28. Takahashi, E. *J. geophys. Res.* **91,** 9367–9382 (1986).
29. Ito, E. & Takahashi, E. *Nature* **328,** 514–517 (1987).
30. Herzberg, C. *J. Geophys. Res.* **95,** 15 779–15 803 (1990).
31. Herzberg, C. T. *Earth planet. Sci. Lett.* **67,** 249–260 (1983).
32. White, R. S. in *Oceanic and Continental Lithosphere, J. Petrol. spec. Lithosphere issue* (ed. Menzies, M. A. & Cox, K. G.) 1–10 (1988).
33. Jordan, R. H. in *Oceanic and Continental Lithosphere, J. Petrol. spec. Lithosphere issue* 11–38 (1988).
34. Galer, S. J. G. *Earth planet. Sci. Lett.* **105,** 214–228 (1991).
35. Ito, E., Harris, D. M. & Anderson, A. T. *Geochim. cosmochim. Acta* **47,** 1613–1624 (1983).
36. Tatsumi, Y. *J. geophys. Res.* **94,** 4697–4707 (1989).
37. Davies, J. H. & Stevenson, D. H. *J. geophys. Res.* **97,** 2037–2070 (1992).
38. Peacock, S. M. *Phil. Trans. R. Soc.* A**335,** 341–353 (1991).
39. Green, D. H. *Earth planet. Sci. Lett.* **19,** 37–53 (1973).
40. Wyllie, P. J. in *Magmatic Processes: Physiochemical Principles,* Geochem. Soc. spec. publ. 1 (ed. Mysen, B. O.) 107–119 (1987).
41. Liu, L. *Phys. Earth planet. Inter.* **49,** 142–167 (1987).
42. Kanzaki, M. *Phys. Earth planet. Inter.* **66,** 307–319 (1991).
43. Ito, E. & Weidner, D. *Geophys. Res. Lett.* **13,** 464–466 (1986).
44. Eggler, D. H. *Carnegie Inst. Washington Yb.* **72,** 464–467 (1973).
45. Hodges, F. N. *Carnegie Inst. Washington Yb.* **72,** 495–497 (1973).
46. Kato, T. & Kumazawa, M. *J. geophys. Res.* **91,** 9351–9355 (1986).
47. Stolper, E., Walker, D., Hager, B. H. & Hays, J. F. *J. geophys. Res.* **86,** 6261–6271 (1981).
48. Yamamoto, K. & Akimoto, S. *Am. J. Sci.* **277,** 288–312 (1977).
49. Liu, L. G. *Tectonophys.* **164,** 41–48 (1989).
50. Tera, F. *et al. Geochim. cosmochim. Acta* **50,** 1535–1550 (1986).
51. Kesson, S. E. & Ringwood, A. E. *Chem. Geol.* **78,** 83–96 (1989).
52. Akimoto, S. & Akaogi, M. *Phys. Earth planet. Int.* **23,** 268–275 (1980).
53. Ringwood, A. E. & Major, A. *Earth planet. Sci. Lett.* **2,** 130–133 (1967).
54. Ahrens, T. J. *Nature* **342,** 122–123 (1989).
55. Liu, L. *Phys. Earth planet. Int.* **42,** 255–262 (1986).
56. Luth, R. W. *EOS* **72,** 199 (1991).
57. Gasparik, T. *J. geophys. Res.* **95,** 15 751–15 769 (1990).
58. Meade, C. & Jeanloz, R. *Nature* **251,** 68–72 (1991).
59. Kushiro, I., Syono, Y. & Akimoto, S. I. *Earth planet. Sci. Lett.* **3,** 197–203 (1967).
60. Sudo, A. & Tatsumi, Y. *Geophys. Res. Lett.* **17,** 29–32 (1990).
61. Heubner, J. S. & Papike, J. J. *Am. Miner.* **55,** 1973–1992 (1970).
62. Trønnes, R. G., Takahashi, E. & Scarfe, C. M. *EOS* **69,** 1510 (1988).
63. Trønnes, R. G. *EOS* **71,** 1587 (1990).
64. Modreski, P. T. & Boettcher, A. L. *Am. J. Sci.* **273,** 385–414 (1973).
65. Foley, S. *Geochim. cosmochim. Acta* **55,** 2689–2694 (1991).
66. Hart, R., Dymond, J. & Hogan, L. *Nature* **278,** 156–159 (1979).
67. O'Nions, R. K. & Oxburgh, E. R. *Nature* **306,** 429–431 (1983).
68. Jochum, K. P., Hofmann, A. W., Ito, E., Seufert, H. M. & White, W. M. *Nature* **306,** 431–436 (1983).
69. Navon, O., Hutcheon, I. D., Rossman, G. R. & Wasserburg, G. J. *Nature* **335,** 784–789 (1988).
70. Menzies, M. A. *Nature* **335,** 769–770 (1988).
71. Bailey, D. K. *Nature* **296,** 525–530 (1982).
72. Harlow, G. E. & Veblen, D. R. *Science* **251,** 652–655 (1991).
73. Bell, D. R. & Rossman, G. R. *Science* **255,** 1391–1396 (1992)
74. Smyth, J. R., Bell, D. R. & Rossman, G. R. *Nature* **351,** 732–735 (1991).
75. Young, T. E., Green, H. W., Hofmeister, A. M. & Walker, D. *EOS* **72,** 144 (1991).
76. Clemens, J. D. & Vielzeuf, D. *Earth planet. Sci. Lett.* **86,** 287–306 (1987).
77. Ryabchikov, I. D. & Boettcher, A. L. *Am. Miner.* **65,** 915–919 (1980).
78. Schneider, M. E. & Eggler, D. H. *Geochim. cosmochim. Acta* **50,** 711–724 (1986).
79. Massone, H. J. & Schreyer, W. *Contr. Miner. Petrol.* **96,** 212–224 (1987).
80. Ringwood, A. E. *Chem. Geol.* **82,** 187–207 (1990).
81. Finger, L. W. & Prewitt, C. T. *Geophys. Res. Lett.* **16,** 1395–1397 (1989).
82. Finger, L. W., Hazen, R. M. & Prewitt, C. T. *Am. Miner.* **76,** 1–7 (1991).
83. Kudoh, Y., Finger, L. W., Hazen, R. M., Prewitt, C. T. & Kanzaki, M. *Carnegie Inst. Washington Yb.* **89,** 89–91 (1989).
84. Robie, R. A., Hemingway, B. & Fisher, J. R. *U.S. Geol. Surv. Bull.* **1452,** (1978).
85. Helgeson, H. C., Delaney, J. M., Nesbitt, N. W. & Bird, D. K. *Am. J. Sci.* **278–A** (1978).
86. Anderson, D. L. *Theory of the Earth* (Blackwell, Boston, 1989).

ACKNOWLEDGEMENTS. I thank M. Kanzaki, C. Herzberg and T. Gasparik for communications about their work, Q. Williams, K. Collerson, G. Ranalli and P. Ulmer, for discussions, J. Holloway, S. Peacock and R. Sweeney for comments, B. Buehlmann for final drafting, U. Stidwill for typing and the Schweizerische Nationalfonds for financial support.



# The folding of hen lysozyme involves partially structured intermediates and multiple pathways

**Sheena E. Radford*, Christopher M. Dobson* & Philip A. Evans†**

* Oxford Centre for Molecular Sciences and Inorganic Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QR, UK
† Cambridge Centre for Molecular Recognition and Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, UK

**Analysis of the folding of hen lysozyme shows that the protein does not become organized in a single cooperative event but that different parts of the structure become stabilized with very different kinetics. In particular, in most molecules the $\alpha$-helical domain folds faster than the $\beta$-sheet domain. Furthermore, different populations of molecules fold by kinetically distinct pathways. Thus, folding is not a simple sequential assembly process but involves parallel alternative pathways, some of which may involve substantial reorganization steps.**

UNDERSTANDING how a globular protein folds requires detailed characterization of partially organized intermediates formed during the folding process. Information about such species is becoming available through studies of disulphide formation in oxidative refolding[1,2], protection from hydrogen exchange[3–6], the effects of amino-acid substitutions[7–9], stable partially folded states[10–13] and peptide analogues of folding intermediates[14,15]. Nonetheless, no clear unifying view of the nature of folding mechanisms has yet emerged and uncertainties remain over several fundamental issues.

Here we describe a detailed study of the refolding of hen lysozyme. The thermodynamics and kinetics of folding of this protein under equilibrium conditions accord well with a cooperative two-state model[16,17]. Under conditions far from equilibrium, however, refolding occurs in at least two stages, acquisition of a native-like far ultraviolet (UV) circular dichroism (CD) spectrum preceding formation of the tertiary structure, as monitored by near-UV CD[18,19]. In addition, a very fast phase has been detected by tryptophan absorption measurements[20].

These results suggest transient population of partially folded intermediates but their detailed interpretation requires complementary information about the behaviour of individual residues, such as can be obtained from hydrogen-exchange labelling techniques[5,6,21]. Using a variant of these experiments[3,4], based on competition between the exchange and folding processes, we have recently confirmed the existence of folding intermediates for hen lysozyme and demonstrated that the two structural lobes that characterize the native structure are also distinct folding domains[22]. Here we describe pulsed hydrogen-exchange labelling experiments, using nearly half of the 126 amide hydrogens in the molecule as probes of refolding, and stopped-flow CD measurements in both far- and near-UV

spectral regions. The combination of these data has allowed us to construct a detailed analysis of events during refolding.

## Evidence for partially folded intermediates

Figure 1 shows the results of pulsed labelling experiments obtained for individual amides, grouped according to their distribution in the native state secondary structure. If folding occurred by a simple two-state mechanism all the curves would be coincident. The fact that this is not the case demonstrates unequivocally that partially structured intermediates are populated during folding.

The different labelling curves are not only distinct for the different sets of amides, they are also not monophasic. Each curve was modelled well by a sum of two exponentials (Fig. 1 and Table 1). The rates of the fast phases show no very clear pattern (time constant ($\tau$) = 7 ± 4 ms) but those of the slower phases fall qualitatively into two groups, differing in their average time constant by a factor of about 4. The more rapidly protected group ($\tau$ < 100 ms) comprises amides in the four $\alpha$-helical segments, the $3^{10}$ helix close to the C terminus of the protein, and three amides, Trp 63, Cys 64 and Ile 78, which lie in the loop region in the native enzyme. With the exception of the last group, these structural elements all occur in one of the two lobes of the native conformation, which we shall call the $\alpha$-domain (Fig. 2). In contrast, amides protected more slowly ($\tau$ > 100 ms) are located, with the single exception of Asn 27, in the other structural domain, which we designate the $\beta$-domain, comprising a short double-stranded and a longer triple-stranded $\beta$ sheet, a $3^{10}$ helix and a long loop.

The same qualitative distinction also extends to the amplitudes of the kinetic phases. For amides in the $\alpha$-domain, the fast phase constitutes some 40 ± 12% of the total whereas in the $\beta$-domain the corresponding amplitude is only 24 ± 6%. This compounds the difference in kinetics of the second phase to ensure that protection of the $\beta$-domain is substantially slower overall, thus accounting for the pattern of labelling observed when the refolding of lysozyme was examined by the competition method[22].

## Tertiary interactions and folding domains

The similar protection curves for nearly all amides in the $\alpha$-domain could, in principle, mean that individual helices form as persistent structures independently but simultaneously. However, the coincidence of both kinetic phases makes this seem unlikely. A more plausible explanation is that the protection from exchange monitors stabilization of the rapidly formed helices by side-chain interactions, and hence that these amide probes actually monitor formation of the domain core, albeit perhaps in an embryonic form. This is consistent with the protection of two amides not directly involved in secondary structure, Leu 17 and Tyr 23 which lie in the loop linking helices

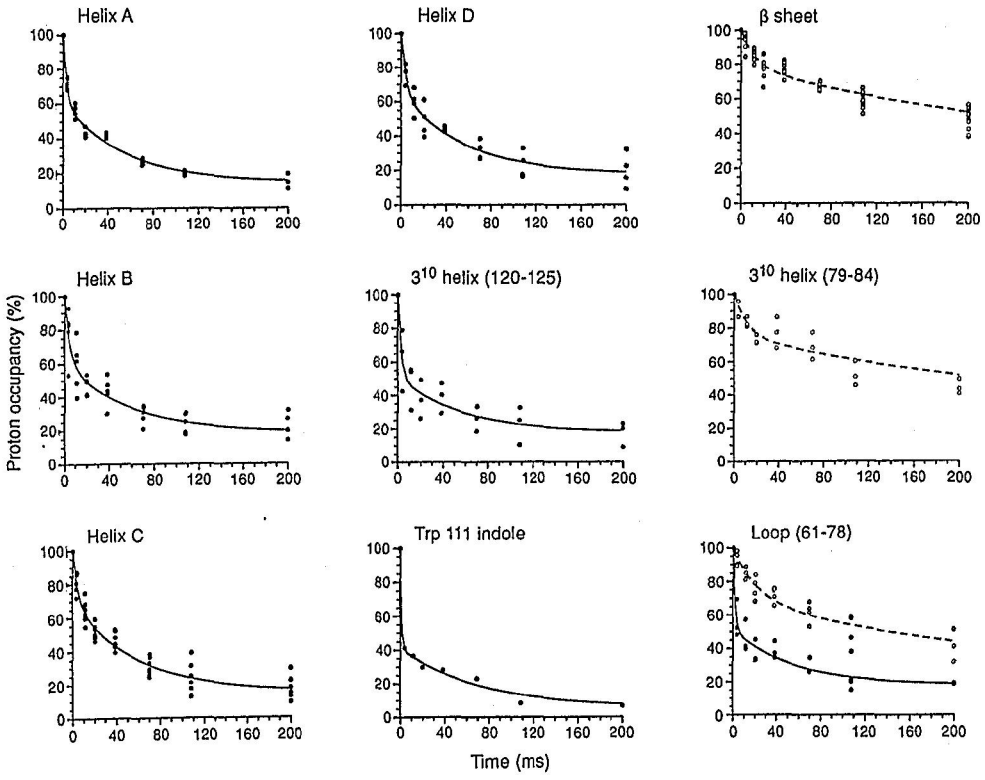


FIG. 1 Time courses for the protection of amides from exchange during the refolding of hen lysozyme. The curve drawn represents the average of a two exponential fit to the time course for individual amides in each native secondary structural element. In the case of the loop region, two averages are drawn, one for Trp 63, Cys 64 and Ile 78 and another for the remainder.

METHODS. Experiments were done at 20 °C using a Biologic QFM5 rapid mixing quench flow apparatus. Lysozyme (20 mg ml$^{-1}$) was initially dissolved in 6 M guanidine deuterochloride in $D_2O$ at pH 6.0, leading to complete denaturation and substitution of all exchangeable hydrogens by deuterium. Refolding was initiated by dilution of this solution 10-fold into 20 mM sodium acetate, pH 5.5 in $H_2O$. At the resulting pH of 5.2 the half life for amide exchange is about 16 s (ref. 32) so that negligible labelling occurred during this phase. After variable refolding times (3.5–2,000 ms) the solution was diluted again with a volume 5 times that of the initial protein solution of 0.2 M sodium borate, pH 10.0. This initiated labelling at a pH of 9.5. After 8.4 ms the labelling pulse was terminated by further dilution, with a volume again 5 times that of the initial volume of protein solution, of 0.5 M acetic acid in $H_2O$. The final pH was about 4.0, at which exchange in the native protein of the 49 amides studied is very slow[26]. A sixfold lower protein concentration during refolding had no effect on the observed rates of protection, confirming that intermolecular associations, if present, are not kinetically significant. Protein samples were concentrated and the buffer exchanged for 40 mM deuterated sodium acetate, pH 3.8 in $D_2O$ by ultrafiltration at 4 °C. For the zero time point (100% labelling) a solution of lysozyme in the same buffer and isotopic composition as in the labelling phase of the refolding experiment was heated to 80 °C for 10 min, causing complete exchange of all amides by reversible unfolding. The sample was then treated identically as for the other time points. This made correction for the residual deuterium content of the labelling solution unnecessary. A phase-sensitive COSY spectrum of each sample was recorded on a 500 MHz GE/Nicolet spectrometer at 35 °C. 256 increments over 1,024 complex points and 32 transients were collected. Data were processed using the FTNMR program of Dennis Hare (Hare Research Inc.) on a SUN computer. Final digital resolution was 1.7 Hz per point. Acquisition and processing of each spectrum was identical. The intensities of the C$\alpha$H–NH cross-peaks in each spectrum were taken as the sum of the absolute values of the four phase-sensitive components. These were normalized using the sum of the eight phase-sensitive peaks associated with the Tyr 23 and Tyr 53 C$\delta$H–C$\varepsilon$H cross-peaks. Proton occupancies at individual amide sites were calculated relative to those of the unfolded control sample, for which COSY cross-peak intensities were taken to correspond to 100% proton occupancy.

TABLE 1 Protection of amide hydrogens during the refolding of lysozyme

| Residue | Fast phase Amplitude (%) | Fast phase Time constant (ms) | Slow phase Amplitude (%) | Slow phase Time constant (ms) | Structural context |
|---|---|---|---|---|---|
| 8 | 47.2 | 3.90 | 38.3 | 69.8 | Helix A |
| 10 | 39.5 | 4.41 | 50.6 | 64.3 | |
| 11 | 43.6 | 3.39 | 41.5 | 54.1 | |
| 12 | 33.9 | 2.61 | 51.5 | 53.4 | |
| 13 | 44.5 | 6.33 | 44.0 | 79.1 | |
| 17 | 43.6 | 9.09 | 39.7 | 89.5 | Irregular |
| 23 | 48.5 | 7.76 | 35.9 | 83.3 | |
| 27 | 66.3 | 12.7 | 22.5 | 492 | Helix B |
| 28 | 42.1 | 4.83 | 42.1 | 59.7 | |
| 29 | 58.3 | 2.40 | 28.1 | 59.4 | |
| 31 | 36.9 | 4.39 | 48.6 | 68.7 | |
| 34 | 40.3 | 8.00 | 45.5 | 97.7 | |
| 36 | 24.5 | 6.68 | 53.5 | 53.1 | |
| 37 | 19.4 | 11.5 | 56.8 | 475 | Irregular |
| 38 | 16.6 | 11.2 | 69.7 | 350 | |
| 39 | 18.2 | 9.33 | 56.1 | 296 | |
| 40 | 22.2 | 7.14 | 59.6 | 451 | Small $\beta$-sheet |
| 42 | 26.5 | 8.76 | 66.0 | 531 | Large $\beta$-sheet |
| 44 | 14.2 | 6.29 | 48.2 | 151 | |
| 50 | 24.0 | 10.6 | 40.4 | 188 | |
| 52 | 33.1 | 9.29 | 52.7 | 238 | |
| 53 | 25.6 | 8.80 | 62.4 | 349 | |
| 56 | 17.7 | 15.7 | 63.7 | 375 | |
| 58 | 17.8 | 16.8 | 67.4 | 392 | |
| 61 | 17.5 | 9.29 | 58.5 | 247 | Loop (irregular) |
| 63 | 60.0 | 2.45 | 25.4 | 63.9 | |
| 64 | 43.2 | 3.67 | 37.4 | 60.4 | |
| 65 | 39.7 | 15.7 | 44.7 | 241 | |
| 75 | 23.5 | 23.5 | 63.8 | 354 | |
| 76 | 25.4 | 8.60 | 54.2 | 215 | |
| 78 | 59.2 | 1.87 | 21.0 | 48.8 | |
| 82 | 26.2 | 8.56 | 65.6 | 788 | $3^{10}$ Helix |
| 83 | 29.0 | 9.53 | 52.5 | 282 | |
| 84 | 26.1 | 9.81 | 48.3 | 262 | |
| 92 | 29.6 | 3.54 | 56.6 | 52.8 | Helix C |
| 93 | 34.9 | 9.11 | 35.5 | 76.0 | |
| 94 | 22.2 | 4.93 | 57.7 | 64.0 | |
| 95 | 28.8 | 3.29 | 61.0 | 63.5 | |
| 96 | 19.4 | 1.43 | 74.8 | 53.5 | |
| 97 | 32.7 | 7.08 | 51.7 | 61.6 | |
| 99 | 28.1 | 2.07 | 56.4 | 40.5 | |
| 108 | 34.9 | 3.73 | 47.5 | 87.5 | Helix D |
| 111 | 48.0 | 3.89 | 38.1 | 68.4 | |
| 112 | 28.2 | 3.04 | 47.9 | 49.8 | |
| 115 | 44.7 | 8.13 | 42.7 | 51.6 | |
| 123 | 46.7 | 3.42 | 35.0 | 65.2 | $3^{10}$ Helix |
| 124 | 67.0 | 2.00 | 17.8 | 33.3 | |
| 125 | 43.1 | 4.95 | 40.4 | 108 | |

The time courses of change in proton occupancies were fitted to the sum of two exponentials of the form $y = A\,e^{(-k_1 t)} + B\,e^{(-k_2 t)} + C$ (where $A$ and $B$ are the fractional amplitudes of the two phases and $K_1$ and $K_2$ are their rate constants; $C$ is the apparent fractional amplitude of a third phase too slow to be followed in our experiments). The last column gives the secondary structural context of each amide in the native state.

A and B, with kinetics indistinguishable from those of amides in the helices (Table 1).

It is clear that the $\alpha$-domain acquires a stabilizing core when, in most molecules, the $\beta$-domain has not become fixed in a stable conformation. To investigate further the structure of the $\alpha$-domain at this stage we have studied the protection of tryptophan indole NHs, two of which exchange slowly enough in the native enzyme to be used as probes of refolding[23]. The kinetics of protection of Trp 111 resemble closely those of the main chain amides of the $\alpha$-domain (Fig. 1). In the native structure this residue is in the D-helix and the side-chain NH is hydrogen bonded to the side-chain oxygen of Asn 27, in the B helix[24]. Thus, its protection may monitor 'docking' of these two helices; at the very least it suggests that the tryptophan side-chain becomes excluded from solvent in a well developed hydrophobic core on this time scale. Similar behaviour was observed qualitatively for the indole NH of Trp 28 which is also buried in the core of the $\alpha$-domain.

The amide NH of Asn 27, which forms a non-helical hydrogen-bond to the carbonyl oxygen of Tyr 23 in the native structure, is the one marked exception to the general pattern of behaviour in the $\alpha$-domain. Its protection kinetics resemble those of amides in the $\beta$-domain, with a slow phase time constant greater than 400 ms. This interaction may thus become fixed only late in folding, when both domains are organized. Therefore, although in most molecules the $\alpha$-domain folds independently, at least some details of its tertiary interactions form more slowly.

Within the $\beta$-domain the situation is more complicated. With the exceptions of Trp 63, Cys 64 and Ile 78, there is a clear qualitative distinction between the protection kinetics of amides in this domain and in the $\alpha$-domain but the slow phase time constants for individual amides vary between 150 and nearly

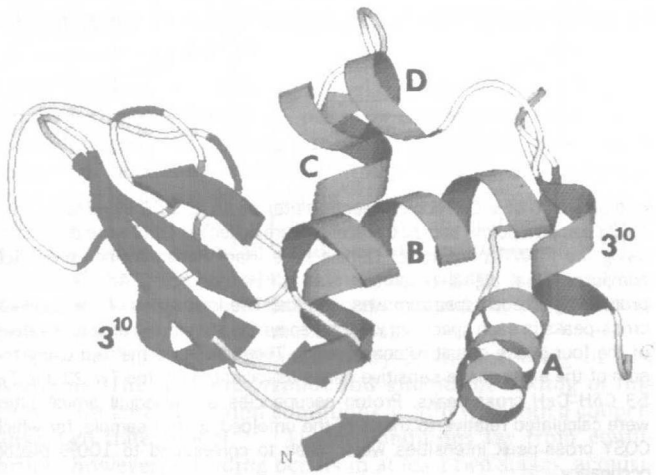


FIG. 2. Schematic view of the native structure of hen lysozyme. Elements of secondary structure in the $\alpha$-domain are coloured red, as are three amides protected with similar kinetics, Trp 63, Cys 64 and Ile 78. Elements of secondary structure that are in the $\beta$-domain are coloured blue. White regions represent mainly surface amides for which there is no information. The four $\alpha$ helices (A–D) and the two $3^{10}$ helices are labelled. The diagram was produced using the program MolScript[43].

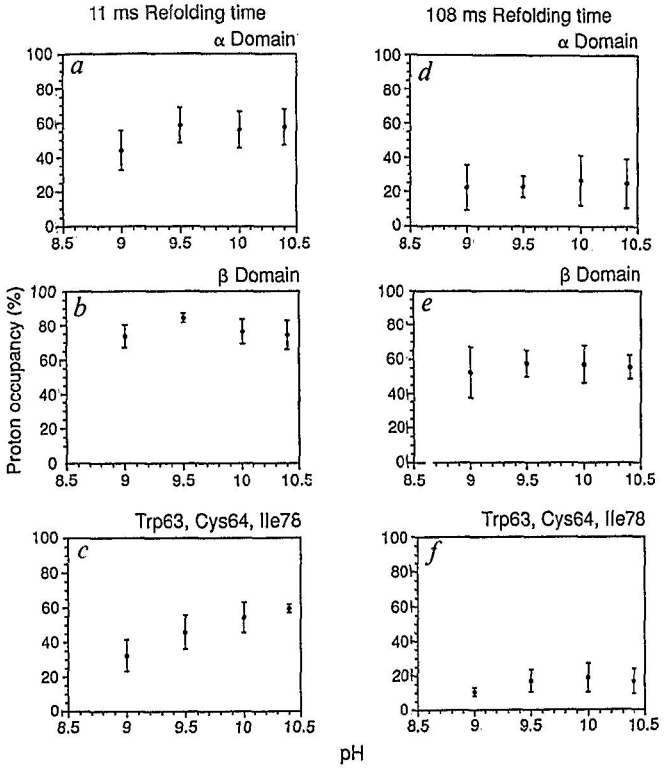


FIG. 3 The pH dependence of the extent of labelling during an 8.4 ms pulse commencing 11 ms and 108 ms after the initiation of folding. Exchange in proteins usually follows EX2 kinetics, in which the intrinsic rate of the chemical exchange step is slow compared with the rate of the structural fluctuations which expose the amide transiently to the solvent[44,45]. In this situation, if the partial protection achieved in the fast phase reflects a single population of a marginally stable intermediate, then the extent of labelling would be predicted to increase substantially as the pH of the labelling pulse is raised[27,39]. If, by contrast, the mechanism of exchange is EX1, where the overall rate is limited by the rate of the fluctuations that expose the amide to solvent water, the pH dependence of the extent of labelling, if any, would not be predictable in a straightforward manner. This ambiguity can be resolved by increasing the duration, rather than the pH, of the labelling pulse[5,6]. This would result in increased labelling of any marginally stable intermediates regardless of the mechanism of exchange. In this figure the average proton occupancies for amide hydrogens in the $\alpha$-domain (*a*, *d*), the $\beta$-domain (*b*, *e*) and the subset containing Trp 63, Cys 64 and Ile 78 (*c*, *f*) are shown as a function of the pulse pH. Vertical bars represent one standard deviation from the average plotted. The pH of the labelling pulse was varied between 9.0 and 10.4 by changing the pH of the exchange buffer. For these experiments the quench buffer was 1 M acetic acid in $H_2O$ solution. Acquisition, processing and analysis of NMR spectra were done as described in the legend to Fig. 1. Control experiments in which a sample of the native protein (90% deuterated at amide sites) was exposed to the different labelling pulses showed no increase in proton occupancy at the 49 sites considered here, confirming that these amides are stable to exchange in the native enzyme under these conditions.

800 ms (Table 1). This implies that the stabilizing tertiary interactions in this domain are not formed in a single cooperative step but that a number of individual or sequential assembly processes are involved.

Protection of the three amides, Trp 63, Cys 64 and Ile 78, which lie in the long loop at the interface between the $\alpha$- and $\beta$-domains cannot be rationalized by examination of the native structure; Cys 64 is not hydrogen bonded and Ile 78 forms only a long tertiary hydrogen-bond[24]. Thus, their protection cannot be associated with native secondary structure but, presumably, with exclusion from solvent by surrounding side-chains[25,26]. These interactions need not be native-like. There are two disulphide bridges close by (64–80 and 76–94), the latter linking the $\alpha$- and $\beta$-domains. Early clustering of these residues in cooperation with formation of the $\alpha$-domain core could explain the similarity in their protection kinetics.

## Multiplicity of events in folding

There are two fundamentally different possible explanations for biphasic kinetics in the labelling experiment: the fast phase leads either to partial protection of a particular amide in all molecules or to complete protection but in only a proportion of molecules. The first case implies formation of a marginally stable intermediate, followed by slower transformation into a stable, presumably native-like, structure. The second possibility involves parallel pathways such that a proportion of the protein folds rapidly to a structure that completely protects the amide concerned from exchange, whereas the remainder follows a route in which protective structure is acquired more slowly.

These possibilities can be distinguished by varying the duration and the intensity of the labelling pulse[5,6,27]. The results (Fig. 3) indicate that the labelling of the majority of amides in both domains is essentially independent of the pulse pH. In addition, no significant change in the extent of labelling occurs when the pulse length is doubled or trebled (at refolding times of 22 and 108 ms, respectively). These results are not compatible with a simple sequential folding mechanism; there must be parallel folding pathways, some leading to protection of individual amides more rapidly than others.

The only exceptions to the observed insensitivity to the labelling conditions are the amides of Trp 63, Cys 64 and Ile 78 (Fig. 3). This suggests that rapidly formed structure protects these amides only partially, full protection being acquired in the slower phase. However, the sensitivity to the pulse conditions is weaker than expected for a single marginally stable species. Instead, there must be a heterogeneous population, with varying levels of protection.

## Native and non-native structure

The sequence of events detected in the far-UV CD (225 nm) is characterized by at least three kinetic phases (Fig. 4). A large negative ellipticity is acquired within the dead time of the experiment (2 ms), so that at the earliest measurable time its magnitude is close to that observed for the native enzyme. The CD at this wavelength is generally dominated by peptide groups in helical structure, suggesting that an average helix content close to that of the native enzyme is achieved in this time. Contributions to the CD from aromatic groups and disulphides cannot, however, be ruled out[28].

There follows a remarkable further development of the dichroism at 225 nm, so that the average negative ellipticity becomes greater than that of the native state, reaching a maximum at around 80 ms. The kinetics of this phase do not coincide closely with any of the phases of exchange protection we have observed. It seems likely that the excursion in the CD reflects the transient formation of some kind of non-native interactions although we cannot rule out the possibility that it arises from asynchronous development of effects that tend to cancel out in the native protein spectrum. Return of the ellipticity to its value in the native state then occurs in a third phase ($\tau \sim 300$ ms) which is much slower than protection of $\alpha$-helical amides but is comparable to the slow phase of protection in the $\beta$-domain.

CD experiments were also done in the near-UV region. The latter monitors aromatic groups in specific orientations fixed through tertiary interactions, so that spectra of denatured proteins, even relatively compact 'molten globule' states, are characterized by loss of nearly all intensity at these wavelengths[17,18]. The time development at 289 nm is quite different from that in the far-UV (Fig. 4). There is virtually no change within the dead time of the experiment and the entire kinetic amplitude can be

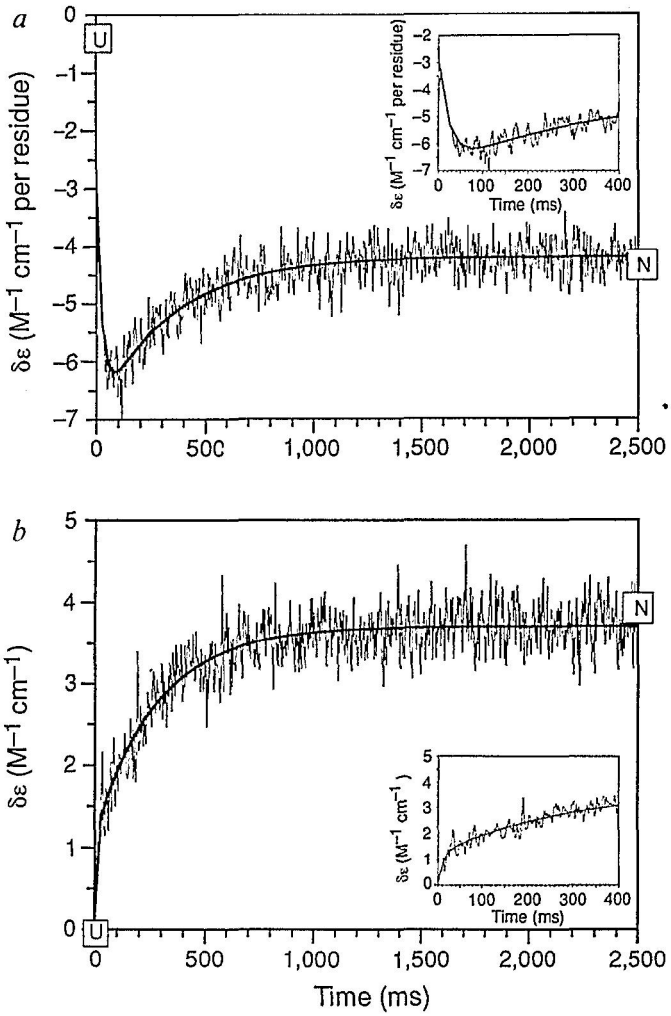


FIG. 4 Refolding kinetics of lysozyme as monitored by stopped flow CD at *a*, 225 and *b*, 289 nm. The refolding conditions were as described in the legend to Fig. 1, except that the final folding protein concentrations were 0.2 and 1.0 mg ml$^{-1}$, respectively. The experiments were done using a Jobin-Yvon CD6 circular dichrograph equipped with a Biologic SFM3 stopped-flow module. The dead time was about 2 ms. In each case the data have been fitted to a sum of two exponentials. The insets show an expansion of the plots during the first 400 ms of folding.

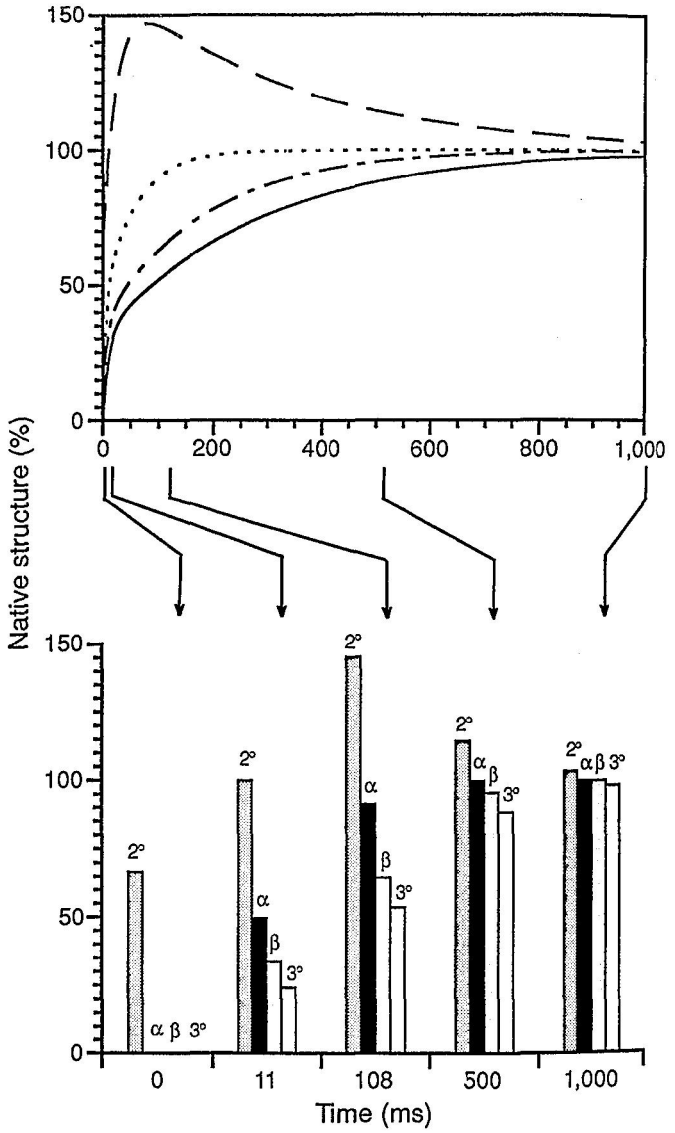


FIG. 5 Summary of events during the refolding of lysozyme. Formation of the native enzyme is virtually complete in 2 s (ignoring the very slow phase that was not recorded in our experiments) and each parameter is therefore assigned a value of 100% at this point. The curves shown are the fits to the near (289 nm; solid line) and the far (225 nm; dashed line) UV CD data and protection of the $\alpha$- (dotted line) and the $\beta$-domains (dashed and dotted line), respectively. The latter two curves represent the averages of two exponential fits to the time courses for individual amides in each folding domain. For various different folding times the magnitude of each parameter as a percentage of that observed in the native protein is also shown: $CD_{225}$, Ellipticity at 225 nm; 3°, ellipticity at 289 nm; $\alpha$, protection of amides in the $\alpha$-domain; $\beta$, protection of amides in the $\beta$-domain.

fitted to a sum of two exponentials, the amplitudes and time constants being $32 \pm 2\%$, $\tau = 10 \pm 2$ ms, and $68 \pm 2\%$, $\tau = 285 \pm 8$ ms. These parameters resemble those observed for protection of amides in the $\beta$-domain (Table 1).

The dichroism at 289 nm is associated principally with the six tryptophan residues of lysozyme[29]. Of these, residues 62 and 63 are in an irregularly structured part of the $\beta$-domain, close to the domain interface, and the others are all in the $\alpha$-domain. This is interesting in view of the apparent synchrony of their forming fixed tertiary contacts and the protection of amides in the $\beta$-, rather than the $\alpha$-domain. Thus, although the $\alpha$-domain can fold more rapidly than, and therefore independently of, the $\beta$-domain, the results suggest that its tertiary structure does not become fully organized until cooperative interactions spanning the two domains are established.

## Folding pathways

The earliest event observed in refolding is the acquisition within 2 ms of a far-UV CD of similar intensity to that of the native protein. This almost certainly reflects formation of a large amount of $\alpha$-helical structure, at least some of it presumably native-like, yet no protection from hydrogen exchange is apparent at this stage (Fig. 5). This suggests that although the average distribution of backbone conformations has become far from random, specific structural elements remain extremely labile. Similar behaviour has been observed in other early folding intermediates[30] and denatured states[26,31,32]. This result is consistent with the near-UV CD which detects no specific interactions involving aromatic chromophores at this time. A change in the environment of these groups on this time scale has, however, been observed by stopped-flow absorption studies[18,19], suggesting that these rapid folding events include some kind of hydrophobic collapse as well as helix formation.

In the next phase there is rapid protection of $\alpha$-domain amides in about 50%, and of $\beta$-domain amides in about 30%, of molecules. The protection data cannot reveal directly whether $\beta$-domain assembly occurs independently or in conjunction with that of the $\alpha$-domain in individual molecules in this phase.

However, the presence of a phase with a similar time constant in the near-UV CD which, as discussed above, may reflect the formation of cooperative structure spanning both domains, suggests that roughly synchronous structure formation in both domains occurs in at least a substantial proportion of these molecules.

The remainder of the population folds more slowly and the greater amplitude of the fast phase and higher rate of the slower phase of protection for $\alpha$-domain amides mean that in most molecules this domain becomes substantially folded before formation of a stable $\beta$-domain. In the resulting intermediates the pH dependence of labelling shows that for most amides in the $\alpha$-domain the protection factors are in excess of 500 whereas those in the $\beta$-domain are less than 20. However, the slower development of near-UV CD intensity and of protection of at least one amide, Asn 27, in the $\alpha$-domain suggest that it is not fully native-like but has some characteristics of a molten globule at this intermediate stage. The pattern of protection has some resemblance to that observed in the stable molten globule state of the homologous protein guinea-pig $\alpha$-lactalbumin. In that case, fewer than 20 amides have protection factors greater than 10; again these occur in the $\alpha$-, rather than the $\beta$-domain (C.-L. Chyan, C. Wormald, C.M.D., P.A.E. and J. Baum, manuscript in preparation). Similar, marginal protection is also characteristic of other stable partially folded proteins[11,12,32], suggesting that these are, in general, more labile structures than the kinetic intermediate observed in the present experiment.

The negative ellipticity in the far-UV CD, already substantially developed during the dead time, continues to grow during the first 80 ms or so of refolding, becoming substantially greater than that in the native enzyme spectrum. The origin of this signal cannot be determined from the present data but it is likely that it arises from some kind of non-native interactions. Reconstruction of CD spectra at these refolding times can provide further insight into this issue (M. E. Goldberg, personal communication). A similar overshoot has also been observed during the folding of the all-$\beta$ protein, $\beta$-lactoglobulin[33]. Return of the ellipticity towards its native value occurs at a similar rate to that of amide protection in the $\beta$-domain and the development of the near-UV CD; these are the slowest phases we observe (Fig. 5). This suggests that although the $\alpha$-domain can fold to a large extent independently, in most of the population the conformation of the remainder of the molecule, though not necessarily wholly disordered, is far from native-like until it becomes fixed in the cooperative tertiary structure.

A major unresolved issue is why different molecules fold by kinetically distinct pathways. This could reflect heterogeneity of the unfolded state[20,34,35]; for example native-like X-Pro peptide bond isomers may be required for folding[5,27,36-39]. It is possible that this could indeed account for the lack of protection in around 17% of molecules even after 2 s of refolding, but neither the rates nor amplitudes of the distinct phases resolved in our experiments are consistent with proline isomerization. The multiple pathways could reflect other heterogeneity in the denatured state or they may arise from an initial rapid collapse of the protein, leading to various intermediates, some amenable to further folding whereas others need first to reorganize, potentially a relatively slow process in a partially condensed state. A similar observation of parallel refolding pathways has been made for cytochrome *c* (ref. 40).

The four disulphide bridges persist in the denatured state and might have a significant influence on the refolding mechanism, either through their isomerism in the denatured state or in intermediates, or through constraints they place on conformational freedom during refolding. In $\alpha$-lactalbumin persistence of substantial residual structure in its molten globule state is consistent with a number of different disulphide pairings[41], suggesting that at least some intermediates can form independently of particular cross-links. Further experiments will be needed to assess the significance of disulphide bonds in the refolding of lysozyme.

These results have enabled us to develop a rather detailed picture of the folding pathway of hen lysozyme. CD and exchange labelling experiments have proved powerful and complementary structural probes, revealing, for example, the very rapid formation of helices before specifically stabilized interactions are detectable. The refolding kinetics are complex, in particular because there is evidently more than one folding pathway. An important general feature, not observed in other proteins studied by pulsed labelling methods[5,6,21,42], is the division of the protein into distinct folding domains which, in most molecules at least, are stabilized asynchronously. □

Received 27 February; accepted 19 June 1992.


1. Creighton, T. E. *Prog. Biophys. molec. Biol.* **33,** 231–297 (1978).
2. Weissman, J. S. & Kim, P. S. *Science* **253,** 1386–1393 (1991).
3. Schmid, F. X. & Baldwin, R. L. *J. molec. Biol.* **135,** 199–215 (1979).
4. Roder, H. & Wüthrich, K. *Proteins Struct. Funct. Genet.* **1,** 34–42 (1986).
5. Udgaonkar, J. B. & Baldwin, R. L. *Nature* **335,** 694–699 (1988).
6. Roder, H., Elöve, G. A. & Englander, S. W. *Nature* **335,** 700–704 (1988).
7. Matouschek, A., Kellis, J. T., Serrano, L., Bycroft, M. & Fersht, A. R. *Nature* **346,** 440–445 (1990).
8. Hughson, F. M., Barrick, D. & Baldwin, R. L. *Biochemistry* **30,** 4113–4118 (1991).
9. Klemm, J. D., Wozniak, J. A., Alber, T. & Goldenberg, D. P. *Biochemistry* **30,** 589–594 (1991).
10. Baum, J., Dobson, C. M., Evans, P. A. & Hanley, C. *Biochemistry* **28,** 7–13 (1989).
11. Hughson, F. M., Wright, P. E. & Baldwin, R. L. *Science* **249,** 1544–1548 (1990).
12. Jeng, M.-F., Englander, S. W., Elove, G. A. Wand, J. A. & Roder, H. *Biochemistry* **29,** 10433–10437 (1990).
13. Harding, M. M., Williams, D. H. & Woolfson, D. N. *Biochemistry* **30,** 3120–3128 (1991).
14. Oas, T. G. & Kim, P. S. *Nature* **336,** 42–47 (1988).
15. Staley, J. P. & Kim, P. S. *Nature* **334,** 685–688 (1990).
16. Tanford, C. *Adv. Prot. Chem.* **24,** 1–95 (1970).
17. Dobson, C. M. & Evans, P. A. *Biochemistry* **23,** 4267–4270 (1984).
18. Kuwajima, K., Hiraoka, Y., Ikeguchi, M. & Sugai, S. *Biochemistry* **24,** 874–881 (1985).
19. Ikeguchi, M., Kuwajima, K., Mitani, M. & Sugai, S. *Biochemistry* **25,** 6965–6972 (1986).
20. Kato, S., Okamura, M., Shimamoto, N. & Utiyama, H. *Biochemistry* **20,** 1080–1085 (1981).
21. Bycroft, M., Matouschek, A., Kellis, J. T. Serrano, L. & Fersht, A. R. *Nature* **346,** 488–490 (1990).
22. Miranker, A., Radford, S. E., Karplus, M. & Dobson, C. M. *Nature* **349,** 633–636 (1991).
23. Wedin, R. E., Delepierre, M., Dobson, C. M. & Poulsen, F. M. *Biochemistry* **21,** 1098–1103 (1982).
24. Handoll, H. thesis, Univ. Oxford (1985).
25. Pederson, T. G. *et al. J. molec. Biol.* **218,** 413–426 (1991).
26. Radford, S. E., Buck, M., Topping, K. D., Evans, P. A. & Dobson, C. M. *Proteins Struct. Funct. Genet.* (in the press).
27. Udgaonkar, J. B. & Baldwin, R. L. *Proc. natn. Acad. Sci. U.S.A.* **87,** 8197–8201 (1990).
28. Manning, M. C. & Woody, R. W. *Biochemistry* **28,** 8609–8613 (1989).
29. Goux, W. J. & Hooker, T. M. Jr *Biopolymers* **19,** 2191–2208 (1980).
30. Elöve, G. A., Chaffotte, A. F., Roder, H. & Goldberg, M. E. *Biochemistry* (in the press).
31. Robertson, A. D. & Baldwin, R. L. *Biochemistry* **30,** 9907–9914 (1991).
32. Jeng, M.-F. & Englander, S. W. *J. molec. Biol.* **221,** 1045–1061 (1991).
33. Sugawara, T., Kuwajima, K. & Sugai, S. *Biochemistry* **30,** 2698–2706 (1991).
34. Ropson, I. J., Gordon, J. I. & Frieden, C. *Biochemistry* **29,** 9591–9599 (1990).
35. Kuwajima, K., Garvey, E. P., Finn, B. E. Matthews, C. R. & Sugai, S. *Biochemistry* **30,** 7693–7703 (1991).
36. Brandts, J. F., Halvorson, H. R. & Brennan, M. *Biochemistry* **14,** 4953–4963 (1975).
37. Kato, S., Shimamoto, N. & Utiyama, H. *Biochemistry* **21,** 38–43 (1982).
38. Schmid, F. X. *Biochemistry* **22,** 4690–4696 (1983).
39. Jackson, S. E. & Fersht, A. R. *Biochemistry* **30,** 10436–10443 (1991).
40. Elöve, G. A. & Roder, H. *Protein Refolding, ACS Symp. Series* **470,** 50–63 (1991).
41. Ewbank, J. J. & Creighton, T. E. *Nature* **350,** 518–520 (1991).
42. Briggs, M. S. & Roder, H. *Proc. natn. Acad. Sci. U.S.A.* **89,** 2017–2021 (1992).
43. Kraulis, P. J. *J. appl. Crystallogr.* **24,** 946–950 (1991).
44. Hvidt, A. A. & Nielsen, S. O. *Adv. Prot. Chem.* **21,** 287–386 (1966).
45. Englander, S. W. & Kallenbach, N. R. *Q. Rev. Biophys.* **16,** 521–655 (1984).


ACKNOWLEDGEMENTS. We thank S. G. Waley, M. E. Goldberg, R. W. Broadhurst and A. Miranker for discussions and A. Berry for help with molecular graphics. We acknowledge support by the Royal Society, the Nuffield Foundation, the Isaac Newton Trust and the SERC and MRC through the Oxford Centre for Molecular Sciences and the Cambridge Centre for Molecular Recognition. S.E.R. is a Royal Society 1983 University Research Fellow.