Proline scanning mutagenesis of a molten globule reveals non-cooperative formation of a protein's overall topology

Brenda A. Schulman[1,2] and Peter S. Kim[1]

Small proteins generally fold cooperatively: disruption of significant parts of the folded structure leads to unfolding of the rest of the protein. We show here, using proline scanning mutagenesis, that the native-like tertiary fold of the α-lactalbumin (α-LA) molten globule is formed by the non-cooperative assembly of its constituent helices. In contrast to the drastic destabilizing effects of proline substitutions in cooperatively folded proteins, proline mutations in the molten globule appear to cause only individual helices to unfold, without significantly influencing the other helices or the overall topology. Thus, the key determinants of a protein's overall fold may not be of the all-or-none type.

[1]Howard Hughes Medical Institute, Whitehead Institute for Biomedical Research, and Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA

[2]Current address: Massachusetts General Hospital Cancer Center, Laboratory of Molecular Oncology Building 149, 13th Street, Charlestown, Massachusetts 02129, USA

One of the major problems in protein folding is to identify the critical features of a protein's primary sequence that code for its native overall fold. Molten globules provide simplified model systems for addressing this question. Although the term 'molten globule' has been used to describe a wide variety of protein forms, classic molten globules, of which the A-state of α-lactalbumin (α-LA) is the paradigm[1], have the following characteristics: (i) a near-native level of secondary structure; (ii) compactness, with dimensions slightly larger than the native protein; and (iii) a lack of extensive fixed tertiary interactions, as judged by a number of criteria including a broad thermal transition and poor NMR chemical shift dispersion[1–4].

Several studies [5–9] indicate that the molten globule of α-LA has a native-like overall fold, even though it lacks the extensive fixed tertiary interactions that are characteristic of native proteins[5,10]. Thus, molten globules are likely to represent the simplest form of a protein in which there is substantial information transfer from the primary sequence to the three-dimensional fold. Detailed knowledge of the structures of molten globules is essential for understanding why they have a native-like fold. However, it is difficult to obtain detailed three-dimensional structural information about molten globules because they lack fixed packing interactions. Indeed, no high-resolution structure has been solved for any classic molten globule.

Here we use proline-scanning mutagenesis to probe the structure and folding of the α-LA molten globule. In this approach, individual residues are changed, one at a time, to proline. Analogous to alanine-scanning mutagenesis, which identifies side-chains that contribute to protein–protein interactions[11], proline-scanning mutagenesis is aimed at identifying regions of the backbone that specify a protein's fold. Proline substitutions are chosen because they unfold or greatly destabilize native proteins and structured peptides when inserted in the middle of secondary structures[12–21]. Because molten globules have a high degree of secondary structure, we reasoned that proline mutations would be useful probes for structure and folding.

Prolines unfold individual helices

Proline substitutions were targeted to residues that are helical in the native protein (Fig. 1*a*) because numerous studies indicate that the helices folded in the α-LA molten globule correspond to the same, or nearly the same, residues that are helical in the native protein[1,5,9,10,22–24]. Proline-sensitive regions were identified by characterizing the circular dichroism (CD) spectra of proline mutants of a variant of α-LA termed [28–111]. [28–111] contains the single disulphide bond between cysteines 28 and 111, with all other cysteines mutated to alanines, and is a monomeric molten globule that retains the characteristic features of the acid-induced α-LA molten globule, even near neutral pH[7]. CD studies are facilitated by the use of a single disulphide variant because formation of a particular disulphide bond is ensured, and the effects of mutations are not obscured by formation of non-native disulphide isomers.

A number of sites are sensitive to proline mutations (Fig. 1*b*,*c*). The loss of overall helix-content observed in the mutants appears to be specific for proline, as alanine substitutions at proline-sensitive sites have no significant effect (Fig. 1*c*). Remarkably, in no case do the proline mutations cause complete unfolding of the molten globule. This observation is surprising because the destabilizing effect of proline substitutions in native proteins (2.5–5.7 kcal mol^{-1})[18–21] is comparable to the total free

energy estimated to stabilize the molten globule of human α-LA[24,25]. Thus, if the molten globule were folded in an all-or-none cooperative manner, a proline mutation that disrupted a single helix would be expected to result in complete unfolding.

The small decrease in the helical CD signal observed in the proline mutants is roughly of the magnitude expected if only the helix containing the mutation, and not the entire protein, unfolds on proline substitution (Fig. 2*a*). Indeed, the decrease in helix content is similar for proline mutations at a number of different sites within each individual native helix (Fig. 1*c*). Most importantly, the helix content is almost unchanged when a second proline substitution is made within a single helix (Fig. 1*c*, hatched bars). We conclude that, to a good approximation, proline mutations cause only individual helices within this molten globule to unfold.

Our results indicate that the A-, B-, D- and 3_{10}-helices are folded in the molten globule formed by human α-LA, but that the C-helix is not (Figs 1*c*, 2*a*). These results agree with and extend an earlier hydrogen exchange study, which indicated that the A- and B-helices, but not the C-helix, were folded[24]. The structure of native α-LA[26] suggests that the helices that are folded in the molten globule form a subdomain[27]: there is an extensive network of van der Waals contacts between the A-, B-, D- and 3_{10}-helices in the structure of native α-LA, but these helices make only a small number of contacts to the C-helix and β-sheet domain (Fig. 2*b*). Although one might expect cooperative folding of subdomains, our observation that proline mutations cause only individual helices to unfold demonstrates that this subdomain is not formed in an all-or-none manner.

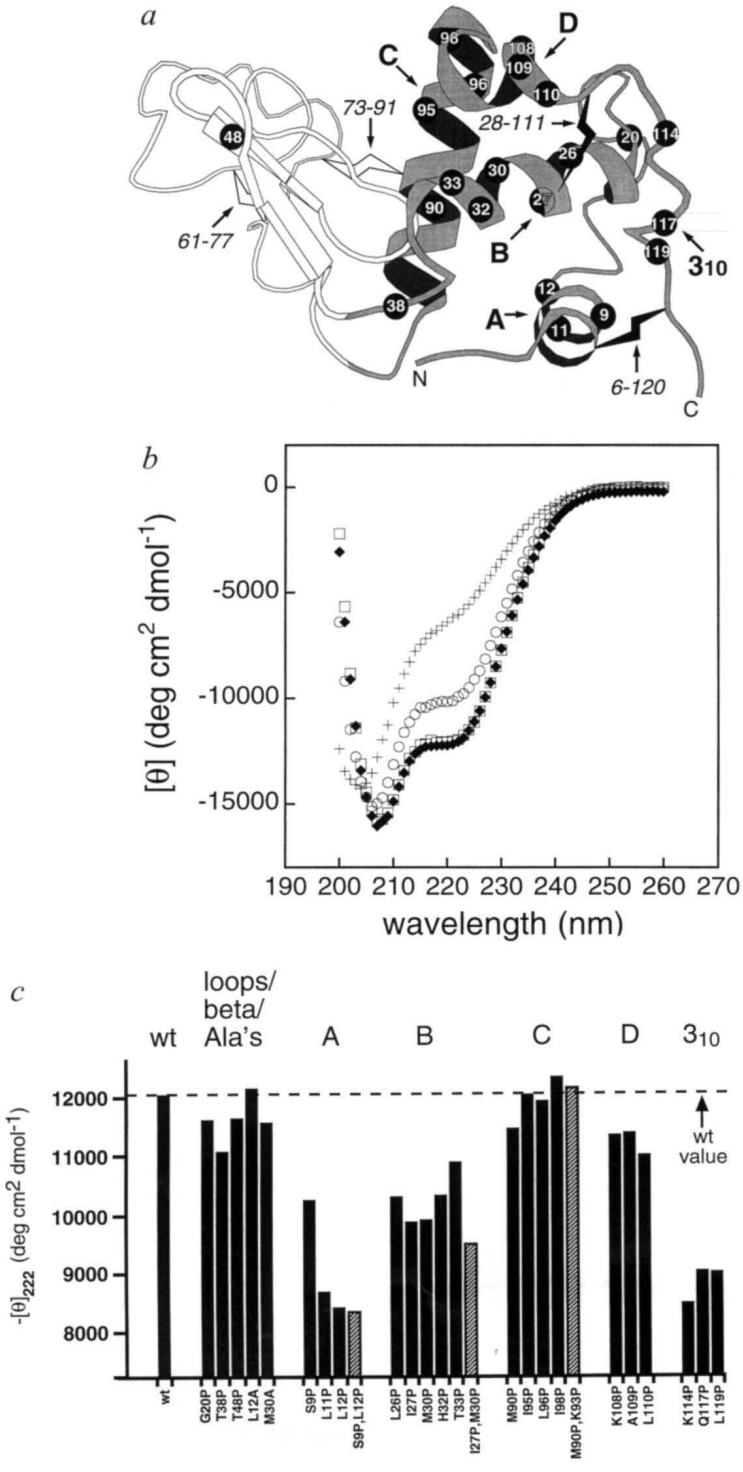

Fig. 1 Proline scanning mutagenesis of the α-LA molten globule. *a*, Sites of proline substitutions are shown on a schematic representation[58] of the structure of native human α-LA[26]. The helices are labelled 'A', 'B', 'C', 'D' and '3_{10}'; the disulphide bonds are identified by the cysteine residues they bridge; and the region of α-LA shown previously to contain the essential elements of the molten globule is shaded[5]. *b*, Far-UV CD spectra of [28–111] (filled diamonds) and some proline mutants. Some sites are insensitive to proline mutations (L96P; squares), whereas others are sensitive to prolines (M30P; circles). Substantial reduction in helix-content requires multiple mutations (L12P, M30P, L119P; plus signs). *c*, Effects of proline mutations on overall helix-content, as measured by CD at 222 nm. The bars are grouped according to the sites of mutation: 'wt' refers to the value of $[\theta]_{222}$ in wild-type [28–111]; 'loops/beta/Ala's' refers to mutations in loops, the β-sheet domain or alanine mutations in proline-sensitive sites; and the other labels refer to the helices containing the mutations. Hatched bars indicate that the effects of two proline mutations within a region corresponding to a single helix in the native structure are similar to the effects of a single mutation.

Prolines and the native-like fold

The ability to use proline mutations to knock out the helices in the molten globule, one at a time, allows us to evaluate the contribution of each helix toward forming the native-like overall fold of the molten globule. The preference of the polypeptide to form a native-like topology is reflected by the equilibrium distribution of disulphide pairings in the two-disulphide variant of α-LA, termed α-LA(α), with only the disulphides in the helical domain[6]. There are three possible ways to form two-disulphide pairings in α-LA(α), and the equilibrium distribution of the three different disulphide isomers reflects the extent to which the chain prefers a particular backbone topology (Fig. 3*a*). Wild-type α-LA(α), which has the characteristics of the acid-induced molten globule of α-LA, has over a 25-fold greater preference for

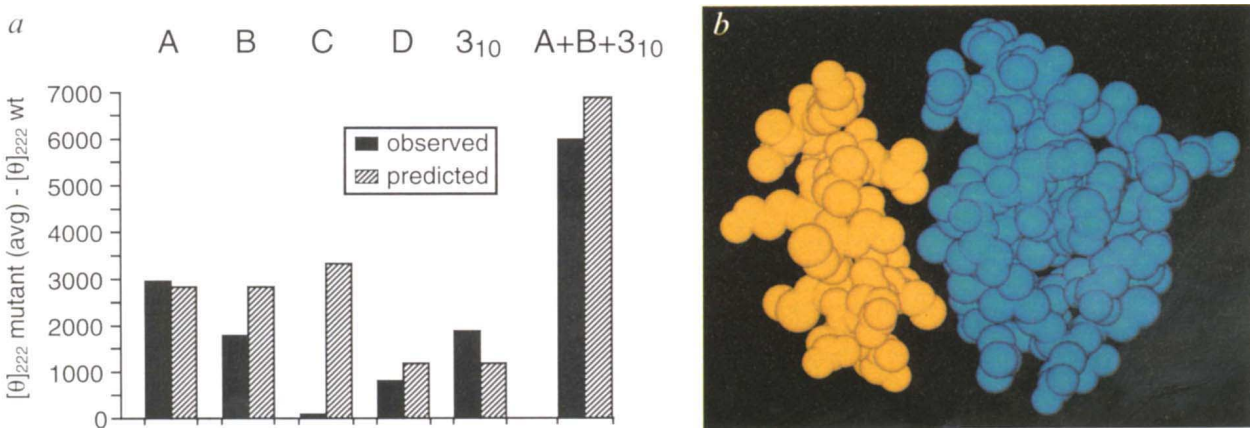

Fig. 2 Proline substitutions cause individual helices within a subdomain to unfold. *a*, Comparison of the observed changes in helix-content for proline mutations (black bars) and the changes expected for disrupting only individual helices (hatched bars). The helices containing mutations are indicated above the bars. The magnitude of the effects of proline mutations indicates that only individual helices (A, B, D and 3_{10}) unfold with proline substitutions. For the A-, B-, C- and D-helices, the observed change in $[\theta]_{222}$ is averaged for all of the mutations within each helix. For the 3_{10}-helix, L119 was mutated to proline in a background with W118 mutated to leucine, because W118 may make an aromatic contribution to $[\theta]_{222}$[59–61]. The effects of the L119P mutation are different in the wild-type and W118L backgrounds (compare Fig. 1*c* with Fig. 2*a*), whereas the effects of a different proline mutation (M30P) are roughly the same in the two backgrounds (data not shown). Therefore, for comparison, we use the results for the 3_{10}-helix obtained in the W118L background. *b*, Space-filling model of the helices and disulphide bonds in the helical domain of native α-LA. The helices folded in the molten globule (blue) comprise a subdomain in the native structure[26]. The C-helix (yellow), which is insensitive to proline mutations in the molten globule, is not part of this subdomain. For clarity, the orientation of α-LA is rotated slightly relative to Fig. 1*a*.

forming native disulphide bonds under native conditions than under denaturing conditions[6].

Remarkably, none of the individual helices within the subdomain identified by proline-scanning mutagenesis is absolutely required by the remainder of the protein for forming a native-like overall fold: introduction of prolines into the A-, B-, D- or 3_{10}-helices does not abolish the preference for forming native disulphide bonds, although in some cases this preference is reduced slightly (Fig. 3*b*). Mutating additional helices reduces further the preference for forming native disulphide bonds, and the effects of multiple proline mutations are additive (Fig. 3*b*). The simultaneous introduction of prolines into the A-, B- and D-helices reduces the fraction of molecules that form native disulphide pairings from the wild-type value of 85% to 16%. Interestingly, this latter value is still significantly larger than the fraction predicted with a random-walk model (2%) or observed for α-LA(α) in denaturing conditions (~3%)[6].

Prolines may diminish collapse

In addition to indicating the specificity for forming the native-like fold, the populations of the different disulphide isomers are also likely to reflect the degree of collapse of the polypeptide. The disulphide bond pairings that are possible in α-LA(α) are: (i) the native arrangement ([6–120; 28–111], where the numbers in brackets refer to cysteine residues that are disulphide bonded); (ii) the non-native 1 (NN1) arrangement ([6–28; 111–120]); and (iii) the non-native 2 (NN2) arrangement ([6–111; 28–120]). A previous study of the effective concentrations of each possible disulphide bond, in native buffer and in buffer containing 6 M guanidine-HCl, suggests that formation of the native and NN2 disulphide bonds is enhanced by collapse, whereas formation of the NN1 disulphide bonds is not[7].

Briefly, in native buffer, without any proline mutations, the effective concentrations of the NN1 and NN2 disulphide bonds are comparable. However, the effective concentrations of the NN1 disulphides are the same in native and guanidine buffers, whereas the effective concentrations of the NN2 disulphide bonds in native buffer are 10–100-fold greater than in guanidine buffer. This difference arises presumably because the NN1 disulphides are between cysteines that are nearest neighbours in the sequence (6–28 and 111–120), and thus are not enhanced by collapse. In contrast, the enhancement in the effective concentrations of the NN2 disulphides probably arises from the collapse of the polypeptide chain in aqueous buffer, which brings together the cysteines (6–111 and 28–120) that are far apart in the protein sequence. (The effective concentrations of the native disulphide bonds are enhanced ~100–1000-fold.)

In the equilibrium disulphide-exchange experiments with the proline mutants, decreases in the population of the native disulphide bonded isomer are accompanied by correspondingly much greater increases in the NN1 isomer than the NN2 isomer (compare the left and right panels in Fig. 3*a*, for example). Because the NN1 and NN2 isomers should be formed with similar frequencies in a nonspecifically collapsed polypeptide, the preferential accumulation of the NN1 isomer in the proline mutants, with little change in NN2, suggests that the loss of specificity is accompanied by decreases not only in secondary structure, but also in polypeptide collapse.

Forming a protein's overall fold

While the nature of the folding transition of molten globules is currently a matter of lively debate[2,28–36] the denaturant-induced unfolding of the α-LA molten globule is thought not to be two-state[37]. Our results indicate that the native-like fold of the helical domain of α-LA is

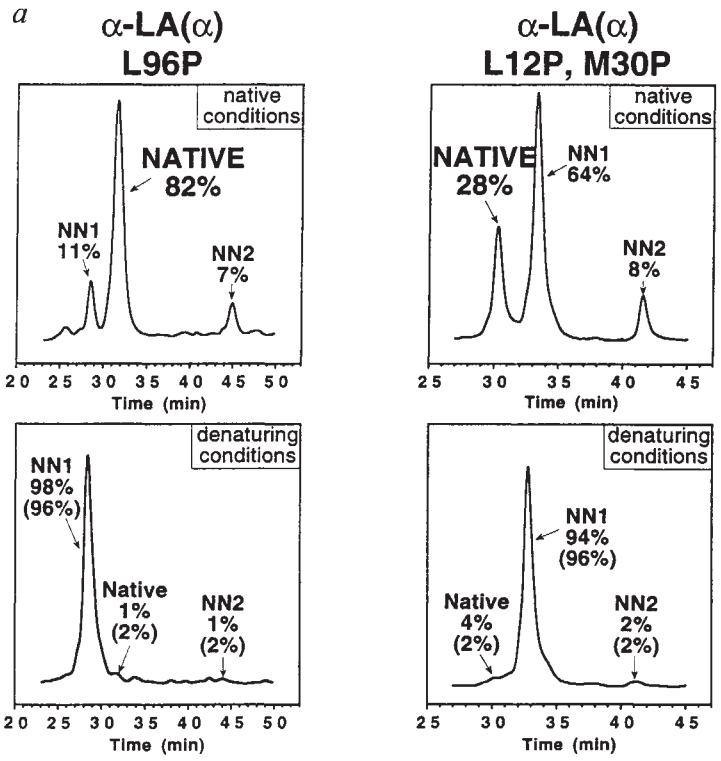

Fig. 3 The effects of knocking out individual helices, with proline mutations, on forming the native-like overall fold (assayed by native disulphide bond formation)[5,6]. *a*, HPLC analyses of equilibrium disulphide bond formation in proline mutants of α-LA(α) under native conditions (upper chromatograms) and in 6 M guanidine-HCl (lower chromatograms). The native species is [6–120; 28–111], where the numbers in brackets refer to cysteine residues that are disulphide bonded. 'NN1' refers to the non-native species [6–28; 111–120] and 'NN2' refers to the non-native species [6–111; 28–120]. The mutation L96P has no significant impact on native disulphide bond formation, whereas the double mutant L12P M30P is compromised in its ability to form the native disulphide bonds. Note that the elution order for the disulphide isomers is different in the two mutants. Under denaturing conditions, for all of the proline mutants, the ratios of the disulphide species are similar to those predicted from a random-walk model of the polypeptide[5]. *b*, The effects of proline mutations in individual helices ('1P'), 2 helices ('2P's') and 3 helices ('3P's') on native disulphide bond formation. The observed frequencies of forming native disulphide bonds (black bars) are compared to the frequencies expected if the energetic consequences of mutating multiple helices were additive (hatched bars). The helices containing mutations are indicated: 'A' refers to mutation of L12, 'B' of M30, 'C' of L96, 'D' of A109, and '3_{10}'of L119. An alanine substitution in the B-helix (M30A) has no significant effect on rearrangement (data not shown). The predicted effects of multiple mutations were estimated by multiplying (% native bonds) observed for the constituent individual mutations.

assembled in the absence of a cooperative all-or-none transition. It appears that the α-LA polypeptide has a significant preference for a native-like overall fold, even in the absence of substantial secondary structure formation. Hydrophobic collapse of the polypeptide chain is likely to be an important global feature for folding[38], and our results suggest that a decrease in the extent of polypeptide collapse accompanies a decrease in the preference for forming a native-like topology.

The inherent preference for a native-like overall fold in the collapsed α-LA polypeptide is enhanced by the formation of individual helices. These helices form independently of one another: individual helices unfold on proline mutagenesis and no single helix is required to form a native-like overall fold (Fig. 4).

In contrast to our observations that the α-LA molten globule folds non-cooperatively, the molten globule forms of apomyoglobin and cytochrome *c* (cyt *c*) have been suggested to fold cooperatively[39,40]. It is possible that the α-LA molten globule is more primitive (that is, may represent an earlier stage on the folding pathway) than other molten globules, even though it has a native-like fold[5]. Alternatively, the apparently conflicting observations may be reconciled by differences in the methods used to denature the proteins. Numerous studies indicate that proteins denatured in aqueous solution are more structured than proteins unfolded in guanidine or urea[41–46]. In our studies, we have used proline mutations to denature the α-LA molten globule in aqueous solution. On the other hand, in the studies that report cooperative folding, the apomyoglobin and cyt *c* molten globules were unfolded by denaturant. The apomyoglobin molten globule may unfold non-cooperatively when denatured with acid in aqueous solution[47]. Moreover, the transition between the more folded and less folded molten globule forms of apomyoglobin involves the non-cooperative addition of another helix[34]. Nonetheless, for the α-LA molten globule, high-resolution NMR studies indicate that urea-induced unfolding is not cooperative, even though low-resolution CD studies of the same transition give rise to sigmoidal-shaped curves (ref. 37; B.A.S., P.S.K., C.M. Dobson and C. Redfield, unpublished results).

It is striking that a protein's overall three-dimensional fold, attained early and in the absence of extensive side-chain packing, is formed in a non-cooperative manner. It will be important to understand how a protein's fold is determined non-cooperatively: properties such as hydrophobic/hydrophilic patterning, side-chain volumes and secondary structure propensities are good candidates for investigation[5,6,48–52]. The relationship between

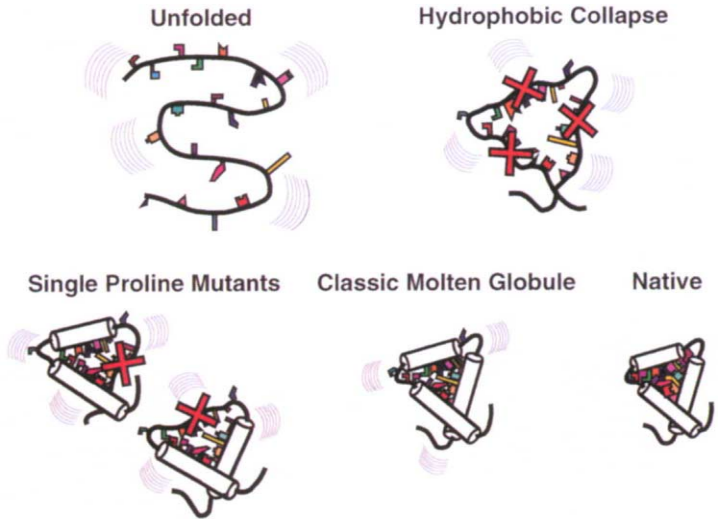

Fig. 4 Non-cooperative formation of an overall native fold. Hypothetical structures are shown for the unfolded, collapsed (three-proline mutant), collapsed with some secondary structures (single-proline mutants), classic molten globule and native stages of folding.

the non-cooperative folding observed here and the cooperative folding events described traditionally is likely to involve later consolidation of side-chain packing[53].

Methods

Production of α-LA variants. Mutations were made as described previously[6], and were verified by restriction digests and sequencing the entire α-LA gene. Proline mutants were expressed, purified from inclusion bodies by ion exchange chromatography, refolded, and purified by reversed-phase HPLC as described previously[6,7] except that NaCl was excluded from the ion exchange wash buffer and mutants of α-LA(α) were HPLC-purified in reduced form.

Circular dichroism (CD) spectroscopy and sedimentation equilibrium. CD spectra were recorded in a 1 mm pathlength cell at protein concentrations of ~25 μM at 4 °C. As a standard, $[\theta]_{222}$ was also measured for fresh samples of [28–111] at the same time as each variant, and the variation was <5%. All protein concentrations were determined by tryptophan, tyrosine and cystine absorbance[54]. An Aviv Model 60DS CD spectrometer equipped with a thermoelectric temperature controller was used. For [28–111] and five variants (L12P, M30P, L96P, A109P, and L119P), sedimentation equilibrium was performed with a Beckman XL-A analytical ultracentrifuge using an An-60 Ti rotor and Beckman epoxy six-sector cells. Protein solutions were dialyzed overnight *versus* 10 mM Tris, 0.5 mM EDTA, pH 8.5. Data were collected at 24 and 28 kr.p.m. at 4 °C for three initial concentrations (15, 40 and 100 μM). The density of the buffer and the partial specific volumes were calculated with constants from Laue *et al.*[55]. Data sets were processed individually using the curve-fitting routine in Kaleidagraph (Abelbeck software). Molecular weights were all within 10% of those calculated for ideal monomers, and the fits displayed random residuals.

Estimation of $[\theta]_{222}$ values for native-like helices. In order to estimate the decrease in $-[\theta]_{222}$ expected for knocking out a single native-like helix, the helices in the native structure are defined as follows[26]: 'A' (the A α-helix plus a contiguous turn of 3_{10}-helix) spans residues 5–16; 'B' spans residues 23–34; 'C' spans residues 86–99; 'D' spans residues 105–109; and '3_{10}' spans residues 115–119. The predicted decreases in $-[\theta]_{222}$ were calculated by assuming a value of −29,400 deg cm^2 dmol^{-1} for residues in short α-helices[56] and taking account of all the 123 residues in α-LA. As an example, the B-helix contains 12 residues of helix, so the value of $[\theta]_{222}$ contributed by the B-helix is estimated as: (12/124)×−29,400=−2845 deg cm^2 dmol^{-1}.

Disulphide exchange. Equilibrium disulphide exchange experiments were performed at room temperature in an anaerobic chamber (Coy Laboratory Products). Native buffer is 10 mM Tris, 0.5 mM EDTA, pH 8.5; denaturing buffer is 6 M guanidine-HCl, 10 mM Tris, 0.5 mM EDTA, pH 8.5. Exchange was initiated by adding reduced, HPLC-purified variants of α-LA(α) to exchange buffers containing 90 μM reduced glutathione and 20 μM oxidized glutathione. The final protein concentration was 5 μM. Exchange was quenched with 10% formic acid (v/v) after 24, 36 and 48 h of equilibration and analyzed by reversed-phase HPLC, as described previously[6]. Time-courses indicate that disulphide exchange reached equilibrium for every variant by 36 h, after which no further changes were observed. Peak areas are reproducible to within ±2%. For wild-type α-LA(α) and five mutants (L96P; L12P M30P; L12P A109P; M30P A109P and L12P M30P A109P), exchange was also initiated from at least one purified oxidized species and the peak areas are reproducible to the same limits. All three major species (corresponding to the native plus the two non-native isomers) were completely oxidized, as assayed by the lack of reaction with Ellman's reagent [5, 5'-dithio-bis(2-nitrobenzoic acid)] in 6 M guanidine-HCl[57].

Assignment of disulphide isomers. Three different disulphide isomers are possible for α-LA(α) and three different species are observed. The non-native isomer with disulphides between cysteines 6–28 and 111–120 (NN1) was identified readily because it is the predominant species (>96%) at equilibrium in strongly denaturing conditions, as predicted by a random-walk model[5]. The other non-native species, with disulphides between cysteines 6–111 and 28–120 (NN2), and the native species (6–120; 28–111), were assigned by proteolysis with endoproteinase-Glu-C (Endo-Glu), which cleaves after Asp and Glu residues, and mass spectrometry following purification by reversed-phase HPLC. Briefly, HPLC-purified native α-LA(α) or NN2 was dissolved in 50 mM sodium phosphate (pH 7.5) at ~1 mg ml^{-1}. Endo-Glu (Sigma) was dissolved in water at ~1 mg ml^{-1}. 1/10th volume of endo-Glu was added to the different isomers of variants of α-LA(α), and digestion proceeded for 2–3 h at room temperature. Under these conditions, the fragments useful for disulphide assignment reflect cleavage after Glu 7, Glu 25, Glu 43, Asp 97, Glu 113, Glu 116 and Glu 121, except in variants containing the L119P mutation, in which there is not substantial cleavage after Glu 121. The digests were desalted by reversed-phase HPLC with a C18 column (Vydac) and a fast gradient (4.5–54% acetonitrile in water over 6 min, with 0.1% TFA). Oxidized and reduced digestion products were analyzed directly after HPLC by laser-desorption mass spectrometry (Voyager Elite, PerSeptive Biosystems). Programs that list all possible cleavage products and their corresponding masses (kindly provided by Zheng-yu Peng), allowed identification of the fragments with masses that are unambiguous for a particular disulphide isomer of each mutant. In all cases, the observed mass was within 2 relative molecular mass units (M_r) of the mass expected for the fragment. For the native isomer, the 6–120 disulphide bond was distinguished by the mass of a fragment spanning the N-terminal methionine (residue 0) to residue 7, disulphide-bonded to a fragment corresponding to residues 117–121 [denoted (0–7)-SS-(117–121)]. For mutants with L119P, the mass distinguishing 6–120 corresponds to (0–7)-SS-(117–123). The 28–111 disulphide bond was distinguished by two masses, corresponding to (26–43)-SS-(98–113) and (26–43)-SS-(98–116). In the reduced samples

of the native isomers, the masses of these fragments disappear and new masses appear, corresponding to the individual fragments (26–43), (98–113), and (98–116). For the NN2 isomer, the 6–111 disulphide bond is distinguished by two masses, corresponding to disulphide-bonded fragments (0–7)-SS-(98–113) and (0–7)-SS-(98–116). The 28–120 disulphide bond is distinguished by the mass of the disulphide-bonded fragment (26–43)-SS-(117–121).

Received 5 April, accepted 7 June 1996.

Acknowledgements
We thank P. Harbury, D. Minor, Z-y. Peng, J. Weissman and L. Wu for exciting discussions which led to some of the ideas presented in this manuscript, and P. Murray and members of the Kim lab for other helpful comments. B.A.S. was supported by a National Institutes of Health predoctoral training grant. This work was supported by the Howard Hughes Medical Institute.

1. Kuwajima, K. The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. *Proteins* **6**, 87–103 (1989).
2. Ptitsyn, O.B. The molten globule state. in *Protein Folding* (ed. Creighton, T.E.) 243–300 (W. H. Freeman and Co., New York, 1992).
3. Dobson, C.M. Solid evidence for molten globules. *Curr. Biol.* **4**, 636–40 (1994).
4. Eliezer, D., Jennings, P.A., Wright, P.E., Doniach, S., Hodgson, K.O. & Tsuruta, H. The radius of gyration of an apomyoglobin folding intermediate. *Science* **270**, 487–8 (1995).
5. Peng, Z.-y. & Kim, P.S. A protein dissection study of a molten globule. *Biochemistry* **33**, 2136–41 (1994).
6. Wu, L.C., Peng, Z.-y. & Kim, P.S. Bipartite structure of the α-lactalbumin molten globule. *Nature Struct. Biol.* **2**, 281–86 (1995).
7. Peng, Z.-y., Wu, L.C. & Kim, P.S. Local structural preferences in the α-lactalbumin molten globule. *Biochemistry* **34**, 3248–52 (1995).
8. Polverino de Laureto, P., De Filippis, V., Di Bello, M., Zambonin, M. & Fontana, A. Probing the molten globule state of alpha-lactalbumin by limited proteolysis. *Biochemistry* **34**, 12596–604 (1995).
9. Wilson, G. *et al.* Vibrational Raman optical activity of alpha-lactalbumin: comparison with lysozyme, and evidence for native tertiary folds in molten globule states. *J. Mol. Biol.* **254**, 747–60 (1995).
10. Alexandrescu, A.T., Evans, P.A., Pitkeathly, M., Baum, J. & Dobson, C.M. Structure and dynamics of the acid-denatured molten globule state of alpha-lactalbumin: a two-dimensional NMR study. *Biochemistry* **32**, 1707–1718 (1993).
11. Cunningham, B.C. & Wells, J.A. High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science* **244**, 1081–1085 (1989).
12. Richardson, J.S. & Richardson, D.C. Amino acid preferences for specific locations at the ends of alpha helices. *Science* **240**, 1648–52 (1988).
13. Strehlow, K.G., Robertson, A.D. & Baldwin, R.L. Proline for alanine substitutions in the C-peptide helix of ribonuclease A. *Biochemistry* **30**, 5810–4 (1991).
14. Minor, D., Jr. & Kim, P.S. Measurement of the beta-sheet-forming propensities of amino acids. *Nature* **367**, 660–663 (1994).
15. Smith, C.K., Withka, J.M. & Regan, L. A thermodynamic scale for the beta-sheet forming tendencies of the amino acids. *Biochemistry* **33**, 5510–5517 (1994).
16. Minor, D., Jr. & Kim, P.S. Context is a major determinant of beta-sheet propensity. *Nature* **371**, 264–267 (1994).
17. Wood, S.J., Wetzel, R., Martin, J.D. & Hurle, M.R. Prolines and amyloidogenicity in fragments of the Alzheimer's peptide beta/A4. *Biochemistry* **34**, 724–30 (1995).
18. O'Neil, K.T. & De Grado, W.F. A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science* 250, 646–651 (1990).
19. Horovitz, A., Matthews, J.M. & Fersht, A.R. Alpha-helix stability in proteins. II. Factors that influence stability at an internal position. *J. Mol. Biol.* **227**, 560–568 (1992).
20. Sauer, U.H., San, D.P. & Matthews, B.W. Tolerance of T4 lysozyme to proline substitutions within the long interdomain alpha-helix illustrates the adaptability of proteins to potentially destabilizing lesions. *J. Biol. Chem.* **267**, 2393–2399 (1992).
21. Blaber, M., *et al.* Determination of alpha-helix propensity within the context of a folded protein. Sites 44 and 131 in bacteriophage T4 lysozyme. *J. Mol. Biol.* **235**, 600–624 (1994).
22. 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 alpha-lactalbumin. *Biochemistry* **28**, 7–13 (1989).
23. Chyan, C.L., Wormald, C., Dobson, C.M., Evans, P.A. & Baum, J. Structure and stability of the molten globule state of guinea-pig alpha-lactalbumin: a hydrogen exchange study. *Biochemistry* **32**, 5681–5691 (1993).
24. Schulman, B.A., Peng, Z.-y., Redfield, C., Dobson, C.M. & Kim, P.S. Different subdomains are most protected from hydrogen exchange in the molten globule and native states of α-lactalbumin. *J. Mol. Biol.* **253**, 651–657 (1995).
25. Nozaka, M., Kuwajima, K., Nitta, K. & Sugai, S. Detection and characterization of the intermediate on the folding pathway of human alpha-lactalbumin. *Biochemistry* **17**, 3753–3758 (1978).
26. Acharya, K.R., Ren, J.S., Stuart, D.I., Phillips, D.C. & Fenna, R.E. Crystal structure of human alpha-lactalbumin at 1.7 A resolution. *J. Mol. Biol.* **221**, 571–581 (1991).
27. Siddiqui, A.S. & Barton, G.J. Continuous and discontinuous domains: an algorithm for the automated generation of reliable protein domain definitions. *Prot. Sci.* **4**, 872–884 (1995).
28. 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).
29. Yutani, K., Ogasahara, K. & Kuwajima, K. Absence of the thermal transition in apo-alpha-lactalbumin in the molten globule state. A study by differential scanning microcalorimetry. *J. Mol. Biol.* 228, 347–350 (1992).
30. Jennings, P.A. & Wright, P.E. Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science* **262**, 892–8926 (1993).
31. Kreimer, D.I., Szosenfogel, R., Goldfarb, D., Silman, I. & Weiner, L. Two-state transition between molten globule and unfolded states of acetylcholinesterase as monitored by electron paramagnetic resonance spectroscopy. *Proc. Natl. Acad. Sci. USA* **91**, 12145–12149 (1994).
32. Chan, H.S., Bromberg, S. & Dill, K.A. Models of cooperativity in protein folding. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **348**, 61–70 (1995).
33. Ptitsyn, O.B., Bychkova, V.E. & Uversky, V.N. Kinetic and equilibrium folding intermediates. *Philos. Trans. R. Soc. Lond. BB iol. Sci.* **348**, 35–41 (1995).
34. Kiefhaber, T. & Baldwin, R.L. Intrinsic stability of individual alpha helices modulates structure and stability of the apomyoglobin molten globule form. *J. Mol. Biol.* **252**, 122–132 (1995).
35. Loh, S.N., Kay, M.S. & Baldwin, R.L. Structure and stability of a second molten globule intermediate in the apomyoglobin folding pathway. *Proc. Natl. Acad. Sci. USA* **92**, 5446–5450 (1995).
36. Marmorino, J.L. & Pielak, G.J. A native tertiary interaction stabilizes the A state of cytochrome c. *Biochemistry* **34**, 3140–3143 (1995).
37. Shimizu, A., Ikeguchi, M. & Sugai, S. Unfolding of the molten globule state of alpha-lactalbumin studied by 1H NMR. *Biochemistry* **32**, 13198-203 (1993).
38. Dill, K.A. Dominant forces in protein folding. *Biochemistry* **29**, 7133–55 (1990).
39. Kay, M.S. & Baldwin, R.L. Packing interactions in the apomyoglobin folding intermediate. *Nature Struct. Biol.* **3**, 439–45 (1996).
40. Bai, Y., Sosnick, T.R., Mayne, L., & Englander, S.W. Protein folding intermediates: native-state hydrogen exchange. *Science* **269**, 192–197 (1995).
41. Dill, K.A. & Shortle, D. Denatured states of proteins. *Annu. Rev. Biochem.* **60**, 795–825 (1991).
42. Dobson, C.M. Unfolded proteins, compact states and molten globules. *Curr. Opin. Struct. Biol.* **2**, 6–12 (1992).
43. Shortle, D. Denatured states of proteins and their roles in folding and stability. *Curr. Opin. Struct. Biol.* **3**, 66–74 (1993).
44. Lumb, K.J. & Kim, P.S. Formation of a hydrophobic cluster in denatured bovine pancreatic trypsin inhibitor. *J. Mol. Biol.* **236**, 412–420 (1994).
45. Agashe, V.R., Shastry, M.C.R., & Udgaonkar, J.B. Initial hydrophobic collapse in the folding of barstar. *Nature* **377**, 754–757 (1995).
46. Oliveberg, M. & Fersht, A. Thermodynamics of transient conformations in the folding pathway of barnase: reorganization of the folding intermediate at low pH. *Biochemistry* **35**, 2738–2749 (1996).
47. Hughson, F.M., Barrick, D. & Baldwin, R.L. Probing the stability of a partly folded apomyoglobin intermediate by site-directed mutagenesis. *Biochemistry* **30**, 4113–4118 (1991).
48. Minor, D.L. Jr. & Kim, P.S. Context-dependent secondary structure formation of a designed protein sequence. *Nature* **380**, 730–734 (1996).
49. Lau, K.F. & Dill, K.A. Theory for protein mutability and biogenesis. *Proc. Natl. Acad. Sci. USA* **87**, 638–642 (1990).
50. Behe, M.J., Lattman, E.E. & Rose, G.D. The protein-folding problem: the native fold determines packing, but does packing determine the native fold? *Proc. Natl. Acad. Sci. USA* **88**, 4195–4199 (1991).
51. Bowie, J.U., Luthy, R. & Eisenberg, D. A method to identify protein sequences that fold into a known three-dimensional structure. *Science* **253**, 164–170 (1991).
52. Xiong, H., Buckwalter, B.L., Shieh, H.M. & Hecht, M.H. Periodicity of polar and nonpolar amino acids is the major determinant of secondary structure in self-assembling oligomeric peptides. *Proc. Natl. Acad. Sci. USA* **92**, 6349–6353 (1995).
53. Ptitsyn, O.B. How does protein synthesis give rise to the 3D-structure? *FEBS Letts* **285**, 176–181 (1991).
54. Edelhoch, H. Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry* **6**, 1948–1954 (1967).
55. Laue, T.M., Shah, B.D., Ridgeway, T.M. & Pelletier, S.L. Computer-aided interpretation of analytical sedimentation data for proteins. in *Analytical Ultracentrifugaton in Biochemistry and Polymer Science* (eds S.E. Harding, A.J. Rowe & J.C. Horton) 90–125 (The Royal Society of Chemistry, Cambridge, 1992).
56. Chen, Y.H., Yang, J.T. & Chau, K.H. Determination of the helix and beta form of proteins in aqueous solution by circular dichroism. *Biochemistry* **13**, 3350–3359 (1974).
57. Ellman, G.L. Tissue sulphydryl groups. *Arch. Biochem. Biophys.* **82**, 70–77 (1959).
58. Priestle, J.P. RIBBON: A stereo cartoon drawing program for proteins. *J. Appl. Crystallogr.* **21**, 572–576 (1988).
59. Woody, R.W. Circular Dichroism of Peptides. in *The Peptides: Analysis, Synthesis, Structure* (ed. V.J. Hruby) 15–114 (Academic Press, Orlando, 1985).
60. Manning, M.C. & Woody, R.W. Theoretical study of the contribution of aromatic side chains to the circular dichroism of basic bovine pancreatic trypsin inhibitor. *Biochemistry* **28**, 8609–8613 (1989).
61. Chakrabartty, A., Kortemme, T., Padmanabhan, S. & Baldwin, R.L. Aromatic side-chain contribution to far-ultraviolet circular dichroism of helical peptides and its effect on measurement of helix propensities. *Biochemistry* **32**, 5560–5565 (1993).