The Refolding of Human Lysozyme: A Comparison with the Structurally Homologous Hen Lysozyme†

Shaun D. Hooke, Sheena E. Radford, and Christopher M. Dobson*

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

Received November 18, 1993; Revised Manuscript Received February 9, 1994®

ABSTRACT: Pulsed hydrogen exchange labeling has been used in conjunction with circular dichroism in the near and far UV to study the refolding of human lysozyme from its guanidinium chloride denatured state. Human lysozyme differs in sequence by 51 residues and one insertion from the hen protein, which has previously been studied under identical conditions by similar methods [Radford, S. E., Dobson, C. M., & Evans, P. A. (1992) *Nature 358*, 302–307]. The two proteins show marked differences in their folding kinetics. First, the overall rate of refolding of human lysozyme is 4-fold faster than that of the hen protein. Second, although protection of amides in the α-domain develops faster than that of amides in the β-domain in both proteins, unlike hen lysozyme stabilization of the secondary structural elements of the α-domain in human lysozyme does not occur in a fully cooperative manner. Rather, amide hydrogens in two α-helices located near to the N-terminus and in the 3_{10} helix close to the C-terminus of the protein are protected from exchange significantly faster than those in the remaining two α-helices in the α-domain of the protein. Third, stopped flow CD measurements show that both proteins develop extensive secondary structure during the dead time of these experiments (ca. 2 ms); this is accompanied by formation of tertiary interactions, probably involving tryptophan residues, only in the human enzyme. These results suggest that although the fundamental folding process is similar in the two proteins, human lysozyme differs in that it forms a stable subdomain involving the two N-terminal α-helices and the C-terminal 3_{10} helix in the first few milliseconds of folding, and that at least some tryptophan residues are ordered before the formation of the native state. This indicates that the details of the folding of homologous proteins may differ as a consequence of amino acid substitutions and suggests that the study of mutant and variant proteins can provide clues as to the determinants of folding.

It has been accepted for many years that all the information required to direct the folding of at least most proteins to their native states is inherent in their amino acid sequences (Anfinsen, 1973). Despite this, the nature of protein folding mechanisms and the manner in which the compact native state is achieved are still not well understood. In recent years much effort has been directed into investigating the specific kinds of interactions likely to be important in driving the folding reaction (Dill, 1990; Creighton, 1992a; Dobson, 1993, and references therein). It is now evident from a wide range of experiments that, at least for many proteins, specific pathways of folding are involved. Furthermore, it is clear that a limited number of transient intermediates may be critical in determining the nature of such pathways. Characterizing these intermediates will be of considerable value in understanding the events and the dominant forces involved in protein folding (Dill, 1990; Baldwin, 1990; Kim & Baldwin, 1990).

Many techniques have been used to gain information about the kinetics of formation and the structural features of such folding intermediates (Creighton, 1992a). Their transient existence and frequently low populations, however, make this a demanding task. One approach has been to study species trapped during the folding process; of particular interest here has been the exploitation of methods for blocking free thiol residues prior to disulfide bond formation in oxidative refolding reactions (Creighton, 1992b; Weissman & Kim, 1992). An alternative strategy has been to use stopped flow methods and derivatives of these to follow the folding process directly or indirectly (Evans & Radford, 1994). Techniques utilized include CD, fluorescence, and pulsed hydrogen exchange labeling in combination with 2D NMR and recently electrospray ionization mass spectrometry (Miranker et al., 1993). While the optical methods provide information about the global events in folding, pulse labeling in combination with NMR gives information about intermediates at the individual residue level. Using this latter approach much information has been gained about the folding of a number of small proteins (Baldwin, 1993). Studies of this type have been complemented by other methods, particularly the use of site-directed mutagenesis to investigate the energetics of folding intermediates, and by inference their structural characteristics (Fersht et al., 1992; Matthews & Zitzewitz, 1993), and by the investigation of stable partially folded species (Kuwajima, 1989; Ptitsyn et al., 1990; Buck et al., 1993).

The folding pathway of hen lysozyme has been studied in detail by a combination of the physical methods discussed above (Dobson et al., 1994). The folding of this protein has been found to be complex and appears to involve parallel folding pathways rather than a simple sequential scheme (Radford et al., 1992a; Miranker et al., 1993). The first step in folding is believed to be collapse to a state which contains extensive, although fluctuating, secondary structure but few of the persistent tertiary contacts required to stabilize units of secondary structure units against amide hydrogen exchange. Following this initial collapse the majority of molecules fold to a well-defined intermediate in which one domain in the native structure (Figure 1), consisting of the four α-helices

† S.E.R. is a Royal Society 1983 University Research Fellow. The research of C.M.D. is supported in part by the award of an International Scholarship by the Howard Hughes Medical Foundation.

* To whom correspondence should be addressed.

® Abstract published in *Advance ACS Abstracts,* April 1, 1994.

0006-2960/94/0433-5867$04.50/0

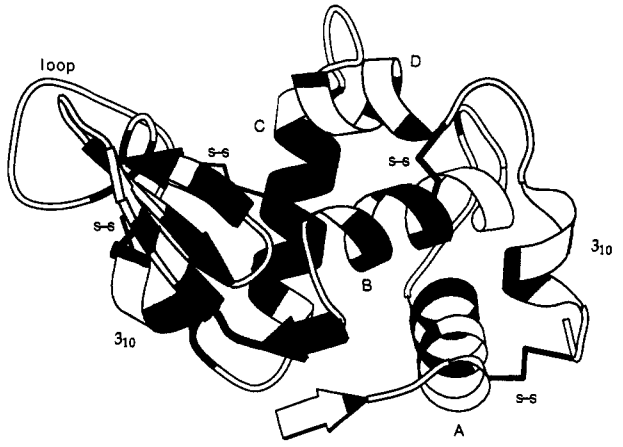

FIGURE 1: Schematic diagram of the native structure of human lysozyme. Shaded areas represent residues with amide protons which are protected from exchange with the solvent in the native state and are used as probes in the pulse labeling experiments. Disulfide linkages (Cys^6–Cys^{128}, Cys^{30}–Cys^{116}, Cys^{65}–Cys^{81}, and Cys^{77}–Cys^{95}) are labeled s–s. These remain intact during the refolding experiments. The figure was drawn using the program MolScript (Kraulis, 1991).

```
                    5         10        15        20
HUMAN LZ    K V F E R C E L A R T L K R L G M D G Y
HEN LZ      K V F G R C E L A A A M K R H G L D N Y

                   25        30        35        40
HUMAN LZ    R G I S L A N W M C L A K W E S G I N T
HEN LZ      R G Y S L G N W V C A A K F E S N F N T

                   45        50        55        60
HUMAN LZ    R A T N Y N A G D R S T D Y G I F Q I N
HEN LZ      Q A T N R N T - D G S T D Y G I L Q I N

                   65        70        75        80
HUMAN LZ    S R Y W C N D G K T P G A V N A C H L S
HEN LZ      S R W W C N D G R T P G S R N L C N I P

                   85        90        95       100
HUMAN LZ    C S A L L Q D N I A D A V A C A K R Y V
HEN LZ      C S A L L S S D I T A S V N C A K K I V

                  105       110       115       120
HUMAN LZ    R D P Q G I R A W V A W R N R C Q N R D
HEN LZ      S D G N G M N A W V A W R N R C K G T D

                  125       130
HUMAN LZ    V R Q Y V Q G C G V
HEN LZ      V Q A W I R G C R L
```

FIGURE 2: Alignment of human and hen lysozyme sequences (Imoto et al., 1972). Nonidentical residues are given in bold type and are underlined if greater than 85% of the surface area of the residue is buried (Kabsch & Sander, 1983; Rose & Dworkin, 1989).

(A–D) and a C-terminal 3_{10} helix (the α-domain), is persistently structured whereas the other domain consisting of a triple-stranded antiparallel β-sheet, a short double-stranded β-sheet, a 3_{10} helix, and a long loop (the β-domain) is still relatively unstable. The protection of amides in the α-domain appears to be highly cooperative. Similarly, protection of the amides in the β-domain, while slower, is also cooperative to a high degree (Miranker et al., 1991; Radford et al., 1992a).

In this paper we describe pulse labeling studies combined with 2D ^{1}H NMR and complementary stopped flow CD measurements designed to investigate the folding mechanism of human lysozyme. This protein is homologous in sequence to hen lysozyme differing by 51 out of 129 residues, of which only 27 include a change in charge and/or a large change in size. The remaining changes are of a relatively conservative nature (Figure 2). In addition, human lysozyme has an insertion of a glycine residue at position 48 relative to the sequence of the hen protein. The X-ray structures of both proteins are available at high resolution and are closely similar to each other (Artymiuk & Blake, 1981; Handoll, 1985). Furthermore, the NMR spectra of both proteins are assigned (Redfield & Dobson, 1988, 1990). This has provided an opportunity to compare the refolding of homologous proteins and to explore further the details of the refolding pathways of this class of lysozymes by analyzing the differences in their behavior.

MATERIALS AND METHODS

Materials. Human lysozyme was isolated from a sample of freeze dried urine of a leukemia patient; we acknowledge the generous gift from Dr. E. Osserman of this material. Following extensive dialysis against water at pH 3.8, the protein was purified by ion exchange chromatography (Bio-Rex 70 resin, 200–400 mesh). The protein was loaded on to a column preequilibrated to pH 5.5 with 50 mM sodium phosphate buffer. The column was then washed with 50 mM sodium phosphate buffer at pH 7.1 to remove impurities. Human lysozyme was eluted with the same buffer, but containing additionally 1 M NaCl, pH 7.1, and the eluted protein was dialyzed extensively against water at pH 3.8 and lyophilized. All operations were performed at 4 °C. The protein migrated as a single band on a SDS denaturing gel and was shown to be >99% pure by N-terminal sequencing and electrospray mass spectrometry. Guanidinium chloride (GuHCl) (>99%) was obtained from Sigma Chemical Co. and was deuterated where necessary by three repeated cycles of dissolution in D_2O (99.9%) followed by lyophilization. All other chemicals were analytical grade.

Pulse Labeling Experiments. A Biologic quench flow QFM5 module was used to dilute rapidly a solution of human lysozyme in 6 M GuDCl/D_2O (1 volume) with 20 mM sodium acetate in H_2O, pH 5.5 (10 volumes), to initiate the refolding reaction. The intrinsic rate of amide hydrogen exchange at 20 °C at the final refolding pH of 5.3 is approximately 10 s (Bai et al., 1993), and thus no significant exchange of amide deuterons occurs over the time course of folding (0–500 ms) monitored in this phase. At different times following the initiation of refolding, the pH of the refolding medium was jumped to 9.3 by dilution with 200 mM sodium borate buffer at pH 10.0 (5 volumes); this labeling pulse was applied for 8 ms. As the intrinsic half-life for amide exchange at this pH and temperature is approximately 1 ms (Bai et al., 1993), unprotected amide deuterons are fully exchanged for protons in this phase. Following this pulse, exchange was quenched by dilution with 0.5 M ethanoic acid (5 volumes) to reduce the pH to approximately 4.3, at which pH the refolding reaction was allowed to proceed to completion. Eight time points between 3.5 and 500 ms were taken. The zero time approximation was made by thermally denaturing a solution of human lysozyme containing the correct H/D ratio (i.e., 1 volume deuterated denatured protein solution with 15 volumes of H_2O) for 15 min at 80 °C at pH 3.8. All amides exchange under these conditions to the equilibrium H/D ratio, thereby simulating the situation which would occur if a labeling pulse was applied at the onset of folding.

Stopped-Flow CD. Refolding was monitored by stopped flow CD at 225, 270, and 293 nm using a Jasco J720 spectropolarimeter following mixing in a Biologic SFM3 module. In all cases the protein was refolded from 6 M GuHCl solution by an 11-fold dilution into 20 mM sodium acetate buffer at pH 5.5 as in the hydrogen exchange experiments. The protein concentration in the refolding buffer was reduced for experiments in the far UV because of intense absorption at 225 nm; the final refolding protein concentrations were 0.23 mg mL^{-1} for experiments at 225 nm and 1.8 mg mL^{-1}

for experiments at 270 and 293 nm. All experiments were performed at 20 °C.

NMR Spectroscopy and Quantitation of NMR Spectra. Phase-sensitive COSY spectra were recorded for each sample on a Nicolet GE500 or GE Ω500 NMR spectrometer. Each free induction decay consisted of 1024 complex points. Thirty-two scans were collected with a total of 256 increments in the t_1 dimension; the total acquisition time was 4.4 h. All spectra were recorded under identical conditions at 35 °C. Data were processed using FTNMR (Hare Research Inc.) on a SUN computer. The intensity of the CαH–NH cross peaks was taken as the absolute sum of the heights of the four phase-sensitive components. These were then normalized in each case to a standard set of eight phase-sensitive peaks from the aromatic region of the spectrum. Fractional exchange was calculated by comparing the normalized cross peak intensities of individual spectra to those of the zero time point which was defined as having a proton occupancy of 100%. Both the CD and pulse labeling data were fitted using nonlinear regression methods on a Macintosh SE/30 computer with Kaleidagraph 2.1.3.

RESULTS

The number of amide hydrogens which could be followed quantitatively in the pulse labeling experiments for human lysozyme (47) is similar to the number examined previously for hen lysozyme (49) (Radford et al., 1992a). The protected amides are largely the same in both proteins and are distributed throughout the structure (Figure 1), being located predominantly in regions of secondary structure in the native enzyme. Some five protected amides in human lysozyme (residues Ala42, Ile56, Cys65, Gln86, and Ala92) and six residues in hen lysozyme (Ile55, Leu56, Asn59, Cys64, Trp111, and Arg125) are not reported to form hydrogen bonds in the native state (Artimuik & Blake, 1991; Pedersen et al., 1991), and protection of these in the native structure is presumably attributable predominantly to burial from the solvent (Radford et al., 1992b).

Figure 3 shows representative data for hydrogen exchange pulse labeling experiments for human lysozyme. The average data for individual secondary structure elements are shown in Figure 4 and are compared with average data for identical experiments on hen lysozyme (Radford et al., 1992a). Despite the similarities in the folding kinetics of the two enzymes, a closer examination of the data in Figure 4 reveals that the behavior of the human protein is, in fact, quite distinct from that of hen lysozyme. First, overall protection of the amides occurs much more rapidly in human lysozyme; more than 70% of the complete protection is observed for all amides in the human protein in about 100 ms compared with a time of around 400 ms for the same extent of protection in the hen protein (Radford et al., 1992a). Second, amides in two of the α-helices (A and B) in the N-terminal region of the human protein as well as one amide proton in the loop between these helices and two amides in the C-terminal 3_{10} helix are protected from exchange very rapidly indeed; more than 70% of complete protection for these amides occurs within the dead time (3.5 ms) of the experiment. By contrast, amides in the C and D helices of the human protein are protected more slowly; only 40% protection occurs in the dead time, and only after 70 ms have these residues gained the same degree of protection as that observed in the A and B helices after only 3.5 ms. Three other residues (Trp64, Cys65, Ile79), which lie in the long loop in the β-domain, appear to behave in a manner indistinguishable from that of amides in the C and D helices. These three amides displayed anomalous behavior in the hen protein and were also protected at a rate similar to amides in the α-domain (Radford et al., 1992a). Amides of the remaining residues in the β-domain which include residues in both the triple and double-stranded β-sheets, a second 3_{10} helix, and three additional residues in the long loop (Arg62, Asn66, and Cys77), are protected more slowly than the four α-helices; more than 100 ms is needed for 70% protection to be achieved.

Fitting the experimental exchange data for human lysozyme to a kinetic model is complex because of the fast protection rates. For the most rapidly protected amides (from the A and B helices and the C-terminal 3_{10} helix) protection is virtually complete in the dead time of the experiment, and only an upper limit can be placed on the time constant for protection (Table1, Supplementary Material). For amides showing intermediate rates of protection (from the C and D helices and residues Trp64, Cys65, and Ile79 of the long loop) the kinetics are clearly at least biphasic; the fast phase (approximately 40% of the total amplitude) was largely complete in the dead time of the experiment, and only estimates of the rates of protection of these amides can be obtained. The slow phase, which accounted for the remaining approximately 40% amplitude could, however, be fitted to a single exponential with a time constant, on average, of 50 ms. The protection behavior of the slowest group of amides, all of which are located in the β-domain, gave good fits to double-exponential functions, with average time constants and amplitudes of 3 ms, 20% and 60 ms, 60%. The latter is similar to the slow time constant for the protection of amides in the C and D helices, although this phase is much larger in amplitude for amides in the β-sheet domain.

Two of the five indole hydrogens (Trp109 and Trp112) of human lysozyme are sufficiently protected against exchange in the native state to be monitored in the pulse labeling experiments and hence can be used as probes for the formation of tertiary structure during folding. The indole hydrogen of Trp112 is largely protected in the dead time of the experiment and most closely resembles the protection kinetics of amides in the A and B helices and the C-terminal 3_{10} helix. Trp109, by contrast, appears to become protected on a slower time scale, which resembles more closely that of the amides in the C and D helices and in the β-domain.

The biphasic protection kinetics observed for all the hydrogen protection data can arise from two distinct situations. In the first, all of the amides could be partially protected from exchange in a single kinetic refolding step, and later kinetic steps represent the stabilization of this structure. The second possibility involves two populations of molecules refolding at different rates to a fully protected state. These two situations can be distinguished by variation of the pH of the labeling pulse (Roder et al., 1988; Udgaonkar & Baldwin, 1988, 1990). No significant dependence of the extent of labeling of the majority of amides in human lysozyme could be detected as the pH was varied from 9.3 to 10.3. There is no evidence, therefore, that states with partial protection contribute to the observed folding kinetics. Furthermore, it indicates that the distinct behavior of the A and B helices and the C-terminal 3_{10} helix relative to the rest of the α-domain does not arise from different marginal stabilities of the individual helices in a partially folded state, but it must arise from noncooperative events in the formation of highly protective structure in this domain.

Stopped flow CD experiments were carried out to complement the pulse labeling data by investigating the formation

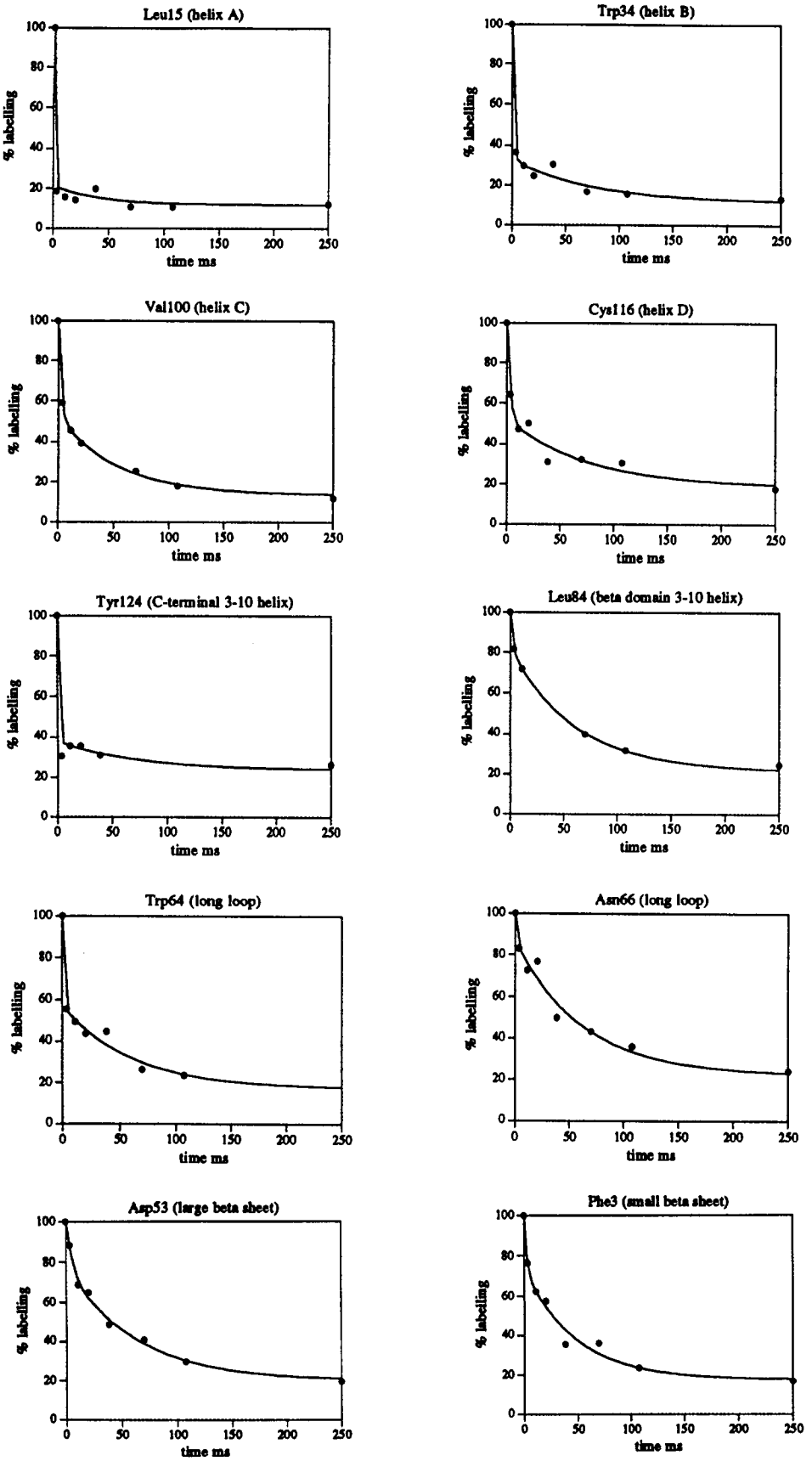

FIGURE 3: Time courses for the protection of amides from exchange during the refolding of human lysozyme. The extent of labeling is plotted against the time after the initiation of folding. One residue is shown as a representative of each secondary structure element in the native structure. The curve drawn in each case is the best double-exponential fit to the time course of each individual amide.

of secondary structure (far-UV CD) and tertiary structure (near-UV CD) during the folding process of human lysozyme. Nearly 50% of the ellipticity change at 293 nm arising predominantly from tryptophan residues occurs within the dead time of the experiment, and the remainder of the signal evolves within about 200 ms (Figure 5a). This indicates that

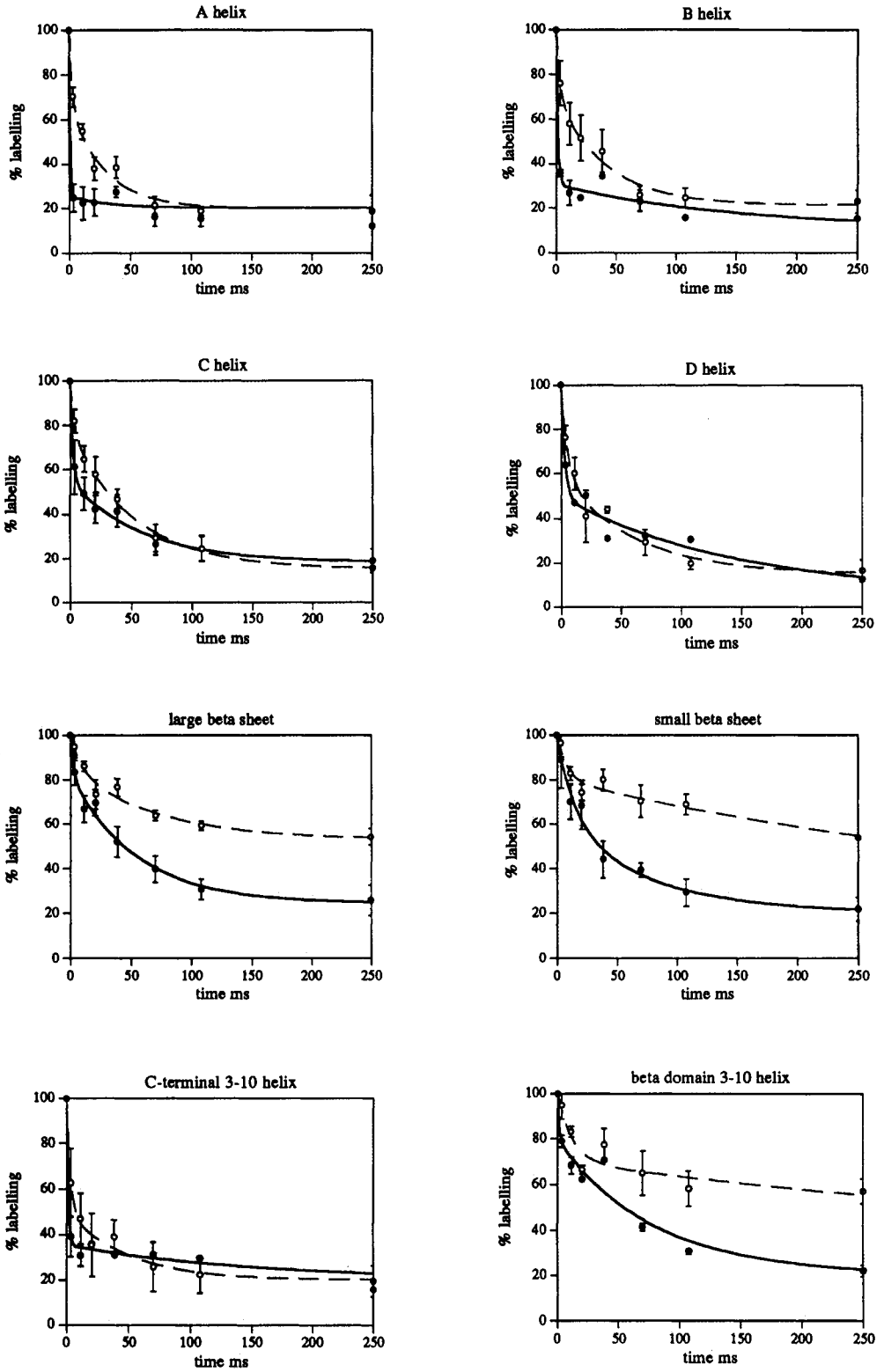

FIGURE 4: Comparison of the time courses for protection of secondary structure units from exchange during refolding of hen and human lysozymes. Pulse labeling data for both human (closed circles) and hen lysozymes (open circles) (Radford et al., 1992a) have been averaged over each secondary structure unit. Error bars represent one standard deviation from the average for the amides considered in each secondary structure unit. Curves represent the best double-exponential fit to the averaged data set.

immobilization of a proportion of the tryptophan residues takes place faster than formation of the fully native structure as judged by near-UV CD and at least as fast as protection of the amides in the A and B and C-terminal 3_{10} helices. The near-UV CD spectrum of human lysozyme is characterized by a large negative ellipticity at 270 nm, in addition to the positive band at 293 nm; the latter which is thought to arise predominantly from tyrosine residues (Beychock et al., 1971). The kinetics of refolding measured at this wavelength are much simpler than those measured at 293 nm and are characterized by a single-exponential function with a time constant of about 80 ms (Figure 5c). This is close to the rate of protection of amides in the slow phase of folding and presumably can be associated with the formation of the fully native protein. In the far-UV region of the CD spectrum, a near native-like CD signal is formed during the dead time of

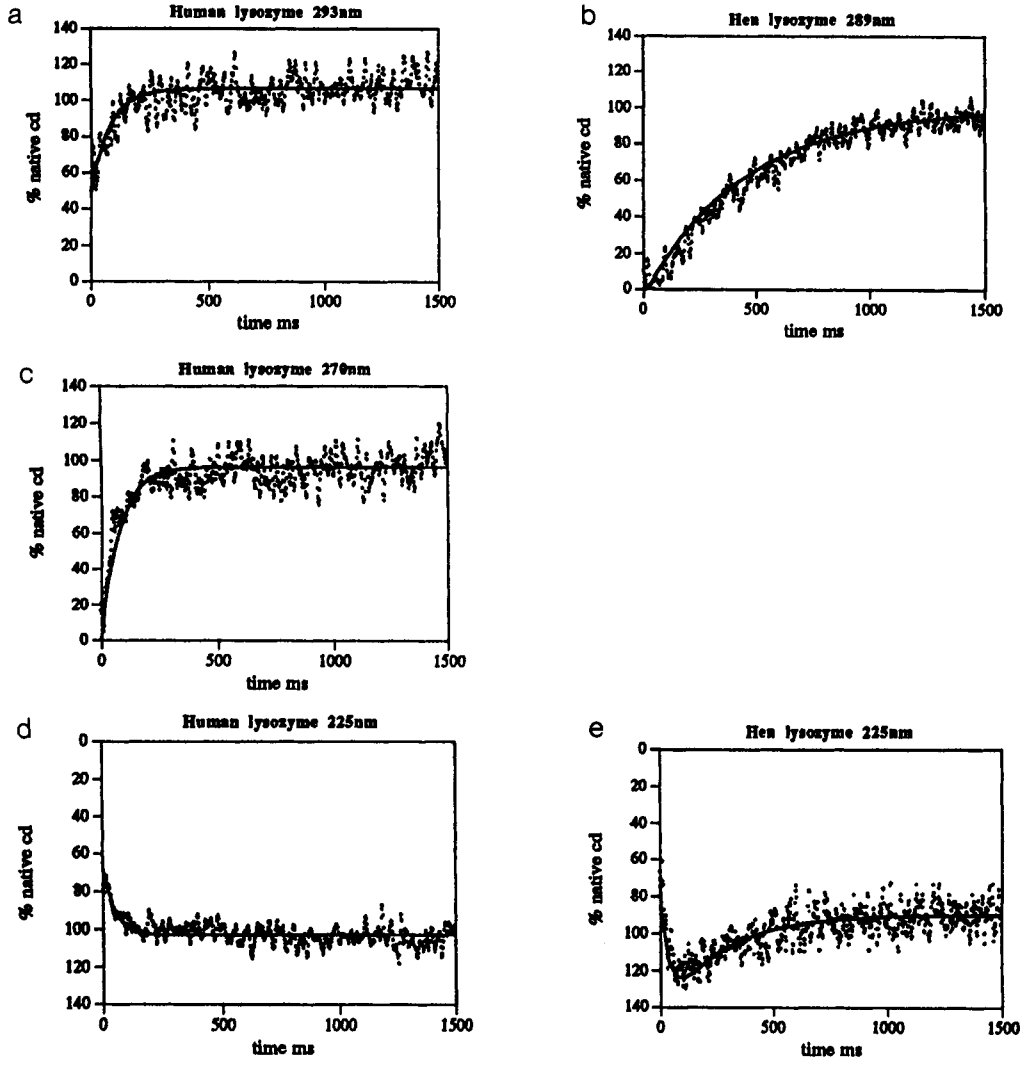

FIGURE 5: Refolding kinetics of human and hen lysozymes monitored by stopped flow CD in (a, b, c) the near-UV and (d, e) the far-UV. Data have been normalized such that the value of the denatured state is 0 and that of the native state is 100%.

the experiment (Figure 5d). A native-like signal then develops in a single subsequent minor phase with kinetics close in rate to those of the slower steps observed in the amide protection experiments and in the near-UV CD.

DISCUSSION

Both the CD and pulse labeling experiments show that hen and human lysozymes differ significantly in their folding kinetics, human lysozyme being both faster and less cooperative in its behavior. Nevertheless, there are important similarities in the refolding of both proteins. In both cases folding to a persistently structured state able to protect against hydrogen exchange occurs predominantly in the α-domain before the β-domain and is cooperative within individual elements of secondary structure. In addition, there is evidence for parallel pathways, in which a proportion of the molecules become protected before the remainder. In both proteins a very slow folding population of around 20% of molecules is observed for all amides; this has been attributed to a requirement for *cis-trans* proline isomerism in the folding of the hen enzyme (Kato et al., 1982) although site-directed mutagenesis experiments suggest that this may not be the origin of the slow phase in the human protein (Herning et al., 1991). It seems likely, therefore, that the fundamental character of the folding process is preserved but that sequence differences between the human and hen proteins cause significant changes in the stability and structures of the intermediates sampled on the folding pathway.

One of the most striking findings in this study is that folding of human lysozyme differs from that of hen lysozyme in that the α-domain does not fold to a protective state in a single cooperative step. In the hen protein, formation of persistent interactions in the hydrophobic core associated with the α-domain occurs on a time scale similar to the stabilization of elements of secondary structure in this domain as indicated by the kinetics of protection of the indole hydrogen of Trp111 [which does not form any specific hydrogen bond, but is presumably protected from exchange by burial from the solvent (Radford et al., 1992a,b)] and by fluorescence quenching of tryptophan residues (Itzhaki et al., 1994). The indole hydrogen of Trp112 in human lysozyme is also rapidly protected from exchange but because of the loss of cooperativity of the α-domain the interpretation of the behavior of this residue is particularly significant. Thus, although Trp112 is located in the D-helix, its indole hydrogen is protected from exchange much faster than the amides of this helix. In fact, its protection kinetics are very similar to those of the amides in the A and B helices and in the C-terminal 3_{10} helix. If the origin of protection in the intermediate is the same as in the native state, this suggests that the hydrophobic core involving the side chain of Trp112 is formed in the intermediate prior to the stabilization of the helix in which it is located. Interest-

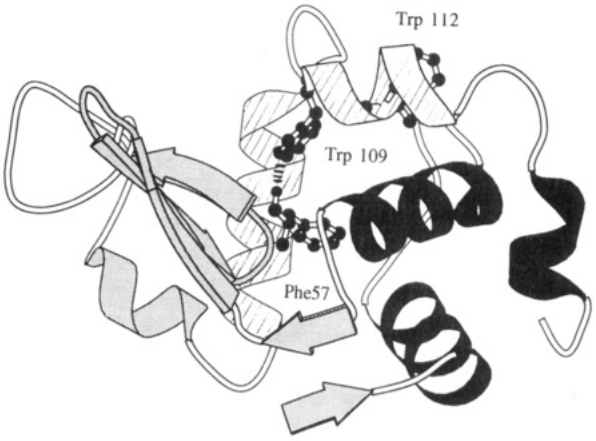

FIGURE 6: Schematic diagram of the native state of human lysozyme showing the two tryptophan indole probes followed in the pulse labeling experiment. Both Trp109 and Trp112 are shown to be buried in the hydrophobic core of the α-domain. The indole proton of Trp109 forms a hydrogen bond (dashed line) with the carbonyl group of Phe57. Areas shaded black represent those secondary structure elements believed to be involved in the initial subdomain. Striped areas represent regions protected more slowly while those secondary structure elements which have the lightest shading protect late on the folding pathway. Areas in white represent regions about which there is no information. The four native disulfide bonds are omitted from this diagram for clarity.

ingly, the indole hydrogen of Trp109 which lies only three residues from Trp112 is protected more slowly than the indole hydrogen of Trp112, with a rate more closely resembling that of the amides in the β-domain or in the C and D helices. The indole hydrogen of Trp 109 forms a hydrogen bond to the carbonyl group of Phe 57, which lies in the β-domain of the native protein, in the turn between the second and third strands of the β-sheet (Figure 6), thereby acting as a probe for the stabilization of interdomain interactions in the refolding molecules. This step clearly occurs subsequent to the protection of the indole hydrogen of Trp112 and presumably the formation of the hydrophobic core of the α-domain.

In both the hen and human proteins a large part of the far-UV CD signal develops rapidly, within the dead time (ca. 2 ms) of the experiment. In the case of the hen protein, reconstruction of the CD spectrum of this early intermediate has suggested that a native-like complement of secondary structure is already formed in this phase (Chaffotte et al., 1992). The slower steps in hen lysozyme (Figure 5e) involve the development of a signal apparently greater than that of the native state, and its subsequent loss on a slower time scale close to the rate of formation of the functional native protein (Dobson et al., 1994). These later phases may be associated with the existence of nonnative-like interactions in intermediates populated at these times, presumably involving disulfide bonds or possibly aromatic residues (Chaffotte et al., 1992). Its absence in human lysozyme [and indeed in hen lysozyme under other conditions (Chaffotte et al., 1992; C. M. Dobson, S. J. Eyles, S. E. Radford, P. A. Evans, and S. Gladwin, unpublished data)] suggests that it is not associated with interactions crucial for folding. Indeed, the faster folding rate of human lysozyme might indicate that such interactions do not exist or are reduced in importance in this protein compared with the hen protein. Interestingly, however, there is still a slow step in the development of the native far-UV CD signal in the human enzyme suggesting that at least for some of the molecules formation of secondary structure or rearrangement of tertiary interactions, perhaps involving disulfide bonds or aromatic residues, is rate limiting.

In the case of hen lysozyme the CD signal in the near-UV region develops much more slowly than in the far-UV CD. The data fit to a single exponential of time constant about 350 ms (Figure 5b). This is close to the time constant of the slowest phase of amide protection and has been shown to correspond to the formation of the fully functional native state indicating that the immobilization of tryptophan residues occurs in a single step as the native structure forms (Itzhaki et al., 1994). The intermediates in hen lysozyme, by inference therefore, do not have structures in which the tryptophan residues are highly ordered. In this regard, the data on human lysozyme are particularly interesting. At 270 nm, where the CD signal arises predominantly from tyrosine residues, the evolution of the ellipticity is well described by a single exponential of time constant (80 ms) consistent with the fixing of these residues in a unique environment as the native state is formed. At 293 nm, however, where tryptophan residues are thought to contribute predominantly to the signal, approximately 50% of the total CD change occurs within the dead time (ca. 2 ms) of the experiment. The remainder fits to a single-exponential curve of time constant ca. 80 ms. Two of the five tryptophans (28 and 112) are located at the center of the hydrophobic core in the α-domain of the enzyme, and Trp109 and Trp64 are largely buried in the α-domain and β-domain, respectively [i.e., less than 30% surface exposed (Kabsch & Sander, 1983; Rose & Dworkin, 1989)], while Trp34 is rather solvent exposed. It is tempting to suggest, therefore, that the rapid development of 50% of the native like CD signal at 293 nm reflects the formation of the core of the α-domain of the protein which appears to immobilize at least some of these tryptophan residues to an extent sufficient to generate a signal in the near-UV CD. The remaining CD signal develops on the same time scale as protection of the C and D helices and the β-domain indicating that the remaining tryptophans do not become fixed until the native structure forms. Thus, although we cannot be sure of the relative time scale of events within the dead time of folding, the earliest detectable intermediate appears to have developed in less than 2 ms a well-defined hydrophobic core in the α-domain, with the A and B helices and the C-terminal 3_{10} helix forming a particularly stable subdomain. In the slower steps, for the majority of molecules, the C and D helices pack on to this subdomain, and this is followed by protection of amides in the β-domain and formation of the native state. As with the hen protein, however, the apparent existence of multiple pathways indicates that, for some molecules at least, structuring of the α-domain may occur cooperatively and concomitantly with formation of the β-domain (Radford et al., 1992a).

With these experimental results, it is possible to speculate as to the origin of the differences between the folding behavior of hen and human lysozymes. First, the possibility was examined that the more rapid folding of the A and B helices and the C-terminal 3_{10} helix of the human protein might arise from a greater intrinsic helical propensity. Several models for predicting secondary structure propensity were used to scan the amino acid sequences of both lysozymes. The results for one of these employing a scale of helical propensities derived experimentally by host–guest experiments (Wojcik et al., 1990) is shown in Figure 7. Other models based on statistical probabilities of finding a particular amino acid in helices of proteins (Chou & Fasman, 1978) and the effect on the stability of a protein upon substituting amino acids involved in its helices (Horovitz et al., 1992) gave similar results. The figure clearly shows that regions which are helical in the native state have, in general, the highest propensity for helix formation. There

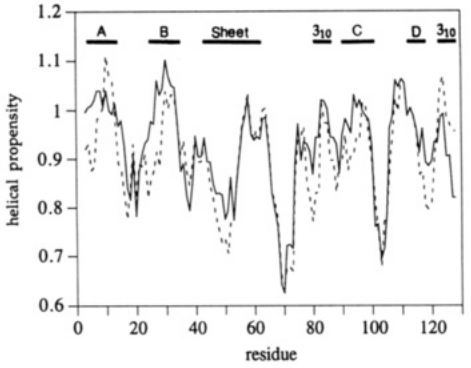

FIGURE 7: Helical propensity profiles for human (solid line) and hen (dashed line) lysozymes. Helical propensity (Wojcik et al., 1990) has been smoothed over five residues and is plotted against sequence number. The position of secondary structure elements is also indicated.

are, however, very few differences in the helical propensity profiles of the two enzymes. One of the more prominent differences, which is consistently observed with all the models tested, is the apparently slightly higher helical propensity of the B helix of human lysozyme, particularly in the N-terminal region. This can be attributed at least in part to the substitution of Gly26 in hen lysozyme by alanine in human lysozyme (O'Niel & Degrado, 1990; Horovitz et al., 1992). Other factors which affect the intrinsic helicity of polypeptide fragments such as the presence of N- and C-caps (Serrano & Fersht, 1989) and the possibility of favorable interaction of terminal residues with the helix dipole (Sali et al., 1988) are also highly conserved within the structures of the two enzymes and are unlikely to be critical determinants of the folding kinetics.

Although no major differences in the predicted helical propensities were found between hen and human lysozymes, an analysis of their native structures in terms of their tertiary packing shows that amino acid substitutions between the two enzymes lead to differences in contacts between individual secondary structural units, especially in the α-domain, which could influence the folding process. Analysis of the amino acid substitutions which occur between the hen and human proteins indicates that 33 out of the 51 substitutions are on the surface of the protein, and 18 in the interior (Figure 2). Although we cannot be sure that surface residues are not responsible for the differences between the folding of the two proteins, we have concentrated here on the internal substitutions. In order to speculate as to why the A and B helices and the C-terminal 3_{10} helix are stabilized independently, at least in the majority of molecules, in the folding pathway of human lysozyme, a comparison of the contacts between these helices was made. Figure 8 shows the three nonconserved buried residues at the A helix–B helix interface in the two proteins. The nature of these substitutions (Met12, Val29, and Phe38 in hen lysozyme and Leu12, Met29, and Tyr38 in human lysozyme) is, however, complementary such that the overall packing of the side chains in this interhelical region is apparently preserved. It therefore seems unlikely that the interactions between these helices differ significantly between the two proteins.

Despite the conclusions of this simple analysis, more detailed theoretical studies reveal some interesting correlations with the experimental results reported here. First, methods using a knowledge-based force field which considers the role of local interactions in the absence of long-range contacts (Rooman & Wodak, 1992) predicts that helix A in both hen and human lysozymes has a well-defined helical conformation which is much lower in energy than other calculated structures for this region. This indicates that its formation in the absence of tertiary contacts might be an important early folding event, a prediction in accord with our experimental findings. Second, a model which considers the surface area buried upon association of preformed fragments of secondary structure to be of critical importance (Chelvanayagam et al., 1992) predicts that the association of the A and B helices in hen lysozyme is highly favorable, followed by the C-terminal 3_{10} helix and then by the D and C helices. This predicted sequence of events is consistent with that observed in the pulsed hydrogen exchange experiments on human lysozyme. This suggests that burial of surface area is, therefore, a crucial determinant of the folding pathway.

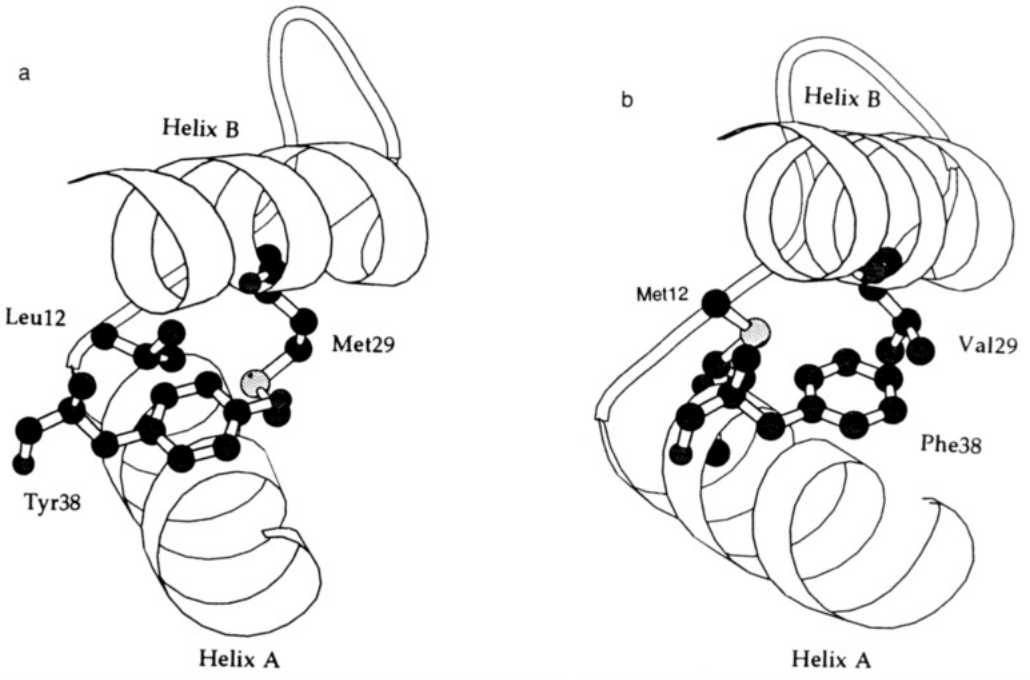

FIGURE 8: Schematic diagram showing the packing of the three nonconserved residues between helices A and B in (a) human and (b) hen lysozymes. The figure was draw using the program MolScript (Kraulis, 1991).

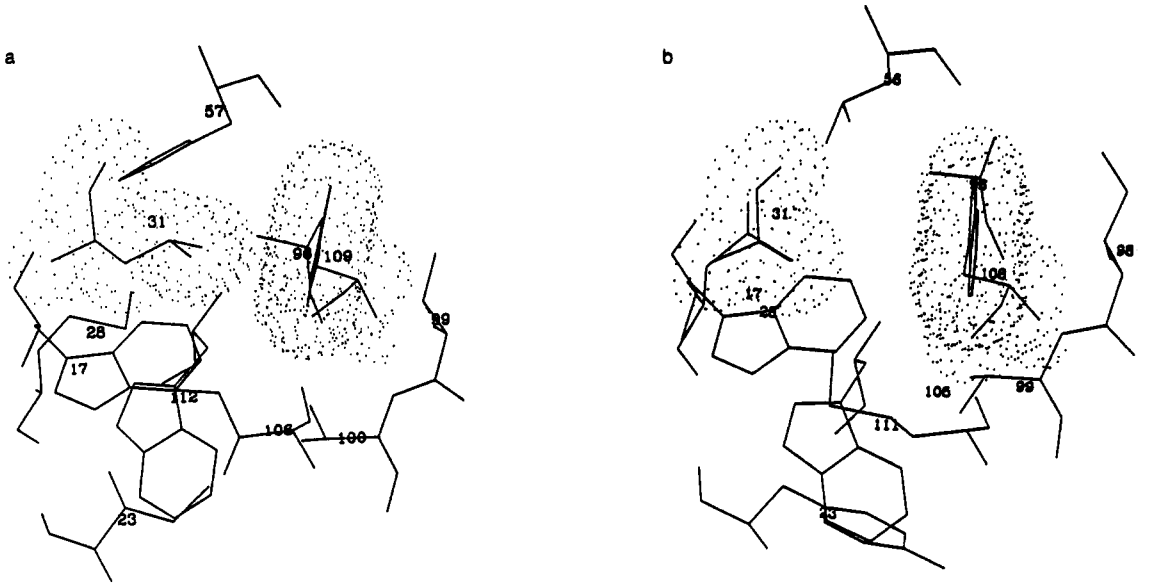

FIGURE 9: Hydrophobic core in the α-domain of (a) human and (b) hen lysozyme. The shaded area indicates the calculated Van der Waals surface areas of Leu31 and Trp109 in (a) and Ala31 and Trp108 in (b).

The refolding studies described here and elsewhere have provided strong evidence that formation of a hydrophobic core is amongst the first events on the folding pathway of a protein (Matouschek et al., 1992; Itzhaki et al., 1994). Figure 9 shows the hydrophobic residues involved in the core in the α-domain of hen and human lysozyme. Although the total hydrophobicity is, in general, conserved in the two proteins, there are several substitutions which cause subtle differences in the packing of residues in the hydrophobic core. The most interesting of these is the change of Ala31 in hen lysozyme to Leu31 in human lysozyme. Figure 9 shows that the packing of this residue against Trp108/109 is more efficient in human lysozyme; efficiency of packing in the cores of proteins has been shown to affect greatly the overall stability of a protein (Sandberg & Terwilliger, 1989, 1991; Serrano et al., 1992). If the net hydrophobic interactions in the early folding intermediate of human lysozyme are indeed greater than those in the hen intermediate, then this might explain why stabilization of this core occurs more rapidly in the human protein. If this core stabilizes elements of secondary structure in early intermediates, the extremely rapid development of this feature could explain the overall increase in the rate of folding of human lysozyme, and the lack of complete cooperativity within the α-domain. Interestingly, the mutation of Ala31 to Val31 has been carried out in hen lysozyme and was found to have a sufficiently large effect on the stability or folding behavior to prevent the mutant protein from being successfully expressed (Imoto et al., 1987).

Given the complexity of the problem, we have at present only attempted to explore global distinctions between the native protein structures and the role of specific factors including predictions of helical propensity and hydrophobicity patterns. Other factors may well be important; for example, it is possible that the slow steps in folding may involve structural rearrangements, perhaps resulting from a variety of topologies distinct from that of the native protein resulting from the coalescence of previously formed elements of secondary structure. Whatever the origin of the difference in the folding behavior of hen and human lysozymes, however, these results suggest that comparison of wild-type proteins from different species can be extremely valuable as an indication of key events in folding and to identify the role of interactions involving side chains in the folding process. By examination of a range of variants or specific mutants, it might then be possible to dissect the folding process of a given topology into a series of distinct steps and the influence of specific factors on the folding process then can more directly be established.

ACKNOWLEDGMENT

We thank Tony Willis for performing N-terminal sequence analysis, Carol Robinson for analysis of the protein by electrospray mass spectrometry, Dr. P. Argos for providing data for the loss of external surface area during folding, and Matthias Buck for performing the initial work on this project and for valuable discussions. The Oxford Centre for Molecular Sciences is funded jointly by the Science and Engineering Research Council and the Medical Research Council of the U.K.

SUPPLEMENTARY MATERIAL AVAILABLE

One table detailing the protection of amides during the refolding of human lysozyme (2 pages). Ordering information is given on any current masthead page.

REFERENCES

Anfinsen, C. B. (1973) *Science 181*, 223–230.
Artimuik, P. J., & Blake, C. C. F. (1981) *J. Mol. Biol. 152*, 737–762.
Bai, Y., Milne, J. S., Mayne, L., & Englander, S. W. (1993) *Proteins: Struct., Funct., Genet. 17*, 75–86.
Baldwin, R. L. (1990) *Nature 346*, 409–410.
Baldwin, R. L. (1993) *Curr. Opin. Struct. Biol. 3*, 84–91.
Beychock, S., Halper, J. P., Latovitzki, N., & Berstein, H. (1971) *Proc. Natl. Acad. Sci. U.S.A. 68*, 517–522.
Buck, M., Radford, S. E., & Dobson, C. M. (1993) *Biochemistry 32*, 669–678.
Chaffotte, A. F., Guillou, Y., & Goldberg, M. E. (1992) *Biochemistry 31*, 9694–9702.
Chelvanayagam, G., Reich, Z., Bringas, R., & Argos, P. (1992) *J. Mol. Biol. 227*, 901–916.
Chou, P. Y., & Fasman, G. D. (1978) *Annu. Rev. Biochem. 47*, 251–276.
Creighton, T. E. (1992a) *Protein Folding*, W. H. Freeman and Co., New York.
Creighton, T. E. (1992b) *Bioessays 14*, 195–199.
Dill, K. (1990) *Biochemistry 29*, 7134–7155.

Dobson, C. M., et al. (1993) *Curr. Opin. Struct. Biol. 3*, 57–118.
Dobson, C. M., Evans, P., & Radford, S. E. (1994) *Trends Biochem. Sci. 19*, 31–37.
Evans, P. A., & Radford, S. E. (1994) *Curr. Opin. Struct. Biol. 4*, 100–106.
Fersht, A. R., Matouschek, A., & Serrano, L. (1992) *J. Mol. Biol. 224*, 771–782.
Handoll, H. (1985) D.Phil. Thesis, Oxford University.
Herning, T., Yutani, K., Taniyama, Y., & Kikuchi, M. (1991) *Biochemistry 30*, 9882–9891.
Horovitz, A., Matthews, J. M., & Fersht, A. R. (1992) *J. Mol. Biol. 227*, 560–568.
Imoto, T., Johnson, L. N., North, A. C. T., Phillips, D. C., & Rupley, J. A. (1972) *The Enzymes* (Boyer, P. D., Ed) 3rd ed., Vol. 7, pp 665–868, Academic, New York.
Imoto, T., Yamada, H., Yasukochi, T., Yamada, E., Ito, Y., Ueda, T., Nagatani, H., Miki, T., & Horiuchi, T. (1987) *Protein Eng. 1*, 333–338.
Itzhaki, L. S., Radford, S. E., Dobson, C. M., & Evans, P. A. (1994) *Biochemistry* (in press).
Kabsch, W., & Sander, C. (1983) *Biopolymers 22*, 2577–2637.
Kato, S., Shimamoto, N., & Utiyama, H. (1982) *Biochemistry 21*, 38–43.
Kim, P. S., & Baldwin, R. L. (1990) *Annu. Rev. Biochem. 59*, 631–660.
Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946–950.
Kuwajima, K. (1989) *Proteins: Struct., Funct., Genet. 6*, 87–103.
Matouschek, A., Serrano, L., & Fersht, A. R. (1992) *J. Mol. Biol. 224*, 819–835.
Matthews, C. R., & Zitzewitz, J. A. (1993) *Curr. Opin. Struct. Biol. 3*, 594–600.
Miranker, A., Dobson, C. M., Radford, S. E., & Karplus, M. (1991) *Nature 349*, 633–636.
Miranker, A. D., Robinson, C. V., Radford, S. E., Aplin, R. T., & Dobson, C. M. (1993) *Science 262*, 896–900.
O'Niel, K. T., & Degrado, W. F. (1990) *Science 250*, 646–651.
Pedersen, T. G., Sigurskjold, B. W., Andersen, K. V., Kjaer, M., Poulsen, F. M., Dobson, C. M. ,& Redfield, C. (1991) *J. Mol. Biol. 218*, 413–426.
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. (1992a) *Nature 358*, 302–307.
Radford, S. E., Buck, M., Topping, K. D., Dobson, C. M., & Evans, P. A. (1992b) *Proteins 14*, 237–248.
Redfield, C., & Dobson, C. M. (1988) *Biochemistry 27*, 122–136.
Redfield, C., & Dobson, C. M. (1990) *Biochemistry 29*, 7201–7214.
Roder, H., Elove, G. A., & Englander, S. W. (1988) *Nature 335*, 700–704.
Rooman, M. J., & Wodak, S. J. (1992) *Biochemistry 31*, 10239–10249.
Rose, G. D., & Dworkin, J. E. (1989) *Prediction of Protein Structure and the Principles of Protein Conformation* (Fasman, G. D., Ed.) pp 625–634, Plenum press New York.
Sali, D., Bycroft, M., & Fersht, A. R. (1988) *Nature 335*, 740–743.
Sandberg, W. S., & Terwilliger, T. C. (1989) *Science 245*, 54–57.
Sandberg, W. S., & Terwilliger, T. C. (1991) *Trends Biotechnol. 9*, 59–63.
Serrano, L., & Fersht, A. R. (1989) *Nature 342*, 296–299.
Serrano, L., Kellis, J. T., Cann, P., Matouschek, A., & Fersht, A. R. (1992) *J. Mol. Biol. 224*, 783–804.
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.
Weissman, J. S., & Kim, P. S. (1991) *Science 253*, 1386–1393.
Weissman, J. S., & Kim, P. S. (1992) *Science 256*, 111–114.
Wojcik, J., Altman, K. H., & Scheraga, H. A. (1990) *Biopolymers 30*, 121–134.