Bipartite structure of the α-lactalbumin molten globule

Lawren C. Wu, Zheng-yu Peng and Peter S. Kim

Molten globules are thought to be general intermediates in protein folding. Apparently conflicting studies have failed to clarify whether one of the best characterized molten globules, that of α-lactalbumin, resembles an expanded native-like protein or a nonspecific collapsed polypeptide. Here we show that the molten globule properties of α-lactalbumin are largely confined to one of its two domains. The α-helical domain forms a helical structure with a native-like tertiary fold, while the β-sheet domain is largely unstructured. Molten globules thus possess a native-like backbone topology, but this topology does not necessarily encompass the entire polypeptide chain. Our studies indicate that molten globules provide an approximate solution to, and considerable simplification of the protein folding problem.

Howard Hughes Medical Institute, Whitehead Institute for Biomedical Research, Department of Biology, Massachusetts Institute of Technology, Nine Cambridge Center, Cambridge, Massachusetts 02142, USA

Correspondence should be addressed to P.S.K.

Molten globules are partially folded forms of proteins which are thought to be general protein folding intermediates and may also be important in the folding and processing of proteins *in vivo*[1–6]. Nonetheless, despite numerous studies, the structure of molten globules and their significance to protein folding remain unclear. The best characterized molten globule is that of α-lactalbumin[7–15] (α-LA), a protein comprised of two domains. Studies of the molten globule of α-LA have failed to resolve whether molten globules resemble expanded native-like proteins or nonspecific collapsed polypeptides. A model of the isolated α-helical domain of α-LA (α-Domain) in the molten globule form shows a strong preference for the native disulphide pairings[15], indicating that molten globules have a native-like backbone topology. However, the molten globule of intact α-LA (with all eight cysteines present) fails to show a preference for the native disulphide pairings[13,14], leading to the contradictory conclusion that molten globules have no preferred backbone topology. Resolution of this issue has important consequences for understanding protein folding.

In order to clarify this apparent discrepancy, we have produced and studied two variants of human α-LA (Fig. 1*a*), allowing us to probe the conformational preferences of each domain in the molten globule form. The species denoted α-LA(α) contains the two disulphide bonds in the α-helical domain, 6–120 and 28–111, but lacks the β-sheet domain and the interdomain cysteines, which are replaced by alanines. Conversely, α-LA(β) contains the 61–77 disulphide bond in the β-sheet domain and the interdomain 73–91 disulphide bond, with the cysteines in the α-helical domain replaced by alanines.

Structural characterization

Both α-LA(α) and α-LA(β) are molten globules in the absence of calcium, even at neutral pH. Each variant is a monomer at concentrations below 100 μM (to within ±5 %) as determined by sedimentation equilibrium at pH 8.5 (see Methods). The far-UV (Fig. 1*b*) and near-UV (Fig. 1*c*) circular dichroism (CD) spectra of each variant resemble that of the pH 2 molten globule form of α-LA (referred to as the A-state), and differ significantly from that of native α-LA. Thermal denaturation of each variant lacks a cooperative transition, resembling instead the transition observed for the A-state of α-LA (Fig. 1*d*). Moreover, the near-UV CD signals of α-LA(α) and α-LA(β) are weak compared to those of the native protein, indicating an absence of extensive, tight side-chain packing.

Backbone topology

In order to assess the tertiary topology of the α-helical and β-sheet domains in the molten globule of α-LA, we performed disulphide exchange studies. At equilibrium the populations of disulphide species reflect the probability of forming specific disulphide pairings and thus reflect the backbone topology of the domain. In our experiments, identical populations of disulphide species were achieved regardless of the starting disulphide isomer, indicating that equilibrium was established. Control experiments carried out under denaturing conditions yield equilibrium populations that are in good agreement with a random walk model (Fig. 2).

Under native conditions, rearrangement of α-LA(α) yields an equilibrium population containing predominantly the native disulphide species (Fig. 2*a*). This sug-

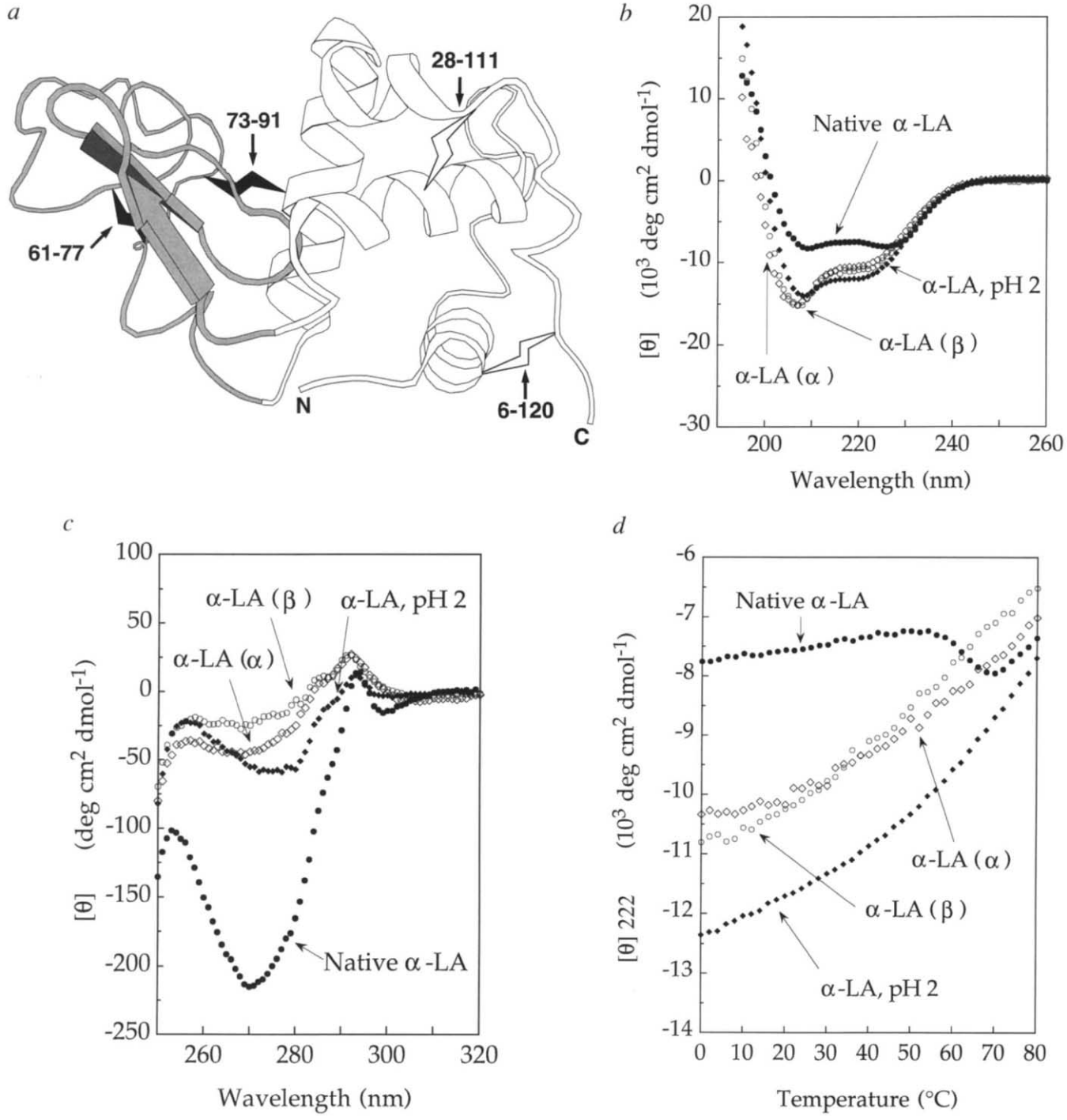

Fig. 1 α-LA(α) and α-LA(β) resemble the molten globule A-state of α-LA and differ from native α-LA. *a*, Schematic representation of human α-LA[34] produced with the program RIBBON[35]. The α-helical domain (white) consists of residues 1–37 and 86–123, and contains four α-helices. The β-sheet domain (shaded) consists of residues 38–85, and contains two short β-strands and several loop structures. α-LA(α) contains the two disulphide bonds in the α-helical domain (white), with the β-sheet domain and interdomain cysteines replaced by alanines. α-LA(β) contains the disulphide bond in the β-sheet domain and the interdomain disulphide bond (black), with the cysteines in the α-helical domain replaced by alanines. *b*, Far-UV circular dichoism spectra (CD). The CD spectra indicate that each variant contains approximately 35% helix, close to the value expected if the helices in the α-helical domain are structured. c, Near-UV CD spectra. *d*, Thermal denaturation monitored at 222 nm.

gests that the α-helical domain of α-LA strongly prefers a native tertiary fold in the molten globule, consistent with previous studies of α-Domain[15], a model for the isolated α-helical domain in the molten globule form. Our results also indicate that the preference of α-Domain for native disulphide pairings is not a result of the three glycine residues in the model that substitute for the β-sheet domain, since α-LA(α) shows the same equilibrium preferences.

In contrast, rearrangement of α-LA(β) under native conditions yields significant quantities of non-native species (Fig. 2*b*). This stongly suggests that the β-sheet domain is predominantly disordered, with only a slight preference for the native disulphide pairings compared with a random walk model (see Fig. 2).

These conclusions regarding the tertiary topology of each domain are confirmed by CD spectroscopy of the non-native two-disulphide species. The non-native isomers in the α-helical domain have significantly reduced helical content (Table 1) as compared to that of the A-state molten globule of α-LA. Moreover, the helical content of each non-native disulphide species is less than that of reduced α-LA(α), indicating that formation of non-native backbone topologies, imposed by non-native disulphide pairings, destabilizes helical structure. Similar observations have been seen with α-Domain[15].

In contrast, the native and non-native disulphide isomers in the β-sheet domain have identical far-UV and near-UV CD spectra (Table 1, Fig. 1). The thermal denaturation transitions for all the β-sheet domain variants are also superimposable (data not shown). These results are consistent with the β-sheet domain being largely unstructured.

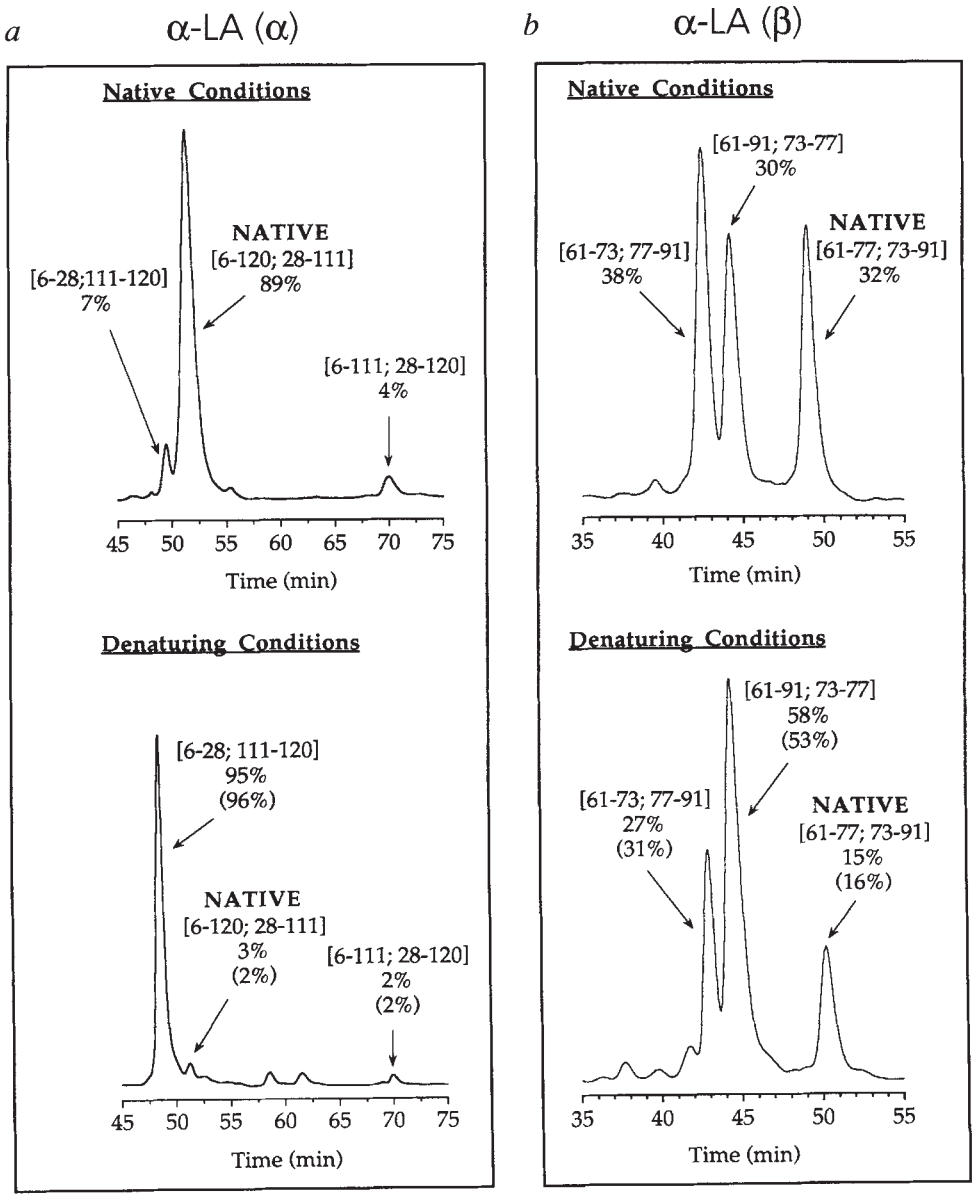

Fig. 2 Disulphide exchange studies of α–LA(α) and α-LA(β): *a*, α-LA(α) under native and denaturing conditions; *b*, α-LA(β) under native and denaturing conditions. The expected equilibrium populations (calculated with a random walk model[36]) for α-LA(α) and α-LA(β) under denaturing conditions are given in parentheses in the figure. The numbers in brackets denote the disulphide-bonded residues in each species.

Bipartite structure

Our results, taken together with previous work[10,11,15] lead to a picture of a bipartite α-LA molten globule in which the α-helical domain resembles an expanded native-like protein and the β-sheet domain is predominantly unfolded. The stark difference in global chain topology between the two domains in the α-LA molten globule provides a likely resolution to the apparent discrepancy between the studies of the α-domain[15] and intact α-LA[13,14] molten globules. Although the α-helical domain of α-LA strongly prefers native disulphide pairings, this preference may well be obscured in the molten globule of intact α-LA (which contains all eight cysteines) by distribution amongst several equally favorable β-sheet domain disulphide pairings. In addition, we have found that, in the α-LA molten globule, formation of the 28–111 disulphide bond is strongly preferred, while the 6–120 disulphide bond is labile[16]. Thus, disulphide exchange involving the 6–120 disulphide bond and the disordered β-sheet domain may further obscure a preference for native disulphide pairings in the α-helical domain.

The bipartite structure we observe is quite different from the previously observed structures of 'highly ordered molten globules'[17,18].

Studies of 'highly ordered molten globules' indicate that these species are comprised of native-like secondary structure elements interspersed locally with disordered loops. The bipartite structure of the α-LA molten globule is, however, a global property of the polypeptide backbone comprising the structural domains of α-LA, irrespective of local structure.

α-LA is structurally homologous to lysozyme. Hydrogen exchange studies indicate that the two domains of lysozyme fold on kinetically distinct time scales, leading to an intermediate containing a predominantly folded α-helical domain and a relatively unstructured β-sheet domain[19,20]. Our studies suggest that the corresponding kinetic folding intermediate of α-LA[21–23] contains a molten globule α-helical domain and an unstructured β-sheet domain.

Table 1 α-Helix content of native and non-native disulphide isomers of α-LA(α) and α-LA(β) as determined by CD spectroscopy

α-LA Species	$[\theta]_{222}$ (deg cm² dmol⁻¹)	Helix Content
α-LA(α)		
[6-120; 28-111] (Native)	-10,300	35%
[6-111; 28-120]	-7,100	24%
[6-28; 111-120]	-4,900	17%
Reduced α-LA(α)	-7,400	25%
α-LA(β)		
[61-77; 73-91] (Native)	-10,800	37%
[61-73; 77-91]	-10,800	37%
[61-91; 73-77]	-10,900	37%
Reduced α-LA(β)	-9,300	32%

The numbers in square brackets denote the disulphide bonded residues in each species. Under the experimental conditions (pH 8.5), the cysteine residues in reduced α-LA(α) and reduced α-LA(β) will be partially ionized. Since the α-helical domain is structured while the β-sheet domain is not, it is possible that the difference in $[\theta]_{222}$ between reduced α-LA(α) and reduced α-LA(β) results from unfavourable interactions associated with burying charged groups in the folded α-helical domain.

Implications for protein folding

Our results indicate that molten globule properties need not encompass the entire polypeptide chain, but can be attained independently by individual domains. This complements the observation that individual domains of multi-domain proteins can fold independently[24,25,37]. One of the major questions of protein folding is how a polypeptide chain can fold quickly and efficiently to a unique structure[26]. Stepwise folding through a molten globule intermediate may deter global misfolding.

Unlike the NMR spectra of 'highly ordered molten globules', which are amenable to structure determination[17,18], the NMR spectra of the A-state of α-LA[1,9,10] and our α-LA variants (data not shown) contain broad resonances and poor chemical shift dispersion. These characteristics indicate an absence of extensive, tight-packing interactions, and suggest that molten globules such as the A-state of α-LA correspond to earlier folding intermediates than 'highly ordered molten globules'. Nevertheless, NMR experiments indicate that native-like elements of secondary structure exist within the α-helical domain of the α-LA molten globule[10,11]. Our results suggest that these secondary structures are arranged with a native-like tertiary fold. Thus, it seems likely that proteins can adopt a native-like tertiary fold early in the folding process, even in the absence of extensive specific tertiary interactions and side-chain packing.

By adopting a native-like backbone topology, molten globules achieve much of the information transfer of the folding process, providing an approximate solution to the protein folding problem. This simplifies considerably the search for a unique folded conformation and reduces the energy barriers for minor structural rearrangements (Fig. 3). Our results sug-

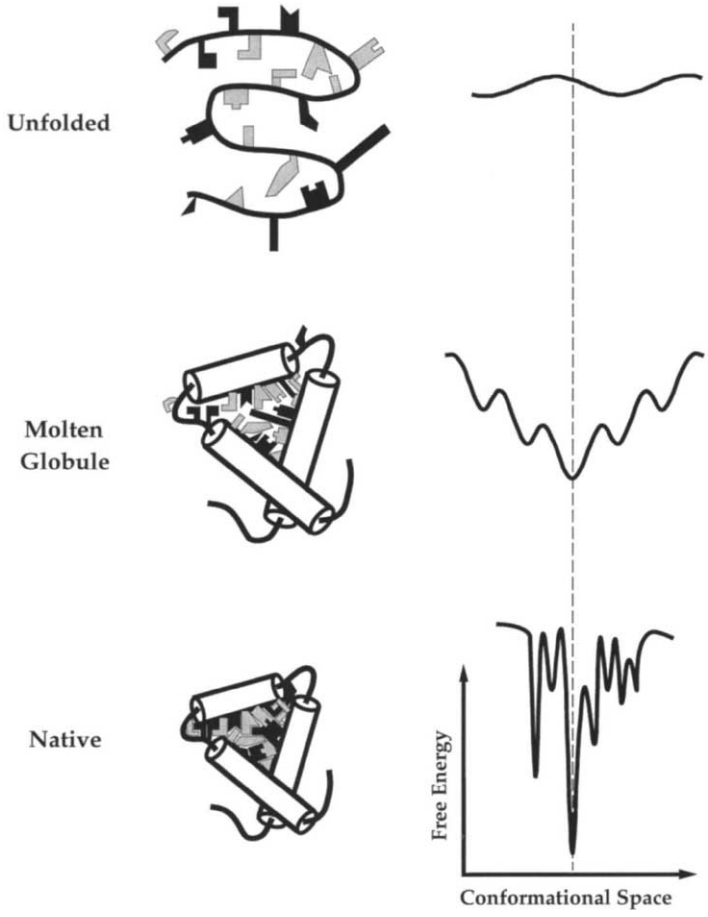

Fig. 3 Schematic diagrams of the free energy landscape for an unfolded polypeptide, the molten globule, and the native protein. Formation of the molten globule with a native tertiary fold greatly simplifies the conformational search for the native state and reduces the energy barriers for minor structural rearrangements.

gest that a key to understanding protein folding is to resolve how a polypeptide chain acquires a native-like backbone topology in the absence of extensive specific tertiary contacts.

Methods

Production of α-LA Variants. α-Lactalbumin variants were produced by oligonucleotide-directed mutagenesis[27] of an expression plasmid for human α-LA in the cloning vector pAED4 (ref. 28), called pALA. The gene was synthesized in eight segments on an Applied Biosystems DNA synthesizer, and ligated into the *Bam*HI/*Nde*I restriction site in pAED4. Mutations were verified by sequencing the entire α-LA gene. Each protein species was expressed in *Escherichia coli* BL21 DE3 pLysS (ref. 29) and purified as described for α-Domain[15]. Protein identity was confirmed by laser desorption mass spectrometry (Finigan Lasermat), and purity was assessed by analytical reverse-phase HPLC. All proteins were completely oxidized, as assayed by the lack of reaction with 5, 5'-dithio-bis(2-nitrobenzoic acid)[30] in 6 M guanidine HCl. Disulphide bonds were assigned by digestion with pepsin[15], followed by analytical reverse-phase HPLC. Peptide fragments with altered elution times upon reduction with dithiothreitol (DTT) were purified and identified by mass spectrometry, N-terminal sequencing, and amino acid analysis. The following fragments were identified for each species. α-LA(α) : (i) Phe 3–Leu 11/ Trp 118–Leu 123 (disulphide 6–120) (ii) Ile 27–Thr 33/Trp 104–Cys 111 (disulphide 28–111). α-LA(β) : (i) Trp 60–Asn 71/Ile 75–Leu 81 (disulphide 61–77) (ii) Trp 60–Asp 82/Ile 89–Ile 95 (disulphides 61–77 and 73–91). All non-native two-disulphide species except α-LA [6–111; 28–120] were prepared from the native two-disulphide species by disulphide exchange in 6 M guanidine HCl, 10 mM Tris, 0.5 mM EDTA, pH 8.5, using 25 μM protein, 226 μM reduced glutathione, and 50 μM oxidized glutathione. α-LA [6-111; 28-120] was prepared by disulphide exchange in 1 M guanidine HCl, 10 mM Tris, 0.5 mM EDTA, pH 8.5. Each species was isolated and purified by reverse-phase HPLC. The following disulphide fragments were identified for each non-native isomer. [61–73; 77–91] : (i) Trp 60–Asp 74 (disulphide 61–73) (ii) Ile 75–Leu 81/Ile 89–Ile 95 (disulphide 77–91). [61–91; 73–77] : (i) Trp 60–Asn 71/ Ile 89–Asp 97 (disulphide 61–91) (ii) Ile 72–Phe 80 (disulphide 73–77). [6–28; 111–120] : (i) Thr 4–Leu 11/Ile 27–Phe 31 (disulphide 6–28) (ii) Leu 105–Leu 123 (disulphide 111–120). [6–111; 28–120] : (i) Phe 3–Ser 9/Trp 104–Glu 113 (disulphide 6–111) (ii) Ile 27–Phe 31/Leu 119–Leu 123 (disulphide 28–120).

Circular dichroism (CD) spectroscopy. CD spectroscopy was performed on an Aviv 62DS spectrometer equipped with a thermoelectric temperature controller. Samples were dissolved in 10 mM Tris, 0.5 mM EDTA, pH 8.5, and spectra were taken at 0 °C. Reduced α-LA(α) and α-LA(β) were dissolved in 10 mM Tris, 0.5 mM EDTA, 1 mM DTT, pH 8.5. The spectra of native α-LA were taken in 10 mM Tris, 1 mM $CaCl_2$, pH 8.5, while the spectra of the pH 2 A-state of α-LA were taken in water with the pH adjusted using HCl. Protein concentrations were determined by absorbance in 6 M guanidine HCl, 20 mM sodium phosphate, pH 6.5, using an extinction coefficient at 280 nm of 23,150 (ref. 31). Far-UV studies were performed in a 1 mm path length cuvette with a 1.5 nm bandwidth. Near-UV and thermal denaturation studies were performed in a 1 cm path length cuvette with a 3.0 nm bandwidth. Data for thermal denaturations were obtained at intervals of 2°, allowing 1.5 minutes equilibration time between measurements, and were >95% reversible with no hysteresis. Helix content was calculated by the method of Chen, *et al.*[32], with a $[\theta]_{222}$ of 29,400 for 100% helix.

Sedimentation equilibrium. Sedimentation equilibrium was performed on a Beckman XL-A analytical ultracentrifuge using an An-60 Ti rotor. Protein solutions were dialyzed overnight against 10 mM Tris, 0.5 mM EDTA, pH 8.5. Protein concentrations of 100, 40, and 15 μM were analyzed in Beckman 6-sector cells at 4 °C. Data were collected at 23 and 27 k.r.p.m. and processed using the XL-A analysis program Origin (Beckman) with a partial specific volume of 0.733, calculated from the amino acid composition[33]. The data for both α-LA(α) and α-LA(β) fit well to a model for an ideal monomer, with no systematic deviation of the residuals. There is no concentration dependence of the observed molecular weight for either variant.

Disulphide exchange. Disulphide exchange studies were performed at room temperature in an anaerobic chamber (Coy Laboratory Products). Native buffer consisted of 10 mM Tris, 0.5 mM EDTA, pH 8.5. Denaturing buffer consisted of 6 M guanidine HCl, 10 mM Tris, 0.5 mM EDTA, pH 8.5. Exchange was initiated in buffers containing 100 μM reduced glutathione, 10 μM oxidized glutathione, and 5 μM protein. Samples were quenched with 10% formic acid (v/v) after approximately 24 h of equilibration and analyzed by reverse-phase HPLC on a C_{18} column (Vydac) with a linear H_2O-acetonitrile gradient (0.09% per minute) containing 0.1% TFA. All peak areas are reproducible to within ±2% and are identical regardless of the starting species.

Received 1 December 1994; accepted 17 February 1995.

Acknowledgements
We thank J.Sakai of the Whitehead Institute and R.Cook of the MIT Biopolymers Laboratory for peptide sequencing and amino acid analysis. We also thank P.Harbury, D. Minor, M.Oakley, B.Schulman, and members of the Kim lab for helpful advice and discussion. This work was supported by grant GM41307 from the National Institutes of Health.

1. Ptitsyn, O.B. The molten globule state. in *Protein Folding* (Ed. Creighton, T.E.) 243–300. (W.H. Freeman and Co., New York, 1992) .
2. Bychkova, V.E. & Ptitsyn, O.B. The molten globule in vitro and in vivo. *Chemtracts — Biochem. molec. Biol.* **4**, 133–163 (1993).
3. Kuwajima, K. The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. *Proteins* **6**, 87–103 (1989).
4. Haynie, D.T. & Friere, E. Structural energetics of the molten globule state. *Proteins* **16**, 115–140 (1993).
5. Christensen, H. & Pain, R.H. Molten globule intermediates and protein folding. *Eur. biophys. J.* **19**, 221–229 (1991).
6. Dobson, C.M. Unfolded proteins, compact states, and molten globules. *Curr. Opin. struct. Biol.* **2**, 6–12 (1992).
7. Kuwajima, K., Nitta, K., Yoneyama, M. & Sugai, S. Three-state denaturation of α-Lactalbumin by guanidine hydrochloride. *J. molec. Biol.* **106**, 359–373 (1976).
8. Dolgikh, D.A. *et al.* α-Lactalbumin: compact state with fluctuating tertiary structure. *FEBS Letts* **136**, 311–315 (1981).
9. Dolgikh, D.A. *et al.* Compact state of a protein molecule with pronounced small-scale mobility: bovine α-Lactalbumin. *Eur. biophys. J.* **13**, 109–121 (1985).
10. Baum., J., Dobson, C.M., Evans, P.A. & Hanley, C. Characterization of a partly folded protein by NMR methods: studies on the molten globule state of guinea pig α-Lactalbumin. *Biochemistry* **28**, 7–13 (1989).
11. Alexandrescu, A.T., Evans, P.A., Pitkeathly, M., Baum, J. & Dobson, C.M. Structure and dynamics of the acid-denatured molten globule state of α-Lactalbumin: a two-dimensional NMR study. *Biochemistry* **32**, 1707–1718 (1993).
12. Xie., D., Bhakuni, V. & Freire, E. Calorimetric determination of the energetics of the molten globule intermediate in protein folding: Apo-α-Lactalbumin. *Biochemistry* **30**, 10673–10678 (1991).
13. Ewbank, J.J. & Creighton, T.E. The molten globule protein conformation probed by disulphide bonds. *Nature* **350**, 518–520 (1991).
14. Creighton, T.E. & Ewbank, J.J. Disulphide-rearranged molten globule state of α-Lactalbumin. *Biochemistry* **33**, 1534–1538 (1994).
15. Peng, Z.-y. & Kim, P.S. A protein dissection study of a molten globule. *Biochemistry* **33**, 2136–2141 (1994).
16. Peng, Z.-y., Wu, L.C. & Kim, P.S. Local structural preferences in the α-Lactalbumin molten globule. *Biochemistry,* **in the press**.
17. Redfield, C., Smith, R.A.G. & Dobson, C.M. Structural characterization of a highly-ordered 'molten globule' at low pH. *Nature struct. Biol.* **1**, 23–29 (1994).
18. Feng, Y., Sligar, S.G. & Wand, A.J. Solution structure of apocytochrome b562. *Nature struct. Biol.* **1**, 30–35 (1994).
19. Miranker, A., Robinson, C.V., Radford, S.E., Alpin, R.T., & Dobson, C.M. Detection of transient protein folding populations by mass spectrometry. *Science* **262**, 896–900 (1993).
20. Radford, S.E., Dobson, C.M. & Evans, P.A. The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature* **358**, 302–307 (1992).
21. Kuwajima, K., Hiraoka, Y., Ikeguchi, M. & Sugai, S. Comparison of the transient folding intermediates in lysozyme and α-Lactalbumin. *Biochemistry* **24**, 874–881 (1985).
22. Ikeguchi, M., Kuwajima, K., Mitani, M. & Sugai, S. Evidence for identity between the equilibrium unfolding intermediate and a transient folding intermediate: a comparitive study of the folding reactions of α-Lactalbumin and lysozyme. *Biochemistry* **25**, 6965–6972 (1986).
23. Gilmanshin, R.I. & Ptitsyn, O.B. An early intermediate of refolding α-Lactalbumin forms within 2 ms. *FEBS Letters* **223**, 327–329 (1987).
24. Wetlaufer, D.B. Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc. natn. Acad. Sci. U.S.A.* **70**, 697–701 (1973).
25. Jaenicke, R. Protein folding: local structures, domain, subunits, and assemblies. *Biochemistry* **30**, 3147–3161 (1991).
26. Levinthal, C. Are there pathways for protein folding? *J. chim. Phys.* **65**, 44–45 (1968).
27. Kunkel, T.A., Roberts, J.D. & Zakour, R.A. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Meths Enz.* **154,** 367–382 (1987).
28. Doering, D.S. *Functional and structural studies of a small f-actin binding protein,* (Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, 1992).
29. Studier, F.W., Rosenberg, A.H., Dunn, J.J. & Dubendorff, J.W. Use of T7 RNA polymerase to direct expression of cloned genes. *Meths Enz.* **185**, 60–89 (1990).
30. Ellman, G.L. Tissue sulphydryl groups. *Arch. Biochem. Biophys.* **82**, 70–77 (1959).
31. Edelhoch, H. Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry* **6**, 1948–1954 (1967).
32. Chen, Y.-H., Yang, J.T., and Chau, K.H. Determination of the helix and β form of proteins in aqueous solution by circular dichroism. *Biochemistry* **13**, 3350–3359 (1974).
33. Laue, T.M., Shah, B.D., Ridgeway, T.M. & Pelletier, S.L. Computer-aided interpretation of analytical sedimentation data for proteins. in *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (eds Harding, S.E. *et al.*) pp. 90–125. (The Royal Society of Chemistry, Cambridge, 1992).
34. Acharya, K.R., Ren, J., Stuart, D.I., Phillips, D.C. & Fenna., R.E. Crystal structure of human α-Lactalbumin at 1.7 Å resolution. *J. molec. Biol.* **221**, 571–581 (1991).
35. Priestle, J.P. RIBBON: A stereo cartoon drawing program for proteins. *J. appl. Crystallogr.* **21**, 572–576 (1988).
36. Kauzmann, W. in *Sulphur in proteins* (Eds Benesch, R. et al.) 93–108. (Academic Press, New York, 1959).
37. Privalov, P.L. Stability of proteins: proteins which do not present a single cooperative system. *Adv. Protein Chem.* **35**, 1–104 (1982).