

Hydrophobic sequence minimization of the α -lactalbumin molten globule

(protein folding/hydrophobic interaction/disulfide bond/proteolysis)

LAWREN C. WU AND PETER S. KIM*

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

Contributed by Peter S. Kim, October 24, 1997

ABSTRACT The molten globule, a widespread protein-folding intermediate, can attain a native-like backbone topology, even in the apparent absence of rigid side-chain packing. Nonetheless, mutagenesis studies suggest that molten globules are stabilized by some degree of side-chain packing among specific hydrophobic residues. Here we investigate the importance of hydrophobic side-chain diversity in determining the overall fold of the α -lactalbumin molten globule. We have replaced all of the hydrophobic amino acids in the sequence of the helical domain with a representative amino acid, leucine. Remarkably, the minimized molecule forms a molten globule that retains many structural features characteristic of a native α -lactalbumin fold. Thus, nonspecific hydrophobic interactions may be sufficient to determine the global fold of a protein.

The molten globule is a general protein-folding intermediate characterized by compact but dynamic structure (1–3) arranged in a native-like tertiary fold (4). The molten globule can be considered a low-resolution protein structure in which the global fold, but not unique native structure, has been established. As such, the molten globule represents a system in which to study the determinants of a protein's overall fold, uncoupled from the determinants of fixed tertiary packing. A key issue is to understand how the molten globule can distinguish between native and nonnative backbone topologies in the apparent absence of extensive fixed tertiary interactions.

Hydrophobic interactions are thought to play a dominant role in protein folding, and the close complementary packing of hydrophobic side chains in a protein's core may help specify a protein's fold (5–7). Mutagenesis studies indicate that specific packing of core side-chain residues exists in the molten globules of cytochrome *c*, apomyoglobin, and staphylococcal nuclease (8–13). In these studies, a correlation is observed between mutational effects on the stabilities of the native and molten globule states. However, the effects of mutations on the molten globule are small in comparison to that on the native state, suggesting that packing interactions may be only partially formed in the molten globule. These results have been interpreted to indicate that some specific native-like packing, as opposed to nonspecific hydrophobic interactions, stabilizes the structure of molten globules (14).

In contrast, other studies suggest that nonspecific hydrophobic interactions may be sufficient to produce tertiary folds. Computer lattice models and structure assessment algorithms based solely on a nonspecific hydrophobic interaction can model some aspects of protein structure, stability, and folding kinetics and can distinguish between native and nonnative folds (15–18). In addition, proteins composed of only glu-

tamine, leucine, and arginine can form helical oligomers, although their detailed structures are unknown (19, 20). Interestingly, proteins with cores of predominantly one amino acid form folded structures with varying degrees of order (21–24). In one case, a well packed native structure is formed, of which a high resolution structure has been determined by x-ray crystallography (23).

Here we use the molten globule formed by the helical domain of α -lactalbumin (α -LA) as a model system to investigate the role of specific hydrophobic amino acids in determining backbone topology. α -LA is a two-domain protein that forms one of the most widely studied molten globules (25–32). Significantly, the α -LA molten globule is highly dynamic and noncooperatively folded and may represent an intermediate that is formed earlier than other well studied molten globules. In particular, NMR spectra of the α -LA molten globule lack significant chemical-shift dispersion (26, 32), suggesting that rigid side-chain packing is minimal. Previous experiments indicate that the molten globule of α -LA is a bipartite structure in which the helical domain adopts a native-like tertiary fold, whereas the β -sheet domain is largely unstructured (30). A recombinant model of the isolated helical domain (α -Domain) is a molten globule that retains a native-like fold (28).

To test the importance of hydrophobic side-chain diversity in determining a protein's fold, we have replaced all of the hydrophobic amino acids in the sequence of α -Domain with the representative residue, leucine. Remarkably, we find that our minimized hydrophobic construct (MinLeu) has many features characteristic of a native-like fold. Thus, nonspecific hydrophobic interactions may be sufficient to establish the native-like architecture of molten globules.

MATERIALS AND METHODS

Protein Production. Full-length α -LA and variants thereof were produced as described previously (30). Mutations were introduced by single-stranded mutagenesis and verified by DNA sequencing. Inclusion bodies of expressed proteins were washed with sucrose and Triton buffers, solubilized and reduced in urea/DTT, and purified as described previously (28, 30). Protein identity was confirmed by mass spectrometry.

Identification of Disulfide Connectivities. Disulfide bonds were assigned by digestion with the specific protease trypsin, followed by mass spectrometry of the proteolysis mixture, by using a slight modification of the procedure of Schulman and Kim (31). Briefly, 100 μ g of protein at a concentration of 1 μ g/ μ l was digested with trypsin at a ratio of 1:20 by weight in 10 mM Tris, pH 8.8, 37°C. Aliquots at 0, 5, 10, and 24 hr were quenched by addition of PMSF to a final concentration of 3 mM. DTT was added to half of each time point sample to reduce all disulfide-linked fragments. The quenched digestion mixtures (oxidized and reduced) were dried and resuspended in α -cyano cinnamic acid matrix for analysis by mass spectrometry on a Voyager Elite MALDI-TOF mass spectrometer

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1997 by The National Academy of Sciences 0027-8424/97/9414314-6\$2.00/0 PNAS is available online at <http://www.pnas.org>.

*To whom reprint requests should be addressed.

(PerSeptive Biosystems, Framingham, MA). Characteristic peptide fragments were identified in the mass spectrum of each disulfide species, thereby facilitating unequivocal disulfide connectivity assignments.

Sedimentation Equilibrium. Sedimentation equilibrium experiments were performed on a Beckman XL-A analytical ultracentrifuge as described previously (28, 30). MinLeu solutions were dialyzed overnight against 10 mM Tris/0.5 mM EDTA, pH 8.8, loaded at initial concentrations of 200, 100, 40, 15, and 5 μ M, and analyzed at 28 and 34 krpm. Data were acquired at three wavelengths per rotor speed and processed simultaneously with a nonlinear least squares fitting routine (Nonlin; ref. 33). Solvent density and protein partial specific volume were calculated according to solvent and protein composition, respectively (34). The data fit well to a model for an ideal monomer ($\pm 10\%$), with no systematic deviation of the residuals. There is no concentration dependence of the observed molecular weight.

CD Spectroscopy. CD spectroscopy was performed with an Aviv 62 DS spectrometer as described previously (30). Proteins were dissolved in 10 mM Tris/0.5 mM EDTA, pH 8.8, to a concentration of 10 μ M, and spectra were acquired at 0°C. Protein concentrations were determined by absorbance at 280 nm in 6 M GuHCl/20 mM sodium phosphate, pH 6.5, with an extinction coefficient calculated based on tryptophan, tyrosine, and cysteine content (35).

Disulfide Exchange. Disulfide exchange studies were performed at room temperature (25–27°C) as described previously (30). Native buffer consisted of 10 mM Tris/0.5 mM EDTA, pH 8.8, and denaturing buffer consisted of 6 M guanidinium isothiocyanate/10 mM Tris/0.5 mM EDTA, pH 8.8.

Proteolysis Mapping of Secondary Structure. Protein samples were subjected to limited proteolysis at room temperature with the nonspecific proteases elastase, proteinase K, and subtilisin. Protein samples were dissolved in 10 mM Tris/0.5 mM EDTA, pH 8.8, to a concentration of 2 mg/ml. Protease was added at a weight ratio of 1:1000, and proteolysis was stopped at 1, 2, 5, 10, and 20 min by the addition of PMSF to 3 mM. Samples were subsequently reduced by incubation with 100 mM DTT. Proteolysis samples were analyzed by reverse-

phase HPLC on a microbore C₁₈ column (Vydac), followed by direct injection into an electrospray mass spectrometer (Finnigan MAT LC-Q). Fragments were assigned by matching masses with possible fragments predicted with a computer algorithm (E. Wolf and P.S.K., unpublished data). Early time points were analyzed to ensure that initial cuts were observed, and complementary fragments were identified. This assignment procedure was verified by N-terminal sequence analysis for some proteolysis samples.

Solvent Accessibility Calculations. Solvent accessibility of side-chain atoms was determined with the areaimol routine in the CCP4 suite of programs (36), with a solvent probe of radius 1.4 Å. Fully buried, partially buried, and exposed methylene groups are defined as those with less than 5 Å², between 5 and 18 Å², and greater than 18 Å² of accessible surface area, respectively.

RESULTS

Minimization of the Helical Domain of α -LA. The helical domain of α -LA consists of residues 1–39 and 81–123 of the wild-type protein. A recombinant model of the isolated helical domain (α -Domain), consisting of residues 1–39 and 81–123 joined by a linker of three glycine residues, has a native-like fold (28). We minimized the hydrophobic amino acid diversity of α -Domain by following a sequence-based rule. An arbitrary cutoff was implemented between histidine and threonine in the hydrophobicity scale of Eisenberg and McLachlan (37). The following residues thus were deemed hydrophobic and were changed to leucine in the sequence of the helical domain: W, M, F, I, L, Y, V, and H. The one proline in the sequence of the helical domain (Pro-24), although a hydrophobic residue by our rule, was left unchanged, because its major structural influence may be through local helix-capping effects rather than side-chain packing (38, 39). A single tryptophan was introduced in the triglycine linker in α -Domain to facilitate protein concentration determination, thereby changing the linker sequence from G-G-G to G-W-G. The resulting minimized molecule is called MinLeu. For convenience, we refer to the amino acids in MinLeu by their positions in the context of full-length α -LA.

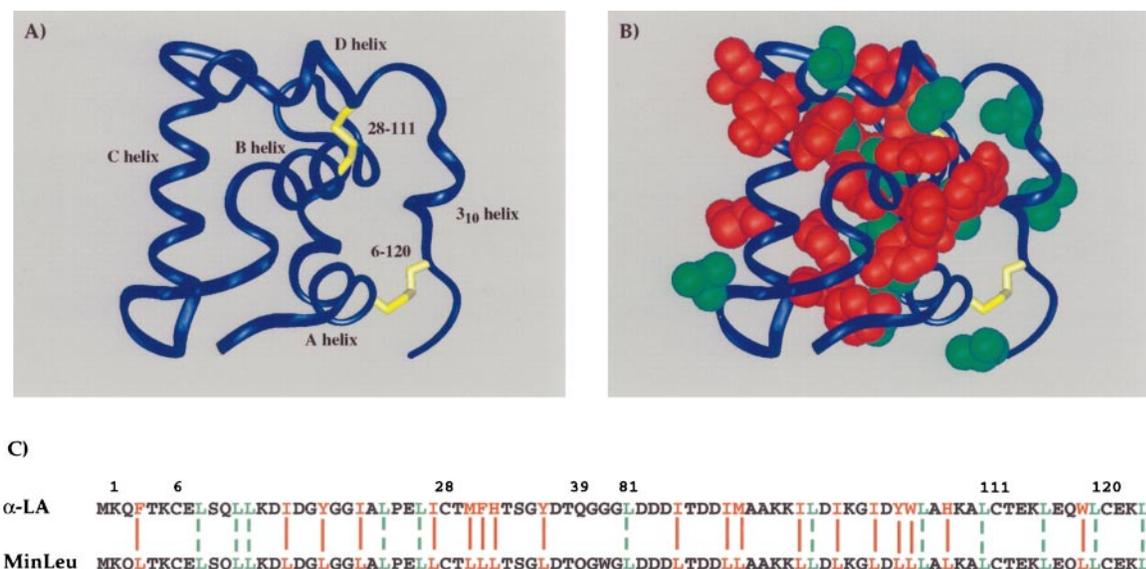


FIG. 1. Hydrophobic sequence minimization of the helical domain of α -lactalbumin. (A) Schematic representation of the helical domain of α -LA based on the structure of full-length human α -LA (55), with helices and disulfide bonds indicated. (B) Hydrophobic side chains in the helical domain of α -LA are highlighted by a space-filling representation. The domain is oriented as in A, with wild-type leucines shown in green and all other hydrophobic side chains, mutated to leucine in the current study, shown in red. (C) Amino acid sequences of the wild-type and minimized helical domains. Residues are numbered in the context of full-length α -LA and colored as in B. Solid red lines align those hydrophobic amino acids in α -LA that are changed to leucine in MinLeu, and dashed green lines align wild-type leucines.

MinLeu corresponds to the helical domain of α -LA minimized with respect to amino acid type, regardless of residue location in the three-dimensional structure of the molecule (Fig. 1). Thus, we have removed sequence-specific hydrophobic interactions from the helical domain of α -LA and replaced them with general hydrophobic interactions mediated by leucines. Notably, our minimization scheme is not biased by decisions based on the native structure. A total of 31 of the 86 amino acid residues in MinLeu are leucine (36%), of which 12 are leucine in the wild-type α -LA sequence.

Lack of Aggregation. Molten globules are prone to aggregation. Moreover, the substantial changes introduced by hydrophobic minimization may yield molecules that associate nonspecifically. Therefore, it is important to determine whether MinLeu is monomeric under the conditions used in this study. We measured the oligomerization state of MinLeu and its variants by sedimentation equilibrium. Significantly, all species studied here are monomeric (Fig. 2). Thus, the structural properties determined in this study are not artifacts of aggregation.

Backbone Topology. We determined whether the minimized hydrophobic sequence adopts a native-like tertiary fold (backbone topology) by performing disulfide exchange experiments (Fig. 3). The four cysteine residues in MinLeu can pair in three possible ways, corresponding to one native and two nonnative structures. The equilibrium distribution of the three disulfide variants of MinLeu reflects the intrinsic propensity of the chain to sample native and nonnative folds. Importantly, the

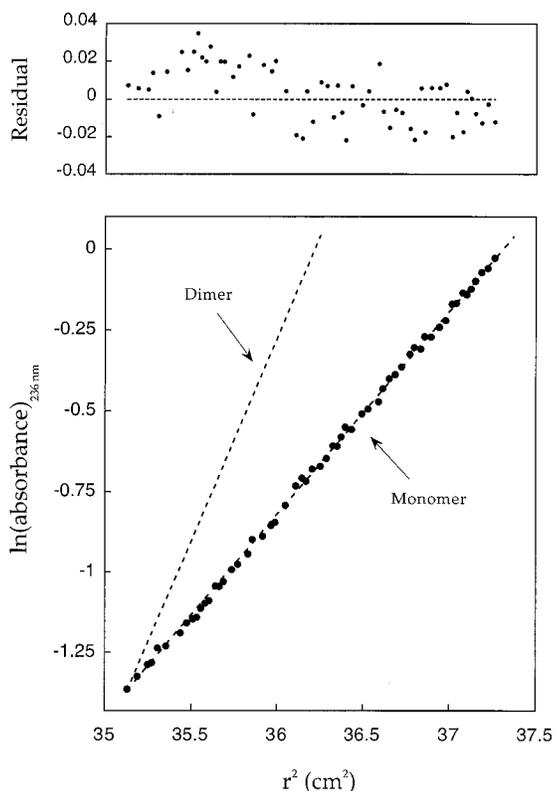


FIG. 2. Sedimentation equilibrium studies of MinLeu and its variants indicate that all species are monomeric. Representative data for native MinLeu (native disulfide bonds) are plotted as $\ln(\text{absorbance})$ against the square of the radius from the axis of rotation. The slope is proportional to the molecular weight: dashed lines indicate calculated slopes for monomeric and dimeric species. Deviations from the calculated values are plotted as residuals (Upper). The composite ratio (of all data sets for a given species) of observed-to-expected molecular weights for native MinLeu, nonnative disulfide isomer [6–28; 111–120], nonnative isomer [6–111; 28–120], and the alanine variant are 1.07, 1.08, 1.10, and 1.10, respectively.

same ratios of disulfide isomers are obtained at extended time intervals and starting from different initial species, confirming that equilibrium was established.

Remarkably, 97% of MinLeu adopts native disulfide pairings. This suggests that the minimized sequence has a strong preference for a native-like fold. For comparison, 90% of the wild-type sequence (α -Domain) adopts native disulfide pairings (28). In contrast, under denaturing conditions only 9% of MinLeu adopts native disulfide pairings. The disulfide distribution of fully denatured MinLeu can be predicted by using a random walk model for an unfolded polypeptide chain (40, 41), and our experimental results are in agreement with the predicted distribution. Thus, the preference of MinLeu for native disulfide bonds under native conditions is the result of structure, as opposed to intrinsic differences in thiol reactivity.

Overall Helix Content. We assessed the helix content of MinLeu with CD spectroscopy to determine whether sequence minimization yields a molecule that still contains secondary structure consistent with a native-like fold (Fig. 4). Native MinLeu (i.e., with native α -LA disulfide pairings) has substantial helix content. The native forms of MinLeu and the full-length α -LA molten globule, which consists of a native-like helical domain and an unstructured β -sheet domain, have similar helix content when normalized for the number of residues in each protein. Nonnative folds, imposed on MinLeu by formation of nonnative disulfide pairings, result in lower overall helix content than in native MinLeu, indicating that the high level of secondary structure in MinLeu is linked to formation of a native-like fold. The spectra of nonnative MinLeu disulfide isomers have striking similarities to those of the corresponding nonnative isomers of the full-length α -LA molten globule (Fig. 4).

Localization of Secondary Structure. To investigate the location of helices and loops in MinLeu, we performed limited proteolysis on the native and nonnative disulfide isomers of MinLeu. Proteolysis is a useful tool with which to assess the location of exposed flexible segments in a protein (42). Three proteases that nonspecifically cleave peptide bonds (elastase, subtilisin, and proteinase K) were chosen to avoid sequence-specific cleavage bias. We identified cleavage sites by reverse-phase HPLC separation and mass spectrometry of proteolysis products, limiting our analysis to initial cuts of the polypeptide chain (Fig. 5). By analyzing initial cleavage sites, we identified regions of the polypeptide chain that are likely to be disordered in the native and nonnative variants of MinLeu.

Proteolysis of native MinLeu occurs in regions of the polypeptide chain corresponding to loops in the native α -LA structure, except for one cleavage site that lies in the region corresponding to the C-helix (Fig. 6). Thus, proteolysis is consistent with the presence of native-like secondary structure, except for the C-helix. Significantly, in the wild-type α -LA molten globule, the A-, B-, D-, and 3_{10} -helices, but not the C-helix, are structured, as assessed by NMR measurements of hydrogen exchange and proline-scanning mutagenesis studies (26, 31, 43). Thus, one would predict that, even for the wild-type α -Domain sequence, cleavage may occur in the region corresponding to the C-helix.

In contrast, the proteolysis profiles of the nonnative variants of MinLeu show additional cuts inconsistent with native-like helices, specifically in regions of the polypeptide chain encompassing, in the native structure, the B- and 3_{10} -helices (both nonnative species) and the D-helix (one nonnative species). Notably, the B-helix is substantially buried in the native structure of α -LA. Proteolysis in the middle of this helix, as observed in one nonnative species, is highly inconsistent with a native-like fold. It is striking that regions corresponding to native helices are cleaved in nonnative variants of MinLeu but not in the native species. Thus, in native MinLeu the lack of proteolysis in these regions likely is because of the presence of protective structure, as opposed to the lack of a cleavage site.

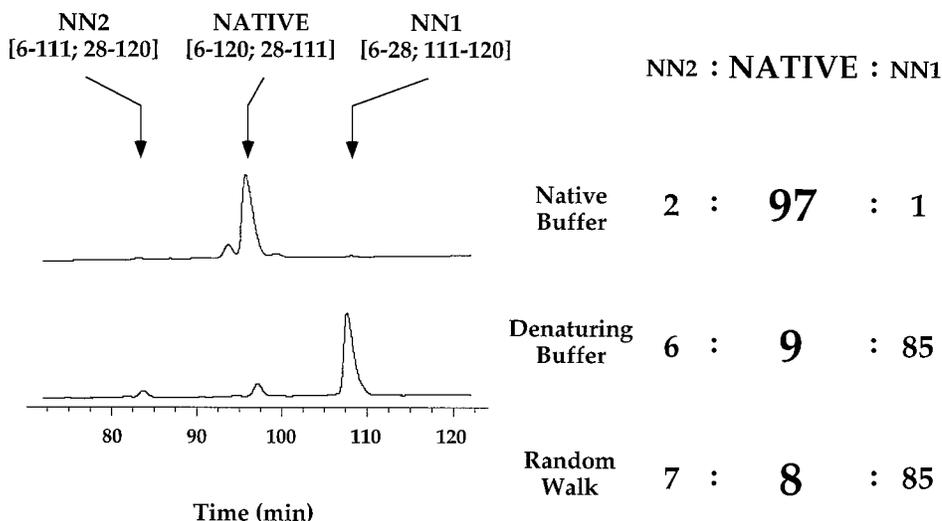


FIG. 3. Disulfide exchange studies indicate that MinLeu has a strong preference for a native-like fold. Reverse-phase HPLC chromatograms of equilibrium disulfide populations under native and denaturing (6 M guanidinium isothiocyanate) conditions are shown, with each disulfide species indicated. The relative populations of disulfide species under native and denaturing conditions, as well as that calculated for an unfolded polypeptide chain by using a random walk model, are indicated. The same results ($\pm 4\%$) were obtained at 4°C as at room temperature.

Our results indicate that proteolysis can differentiate between native and nonnative structures and confirms that formation of nonnative disulfide pairings disrupts native-like secondary structure.

Structure in the Absence of Disulfide Constraints. To examine the structure of the MinLeu sequence in the absence of topological constraints imposed by disulfide bonds, we constructed a variant of MinLeu in which all four cysteines are changed to alanine. This alanine variant can freely sample all possible structures in the absence of disulfide crosslinks. Importantly, the CD spectrum of the alanine variant is very similar to that of native MinLeu, as opposed to nonnative structures (Fig. 4), indicating that native-like helix content is intrinsic to the MinLeu sequence and is not imposed by disulfide bonds. Moreover, the alanine variant of MinLeu has a proteolysis profile similar to that of native MinLeu (Fig. 5). Thus, the native-like structure in MinLeu exists even in the absence of disulfide bonds.

DISCUSSION

Hydrophobic Amino Acid Diversity. Our results suggest that specific hydrophobic amino acid diversity may not be necessary to determine a protein’s overall fold, as represented by the molten globule. Although we have substituted all of the hydrophobic amino acids in the helical domain of α -LA with leucine, the resulting molecule retains many features consistent with a native-like fold. MinLeu has a strong preference for a native-like backbone topology, as determined by disulfide exchange studies. In addition, MinLeu is monomeric and has a helix content similar to that present in the helical domain of the full-length α -LA molten globule (both MinLeu and the α -LA molten globule lack a well formed C-helix). Nonnative forms of MinLeu have CD spectra similar to the analogous nonnative forms of the full-length α -LA molten globule. Finally, proteolysis studies of MinLeu show protection patterns consistent with native-like secondary structure. Thus, nonspecific hydrophobic interactions may play a significant role in determining the global fold of a protein.

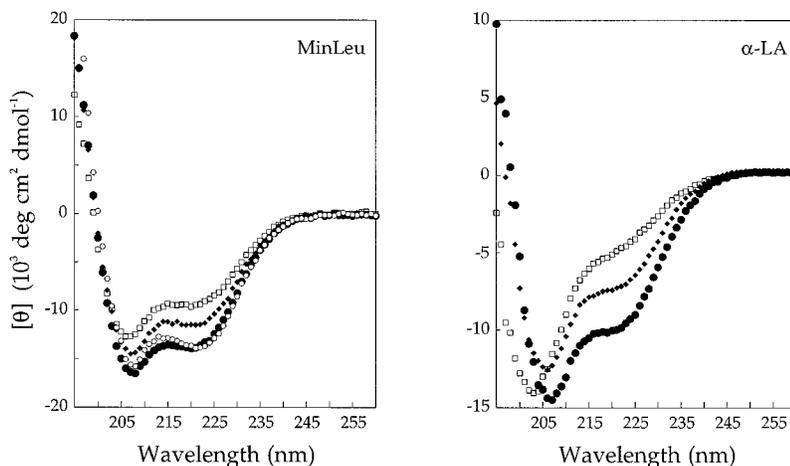
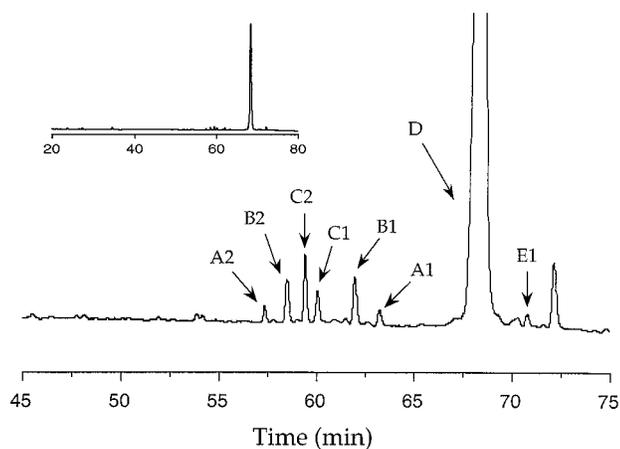


FIG. 4. Circular dichroism spectra of native, nonnative, and alanine variants of MinLeu (Left) and the full-length α -LA molten globule (Right). The native species (native disulfide bonds) is depicted by large, solid circles. The two nonnative disulfide variants [6–111; 28–120] and [6–28; 111–120] are depicted by small, solid diamonds and open squares, respectively. The alanine variant of MinLeu is depicted by small, open circles. CD comparisons are made between analogous MinLeu and full-length α -LA species because the isolated wild-type helical domain (α -Domain) is less structured than the helical domain in the context of the full-length molten globule. $[\theta]_{222}$ of α -Domain is less than that expected from the full-length α -LA molten globule, normalized to the length of α -Domain (28, 30).



Peak	Fragment	Predicted Mass	Observed Mass
A2	D82-End	4605	4604
B2	Link3-End	4775	4775
C2	Link1-End	5018	5018
C1	M0-Q39	4223	4223
B1	M0-Link2	4467	4466
A1	M0-L81	4637	4636
D	Uncut	9223	9223
E1	M0-E121	8979	8980

FIG. 5. Proteolysis of the alanine variant of MinLeu with proteinase K. Products of an early digestion time point, separated by reverse-phase HPLC, are shown. The region of the HPLC chromatogram from 45 to 75 min is expanded, with the full chromatogram shown in the *Inset*. Initial cuts were identified as those cleavage sites that create two complementary digestion products spanning the entire MinLeu sequence. It is possible that complementary fragments are generated from secondary digestion of larger initial fragments. This is unlikely, however, because we have analyzed very early digestion time points in which we do not observe other secondary digestion fragments without a complementary product. The identities of individual species, their predicted masses, and their masses measured by mass spectrometry, are indicated. Amino acid residues are numbered as in the context of full-length α -LA, with the N-terminal methionine designated as residue zero, and the linker residues designated as Link1, -2, and -3.

Given that MinLeu is a molten globule, our characterization of MinLeu is limited to low-resolution structural assays. Thus, we have not obtained high-resolution NMR or x-ray crystallographic structural data that would describe the structure of MinLeu more definitively. Our proteolysis data indicate that some elements of native structure are not present in MinLeu, because the carboxyl-terminal half of the region corresponding to the C-helix in the native structure is sensitive to protease digestion. In addition, MinLeu, although similar in helix content to that expected based on the spectrum of the full-length α -LA molten globule, is more helical than the isolated wild-type α -Domain. It should be noted, however, that α -Domain is less structured in isolation than in the context of the full-length α -LA molten globule (see Fig. 4 legend). Additional experiments, including the possibility of obtaining residue-specific information by multidimensional NMR techniques as was recently achieved for the α -LA molten globule (32), may further establish the degree of native structure that is present in MinLeu.

Specific Side-Chain Packing. Partially folded structures can range in orderliness from species with noncooperative thermal denaturations and poor NMR chemical shift dispersion, such as the α -LA molten globule, to highly ordered structures that have cooperative thermal denaturations and substantial NMR chemical shift dispersion (44, 45). The role of hydrophobic side-chain diversity in stabilizing and specifying this range of

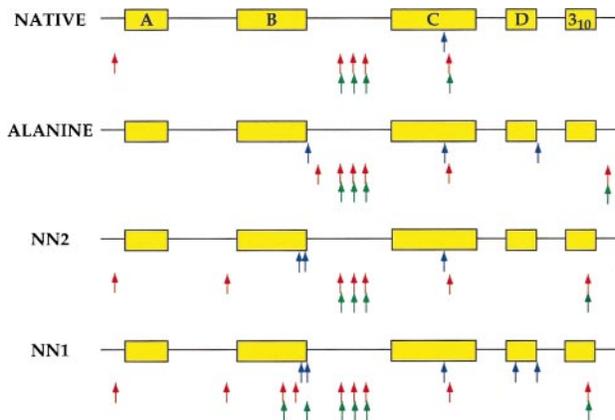


FIG. 6. Summary of proteolysis patterns for the native, nonnative, and alanine variants of MinLeu. The polypeptide chain is represented schematically, with the locations of each native helix highlighted in yellow. Proteolysis sites for each variant of MinLeu are indicated by rows of arrows beneath the polypeptide chain. Elastase, subtilisin, and proteinase K cuts are shown by blue, red, and green arrows, respectively (*Top to Bottom*).

structures may differ. Our results suggest that specific packing interactions are not required for the formation of a native overall topology by molten globules. Thus, nonspecific hydrophobic interactions may help determine a protein's overall fold early in the folding process. Subsequently, specific side-chain packing may order loosely formed native-like structures arising from nonspecific hydrophobic interactions.

Our studies do not preclude the possibility that molten globules contain specific packing interactions. Indeed, studies of the apomyoglobin, cytochrome *c*, and staphylococcal nuclease molten globules indicate that specific packing of core hydrophobic residues exists in these molten globules (8–13). In addition, we have obtained evidence that some native-like core hydrophobic packing may exist in the α -LA molten globule (46).

One might envision that the multiple leucine residues in the core of MinLeu may adopt a specific packing arrangement, as has been observed in variants of T4 lysozyme (23) and the src SH3 domain (47) that consist of well packed but simplified hydrophobic cores lacking significant amino acid diversity. It should be noted, however, that our leucine mutations have resulted in the net loss of 25 side-chain methylene groups, of which 7 are fully buried, 12 are partially buried, and 6 are exposed, as determined by solvent-accessible surface area calculations (see *Methods*). This type of large loss in buried nonpolar surface area is expected to significantly destabilize well packed native structure to the extent that MinLeu is not expected to fold to a stable native structure, even in the context of full-length α -LA (48, 49). Thus, it is unlikely that there is extensive specific packing of the leucine side chains in the MinLeu molten globule.

Interestingly, the molten globule formed by MinLeu is more stable than that formed by wild-type α -Domain. MinLeu must be subjected to 6 M guanidinium isothiocyanate to be fully unfolded, as observed in disulfide exchange assays. In addition, the effective concentration for formation of the 28–111 disulfide bond is higher in MinLeu than in wild-type α -Domain under native conditions (46). Other proteins with cores composed of only leucine have been observed to be highly stable molten globule-like structures (21, 22). It is possible that our sequence minimization has removed unfavorable interactions present in the wild-type α -LA molten globule or has increased the stability of the molten globule, for example, by increasing overall helix propensity.

Binary Patterning of Amino Acid Types. It is thought that a binary pattern of hydrophobic and hydrophilic amino acid

types may be sufficient to determine a protein's overall fold (7, 16, 50). Studies of libraries of protein sequences indicate that a large combination of appropriately patterned hydrophobic and hydrophilic residues can be tolerated by the four-helix bundle fold (50). Our studies of MinLeu also support the binary patterning hypothesis. If patterning of amino acid types is sufficient to determine a protein's fold, our sequence minimization approach could be extended successfully to other residues, resulting in a minimized model of the helical domain of α -LA consisting of very few amino acid types, representing perhaps hydrophobic, polar, and charged amino acids. In this light, it is interesting that a substantial portion of protein libraries composed of only glutamine, leucine, and arginine can form helical structures with properties similar to natural proteins containing significant amino acid diversity (19, 20), and that the src SH3 domain fold can be encoded largely by only five different amino acids (47).

Alternatively, it is possible that the determinants of a protein's fold lie in specific polar and/or charged amino acids. Our leucine substitutions have not affected either buried polar residues or a substantial portion of the hydrophilic protein surface. Polar residues buried in the cores of proteins have been found to contribute to structural uniqueness and specificity (51, 52), although sometimes they can be replaced by hydrophobic residues without significant structural effects (53). Recent experiments in which multiple residues in the core of the phage 434 cro protein have been replaced by leucine without significant perturbation of the protein structure have been interpreted to indicate that surface residues may play a more active role in protein conformation than assumed traditionally (24), although a large body of data indicates that surface residues play a minor role in determining protein structure (49, 54).

The Role of the Molten Globule in Protein Folding. Our results emphasize how formation of the molten globule can assist protein folding (1–4). Hydrophobic collapse of the polypeptide chain yields a low-resolution native-like structure, thereby greatly reducing the conformational space to be searched by the polypeptide chain. The driving force for formation of such native-like structures may be the distribution of hydrophobic amino acids along the polypeptide chain. Specific packing interactions resulting from hydrophobic side-chain diversity and packing complementarity may stabilize loosely ordered structure but do not appear to be necessary for establishing an overall native-like fold.

We thank James Pang for help with mass spectrometry and Zheng-yu Peng, Brenda Schulman, Pehr Harbury, Martha Oakley, and members of the Kim lab for helpful discussions and comments on the manuscript. This work was supported by the Howard Hughes Medical Institute.

- Ptitsyn, O. B. (1992) in *Protein Folding*, ed. Creighton, T. E. (Freeman, New York), pp. 243–300.
- Kuwajima, K. (1989) *Proteins Struct. Funct. Genet.* **6**, 87–103.
- Dobson, C. M. (1992) *Curr. Opin. Struct. Biol.* **2**, 6–12.
- Peng, Z.-y., Wu, L. C., Schulman, B. A. & Kim, P. S. (1995) *Phil. Trans. R. Soc. Lond. B* **348**, 43–47.
- Ponder, J. W. & Richards, F. M. (1987) *J. Mol. Biol.* **193**, 775–791.
- Richards, F. M. & Lim, W. A. (1993) *Q. Rev. Biophys.* **26**, 423–298.
- Cordes, M. H. J., Davidson, A. R. & Sauer, R. T. (1996) *Curr. Opin. Struct. Biol.* **6**, 3–10.
- Marmorino, J. L. & Pielak, G. J. (1995) *Biochemistry* **34**, 3140–3143.
- Colon, W. & Roder, H. (1996) *Nat. Struct. Biol.* **3**, 1019–1025.
- Colon, W., Elove, G. A., Wakem, L. P., Sherman, F. & Roder, H. (1996) *Biochemistry* **35**, 5538–5549.
- Kay, M. S. & Baldwin, R. L. (1996) *Nat. Struct. Biol.* **3**, 439–445.
- Lin, L., Pinker, R. J., Forde, K., Rose, G. D. & Kallenbach, N. R. (1994) *Nat. Struct. Biol.* **1**, 447–452.
- Carra, J. H., Anderson, E. A. & Privalov, P. L. (1994) *Protein Sci.* **3**, 952–959.
- Ptitsyn, O. B. (1996) *Nat. Struct. Biol.* **3**, 488–490.
- Behe, M. J., Lattman, E. E. & Rose, G. D. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 4195–4199.
- Dill, K. A., Bromberg, S., Yue, K., Fiebig, K. M., Yee, D. P., Thomas, P. D. & Chan, H. S. (1995) *Protein Sci.* **4**, 561–602.
- Huang, E. S., Subbiah, S. & Levitt, M. (1995) *J. Mol. Biol.* **252**, 709–720.
- West, M. W. & Hecht, M. H. (1995) *Protein Sci.* **4**, 2032–2039.
- Davidson, A. R. & Sauer, R. T. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 2146–2150.
- Davidson, A. R., Lumb, K. J. & Sauer, R. T. (1995) *Nat. Struct. Biol.* **2**, 856–864.
- Regan, L. & DeGrado, W. F. (1988) *Science* **241**, 976–978.
- Munson, M., Balasubramanian, S., Fleming, K. G., Nagi, A. D., O'Brien, R., Sturtevant, J. M. & Regan, L. (1996) *Protein Sci.* **5**, 1584–1593.
- Gassner, N. C., Baase, W. A. & Matthews, B. W. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 12155–12158.
- Desjarlais, J. R. & Handel, T. M. (1995) *Protein Sci.* **4**, 2006–2018.
- Dolgikh, D. A., Abatur, L. V., Bolotina, I. A., Brazhnikov, E. V., Bychkova, V. E., Gilmanshin, R. I., Lebedev, Y. O., Semisotnov, G. V., Tikupulo, E. I. & Ptitsyn, O. B. (1985) *Eur. Biophys. J.* **13**, 109–121.
- Alexandrescu, A. T., Evans, P. A., Pitkeathly, M., Baum, J. & Dobson, C. M. (1993) *Biochemistry* **32**, 1707–1718.
- Xie, D., Bhakuni, V. & Friere, E. (1991) *Biochemistry* **30**, 10673–10678.
- Peng, Z.-y. & Kim, P. S. (1994) *Biochemistry* **33**, 2136–2141.
- Kuwajima, K. (1996) *FASEB J.* **10**, 102–109.
- Wu, L. C., Peng, Z.-y. & Kim, P. S. (1995) *Nat. Struct. Biol.* **2**, 281–286.
- Schulman, B. A. & Kim, P. S. (1996) *Nat. Struct. Biol.* **3**, 682–687.
- Schulman, B. A., Kim, P. S., Dobson, C. M. & Redfield, C. (1997) *Nat. Struct. Biol.* **4**, 630–634.
- Johnson, M. L., Correia, J. J., Yphantis, D. A. & Halvorson, H. R. (1981) *Biophys. J.* **36**, 575–588.
- Laue, T. M., Shah, B. D., Ridgeway, T. M. & Pelletier, S. L. (1992) in *Computer-Aided Interpretation of Analytical Sedimentation Data for Proteins*, eds. Harding, S. E., Rowe, A. J. & Horton, J. C. (The Royal Society of Chemistry, Cambridge, U.K.), pp. 90–125.
- Edelhoch, H. (1967) *Biochemistry* **6**, 1948–1954.
- CCP4 (1994) *Acta Cryst.* **D50**, 760–763.
- Eisenberg, D. & McLachlan, A. D. (1986) *Nature (London)* **319**, 199–203.
- Richardson, J. S. & Richardson, D. C. (1988) *Science* **240**, 1648–1652.
- Presta, L. G. & Rose, G. D. (1988) *Science* **240**, 1632–1641.
- Snyder, G. H. (1987) *Biochemistry* **26**, 688–694.
- Kauzman, W. (1959) in *Sulfur in Proteins*, ed. Benesch, R. (Academic, New York), pp. 93–108.
- Fontana, A., de Laureto, P. P., De Felippis, V., Scaramella, E. & Zamboni, M. (1997) *Folding Des.* **2**, R17–R26.
- Schulman, B. A., Redfield, C., Peng, Z.-y., Dobson, C. M. & Kim, P. S. (1995) *J. Mol. Biol.* **253**, 651–657.
- Redfield, C., Smith, R. A. G. & Dobson, C. M. (1994) *Nat. Struct. Biol.* **1**, 23–29.
- Feng, Y., Sliagar, S. G. & Wand, A. J. (1994) *Nat. Struct. Biol.* **1**, 30–35.
- Wu, L. C. (1997) Dissertation (Massachusetts Institute of Technology, Cambridge, MA).
- Riddle, D. S., Santiago, J. V., Bray-Hall, S. T., Doshi, N., Grantcharova, V. P., Qian, Y. & Baker, D. (1997) *Nat. Struct. Biol.* **4**, 805–809.
- Lim, W. A. & Sauer, R. T. (1991) *J. Mol. Biol.* **219**, 359–376.
- Matthews, B. W. (1995) *Adv. Protein Chem.* **46**, 249–278.
- Kamtekar, S., Schiffer, J. M., Xiong, H., Babik, J. M. & Hecht, M. H. (1993) *Science* **262**, 1680–1685.
- Lumb, K. J. & Kim, P. S. (1995) *Biochemistry* **34**, 8642–8648.
- Hendsch, Z. S. & Tidor, B. (1994) *Protein Sci.* **3**, 211–226.
- Waldburger, C. D., Schildbach, J. F. & Sauer, R. T. (1995) *Nat. Struct. Biol.* **2**, 122–128.
- Bowie, J. U., Reidhaar-Olson, J. R., Lim, W. A. & Sauer, R. T. (1990) *Science* **247**, 1306–1310.
- Acharya, K. R., Ren, J., Stuart, D. I., Phillips, D. C. & Fenna, R. E. (1991) *J. Mol. Biol.* **221**, 571–581.