsyn et al., 1990; Semisotnov et al., 1991). ANS is known to bind to hydrophobic surfaces with an accompanying fluorescence increase. Ptitsyn and colleagues find for several proteins investigated that protein refolding initiated in the presence of ANS is accompanied by an initial relatively fast increase in fluorescence intensity, followed by a slower loss of the excess fluorescence intensity (Ptitsyn et al., 1990). The initial increase in ANS fluorescence is attributed to formation of a molten globule state of the protein, which, by virtue of being slightly less compact than the native state, may have accessible hydrophobic binding sites for ANS. The subsequent loss of ANS fluorescence is attributed to the further folding of the molten globule intermediate to the native state, with concomitant squeezing out of the bound ANS (Semisotnov et al., 1991). Thus ANS fluorescence is thought to allow monitoring of both the initial

† This project was supported by a grant to J.W. from the NSF and by the generous donation of equipment by Beckman Instruments, Brookhaven Instruments Corp., Hewlett-Packard, and Newport Corp.

* To whom correspondence should be addressed.

® Abstract published in *Advance ACS Abstracts,* October 15, 1994.

[1] Abbreviations: ANS, 8-anilino-1-naphthalenesulfonate; CD, circular dichroism; D_t, translational diffusion coefficient; GuHCl, guanidine hydrochloride; HEWL, hen egg white lysozyme; QLS, quasielastic light scattering.

0006-2960/94/0433-13382$04.50/0

compaction of an extended chain into a molten globule state as well as of the subsequent further packing of the molten globule into the native state.

These are important experiments because, at this time, they provide the only available information about the kinetics of overall compactness during protein folding. However, such experiments also have some disadvantages. They are indirect; it remains possible that some alternative explanation may obtain. One particular potential problem is that ANS fluorescence might not discriminate between intramolecular compaction and aggregation, which may be a side reaction during folding *in vitro*.

In the course of our work, we have developed a different experiment that is capable of monitoring the time evolution of the hydrodynamic diameter of macromolecules or macromolecular assemblies in solution (Feng et al., 1993). This experiment uses quasielastic light scattering (Chu, 1974; Berne & Pecora, 1976; Bloomfield & Lim, 1978) to measure the translational diffusion coefficient (D_t) of macromolecules in solution, a quantitative measure of compactness. Kinetic information is obtained by coupling the apparatus to a continuous-flow mixing system, allowing one to signal average for as long as necessary, while measuring the D_t corresponding to a particular time in the kinetic history of the reaction or process under investigation. The first-generation instrument that we have constructed allows us to measure D_t in solutions having "ages" (post-mixing) of 1 s or longer. In an earlier study we used this instrument to determine the lifetime of the histone octamer when exposed to physiological ionic conditions, which allowed a test of one model of nucleosome transcription (Feng et al., 1993).

Here we report the application of this instrument to the study of the kinetics of compaction during the refolding of the protein lysozyme. HEWL is chosen as a model system for our study because of the likely and presumed similarity between the kinetic folding of this protein and the stepwise progressive refolding through a molten globule equilibrium intermediate state of the related protein α-lactalbumin (Kuwajima et al., 1985; Ikeguchi et al., 1986). The kinetic intermediate of HEWL was characterized as having spectroscopic properties that are very similar to those of the molten globule intermediate of α-lactalbumin (Kuwajima et al., 1985; Ikeguchi et al., 1986). In other studies, amide proton exchange ^{1}H NMR experiments showed that the secondary structure in the HEWL intermediate was in both structural lobes that characterize the native protein (Miranker et al., 1991; Radford et al., 1992).

We used the continuous-flow QLS method to ask whether the kinetic folding intermediate of HEWL is indeed compact and comparable to the molten globule state of α-lactalbumin. The advantages and the limitations of this methodology, and the results and their implications for models of protein folding, are discussed.

MATERIALS AND METHODS

Protein Samples and Substrate. Three times crystallized hen egg white lysozyme (grade I; lot no. 111H7010) and dry *Micrococcus lysodeikticus* were purchased from Sigma. Ultra pure GuHCl was purchased from J. T. Baker.

All protein solutions (with or without GuHCl) were centrifuged at 41 000 rpm in a Beckman SW-41 rotor at 4 °C for 12−14 h to remove dust particles, aggregates, and any insoluble material. In control studies, the commercial HEWL preparations were further purified as described (Sophianopoulos et al., 1962). The resulting material did not differ detectably from the centrifuged commercial material as judged by Coomassie-stained SDS−PAGE, nor did it behave differently in the experiments described below (data not shown). For these reasons, the remainder of our studies were carried out using the commercial preparations further purified only by the centrifugation step.

GuHCl solutions (5 M) were prepared by measuring appropriate weights of ultra pure GuHCl and dissolving it in a buffer containing 50 mM glycine at pH 1.5 to the desired final volume, followed by filtration through 0.45- or 0.2-μm filters and degassing under vacuum. Solutions of denatured HEWL were prepared by dissolving measured amounts of protein in the filtered, degassed 5 M GuHCl solutions, followed by ultracentrifugation as described above. The protein concentrations were approximately 60, 50, 40, 30, and 20 mg/mL, respectively. The prepared protein solutions were loaded into the continuous flow setup and used immediately. A small fraction of the ~60 mg/mL protein solution was diluted with 5 M GuHCl solution to a final protein concentration of ~3 mg/mL. This solution was used in the difference absorption spectroscopy experiment. The substrate for HEWL, *M. lysodeikticus,* was dissolved in a buffer containing 200 mM sodium acetate at pH 5.2. The concentration was ~0.1 mg/mL before mixing with the lysozyme solution.

Kinetic Difference Absorption Spectroscopy. The kinetics of folding was monitored at three different wavelengths, 250, 292, and 300 nm (Kato et al., 1982), using a Hewlett-Packard diode array spectrometer (Model 8452A). The cuvette and the holder were controlled at 23 ± 0.1 °C. Folding was initiated by manually mixing the denatured HEWL solution with the renaturing buffer using a commercial cuvette-mixing device. For some studies, to reduce the inhomogeneity during manual mixing, we chose a dilution ratio of 15-fold so that a smaller volume of the protein sample could be used. In this case, the renaturing buffer contained an appropriate [GuHCl] to compensate for the higher dilution ratio, so that the final [GuHCl] was 0.5 M, the same condition used in the continuous flow experiments. The final protein concentration was ~0.2 mg/mL. The dead time for manual mixing is ~1 s. The kinetic traces of folding were fitted to a double-exponential function with the program IGOR by WaveMetrics using a nonlinear least-squares fitting algorithm.

QLS Studies of the Equilibrium Folding and Unfolding Transitions. The hydrodynamic diameters of HEWL in the GuHCl-induced equilibrium folding and unfolding transitions at 23 °C were measured by QLS in static (i.e., ordinary, nonflowing) conditions. The protein concentration in the samples for all the measurements was 5 mg/mL. The measurements were made at GuHCl concentrations from 6 to 1 M for the folding transition and 0 to 5 M for the unfolding transition. The samples were prepared from concentrated protein solutions in 6 M GuHCl or in native buffer, for the folding and unfolding transitions, respectively. The native buffer contained 50 mM glycine at pH 1.5. The concentrated protein stock solutions were mixed with appropriate buffers to reach the desired GuHCl concentrations. The protein solutions were then further diluted with buffers containing the final GuHCl concentrations to reach the

protein concentrations of 5 mg/mL. The samples were allowed to equilibrate for 1−2 h and were spun in a Beckman airfuge for 10 min to remove dust particles and aggregates, prior to the QLS measurements.

The instrument used in our QLS studies includes a Coherent Innova-90 argon laser operating at 488 nm; autocorrelation functions are obtained with a Brookhaven Instrument model 2030 128-channel autocorrelator. The channel width was 0.5 μs for samples containing [GuHCl] < 3 M and was 0.7 μs for all the other samples. The laser power was 200 mW, and the accumulation time was typically 100 s. The autocorrelation functions were analyzed using the computer program NNLS [Non-Negatively-constrained Least Squares (Morrison et al., 1985), provided by Brookhaven Instruments] to remove contributions to the autocorrelation functions from the rapidly diffusing GuHCl (Dubin et al., 1973).

Continuous-Flow QLS. Our adaptation of the QLS instrument for continuous-flow studies is described elsewhere (Feng et al., 1993). Refolding was initiated by mixing the denatured protein solutions with the renaturing buffer (containing 0 M GuHCl) in the mixer. The dilution ratio was 10-fold and the final GuHCl concentration was 0.5 M. Delay times obtained in the refolding reactions were estimated using the combined output flow rate of the syringe pumps together with the dead volume in the apparatus (Feng et al., 1993). Data at different delay times were obtained by changing the combined flow rate; the flow rate of each individual pump was varied so that the dilution ratio remained at 10-fold at all different time points. After each experiment, the solution at the output of the mixer and flow cell assembly was collected and analyzed to verify the dilution ratio (see below). The renaturing buffer was degassed under vacuum for 1 h before loading into the reservoir of the continuous flow setup. The protein solutions were prepared using filtered, degassed GuHCl and were centrifuged as described above.

The hydrodynamic diameters of the starting (denatured) protein, and of the native protein, were determined in static (nonflowing) conditions, as described above for the equilibrium folding and unfolding transitions. The concentration of lysozyme was 6 mg/mL for the denatured protein (in 5 M GuHCl), and 2−8 mg/mL for the native.

In our current design, only the flow cell assembly is temperature regulated. Therefore we carried out the experiments with the room temperature controlled at 23 ± 0.25 °C and the flow cell assembly controlled at 23 ± 0.1 °C. The laser power used was typically 500 mW. For each run in flow experiments, the autocorrelation function was recorded after a certain wait period during which the pumps were allowed to stabilize (Feng et al., 1993). The accumulation time was 30 s for each autocorrelation function in flow experiments and was 60 s in static experiments. The typical number of counts per channel for the baseline in an autocorrelation function thus recorded was on the order of 10^5−10^6. The autocorrelation functions were analyzed using either NNLS or the method of cumulants (Koppel, 1972), with software provided by Brookhaven Instruments. When the cumulant fit method was used to obtain the average diameter of HEWL in 0.5 M GuHCl, the first eight channels in the autocorrelation function were excluded in the fitting to reduce the contribution from GuHCl (Dubin et al., 1973).

Physical Data. The viscosity of the protein solution was measured with a homemade Oswald-type viscometer thermostated at 23 ± 0.1 °C. The protein concentration of the solution used in these measurements was 2 mg/mL. The GuHCl concentration was 0.45 M. The same viscosity value was used without correction for other protein concentrations in solutions containing 0.5 M GuHCl. The error in viscosity introduced by the difference in GuHCl concentrations was estimated to be less than 0.1%. The error in viscosity introduced by the difference in protein concentrations can be estimated by using the typical intrinsic viscosity value of 3.5 cm^{-3}/gm (Cantor & Schimmel, 1980) for small globular proteins and was less than 1%. Both of these error estimates are small compared to the typical experimental error of our QLS measurements, ~3−5% in D_t. The index of refraction of the final solution was measured at room temperature using a Abbe-3L refractometer (Milton Roy Co.).

Dilution Ratio. The dilution ratio of the continuous flow experiments was examined by one or both of the following methods: (1) Indices of refraction of several standard solutions made with a known dilution ratio from the concentrated GuHCl solution were measured. The refractive index of the output mixture from the experiments was then measured and compared with the standards. (2) The actual flow rate in an experiment was calculated by the weight of the output solution collected during a known time period and was compared with the sum of the expected output rate for each pump in the experimental setting. These quantities agree with each other, so the dilution ratio can be obtained from the ratio of the output rates of the pumps. These two different methods were found to give comparable results for the dilution ratio.

Enzyme Activity. The enzymatic activity of the lysozyme was measured as described (Wetlaufer et al., 1974). The activity was defined as the decrease in OD of the substrate solution at 600 nm. The activity of native HEWL was measured using protein samples directly dissolved in acetate buffer at pH 5.2. One unit of activity for the native enzyme thus determined corresponded to −0.019 OD_{600}/min. For measurement of the activity of refolded HEWL, samples were prepared as follows: a small portion of the unfolded HEWL prepared for the continuous-flow experiment was manually mixed with glycine buffer at pH 1.5, with a dilution ratio of 13.3. The protein concentration after mixing was typically ~2 mg/mL. The samples were allowed to refold and fully equilibrate for several hours before they were further diluted into acetate buffer at pH 5.2, the optimum pH for lysozyme activity. The protein concentration was ~0.001 mg/mL before mixing with the substrate.

RESULTS

Equilibrium Folding and Unfolding Transitions. The Stokes diameters of HEWL in the folding and unfolding transitions are plotted as a function of [GuHCl] in Figure 1. The transition evidently is reversible. Within experimental error, the Stokes diameters of HEWL in the high and low [GuHCl] regions do not depend on [GuHCl]. The measured Stokes diameters from samples in which the protein molecules were directly dissolved in buffers containing 5 or 0.5 M GuHCl were therefore used as the zero-time and the final,

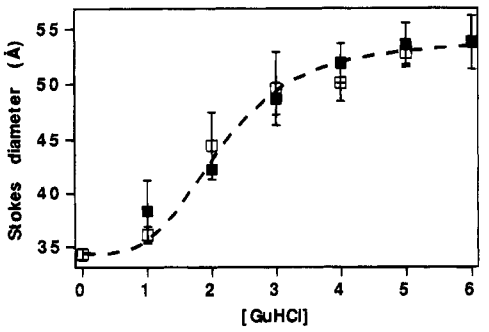

FIGURE 1: Stokes diameter of HEWL measured in equilibrium folding (■) and unfolding (□) transitions as a function of [GuHCl]. The Stokes diameters were obtained from autocorrelation functions recorded in static (i.e., nonflowing) conditions. NNLS was used for fitting the autocorrelation functions (see Discussion). The error bars represent the standard deviations for repeated measurements at each GuHCl concentration. The dashed line (- - -) represents the least squares fit to both transitions. The protein concentration in the samples for all the measurements was 5 mg/mL. The measurements were made in 50 mM glycine, pH 1.5, at GuHCl concentrations from 6 to 1 M for the folding transition and 0 to 5 M for the unfolding transition. The temperature was 23 °C.

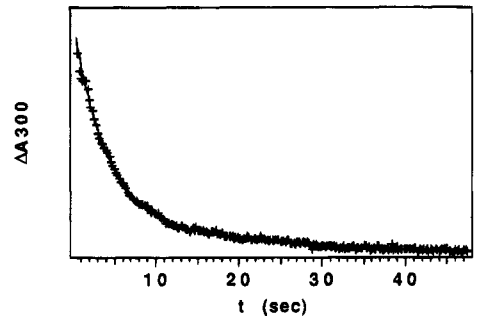

FIGURE 2: Kinetic trace (+) of a difference absorption experiment for the lysozyme refolding reaction monitored at 300 nm. Refolding was initiated by manually diluting the protein denatured in 5 M GuHCl to a final [GuHCl] of 0.50 M and 50 mM glycine, pH 1.5. The solid line (—) is the result of a nonlinear least-squares fit of the trace to a double-exponential model.

equilibrium values, respectively, for the kinetic QLS experiments.

Enzyme Activity. The activity of the refolded lysozyme was measured as described (data not shown) and was found to be 90% that of the native enzyme.

Refolding Kinetics Measured by Difference Absorption Spectroscopy. A typical kinetic trace of an HEWL refolding experiment monitored by difference absorption spectroscopy at 300 nm (Kato et al., 1982) is shown in Figure 2. The results of fitting the kinetic traces are summarized in Table 1. Also listed in Table 1 are the results from experiments performed at 13.3-fold dilution, jumping from 6 to 0.45 M GuHCl. The refolding kinetics of HEWL consists of two phases, the fast phase being the major phase, with a time constant of ~3.4 s in 0.5 M GuHCl. The time constant of the slow phase was determined with a larger error margin due to the small amplitude of this phase. The time constants and the relative amplitudes varied slightly depending on the final conditions. Similar results were obtained at the other wavelengths investigated, but with worse signal to noise ratios (data not shown). Our results are consistent with earlier observations that the refolding of HEWL is biphasic (Kato et al., 1982; Kuwajima et al., 1985; Radford et al., 1992); note that the conditions used in those studies were different from the ones used in our experiments; therefore, the detailed results are somewhat different.

Table 1: Kinetic Parameters for the Refolding of HEWL at 23 °C Measured at 300 nm[a]

final GuHCl concn (M)	τ_F (s)	τ_S (s)	$\frac{A_F}{A_F + A_s}$ (%)
0.45[b]	2.6 ± 0.3	12 ± 5	87 ± 1
0.50[c]	3.4 ± 0.6	50 ± 20	78 ± 6

[a] Error margins represent the standard deviations of repeated measurements. [b] The initial [GuHCl] was 6.0 M; the denatured protein was diluted 13.3-fold into 50 mM glycine, pH 1.5. [c] The initial [GuHCl] was 5.0 M. The denatured protein solution was diluted 15-fold into a solution containing appropriate concentrations of GuHCl and glycine, such that the final conditions were 0.50 M GuHCl and 50 mM glycine, pH 1.5. This recreates the conditions used in the continuous-flow QLS studies.

Continuous-Flow QLS. The shortest delay time achievable in our current first-generation continuous-flow instrument is approximately 1 s. Using the refolding kinetic data obtained from the difference absorption experiments for jumps to 0.5 M GuHCl, we calculate that by 0.9 s (the shortest delay time used for these studies) only 0.185 (18.5%) of the protein has reached the native state. Therefore, data obtained at 0.9-s delay time will be dominated by the nonnative conformation by a factor of 4.4:1, and so we can in principle follow the kinetics of compactness as the majority of the protein undergoes the transition to the native state.

Typical autocorrelation functions and residuals after fitting from the static and the continuous-flow QLS experiments are shown in Figure 3. The quality of the autocorrelation functions is similar in all three cases. The noise level in the autocorrelation functions from the flow experiments was somewhat higher due to the short accumulation time used in order to conserve materials. This higher noise level precluded our use of fitting routines such as NNLS and multiple-exponential fits that are sensitive to noise. However, the residuals from cumulant fits from the flow experiments were similar to those from static experiments in native conditions, indicating that the Stokes diameters obtained from the flow experiments in this way are reliable [see also Feng et al. (1993)].

The dilution ratio for each continuous flow experiment was determined as described (see Materials and Methods) and was set to 10-fold. We carried out the continuous-flow QLS refolding experiments at five different final protein concentrations (5.9, 5.0, 4.2, 2.9, and 2.1 mg/mL) at six delay time points (0.9, 1.8, 2.6, 3.8, 6.4, and 9.0 s). In Figure 4 the measured Stokes diameter of refolding HEWL is plotted versus the final protein concentration at each delay time.

The results of Figure 4 are summarized in Figure 5, in which the averages of Stokes diameters obtained from experiments at the two highest final protein concentrations are plotted as a function of delay time. We chose these averages to represent the size of the intermediates for the following reasons: (1) The Stokes diameters showed no clear dependence on the protein concentration in our experiments when the error margins in the experiments are taken into consideration. In a separate set of static (equilibrium) experiments (not shown), where we could obtain data of higher quality, we determined that there is a slight dependence of the apparent size on the protein concentration, but

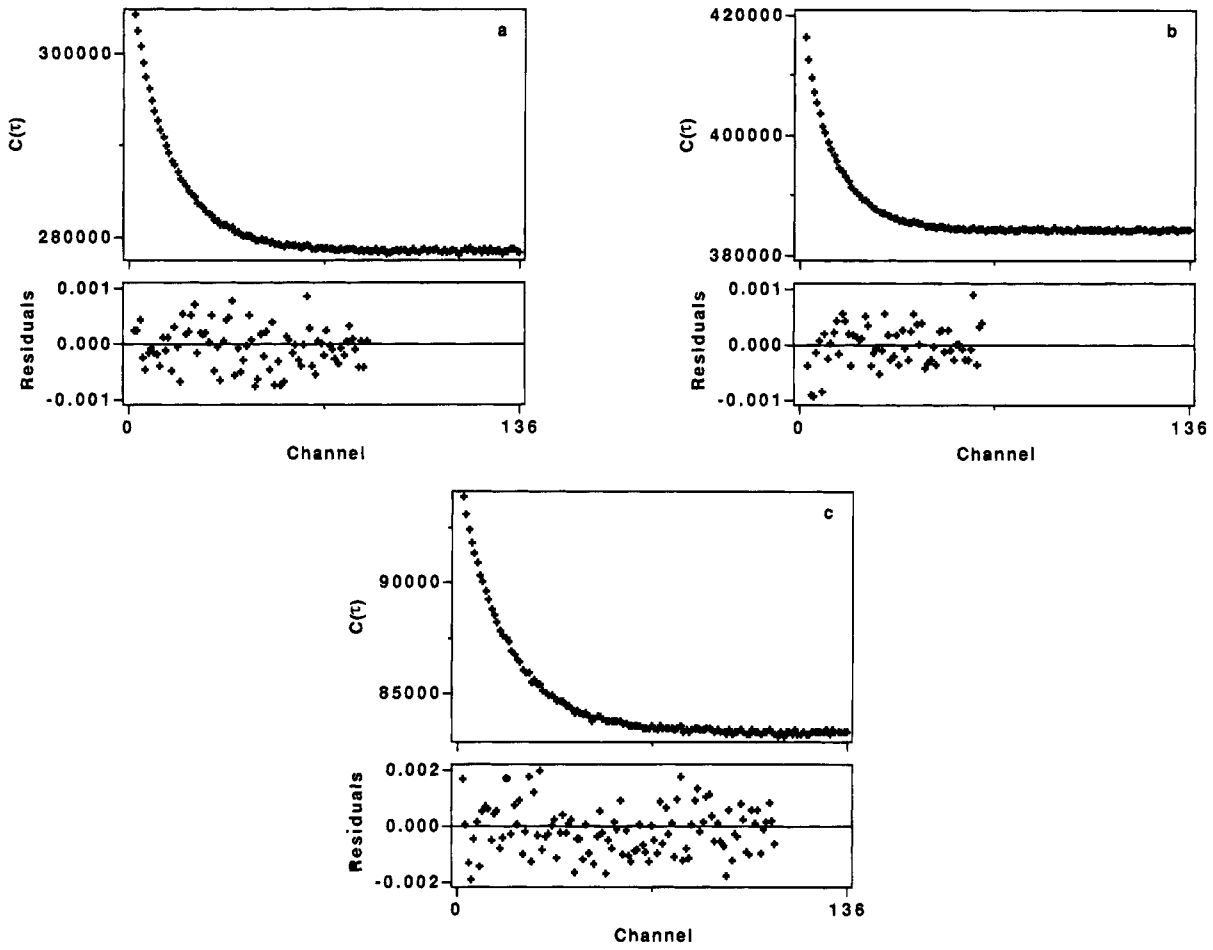

FIGURE 3: Plots of experimental autocorrelation functions [$C(\tau)$] and residuals after fitting versus correlator channel number for (a) native, (b) denatured lysozyme, and (c) the folding intermediate 0.9 s after mixing. Correlation time (τ) is given by the channel number times the channel width ($\Delta\tau$). The $\Delta\tau$'s used were 0.5 μs for a and c and 1 μs for b. The autocorrelation functions were collected under static conditions (not flowing) for a and b and under flow conditions for c. Residuals are defined as the difference between the experimentally determined, normalized, and baseline-subtracted autocorrelation functions [$C^*(\tau)$] and that calculated using the parameters obtained from fitting. Cumulant fits were used for a and c; NNLS was used for fitting b to remove the contribution to the autocorrelation functions from GuHCl (see Discussion). Only the nonnegative points can be used for fitting $\log[C^*(\tau)]$. To avoid systematic bias, only points prior to the channel containing the first negative value in $C^*(\tau)$ were included in the fitting, and therefore the residual plots do not extend over all the correlator channels. The temperature of the experiments was controlled at 23 °C; the incident laser light intensity was between 0.3−0.7 W at 488.0 nm. The buffer used in the experiments was 50 mM glycine at pH 1.5 with 0.5 M GuHCl for a and c, and with 5 M GuHCl for b. The samples for a and b were spun in a Beckman airfuge for 3−10 min before collecting the autocorrelation functions. The sample in c was prepared as described under Materials and Methods. The HEWL concentration was roughly 6 mg/mL for a and b. The accumulation time was 60 s for a and b and was 30 s for c. The differences in counts in $C(\tau)$'s are due to the use of different accumulation times and different incident laser intensities.

that this slight dependence is actually due to the increasing contribution to the autocorrelation functions from the fast-moving GuHCl molecules. (2) The autocorrelation functions from the more concentrated samples have a better signal-to-noise ratio since the signal is proportional to the square of the protein concentration; moreover, these minimize the relative signal from the GuHCl. (3) All of the protein concentrations investigated are <1%, and so all are in the realm typically considered to be dilute. For these dilute samples, signals from aggregates do not dominate the experimental autocorrelation functions.

The results of our experiments show that HEWL acquires a compact conformation within 0.9 s after the initiation of refolding. The average Stokes diameter at 0.9 s appears to be 25% larger than for the native protein measured in static conditions; this can be compared to Stokes diameter of the starting denatured HEWL which is 50% larger than native. The average Stokes diameters of HEWL at the other five longer delay time points appear to be 11% larger than those of the native conformation, considering the experimental error. Possible explanations for the apparent difference in size between the various folding time points and the final equilibrium value are considered below.

DISCUSSION

Data Analysis. For the denatured HEWL samples, the light scattered from the 5 M GuHCl constitutes ~35% of the integrated signal at a protein concentration of 5 mg/mL (Dubin et al., 1973). A cumulant fit is plainly not appropriate in this case. Instead, the Stokes diameter of the denatured protein at equilibrium was obtained by fitting the experimental autocorrelation functions using the computer program NNLS (Morrison et al., 1985), which resolves the experimentally obtained autocorrelation function into contributions from scatterers of varying size. In practice, we obtained two well-resolved peaks, one at a small size, attributable to the GuHCl, and one at a larger size, attributable to the protein. These results are analogous to those obtained earlier (Dubin

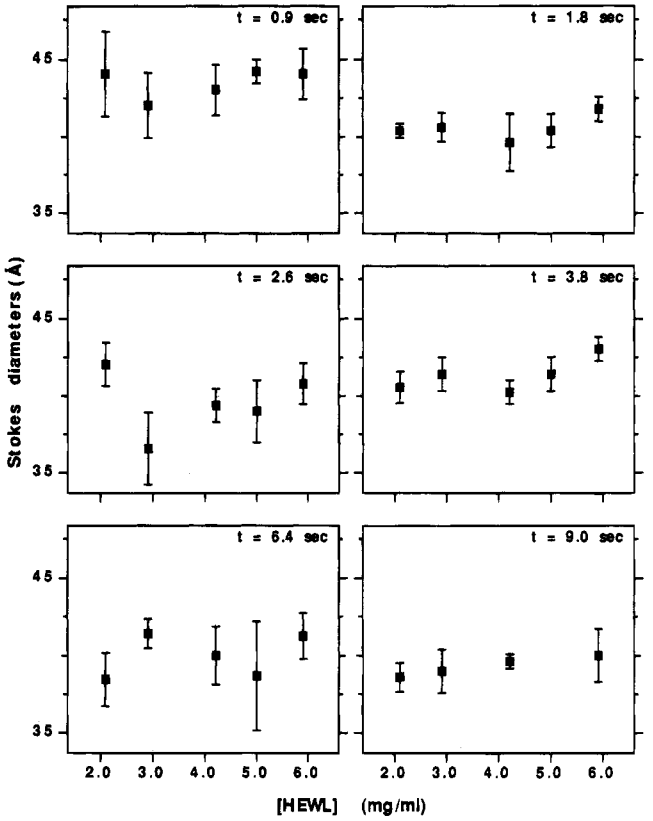

FIGURE 4: Dependence of Stokes diameters (■) on final HEWL concentrations at different delay times after mixing. The Stokes diameters were obtained from cumulant fits (see Discussion) of autocorrelation functions recorded in continuous-flow experiments. The error bars represent the standard deviations for repeated measurements. The experiments were performed at 23 °C, and the final folding conditions were 0.5 M GuHCl and 50 mM glycine, pH 1.5.

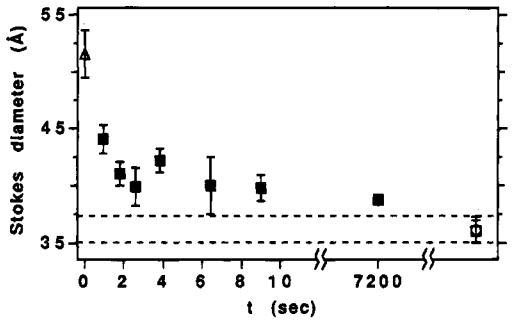

FIGURE 5: Stokes diameters of HEWL during refolding (■) as a function of delay time after mixing. The size of the protein was obtained by averaging the results from experiments in which the final protein concentrations were 5 or 6 mg/mL (Figure 4). The Stokes diameter of HEWL at delay time = 7200 s was obtained in static experiments in which the folding was initiated by manual mixing. For other time points, the autocorrelation functions were recorded in continuous-flow experiments. The open symbols represent equilibrium measurements for the protein in denaturing (△) and the native conditions (□, ○). Cumulant fits were used for (■) and (□); NNLS was used to obtain the results for (△) and (○) (see Discussion). Error bars represent the standard deviations of repeated measurements. The error bars for the native HEWL are extended across the plot (- - -) for ease of comparison.

et al., 1973), although the primary data in that study came from measurement of the photocurrent spectrum rather than its autocorrelation function. In our study, the autocorrelation functions were recorded in static conditions with a long period of integration time to increase the signal-to-noise ratio, and the results from many measurements were averaged. The Stokes diameter that we obtain for HEWL in 5 M GuHCl is comparable to but slightly *smaller* than that reported earlier (Dubin et al., 1973). Importantly, we conclude that our procedure does not *over*estimate the Stokes diameter of unfolded HEWL.

For consistency, NNLS was used for analysis of all of the equilibrium folding and unfolding data obtained for Figure 1. As a further check, the lowest [GuHCl] data (for which the contribution from GuHCl is negligible—see below) were also analyzed by the method of cumulants; the Stokes diameters obtained agreed with those from NNLS to better than the standard deviation between repeated measurements.

The continuous-flow data are noisier than the static (nonflowing) data. This is due at least in part to the shorter signal-averaging times used with the flowing solutions. We found that these experimental autocorrelation functions were not stably fit by NNLS. This is in contrast to our earlier work which involved continuous-flow QLS studies of the histone octamer (Feng et al., 1993). Presumably the much greater molecular weight of the histone octamer compared to HEWL allowed higher quality autocorrelation functions to be obtained with similar data acquisition times.

Since NNLS could not be used to fit the present continuous-flow data, we used the method of cumulants instead (Koppel, 1972), which is much more robust because it does not attempt to analyze for a distribution. In contrast to NNLS, the cumulant fit cannot remove the contributions to the scattering signal from GuHCl, and so, in the present study, it can only be used for the native conditions, ≤0.5M GuHCl. In such conditions, scattering from GuHCl does not contribute significantly to the signal: to reiterate, both NNLS and the cumulant method were used in such conditions, and, with the high quality data obtained from static (i.e., nonflowing) runs, the two methods gave results that are identical within experimental error (see Results). Scattering from the 10-fold lower [GuHCl] is negligible in part because the 10-fold decrease in concentration decreases the corresponding contribution to the autocorrelation function by 10^2, i.e., 100-fold, to $\ll 1\%$ of the protein's contribution; moreover, the refractive increment of the protein is substantially increased compared to its value in 5 M GuHCl, further diminishing the relative contributions from the GuHCl. And, to further suppress possible contributions from the GuHCl, the first several correlator channels were omitted from the cumulant fits (although, in practice, this made little difference).

The use of the cumulant fitting method for analysis of the continuous-flow data therefore restricts application of the continuous-flow method in the present study to native conditions (≤0.5 M GuHCl). Thus we can obtain continuous-flow data for native (never denatured) protein and for *refolding* or *refolded* protein; data for *unfolded* protein are obtained only from static experiments. But, to reiterate, our results for unfolded protein are in good agreement with previously published data; and our results for native protein are in good agreement with previously published data, whether obtained from static *or* flowing experiments [see above and Feng et al. (1993)].

Since we found no dependence of the hydrodynamic diameter of the protein on [GuHCl] in the denaturing region of the folding/unfolding transition, we used the diameter of

HEWL measured in 5 M GuHCl as the starting size for refolding, corresponding to zero time in native conditions.

Aggregation as a Side Reaction during Protein Folding. The Stokes diameters of HEWL from continuous-flow QLS experiments at delay times of 1.8, 2.6, 3.8, 6.4, and 9.0 s after initiation of refolding are virtually indistinguishable from one another to within our experimental error, while all are ~11% larger than the diameter of the native protein molecules obtained in static conditions. It was possible that this difference in size could be due to the molecules undergoing the process characterized by the slow phase (time constant $\tau_s \approx 50$ s). To test this possibility, we carried out static QLS experiments in which refolding was initiated by manual mixing and the samples were allowed to refold to completion ($t \gg \tau_s$) before data acquisition. The Stokes diameter thus obtained is still ~10% larger than that of the native molecules (Figure 5). These observations suggest that HEWL has reached an apparent equilibrium size by 1.8 s after the initiation of refolding and that this *apparent* equilibrium size is slightly larger than that of the native molecules. Since QLS is very sensitive to small amounts of aggregates in the sample, we wished to test the possibility that aggregation of some of the HEWL was responsible for this discrepancy in Stokes diameters. We spun the manually refolded samples in an airfuge ultracentrifuge for 10 min to remove aggregates and then measured the Stokes diameter in static QLS experiments. The size obtained from these samples was restored to that of the native molecules (data not shown). This result supports the hypothesis that aggregation of some of the HEWL is responsible for the increased Stokes diameters apparent in the continuous-flow experiments. We attempted to remove the contribution from aggregates by including an additional exponential term in the fitting of the autocorrelation functions, but we were unable to achieve reliable fits. Moreover, the aggregates may exist as a distribution of particles of different sizes, so that fitting the autocorrelation functions to two exponentials is not appropriate. The noise level in these continuous-flow data precludes the use of NNLS to obtain a size distribution. Consistent with the possibility of a small amount of aggregation accompanying refolding, we observed that refolding did not restore 100% of the expected enzymatic activity.

Therefore one important conclusion of this study is that refolding of HEWL from concentrated GuHCl solutions may be accompanied by a small amount of aggregation. We anticipate that this might be generally true for most proteins. The HEWL concentrations used in this study were in the range of several mg/mL. Presumably, reducing the protein concentration would reduce but not eliminate accompanying aggregation. We suggest that protein refolding experiments that are potentially sensitive to the presence of aggregates should be interpreted cautiously.

Kinetics of Compaction during HEWL Refolding. The most important conclusion of our continuous-flow QLS experiments is that the Stokes diameter of the folding intermediate of HEWL is considerably smaller than that of the denatured protein even at 0.9 s in the refolding reaction. Two undesirable complications must be recognized. First, if small amounts of aggregates are present, these would artifactually *increase* the apparent Stokes diameter; that is, the true Stokes diameter at $t = 0.9$ s would be even smaller than our measured value. Second, and conversely, at $t =$ 0.9 s, ~18% of the protein has already undergone the process characterized by τ_f. If such molecules were more compact than those which have not yet undergone that process, this would misleadingly *decrease* the apparent Stokes diameter. However, such molecules would contribute to the measured average Stokes diameter in proportion to their relative concentration, so the measured size at this time would be misleadingly low by a small amount at most: roughly one-fifth of the already small difference between the measured Stokes diameter and that of the native protein might be attributable to this effect. We conclude that neither of these undesirable complications accounts for our chief observation that HEWL is already substantially compact by 0.9 s; indeed, these problems may tend to cancel each other out.

The Stokes diameter of HEWL at 0.9 s in the refolding reaction appears to be slightly larger than the measured sizes at all the other time points. At this delay time the molecules in the samples exist as a distribution of conformations; most of the molecules at this time have not undergone the process characterized by the fast phase time constant $\tau_f = 3.4$ s. These conformations contain high secondary structure content (Kuwajima et al., 1985) and the secondary structure is present in both domains that are characteristic of the native protein (Miranker et al., 1991; Radford et al., 1992). It is possible that this population of partly folded states may be slightly less compact than the native state. This explanation is consistent with the observation that an equilibrium molten globule intermediate is slightly larger than the native state (Dolgikh et al., 1985; Gast et al., 1986; Ptitsyn et al., 1990). However, we cannot rule out the alternative possibility that there may be more aggregates early in the folding reaction and that these aggregates cause the apparent increase in the average Stokes diameter.

Relation to Other Studies and Implications for the Generality of Molten Globule States. The folding pathways for HEWL and α-lactalbumin were proposed to be similar on the basis of the structural homology between the two proteins (Acharya et al., 1989). In the case of α-lactalbumin, a partially folded intermediate, the molten globule state, exists in equilibrium with the native and the unfolded conformations under mildly denaturing conditions (Dolgikh et al., 1981; Kuwajima et al., 1976). The molten globule was established to be on the folding pathway of α-lactalbumin by spectroscopic methods (Ikeguchi et al., 1986). HEWL does not exist as such an equilibrium intermediate state under mildly denaturing conditions; however, the kinetic intermediate on the folding pathway was shown to have spectral properties similar to those of a molten globule (Ikeguchi et al., 1986; Kuwajima et al., 1985). The results of the present study provide a direct measurement of the size of the kinetic folding intermediates of HEWL and demonstrate that this intermediate is indeed compact.

Recently several proteins were reported to exist as a molten globule either in equilibrium or nonequilibrium folding conditions (Christensen & Pain, 1991; Flanagan et al., 1992; Mayo et al., 1992; Morjana et al., 1993; Peng et al., 1993; Ptitsyn et al., 1990). In some cases, ANS binding experiments and fluorescence energy transfer experiments were used to show that the kinetic intermediates were molten globules (Ptitsyn et al., 1990; Semisotnov et al., 1987; Semisotnov et al., 1991). These experiments are not able to distinguish between aggregates that may exist in the folding reactions and the true folding intermediates. Also, such experiments are subject to concerns that the probes may

disturb the folding pathways. We believe that such studies may be complemented by continuous-flow QLS studies to directly test the hypothesis that the general folding intermediate is compact and globular. One would like to measure the Stokes diameters of the folding intermediates of proteins consisting of different structural motifs, proteins with different functions, and proteins with different folding kinetics.

The detailed structure of the HEWL folding intermediate has been studied by several other techniques (Ikeguchi et al., 1986; Kuwajima et al., 1985; Miranker et al., 1991, 1993; Radford et al., 1992). The results from these studies establish an order of events for the folding of HEWL: the protein acquires a high secondary structure content within tens of milliseconds in the folding reaction (Kuwajima et al., 1985; Miranker et al., 1991). The secondary structure formed is in both structural lobes that are characteristic of the native protein (Miranker et al., 1991; Radford et al., 1992). The protein follows parallel pathways to form the early intermediates (Miranker et al., 1993; Radford et al., 1992): a fraction of the molecules fold both structural domains cooperatively while the rest of the molecules follow a sequential pathway, the folding of α-domain preceding the folding of β-domain. Our new measurements show that the protein becomes compact before the formation of specific tertiary interactions.

The compaction of the polypeptide chains may occur prior to or simultaneously with the formation of secondary structure due to hydrophobic collapse (Allegra et al., 1990; Bryngelson & Wolynes, 1990; Chan & Dill, 1988); the chain can then rearrange within the compact intermediate to the native conformation. Alternatively, the compaction of the polypeptide chain may also occur after the formation of a native-like α-domain, the main driving force being the local interactions. The compact intermediate may serve to stabilize the short stretches of newly formed secondary structure and serve as a core for other parts of the protein to dock onto (Baldwin, 1990). These two folding mechanisms describe very different kinetics for the compaction of the polypeptide chains; they could be distinguished by measuring the Stokes diameters of the early folding intermediates.

Limitations of the Continuous-Flow QLS Experiment. The continuous-flow QLS experiment described here and in our earlier work (Feng et al., 1993) allows direct measurements of hydrodynamic size as a function of time on the seconds time scale, and in this way it represents an advance over what was previously possible. However, we note that this experiment also has limitations that must be recognized.

One significant limitation, discussed above, concerns the effects of accompanying aggregation. QLS studies are particularly sensitive to the presence of aggregates, because of their larger molecular weight. When the aggregates are much larger than the monomeric molecules, they diffuse on a different time scale, and their contributions are in principle easily subtracted from the correlation functions. This will often not be the case, as seen in the present study. We note, however, that this sensitivity to aggregation is not entirely undesirable: it can be important to know if aggregation is occurring.

Another significant limitation of our present first-generation instrument concerns the shortest delay times that are experimentally accessible. The shortest delay time in our experiments is limited by how fast we allow the mixed solution to flow through the flow cell. The flow introduced in our QLS experiments provides a mechanism for the decay of the autocorrelation function, not dependent on diffusion, owing to replacement of the molecules in the scattering volume with new molecules at uncorrelated positions (Feng et al., 1993). If the flow rate is sufficiently high, the decay of the autocorrelation function will characterize the time scale of the replacement of molecules in the scattering volume rather than the diffusion of the molecules of interest. Our experiments at 0.9 s approach the limiting flow rate at which the distortion of the autocorrelation functions is still negligible (Feng et al., 1993). To further reduce the dead time, a new mixer and flow cell assembly having a smaller dead volume is necessary. A secondary benefit of such a smaller device would be the use of proportionally smaller quantities of valuable reagents.

ACKNOWLEDGMENTS

We thank Anna Camarena for her assistance in the early stages of this work and Dr. C. Robert Matthews for substantial comments on an early version of the manuscript.

REFERENCES

Acharya, K. R., Stuart, D. I., Walker, N. P. C., Lewis, M., & Phillips, D. C. (1989) *J. Mol. Biol. 208,* 99−127.

Allegra, G., Ganazzoli, F., Bignotti, F., & Bolognesi, M. (1990) *Biopolymers 29,* 1823−1833.

Anfinsen, C. B., & Scheraga, H. A. (1975) *Adv. Protein Chem. 29,* 205−300.

Baldwin, R. L. (1990) *Nature 346,* 409−410.

Berne, B. J., & Pecora, R. (1976) in *Dynamic Light Scattering*, pp 376, John Wiley & Sons, Inc., New York.

Bloomfield, V. A., & Lim, T. K. (1978) *Methods Enzymol. 48,* 415−494.

Bryngelson, J. D., & Wolynes, P. G. (1990) *Biopolymers 30,* 177−188.

Cantor, C. R., & Schimmel, P. R. (1980) in *Biophysical Chemistry*, pp 643−653, W. H. Freeman and Company, New York.

Chan, H. S., & Dill, K. A. (1988) *J. Chem. Phys. 90,* 492−509.

Christensen, H., & Pain, R. H. (1991) *Eur. Biophys. J. 19,* 221−229.

Chu, B. (1974) in *Laser Light Scattering*, p 317, Academy Press, Inc., Orlando, FL.

Dill, K. A., Fiebig, K. M., & Chan, H. S. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90,* 1942−1946.

Dolgikh, D. A., Gilmanshin, R. I., Brazhnikov, E. V., Bychkova, V. E., Semisotnov, G. V., Venyaminov, S. Y., & Ptitsyn, O. B. (1981) *FEBS Lett. 136,* 311−315.

Dolgikh, D. A., Abaturov, L. V., Brazhnikov, E. V., Bychkova, E. V., Gilmanshin, R. I., Lebedev, Y., Semisotnov, G. V., Tiktopulo, E. I., & Ptitsyn, O. B. (1985) *Eur. Biophys. J. 13,* 109−121.

Dubin, S. B., Feher, G., & Benedek, G. B. (1973) *Biochemistry 12,* 714−720.

Feng, H.-P., Scherl, D. S., & Widom, J. (1993) *Biochemistry 32,* 7824−7831.

Flanagan, J. M., Kataoka, M., Shortle, D., & Engleman, D. M. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89,* 748−752.

Gast, K., Zirwer, D., Welfle, H., Bychkova, V. E., & Ptitsyn, O. B. (1986) *Int. J. Biol. Macromol. 8,* 231−236.

Ikeguchi, M., Kuwajima, K., Mitani, M., & Sugai, S. (1986) *Biochemistry 25,* 6965−6972.

Kato, S., Shimamoto, N., & Utiyama, H. (1982) *Biochemistry 21,* 38−43.

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.
Koppel, D. E. (1972) *J. Chem. Phys. 57,* 4814−4820.
Kuwajima, K. (1989) *Proteins 6,* 87−103.
Kuwajima, K., Nitta, K., Yoneyama, M., & Sugai, S. (1976) *J. Mol. Biol. 106,* 359−373.
Kuwajima, K., Hiraoka, Y., Ikeguchi, M., & Sugai, S. (1985) *Biochemistry 24,* 874−881.
Matthews, C. R. (1993) *Annu. Rev. Biochem. 62,* 653−683.
Mayo, K. H., Barker, S., Kuranda, M. J., Hunt, A. J., Myers, J. A., & Maione, T. E. (1992) *Biochemistry 31,* 12255−12265.
Miranker, A., Radford, S. E., Karplus, M., & Dobson, C. M. (1991) *Nature 349,* 633−636.
Miranker, A., Robinson, C. V., Radford, S. E., Aplin, R. T., & Dobson, C. M. (1993) *Science 262,* 896−900.
Morjana, N. A., McKeone, B. J., & Gilbert, H. F. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90,* 2107−2111.
Morrison, I. D., Grabowski, E. F., & Herb, C. A. (1985) *Langmuir 1,* 496−501.
Peng, X., Jonas, J., & Silva, J. L. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90,* 1776−1780.
Ptitsyn, O. B. (1987) *J. Protein Chem. 6,* 273−297.
Ptitsyn, O. B., Pain, R. H., Semisotnov, G. V., Zerovnik, E., & Razgulyaev, O. I. (1990) *FEBS Lett. 262,* 20−24.
Radford, S. E., Dobson, C. M., & Evans, P. A. (1992) *Nature 358,* 302−307.
Semisotnov, G. V., Rodionova, N. A., Kutyshenko, V. P., Ebert, B., Blanck, J., & Ptitsyn, O. B. (1987) *FEBS Lett. 224,* 9−13.
Semisotnov, G. V., Rodionova, N. A., Razgulyaev, O. I., Uversky, V. N., Gripas, A. F., & Gilmanshin, R. I. (1991) *Biopolymers 31,* 119−128.
Sophianopoulos, A. J., Rhodes, C. K., & Van Holde, K. E. (1962) *J. Biol. Chem. 237,* 1107−1112.