Structure and Stability of the Molten Globule State of Guinea-Pig α-Lactalbumin: A Hydrogen Exchange Study†

Chia-Lin Chyan,‡ Claire Wormald,§ Christopher M. Dobson,§ Philip A. Evans,‖ and Jean Baum*,‡

Department of Chemistry, Rutgers University, Piscataway, New Jersey 08855-0939, Oxford Centre for Molecular Sciences and Inorganic Chemistry Laboratory, University of Oxford, South Parks Road, Oxford, OX1 3QR, England, and Department of Biochemistry, Cambridge University, Tennis Court Road, Cambridge, CB2 1QW, England

Received July 8, 1992; Revised Manuscript Received March 25, 1993

ABSTRACT: A partially folded state of guinea pig α-lactalbumin (the A-state or molten globule state), formed by denaturation at low pH, has been studied using hydrogen exchange methods. The overall distribution of exchange kinetics, measured by 1-D NMR, suggests that fewer than 20 amides in the structure are involved in highly persistent residual structure, although CD results suggest that many other parts of the chain are folded, for a significant proportion of the time, into less stable structural elements. The pH-jump experiments show that some amides that are strongly protected from exchange in the native state become freely accessible in the A-state but that conversely a majority, at least, of those that are slow to exchange in the A-state retain that protection in the native state. This suggests that the persistent structure in the A-state is native-like although the possibility that nonnative like structural elements persist cannot be eliminated. Resonance assignments for key residues in the NMR spectrum of the native state have enabled us to use the pH-jump method also to identify the majority of the most protected amides in the A-state: they are located in two hydrophobic segments, corresponding to the B- and C-helices of the native structure. This strongly suggests that the most persistent structure of the A-state includes these regions. A variety of lines of evidence, including fluorescence quenching data and, most remarkably, very effective protection from exchange of an indole NH in a tryptophan side chain, suggest that some form of hydrophobic core in the helical domain of the native structure persists in the A-state, although without the stereochemical rigidity of the native tertiary structure. The other domain of the native structure, including the β-sheet, appears not to contain structural elements which persist to the same extent in the A-state, emphasizing that the molten golbule is highly heterogeneous, in terms of the stability and specificity of both backbone and side chain interactions.

The low-pH form of α-lactalbumin (the A-state) is one of the best characterized examples of a partially ordered state of a globular protein that is stable at equilibrium (Dolgikh et al., 1985; Baum et al., 1989; Kuwajima, 1989). It has been intensively investigated using a wide spectrum of techniques, and the picture which has emerged from these studies has to a large extent served as a paradigm for the "molten globule" model that is now widely applied to describe both stable and transient partially folded states of many globular proteins (Ptitsyn, 1987; Christensen & Pain, 1991; Dill & Shortle, 1991; Dobson, 1992).

The key features characterizing molten globule states are a substantial content of secondary structure which is presumed to be largely native-like, a compactness which may approach that of the native state, but substantially disordered tertiary interactions. These conclusions have, by and large, been drawn from studies using techniques that generate parameters characteristic of the molecule as a whole. In order to understand the structures of such species in detail, and hence their significance for our understanding of protein stability and folding, it is very important to characterize the properties of individual residues. NMR methods, in particular, have the potential to provide this information and are now being applied with some success to partially folded states of a number of proteins.

One particularly powerful application of NMR to the study of partially folded proteins is as a probe of hydrogen exchange behavior because, in addition to yielding important clues as to the nature of persistent structural elements, it provides information concerning their stability. The key to the success of the technique is that it enables the pattern of exchange kinetics in the partially folded state to be probed after the protein has been refolded into its native state, where NMR spectra are much more readily analyzed. This is the basis of the strategy we have applied to characterizing the A-state of α-lactalbumin, utilizing a pH jump to initiate refolding. In a previous report we demonstrated the persistence of a specific element of residual structure in the A-state using this approach (Baum et al., 1989). Using similar strategies, hydrogen exchange has now also been applied successfully to the characterization of a number of other partially folded states (Hughson et al., 1990; Jeng et al., 1990; Jeng & Englander, 1991; Harding et al., 1991), and a related experiment uses the same principle to probe the structures of intermediates that are transiently populated in the course of refolding (Roder et al., 1988; Udgaonkar & Baldwin, 1988; Miranker et al., 1991). The relationship between such species and equilibrium states is of great interest, and the applicability of hydrogen exchange as a probe of both makes it a particularly attractive comparative tool.

We have been motivated to pursue a detailed investigation of α-lactalbumin for a number of reasons. The wealth of data that have already been obtained on its behavior in the molten

† This work was supported by NIH Grant GM45302 and by U.K. Science and Engineering Research Council.

* To whom correspondence should be addressed.
‡ Rutgers University.
§ University of Oxford.
‖ Cambridge University.

0006-2960/93/0432-5681$04.00/0

globule state means that there should be ample complementary information to ensure that detailed hydrogen exchange data can be interpreted sensibly. In addition, the native state of the protein is now well defined, as a consequence of recent crystal structures of the baboon and human proteins (Acharya et al., 1989, 1991). Our hydrogen exchange studies have mainly concentrated on the guinea pig protein, since it turns out that problems with refolding to the native state limit the utility of the exchange experiment when applied to the human protein (A. T. Alexandrescu, C.-L. Chyan, J. Baum, and C. M. Dobson unpublished data). The fold of α-lactalbumins is closely similar to that of c-type lysozymes (Acharya et al., 1989), to which α-lactalbumins are also homologous in sequence (Hill & Brew, 1975; McKenzie & White, 1991). It comprises two structural lobes, one comprising a substantial, aromatic-rich hydrophobic core bounded by four α-helices and the other being more irregular with less of a defined core but including three strands of antiparallel β-sheet. The protein is stabilizing by binding calcium in its native state but the A-state does not bind the metal ion (Ikeguchi et al., 1986). Interpretation of the hydrogen exchange behavior of the A-state draws heavily on knowledge of the native structure, since a primary goal of the investigation is to elucidate the relationship between the two.

Our purpose in this paper is thus to report a detailed characterization of the exchange behavior of the A-state of α-lactalbumin and to use this to construct, as far as possible, a model to describe its structure and dynamics. From this we can draw general conclusions about the nature of partially folded proteins and, in particular, their significance for our understanding of the folding of the native state.

MATERIALS AND METHODS

Guinea pig α-lactalbumin (GPLA) was purified from whole milk. The pH of the milk was lowered to 4 and the whey prepared by centrifugation at 7000*g*. After exhaustive dialysis against 20 mM sodium phosphate, pH 6.9, the whey was loaded onto a DEAE Sephacel column (Pharmacia) which had been equilibrated with 20 mM sodium phosphate at pH 6.9. After being washed with this buffer, the GPLA was eluted with 100 mM sodium phosphate, pH 6.9. The product was dialyzed against deionized water and lyophilized.

Assignment of the ^{1}H NMR resonances of slowly exchanging amide protons was undertaken using unbuffered solutions of the protein in H_2O and D_2O. NOESY[1] and COSY spectra were obtained under various conditions of pH and temperature, in order to resolve ambiguities arising from resonance overlap.

The total number of amides remaining unexchanged after dissolution in D_2O for various lengths of time was measured by 1-D NMR spectroscopy. The protein was dissolved in D_2O at a known pH and a series of spectra recorded; the time between dissolution and acquisition of the first spectrum was 6.5 min. The experiments were performed at 25 °C at protein concentrations ranging from 0.5 to 1 mM. At the end of each experiment the sample was exposed to a higher temperature (55–60 °C) for at least 120 min, long enough to ensure complete exchange for deuterons of any remaining amide protons. After the temperature was lowered to 25 °C, a final spectrum was recoded, in which the only resonance intensity downfield of 6 ppm was due to nonexchangeable aromatic protons. The integrated area due to these protons was then subtracted from the total area in this region in each of the spectra in the time series, the difference then being a measure of the number of amides remaining at each time.

The hydrogen exchange kinetics of individual amide protons in the A-state at pH 2.0 were measured by pH-jump experiments which were performed in three steps: (a) Fully protonated GPLA was dissolved in H_2O, the pH was lowered to 2.0, and then the protein was lyophilized. (b) The protein was then dissolved in D_2O, giving a solution of the A-state without the need for further pH adjustment. (c) After various lengths of time, aliquots of the protein were restored to the native state by carefully raising the pH to 5.5. The samples were then concentrated by ultrafiltration in an Amicon stir cell (model 8MC), to a final concentration of 3.5–5 mM for 2-D, 0.5–1 mM for 1-D NMR spectroscopy. Spectra (1-D or 2-D DQF-COSY or NOESY) were recorded at 35 °C. A reference spectrum was taken of a sample of freshly dissolved GPLA at pH 5.5, which had not been exposed to pH 2 conditions. Cross peak intensities reflecting 100% ^{1}H occupancy of individual amide sites were obtained from this reference spectrum. Spectral intensities in a given time series were normalized with respect to the intensities of four nonexchangeable aromatic peaks, to ensure comparability. Two criteria were used to select data of sufficient quality to be acceptable for further analysis: (a) the intensity of the amide proton peak in the freshly dissolved protein spectrum had to be at least 15% of that of the aromatic proton reference peak, in order for the signal-to-noise ratio to be sufficient in the pH-jump experiments, and (b) the exchange curve had to be defined by at least four time points.

The theoretical exchange rate of each amide proton in a statistical coil state was calculated from model compound data (Jeng & Englander, 1991), taking into account the effects of neighboring residues (Molday et al., 1972) and temperature (Englander et al., 1979). Predicted overall exchange curves for the protein in a statistical coil state were generated as a sum of the exponentials calculated for each individual amide.

NMR experiments were performed on a Varian VXR500 spectrometer at Rutgers University and a GE/Nicolet 500-MHz instrument at the Oxford Centre for Molecular Sciences. NOESY (Kumar et all, 1980; Otting et al., 1986) and DQF-COSY (Rance et al., 1986) spectra were recorded on the Varian instrument using the method of time-proportional phase incrementation (Redfield & Kunz, 1975; Marion & Wüthrich, 1983). A total of 64 scans of 2K complex data points were collected for each of the 512 t_1 increments of the NOESY experiments and the 1024 t_1 increments of the DQF-COSY experiments. FTNMR and FELIX (Hare Research, Inc.) were used for data processing on a Silicon Graphics Iris workstation. Before Fourier transformation, the time domain data were multiplied by a phase-shifted skewed sine-bell window function in each dimension, and zero filling was used to achieve a resolution of 3.4 Hz/point in the t_1 and t_2 dimensions. NOESY and DQF-COSY spectra were obtained on the GE/Nicolet instrument using the States–Haberkorn–Ruben method (States et al., 1982). Data sets comprised 512 t_1 increments for the COSY and 256 t_1 increments for the NOESY experiments. FIDs were multiplied by trapezoidal and double-exponential multiplication prior to complex Fourier transformation. After zero filling the digital resolution was 3.5 Hz/point.

CD spectra were recorded on an AVIV circular dichroism spectrometer (model 60DS) at Rutgers University. Sample

[1] Abbreviations: NMR, nuclear magnetic resonance; 1-D, one dimensional; 2-D, two dimensional; CD, circular dichroism; COSY, correlation spectroscopy; DQF-COSY, double-quantum filtered correlation spectroscopy; GPLA, guinea pig α-lactalbumin; NOE, nuclear Overhauser enhancement; NOESY, two-dimensional NOE spectroscopy.

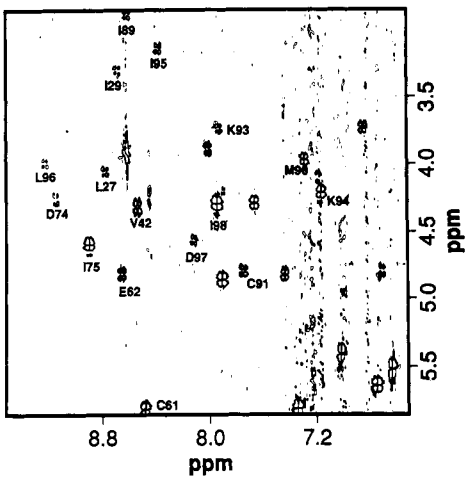

FIGURE 1: Fingerprint region of a DQF-COSY spectrum of guinea-pig α-lactalbumin, freshly dissolved in D_2O at pH 5.5, 35 °C. The cross-peaks identified arise from the most slowly exchanging protons in the molecule.

concentrations of guinea pig α-lactalbumin were determined from the extinction coefficient $\epsilon_{1\%,1cm}$ 17.2 at 280 nm (Harushima & Sugai, 1989). Sample concentrations ranged from 140 to 250 μM for spectra acquired between 250 and 320 nm and from 14 to 25 μM for spectra acquired between 190 and 250 nm. All the spectra were taken at 25 °C. Reference basis sets used to estimate the secondary structures were the basis sets of Greenfield and Fasman (1969).

RESULTS

Assignments of the Nature Protein Spectrum. NMR spectral assignments were made by mapping *J*-coupled spin systems corresponding to individual residues and then elucidating sequential connectivities between them (Wüthrich, 1986). Complete assignment of the spectrum has not so far been possible, because of the large line widths and limited resolution of the spectrum. However, of the 26 slowly exchanging amides giving detectable cross peaks in the fingerprint region of the COSY spectrum of native protein samples freshly dissolved in D_2O (Figure 1), 19 have been identified and these are the focus of interest in the present paper. They comprise residues 26–31 which form part of the B-helix, residues 89–98 which are in the C-helix, residue 42 from the β-sheet, and residues 61–62 and 74–75 which occur in regions of irregular structure. Figure 2 panels a and b show the sequential αN ($d_{\alpha N}$) and NN (d_{NN}) connectivities identified in a NOESY spectrum at pH 6.5. The assignments of residues 89–96 have been described previously (Baum et al., 1989); these have been confirmed in the present study and extended by identification of sequential connectivities to Asp 97 and Ile 98. All of the d_{NN} connectivities from this segment of the protein are seen in Figure 2b, as are $d_{\alpha N}$ connectivities between residues 89–90, 90–91, 92–93, 93–94, 94–95, and 95–96 in Figure 2a.

A second series of residues was identified from a similar analysis of COSY and NOESY spectra as Trp-U-U-Ile-Ile-Phe, where U denotes an unidentified spin system for which two αCH–βCH couplings have been identified but the remainder of the *J*-couplings could not be traced in detail (Redfield & Dobson, 1988). This segment identified as a unique sequence of residues in the protein and assigned to W26-L27-C28-I29-I30-F31. All of the d_{NN} connectivities, as well as $d_{\alpha N}$ connectivities between residues 26–27, 27–28, and 29–30, have been identified for this segment.

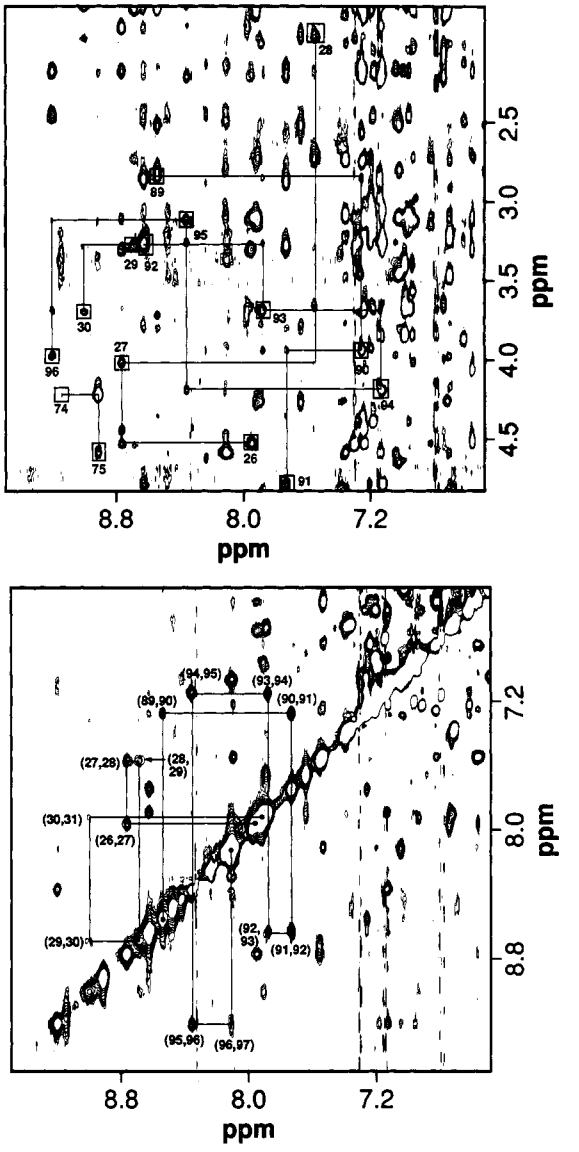

FIGURE 2: NOESY spectrum of guinea pig α-lactalbumin in H_2O at pH 6.5, 35 °C. (a, top) NH–αCH region showing sequential $d_{\alpha N}$ connectivities. The boxes represent the positions of the COSY peaks. (b, bottom) NH–NH region showing sequential d_{NN} connectivities.

A third segment was characterized in H_2O, corresponding to a sequence of residues J-Thr-U-Ala-Ile-Val, where J denotes a residue with an αCH and two βCH but no γCH protons, giving rise to a characteristic AMX spin system (Redfield & Dobson, 1988). This sequence was assigned to residues 37–42. Of these, only the amide of Val 42 is slowly exchanging in D_2O. This region was characterized by αN connectivities between residues 37–38, 39–40, 40–41, and 41–42, an NN connectivity between residues 38 and 39, and a βN connectivity between residues 40 and 41. A fourth sequence of residues, J-Ile-Ser-J, characterized by sequential αN connectivities, was also observed in H_2O. This was assigned to residues 74–77, of which the amides of Asp 74 and Ile 75 are slow to exchange in D_2O.

These assignments above are well based on sequential information. Tentative assignments for residues 61 and 62 were also made by comparing distance constraints derived from NOEs with the crystal structure of baboon α-lactalbumin (Acharya et al., 1989) and by comparison with spectra of hen and human lysozyme (Redfield & Dobson, 1988, 1990). The crystal structure of baboon α-lactalbumin reveals the proximity of the NH of Ile 75 to the αCH of Cys 61, and a medium

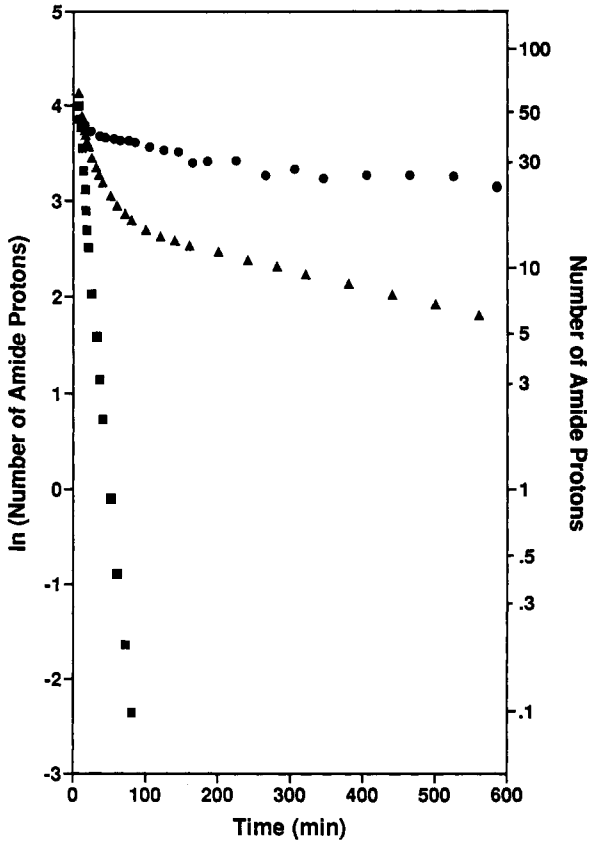

FIGURE 3: Time dependence of the average number of unexchanged amides per molecule in a sample of guinea pig α-lactalbumin after dissolution in D_2O at 25 °C, determined by one-dimensional NMR: (▲) A-state, pH 2.0; (●) native state, pH 5.5; (■) the hypothetical exchange curve predicted for the protein in a statistical coil state at pH 2.0 (Jeng & Englander, 1991; Molday et al., 1972).

strength NOE from Ile 75 NH, additional to its sequential NOE to Asp 74 Hα, could thus tentatively be assigned as involving Cys 61. Confidence in this assignment was reinforced by noting that this is the most downfield-shifted Hα proton resonance; the resonances of the corresponding protons in lysozymes (Hα of Cys 64 of hen and Cys 65 of human) are similarly shifted (Redfield & Dobson, 1988, 1990). A sequential αN connectivity then allowed Glu 62 to be identified in the spectrum.

Bulk Hydrogen Exchange Studies by One-Dimensional NMR. The exchange curve of the native protein observed at pH 5.5 is shown in Figure 3. The intrinsic exchange rate under these conditions is such that exchange from a statistical coil is predicted to be virtually complete within 1 min. Nonetheless, it is clear that the native state exchange curve reaches a virtual plateau where the residual proton intensity tails off much more gradually than the simple model predicts. As many as 30 protons show negligible exchange on the time scale of the experiment. This number accords well with the number of NH–αCH cross peaks observed in a COSY spectrum of the protein recorded immediately after dissolution in D_2O.

The overall exchange profile of GPLA in the A-state at pH 2.0, determined by integration of 1-D NMR spectra, is also shown in Figure 3, together with a hypothetical curve for the protein in a statistical coil state under the same conditions. The intrinsic exchange rate is much slower than at pH 5.5, but again it is apparent that for the majority of the amides the statistical coil model predicts the exchange kinetics rather well, so that these curves are initially almost coincident. There is, however, as for the native state at pH 5.5, a significant subset of protons which exchanges markedly more slowly than anticipated for a statistical coil. This number is less than 20, therefore less than the number that are strongly protected at pH 5.5 in the native state. We have shown previously that aggregation of the protein does not significantly influence hydrogen exchange behavior under these conditions, so that this is evidently an intramolecular effect (Baum et al., 1989). This suggests, therefore, that residual structure in at least part of the A-state is sufficiently persistent to restrict exchange of the backbone amides.

Table I: Kinetic Parameters from Three-Exponential Fits to Bulk Exchange Data

A-state (experimental data)		statistical coil (simulation)	
k^a	A^b	k^a	A^b
0.24	64	0.25	64
0.039	42	0.12	59
0.0018	17	*c*	

[a] Average rate constant (number of protons per minute) and [b] kinetic amplitude (number of protons) characteristic of the three phases giving the best fit of the data to an equation of the form

$$n(t) = A_1 \exp(-k_1 t) + A_2 \exp(-k_2 t) + A_3 \exp(-k_3 t)$$

where $n(t)$ is the average number of unexchanged NH protons per molecule at a time t after dissolution in D_2O. It was assumed that $n(t = 0) = 124$, corresponding to the number of amide and indole NHs in the protein.[c] In the case of the statistical coil simulation, the fitting program did not assign a significant amplitude to a third phase.

In order to describe the exchange behavior in a more quantitative way, we have explored the feasibility of fitting these curves to a sum of exponentials using a nonlinear regression algorithm. In reality each curve is the sum of 124 exponentials, corresponding to the number of backbone amide and indole protons in the molecule, but, in the case both of the statistical coil prediction and of the experimental curve observed for the A-state, it appears that the spread of the individual rate constants is sufficiently limited to allow the curve to be reproduced quite well by a sum of just two or three exponentials. The criterion that was used to determine the minimum number of exponentials was to obtain the best correlation factor and the best fit by visual inspection. For the statistical coil model, the best fit to the curve is described by two exponentials with rate constants and amplitudes listed in Table I. The difference between the exchange curves of the A-state and the statistical coil is that the experimental data for the A-state require at least three exponentials. From Table I it appears that the first phase obtained from the exponential fit to the A-state has a rate that is similar to the rate obtained for the statistical coil model. It is interesting to note, however, that the two phases of the statistical coil model differ only by a factor of 2, whereas the first two phases in the A-state differ by a factor of 6. This suggests that there is a subset of amide protons in the A-state which are marginally protected, with average protection factors of less than 10. In addition, there is a significant slow phase containing 17 ± 3 protons, with an average rate approximately 100-fold smaller than that of the fastest phase. Although it is not possible to fit the native state bulk exchange to the sum of two or three exponentials, we can approximate the number and protection for the slow phase. There are 30 ± 4 amides in the slow phase of the native state with protection on the order of 10^4. The number of protected amides in the A-state is clearly less than in the native state, and although the protection is smaller in the A-state the 17 or so protons are still substantially protected in the A-state and can be associated with structural elements in the partially folded molecule.

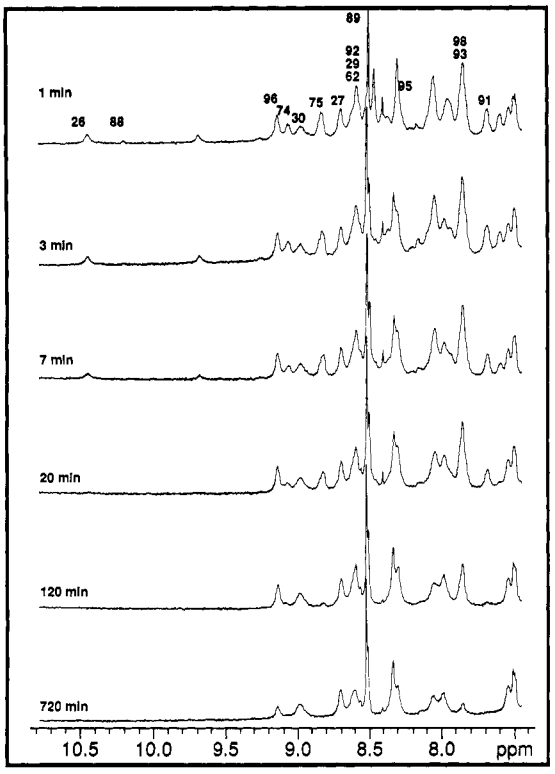

FIGURE 4: Low-field region of the one-dimensional NMR spectra of guinea pig α-lactalbumin in D_2O at pH 5.5, 35 °C, after preincubation at pH 2.0, 25 °C, for the times indicated. Resolved NH protons whose exchange kinetics could be determined from measurements of changing peak heights in these spectra are indicated.

pH-Jump Hydrogen Exchange Studies. More detailed interpretation of the exchange data requires measurement of individual exchange rates. Since the resonances of individual amides are not resolved in the spectrum of the A-state, the exchange of specific amides had to be monitored by an indirect method. The principle of the experiment is to allow exchange to proceed for variable periods of time in the A-state (pH 2.0) and then to restore the protein to its native state by raising the pH carefully to 5.5. In this state a substantial number of the amides that are slow to exchange with solvent give rise to resolved resonances in 1-D and particularly 2-D NMR spectra. The variation in the integrated intensity of these resonances with the duration of exposure to D_2O at pH 2.0 then gives a measure of the exchange kinetics in the A-state.

Resonance intensities in the pH-jump experiments were determined using various approaches. 1-D spectra show the overall time scale of exchange very clearly (Figure 4) and allow the exchange rates to be determined directly for the small number of NHs which give rise to fully resolved resonances. It is apparent that there is substantial variation in the rates at which individual amides exchange. The amide resonance of Asp 88, for example, has disappeared within 3 min and that of Cys 91 within 1 h, while that of Leu 96 still has significant intensity after 12 h. The indole proton of Trp 26 disappears within 30 min.

To follow the exchange of amides which do not give rise to resonances resolved in 1-D NMR spectrum, 2-D DQF-COSY and NOESY spectra were analyzed. The intensities of NH–αCH cross peaks in either of these spectra give a direct measure of the proton population at individual amide NH sites. Figure 5 shows the change in the fingerprint region of the DQF-COSY spectrum as a function of the incubation time in D_2O at pH 2.0 prior to refolding into the native state. Again the differential exchange rates of different amides is apparent. At the shortest times, the most prominent cross peaks arise from amides in regions of β-sheet and irregular structure, reflecting the larger αCH–NH coupling constants for these compared to amides in helical segments. However, as exchange in the A-state progresses, it is evident that the most persistent cross peaks are all from amides which are in the B- and C-helices in the native state. After 90 min all of the resonances except for Ile 75 known to be from regions forming β-sheet and irregular structures in the native state have vanished from the spectrum.

The strings of NH_i–NH_{i+1} correlations in NOESY spectra which are characteristic of α-helices also provide an interesting reflection of the exchange behavior of these segments of the molecule although, since the intensity of a given cross peak depends on the residual population of both of the exchangeable protons contributing to it, interpretation has to be largely qualitative in this case. Figure 6 shows the changes in the NH–NH region of NOESY spectra as a function of incubation time at pH 2.0. The stability of the B- and C-helical segments in the A-state is clearly demonstrated by the persistence of amides in these segments. In particular, residues 26–31, corresponding to the B-helix of the native protein, are most resistant to exchange as evidenced by the fact the cross peaks are still clearly present even after 12 h.

The exchange behavior was quantified by measurement of the intensities of resolved peaks in 1-D spectra or of NH–αCH cross peaks in COSY or NOESY spectra. The results are summarized in Table II, together with the protection factors that are implied by comparison with rates predicted for a statistical coil, taking account of pH, temperature, and sequence effects (Molday et al., 1972; Englander et al., 1979; Jeng & Englander, 1991; Robertson & Baldwin, 1991).

The total population of unexchanged amide protons 30 min after dissolution is estimated to amount to an equivalent of 25 protons per molecule, from the bulk exchange curve (Figure 3). From the rate constants for the 17 amides that were followed in detail by the method described above, we calculate that the amides will contribute an intensity equivalent to 13 ± 2 protons at this time. The residual intensity from the fast statistical coil like amides is calculated to be equivalent to two protons at this time. In addition, there are eight amides that were observed but could not be followed in detail in the pH-jump experiment. Therefore, it appears that the total intensity of 25 protons seen in the bulk experiment at 30 min can be accounted for by the sum of the individual factors described above. Similarly, after 720 min the total unexchanged intensity calculated from a 1-D pH-jump experiment is equivalent to 4 ± 1 protons, and the total unexchanged intensity expected from the bulk experiment is also 4 ± 1. It appears, therefore, that the most protected amides in the A-state can be detected by the pH-jump experiment although we cannot totally eliminate the possibility that some of the amides that are slow to exchange in the A-state are not sufficiently protected in the N-state to be probed by the pH-jump experiment.

In order to investigate this possibility more directly, an additional experiment has been devised, involving a "reverse" pH jump. In this experiment a sample was dissolved in D_2O at pH 5.5 and incubated for 2 h, during which time those protons that exchange rapidly from the native state would be lost from the protein. The pH was then lowered to 2.0 and the residual total amide intensity followed as a function of time by 1-D NMR. This experiment thus provides a direct measure of the overall exchange kinetics in the A-state of just those amides which are sufficiently protected in the native state to be detected by the pH-jump experiment.

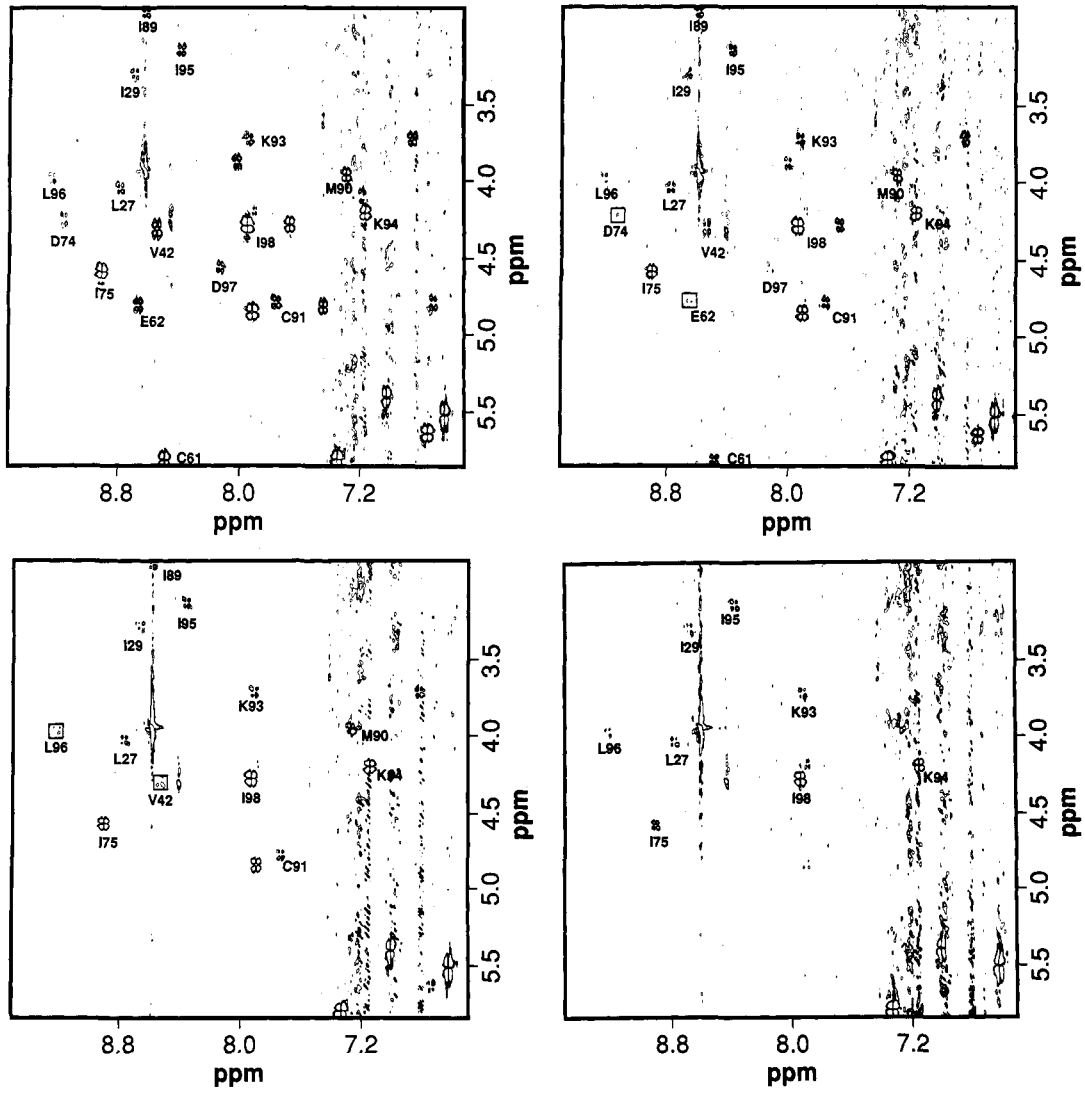

FIGURE 5: Fingerprint region of DQF-COSY spectra of guinea pig α-lactalbumin freshly dissolved in D_2O at pH 5.5, 35 °C (a, top left) and in D_2O at pH 5.5 after preincubation at pH 2.0, 25 °C, for 7 min (b, top right), for 30 min (c, bottom left), and for 90 min (d, bottom right). Square boxes indicate the peaks which can be seen in lower contour levels.

The bulk exchange curve resulting from the "reverse" experiment is shown in Figure 7. It is clear that the curve differs most substantially from the directly measured bulk at pH 2 (Figure 3) in the faster phase, the amplitude having been greatly diminished by the preincubation in the native state. At longer times, the intensity in the two experiments is identical within experimental error; for example, after 500 min the integrated intensity in both cases is 6.1 ± 2.0. In order to examine this in more detail, the data in Figure 7 were fitted to two exponentials. The fast phase contains 24 ± 3 amides with a rate constant of 0.09 min^{-1}. This is close to the value of the slower phase found for the rate calculated for the statistical coil model. The fast phase of the reverse experiment can therefore be attributed to the exchange of those amides protected significantly in the native state but not protected significantly in the A-state. The slow phase of the "reverse" experiment has a rate constant of 1.4×10^{-3} min^{-1}, which is similar to the rate of the slow subset seen in the bulk experiment at ph 2.0. The number of protons in the slow group of the reverse experiment is calculated to be 11 ± 3. Although this is less than the results for the bulk (17 ± 4), it is not significantly different given the approximations made in the analysis. In any case, it confirms that the majority of the amides that are slowly exchanging in the A-state are within the subset that is protected in the native state.

CD Spectra. CD spectra of GPLA in the near- and far-UV regions are shown in Figure 8. The near-UV spectra of the A-state show a loss of intensity relative to the native state spectrum, indicating that the A-state has lost the specific tertiary interactions that are present in the native state. The far-UV CD spectra were used to estimate the secondary structure content of the protein in the native and A-states. Using the deconvolution procedures of Greenfield and Fasman (1969), the estimated α-helix and β-sheet contents are 22% and 24% in the native state and 22% and 20% in the A-state. It is important to note that different deconvolution procedures (Brahms & Brahms, 1980; Yang et al., 1986) result in different estimates for the α-helix and β-sheet content. We choose the method of Greenfield and Fasman (1969) in order to compare our results with previous CD deconvolution of other α-lactalbumin species (Dolgikh et al., 1985; Kuwajima et al., 1985) and because the percentage of α-helix resulting from their method is more similar to the reported crystal structure than other methods. In the native state, the estimate for the α-helical content is a bit low relative to the crystal structure (~30%, Acharya et al., 1989), and the β-sheet is hard to quantify from the structure using standard algorithms as it is substantially distorted. The estimated α-helix and β-sheet content in the A-state is similar to the α-helix and β-sheet content of the native state spectrum.

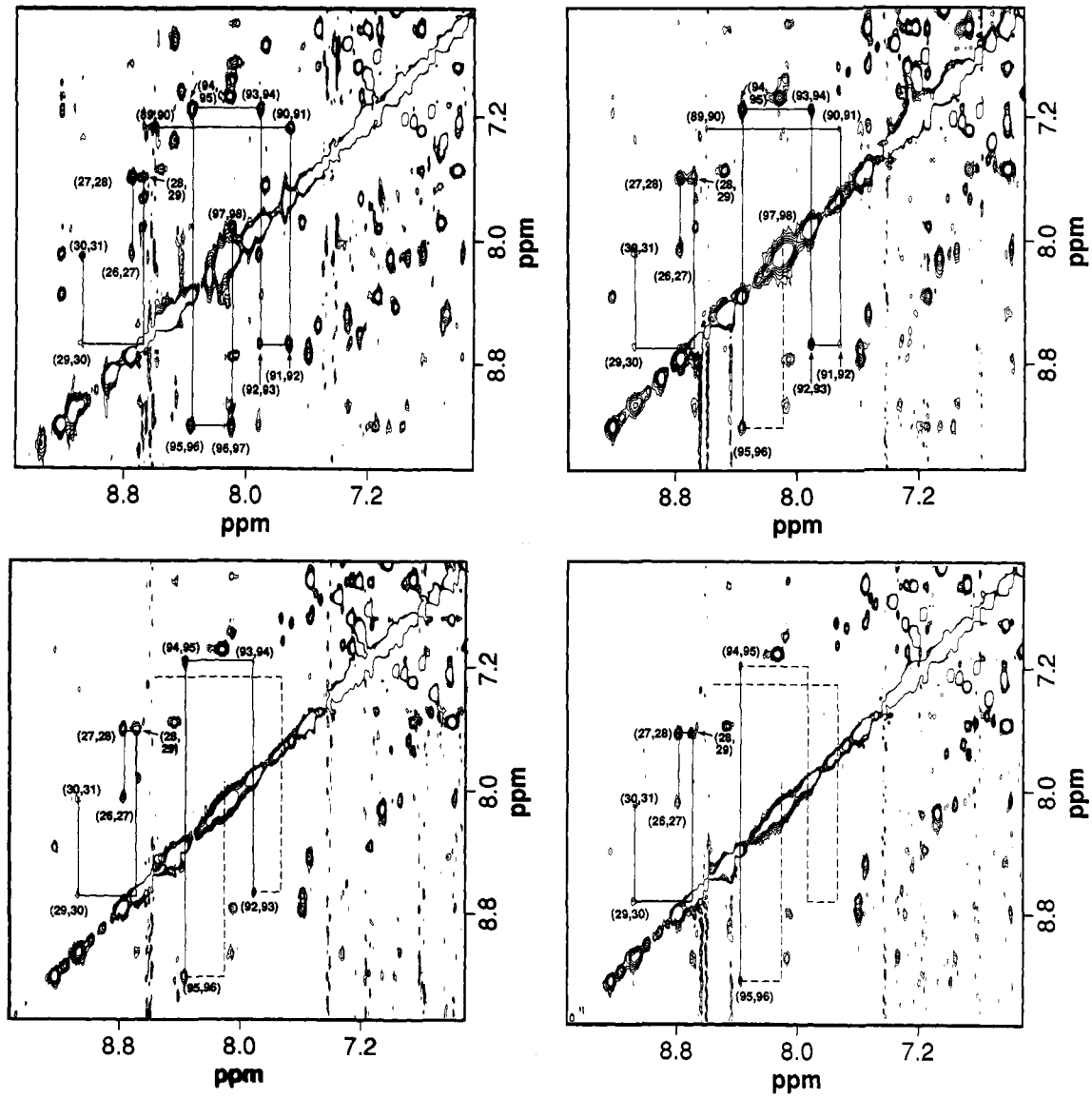

FIGURE 6: Low-field region of the NOESY spectra of guinea pig α-lactalbumin freshly dissolved in D_2O at pH 5.5, 35 °C (a, top left), and in D_2O at pH 5.5 after preincubation at pH 2.0, 25 °C, for 90 min (b, top right), for 6 h (c, bottom left), and for 12 h (d, bottom right).

DISCUSSION

From the bulk hydrogen exchange data, we see that the great majority of the amides in the A-state of α-lactalbumin exchange at rates close to, at least within a factor of 15, of those predicted by the statistical coil model, implying that any protective interactions in which they are involved are of marginal stability. Indeed, protection factors in this range have been observed for a number of residues in denatured lysozyme which, by most criteria, is a highly disordered state (Radford et al., 1992a,b). A small group of fewer than 20 residues in the α-lactalbumin molecule do appear to be involved in structure sufficiently stable to provide more substantial protection from exchange in the A-state. The amides in these segments have a wide range of protection factors up to a maximum of around 190 as seen from the pH-jump experiments. Although, even at the top end of the range, the protection factors are small relative to those measured for amides in the native state, many of which are of the order of 10^4, they do reflect a situation in which the amides are in a structured environment for the great majority of the time. They are, therefore, very significant in terms of our understanding of the average conformation of the molecule, enabling us to map regions of persistent, well-defined structure within the molten globule state. The majority, perhaps all, of the slowest exchanging amides in the A-state are strongly protected in the native state, but less than half of the amides that are strongly protected in the native state are even marginally protected in the A-state, indicating a smaller subset of protected amides in the A-state than in the native state. These hydrogen exchange data suggest that protection is probably associated with structural features that are closely related in the native and A-states. Although it is possible that this is not the case and that there is secondary structure rearrangement, the patterns of protection and the similarity of the A-state and native state CD spectra are consistent with the view that the structure in which the most protected amides are located is similar in the A-state and native state.

Those amides which could be studied by the pH-jump experiment give us specific probes of the extent to which individual regions of the protein molecule form stable structural elements in the A-state. Detailed intepretation of the protection factors has to be limited by the uncertainty in the appropriate "statistical coil" exchange rate. Nonetheless some clear trends can be discerned. Figure 9 shows a schematic view of the native conformation of α-lactalbumin in which the shaded regions indicate amides that could be probed by pH-

Table II: Rate Constants of Hydrogen–Deuterium Exchange of Individual Residues (Guinea Pig α-Lactalbumin at pH 2, 25 °C)

residue	$k_{intrinsic}$ (min^{-1})[a]	$k_{expt'l}$ (min^{-1})[b]	P (protection factor)
W26(indole)	$>1.2 \times 10^{2}$	1.3×10^{-1} (1D)	>920
W26(amide)[c]	1.41×10^{-1}	$<1.9 \times 10^{-3}$ (N)	>70
L27	1.66×10^{-1}	3.0×10^{-3} (C)	55
C28	1.49×10^{-1}	7.9×10^{-4} (N)	190
I29	1.35×10^{-1}	9.8×10^{-4} (A)	140
I30[c]	2.79×10^{-1}	$<1.9 \times 10^{-3}$ (N)	>140
F31[c]	1.95×10^{-1}	$<1.9 \times 10^{-3}$ (N)	>100
I89	2.55×10^{-1}	7.5×10^{-3} (A)	34
M90	2.79×10^{-1}	1.2×10^{-2} (A)	23
C91	1.49×10^{-1}	1.9×10^{-2} (C)	8
K93	1.16×10^{-1}	3.3×10^{-3} (A)	35
K94	8.18×10^{-2}	5.0×10^{-3} (A)	16
I95	1.45×10^{-1}	9.9×10^{-4} (N)	150
L96	2.79×10^{-1}	2.7×10^{-3} (A)	100
D97	3.40×10^{-1}	2.0×10^{-2} (A)	17
I98	2.55×10^{-1}	9.2×10^{-3} (A)	28
V42	2.79×10^{-1}	2.2×10^{-2} (C)	13
C61	1.07×10^{-1}	8.9×10^{-2} (C)	1
E62	1.09×10^{-1}	1.7×10^{-2} (C)	6
D74	1.58×10^{-1}	1.8×10^{-2} (C)	9
I75	2.55×10^{-1}	2.2×10^{-2} (A)	12

[a] The intrinsic rate constant, $k_{intrinsic}$, of each amide proton was calculated from model compound data, taking into account the effects of neighboring residues and temperature (Molday et al., 1972; Englander et al., 1979; Jeng & Englander, 1991). [b] The experimental rate constant, $k_{expt'l}$, of each amide proton was obtained from pH-jump hydrogen exchange experiments. The letter in parentheses indicates whether the exchange rate comes from one-dimensional (1D), NOESY (N), or COSY (C) data only, or whether the rate is the average of the exchange rates obtained from NOESY and COSY (A) data. Two criteria are used in order to establish whether data were used. First, at least four time points were needed to define the exchange curve. Second, the intensity of the amide proton peak had to be at least 15% of the aromatic peak in the freshly dissolved spectrum in order for the signal-to-noise level to be acceptable in the pH-jump experiments. [c] Experimental rate constants represent approximate lower limits based on N–N cross peaks in the NOESY spectra.

jump methods. From Table II it is apparent that the most strongly protected amides (with protection factors of 15 or more) occur in two sequence segments, corresponding to parts of the B and C helices of the native structure. Thus it is very likely that the most stable elements of structure in the A-state include these regions, presumably in a native-like helical conformation. Most of the remaining residues for which amide exchange could be measured in the pH-jump experiment exhibit more limited protection from exchange in the A-state. These residues include Val 42, which in the native protein forms part of the β-sheet, as well as Cys 61, Glu 62, Asp 74, and Ile 75, which are in irregularly structured segments of the same domain in the native state. For these, the absence of more extended groups of protected amides, and the uncertainties in interpreting small protection factors preclude us from interpreting their amide exchange in detailed structural terms. It is notable, however, that the protection factors for these are less than 15 and that values of this magnitude are found even in denatured lysozyme, which does not form a compact or molten globule state.

The small protection factor measured for Val 42 in the A-state could be taken to suggest that the β-sheet, in which this is located in the native structure, has little or no persistence in the A-state. However, exchange of a single amide cannot be taken as strong evidence, as this could just be a local effect. The CD data appear to indicate that β-sheet does persist to some extent, but whether such spectra can be interpreted reliably in the case of a partially folded state is open to question.

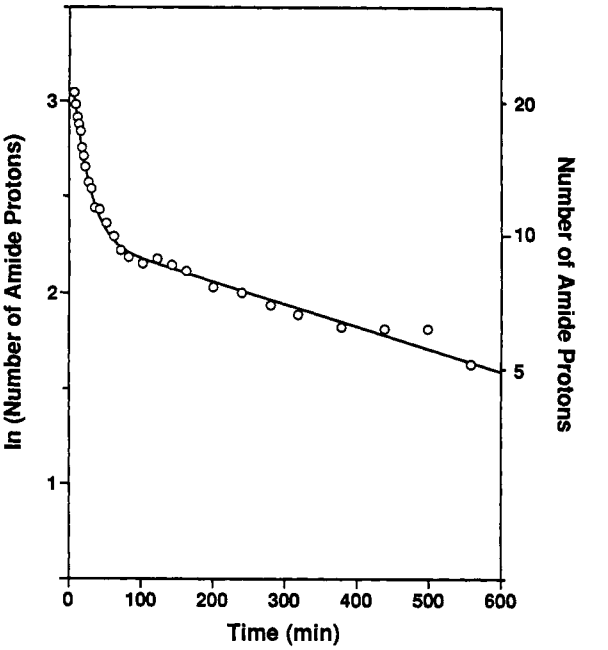

FIGURE 7: Time dependence of the number of unexchanged amide protons per molecule in guinea pig α-lactalbumin dissolved in D_2O at pH 2.0, determined from one-dimensional NMR. Protein was preincubated at pH 5.5, 25 °C, for 2 h before pH was reduced to 2. (O) represents the experimental data, and the line represents the best fit of the data.

The CD and exchange data are, in any case, not necessarily contradictory since the CD could detect structure which is significantly populated but of such marginal stability as to afford little protection from exchange. The fate of the β-sheet in the molten globule therefore remains uncertain. What is clear is that the very slowest exchanging group has been acounted for by amides in the B- and C-helices of the native protein and that the sheet is a much less persistent structure than these. The same applies to the other regions of α-lactalbumin that are α-helical in the native state; the A- and D-helices do not contain any amides that exchange sufficiently slowly in the native state for the pH-jump experiment to be applicable, so that we can conclude about them is that, if they exist at all in the A-state, they must be substantially less able to protect against exchange than the B- and C-helices. The question is what interactions afford the B- and C-helices the stability observed in the A-state. In the native conformation much of the space between the B- and C-helices is occupied by side chains from regions of the sequence that are not themselves part of secondary structural elements. Prominent among these is Trp 104 which, in the native state, is in close contact with residues in both helices, including Ile 95, Leu 96, Ile 29, and Ile 30. Residue 96 was identified in the hydrogen exchange studies as one of the most highly protected residues in the A-state (Figure 10). It is noticeable that the side chain of Trp 104, as well as Tyr 103 and His 107, which are close in the sequence, give rise to ^1H resonances that stand out in the NMR spectrum of the A-state (Alexandrescu et al., 1992, 1993). Unlike the majority of aromatic proton resonances in the A-state, these three resonances experience substantial chemical shift perturbations suggesting close interaction with one another and/or other aromatic resonances. This is consistent with their being involved in a hydrophobic cluster. NOE effects show directly that these aromatic groups, at least, are indeed in close proximity in the A-state (Alexandrescu et al., 1992, 1993). While the present results do not demonstrate it directly, it is

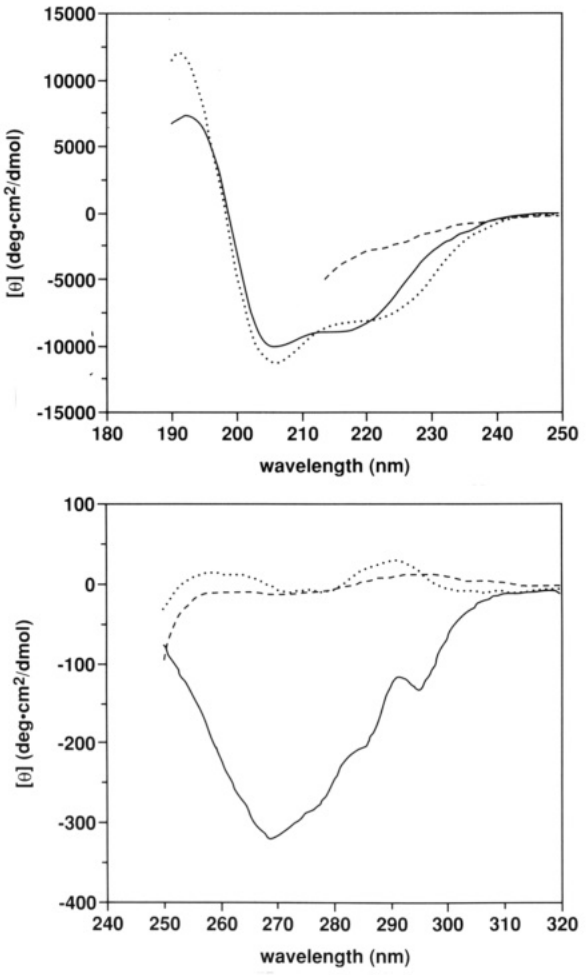

FIGURE 8: (a, top) Far-UV and (b, bottom) near-UV CD spectra of guinea-pig α-lactalbumin. (—) Native state (pH 7); (···) A-state (pH 2); (- - -) unfolded state (in 9 M urea, pH 2). The sample temperature was 25 °C.

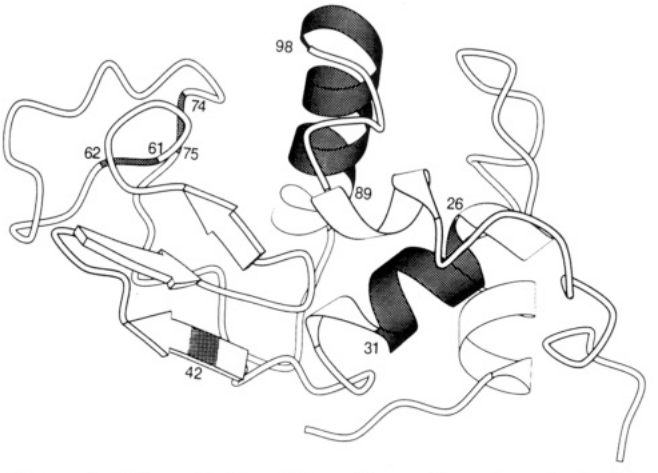

FIGURE 9: Schematic view of the native conformation of α-lactalbumin. The shaded segments are those where we have been able to probe hydrogen exchange of specific amides in the A-state by means of pH-jump experiments. The drawing, which was produced using the program Molscript (Kraulis, 1991), is based on the crystal structure of baboon α-lactalbumin (Acharya et al., 1989), to which the guinea pig protein, used in the present study, is highly homologous.

probable that such a hydrophobic cluster could play a role in stabilizing structure in the A-state.

Further evidence for persistence of some kind of hydrophobic core to the A-state comes from the behavior of tryptophan residues, which can be probed both through hydrogen exchange properties of the indole NH and through intrinsic fluorescence

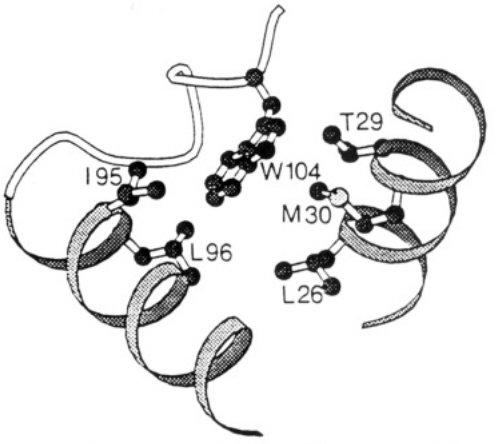

FIGURE 10: Tertiary interactions involving some of the residues forming the B- and C-helices of baboon α-lactalbumin in the native state, as revealed by the crystal structure (Acharya et al., 1989). The key role of Trp 104 in packing between the two helices is apparent. The residues shown are conserved in guinea pig α-lactalbumin, except for Leu 26 and Thr 29, which are replaced by Trp and Ile, respectively. Model building and NOE data suggest that the disposition of the helices is likely to be closely similar in the guinea pig protein (K. R. Acharya and D. I. Stuart, personal communication; Alexandrescu et al., 1992). The drawing was produced using the program Molscript (Kraulis, 1991).

measurements. There are three tryptophans in guinea pig α-lactalbumin, of which Trp 26 and Trp 104 are both deeply buried in the hydrophobic core of the native state. The protection factor, in the A-state, of the indole NH hydrogen of Trp 104 cannot be measured because the native state resonance exchanges too rapidly to be used as a probe of hydrogen exchange for the A-state. The protection factor for exchange of the indole NH hydrogen of Trp 26 is estimated to be more than 900, which is greater than that for any of the individual amide hydrogens. While we need to be somewhat cautious in interpreting this high protection factor result, because the reference "statistical coil" rates of indole protons are different and because the mechanism of exchange is not the same, this clearly suggests that, whether or not native-like interactions are conserved, this residue participates in the A-state in a stable cluster which protects the indole NH through solvent exclusion and/or formation of a tertiary hydrogen bond. Trp 26, in the native structure, is part of the B-helix, and its backbone amide is slowly exchanging. The tryptophan fluorescence of GPLA in both the native and A-states has been assigned principally to Trp 26 and Trp 104 (Sommers & Kronman, 1980). It has been observed that this fluorescence is only marginally more efficiently quenched by small molecule solutes such as iodide in the A-state than it is in the native state, where the structure shows these side chains to be effectively buried. The wavelength of maximal emission is also very similar in the two states (Sommers & Kronman, 1980; L. S. Itzhaki and P. A. Evans, unpublished results). These results strongly support the suggestion from the exchange studies that these residues are substantially excluded from solvent in the A-state.

It seems likely that the existence of hydrophobic interactions plays an important role in stabilizing the persistent structure in the A-state. This is further supported by the observation that peptides corresponding to the isolated B- and C-helices of α-alctalbumin appear to be entirely unfolded in water (D. P. Raleigh, and C. M. Dobson, unpublished results). Similar observations have been made in the case of helical segments that persist in a partly folded state of apomyoglobin (Hughson et al., 1990, 1991). In these cases, at least, some kind of packing interactions, albeit nonspecific compared to those in

the native structure, appear to be necessary to promote stable structure in these partially folded proteins. There do appear to be exceptions to this, however, for example, in the case of cytochrome *c*, at low pH and high ionic strength. Here, the persistence of helical segments has been demonstrated in a state which does not appear to be sufficiently condensed for interactions between helices to be significant (Jeng & Englander, 1991).

The results are thus consistent with the persistence of a rather resilient hydrophobic core to the α-helical domain region of the structure in the A-state of α-lactalbumin. A variety of factors demonstrate, however, that there are substantial differences from the native structure. In particular, it is clear that there is more conformational variability than is the case in the native state. The dynamics are reflected in the chemical shifts of side chain proton resonances which are considerably different from those characteristic of the native state; they are generally closer to the statistical coil position, as expected for an average of a range of sampled conformations (Baum et al., 1989). In addition, the absence of strong signals in the near-UV CD spectrum and the lack of many NOEs between side chain protons suggest that there are unlikely to be many fixed interactions between residues. It seems likely, therefore, that the residual core is substantially disordered such that side chain configurations exist in dynamic equilibrium, leading to significant averageing of CD and NMR effects. This model would also be consistent with experimental data indicating that rapid interconversion of different sets of disulfide bridges can take place in the molten globule state (Ewbank & Creighton, 1991), since these groups are located at the periphery of the core structure.

In summary, the results are reasonably consistent with the conventional "molten globule" model, but an important feature highlighted by the present data is that the stability and specificity of the structure is decidedly heterogeneous. Large regions of the protein sample a significant conformational space, and the relationship to the native state in these regions is unclear. The notion that not only side chain interactions but also a majority of the backbone structure may be fluctuating is an important extension of our understanding of the nature of the molten globule state. This disorder explains the lack of many specific NOEs that would arise from nativelike hydrophobic clusters. Only a limited region, corresponding to the B and C-helices of the native protein, appears to maintain a subset of slowly exchanging amide protons. These results indicate that although different structures of the molten globule state may be interconverting, the B and C helices must be present in effectively all of these different structures to give rise to the high protection factors.

The disordered character of the molten globule state has important implications for our understanding of folding mechanisms, since it means that a secondary structural framework in a partially folded compact intermediate need not be expected to be entirely native-like or indeed to have any unique configuration. It should be noted, however, that protection factors which have been estimated for amide exchange in transient folding intermediates appear often to be much larger than is observed in equilibrium molten globule states (Udgaonkar & Baldwin, 1990; Radford et al., 1992a,b). This suggests that, under the conditions of the refolding experiments, which strongly favor native structure, partially folded states are less structurally heterogeneous and more native-like than they are under mild denaturing conditions such as those employed in the present study. This may be because the conformational changes are too slow to allow significant interchange of conformational states on the time scale (ms) of the lifetime of kinetic intermediates. On the time scale of the hydrogen exchange experiments in the molten globule state, however, conformational interchange may occur and allow hydrogen exchange to take place in amides in all but the most persistent regions of the structure.

ACKNOWLEDGMENT

We thank Drs. Andrei Alexandrescu and Christina Redfield for helpful discussions. We acknowledge support from NIH Grant GM45302 to J. B., the Louis Bevier Graduate Fellowship to C.-L.C., and the SERC and MRC through the Oxford Centre for Molecular Sciences and the Cambridge Centre for Molecular Recognition to C.M.D. and P.A.E.

REFERENCES

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

Acharya, K. R., Ren, J. S., Stuart, D. I., Phillips, D. C., & Fenna, R. E. (1991) *J. Mol. Biol. 221*, 571–581.

Alexandrescu, A. T., Broadhurst, R. W., Wormald, C., Chyan, C.-L., Baum, J., & Dobson, C. M. (1992) *Eur. J. Biochem. 210*, 699–709.

Alexandrescu, A. T., Evans, P. A., Pitkeathly, M., Baum, J., & Dobson, C. M. (1993) *Biochemistry 32*, 1707–1718.

Baum, J., Dobson, C. M., Evans, P. A., & Hanley, C. (1989) *Biochemistry 28*, 7–13.

Brahms, S., & Brahms, J. (1980) *J. Mol. Biol. 138*, 149–178.

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

Dill, K. A., & Shortle, D. (1991) *Annu. Rev. Biochem. 60*, 795–825.

Dobson, C. M. (1992) *Curr. Opin. Struct. Biol. 2*, 6–12.

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

Englander, J. J., Calhoun, D. B., & Englander, W. (1979) *Anal. Biochem. 92*, 517–524.

Ewbank, J. J., & Creighton, T. E. (1991) *Nature 350*, 518–520.

Greenfield, N., & Fasman, G. D. (1969) *Biochemistry 8*, 4108–4116.

Harding, M. M. H., Williams, D. H., & Woolfson, D. N. (1991) *Biochemistry 30*, 3120–3128.

Harushima, Y., & Sugai, S. (1989) *Biochemistry 28*, 8568–8576.

Hill, R. L., & Brew, K. (1975) *Adv. Enzymol. Relat. Areas Mol. Biol. 43*, 411–490.

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

Hughson, F. M., Barrick, D., & Baldwin, R. L. (1991) *Biochemistry 30*, 4113–4118.

Hvidt, A., & Nielsen, S. O. (1966) *Adv. Protein Chem. 21*, 288–386.

Ikeguchi, M., Kuwajima, K., & Sugai, S. (1986) *J. Biochem. (Tokyo) 99*, 1191–1201.

Jeng, M.-F., & Englander, S. W. (1991) *J. Mol. Biol. 221*, 1045–1061.

Jeng, M.-F., Englander, S. W., Elöve, G. A., Wand, A. J., & Roder, H. (1990) *Biochemistry 29*, 10433–10437.

Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946–950.

Kumar, A., Ernst, R. R., & Wüthrich, K. (1983) *Biochem. Bophys. Res. Commun. 95*, 1–6.

Kuwajima, K. (1989) *Proteins: Struct., Funct., Genet. 6*, 87–103.

Kuwajima, K., Hiraoka, Y., Ikeguchi, M., & Sugai, S. (1985) *Biochemistry 24*, 874–881.

Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.
McKenzie, H. A., & White, F. H. (1991) *Adv. Protein Chem. 41*, 173–315.
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.
Otting, G., Widmer, H., Wagner, G., & Wüthrich, K. (1986) *J. Magn. Reson. 66*, 187–193.
Ptitsyn, O. B. (1987) *J. Protein Chem. 6*, 272–293.
Radford, S. E., Dobson, C. M., & Evans, P. A. (1992a) *Nature 358* 302–307.
Radford, S. E., Buck, M., Topping, K. D., Dobson, C. M., & Evans, P. A. (1992b) *Proteins: Struct., Funct., Genet. 14*, 237–248.
Rance, M., Sorensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 95*, 1–6.
Redfield, A. G., & Kunz, S. D. (1975) *J. Magn. Reson. 19*, 250–254.
Redfield, C., & Dobson, C. M. (1988) *Biochemistry 27*, 122–136.
Redfield, C., & Dobson, C. M. (1990) *Biochemistry 29*, 7201–7214.
Robertson, A. D., & Baldwin, R. L. (1991) *Biochemistry 30*, 9907–9914.
Roder, H., Elöve, G., & Englander, S. W. (1988) *Nature 335*, 700–704.
Sommers, P. B., & Kronman, M. J. (1980) *Biophys. Chem. 11*, 217–232.
States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286–290.
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.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
Yang, J. T., Wu, C.-S., & Martinez, H. M. (1986) *Methods Enzymol. 130*, 208–209.