Peptide models of local and long range interactions in the molten globule state of human α-lactalbumin

ARTICLE *in* JOURNAL OF MOLECULAR BIOLOGY · FEBRUARY 1998

Impact Factor: 4.33 · DOI: 10.1006/jmbi.1998.2099 · Source: PubMed

3 AUTHORS, INCLUDING:

Robert Fairman
Haverford College
78 PUBLICATIONS **3,733** CITATIONS

SEE PROFILE

Article No. mb982099 *J. Mol. Biol.* (1998) **283**, 279–291

JMB

Peptide Models of Local and Long-range Interactions in the Molten Globule State of Human α-Lactalbumin

Stephen J. Demarest[1], Robert Fairman[2] and Daniel P. Raleigh[1,3]*

[1]*Department of Chemistry State University of New York at Stony Brook, Stony Brook NY 11794-3400, USA*

[2]*Department of Molecular Cellular, and Developmental Biology, Haverford College Haverford, PA 19041, USA*

[3]*Graduate Program in Biophysics and Graduate Program in Molecular and Cellular Biology, State University of New York at Stony Brook, Stony Brook NY 11794-3400, USA*

**Corresponding author*

α-Lactalbumin, a small calcium-binding protein, forms an equilibrium molten globule state under a variety of conditions. A set of four peptides designed to probe the role of local interactions and the role of potential long-range interactions in stabilizing the molten globule of α-lactalbumin has been prepared. The first peptide consists of residues 20 through 36 of human α-lactalbumin and includes the entire B-helix. This peptide is unstructured in solution as judged by CD. The second peptide is derived from residues 101 through 120 and contains both the D and 3_{10} helices. When this peptide is crosslinked *via* the native 28 to 111 disulfide to the B-helix peptide, a dramatic increase in helicity is observed. The crosslinked peptide is monomeric, as judged by analytical ultracentrifugation. The peptide binds 1-anilinonaphthalene-8-sulphonate (ANS) and the fluorescence emission maximum of the construct is consistent with partial solvent exposure of the tryptophan residues. The peptide corresponding to residues 101 to 120 adopts significant non-random structure in aqueous solution at low pH. Two hydrophobic clusters, one involving residues 101 through 104 and the other residues 115 through 119 have been identified and characterized by NMR. The hydrophobic cluster formed by residues 101 through 104 is still present in a smaller peptide containing only residues 101 to 111 of α-lactalbumin. The cluster also persists in 6 M urea. A non-native, pH-dependent interaction between the Y103 and H107 side-chains that was previously identified in the acid-denatured molten globule state was examined. This interaction was found to be more prevalent at low pH and may therefore be an example of a local interaction that stabilizes preferentially the acid-induced molten globule state.

Keywords: molten globule state; protein folding; α-lactalbumin; peptide structure; denatured state

Introduction

Many proteins are believed to fold *via* a compact intermediate that contains significant secondary structure (Creighton, 1997; Kuwajima, 1989, 1996; Privalov, 1996; Ptitsyn, 1995). These kinetic intermediates are thought to be similar to the equilibrium molten globule state, which can be stabilized under mild non-native conditions (Ikeguchi *et al.*, 1986; Kuwajima *et al.*, 1989), and there is considerable interest in elucidating the interactions that stabilize these states. The classic description of the equilibrium molten globule is a state that is compact and contains a high degree of native-like secondary structure, but lacks the fixed tertiary contacts of the native state (Dolgikh *et al.*, 1981; Kuwajima *et al.*, 1976). Detailed information, such as knowledge of the specific contacts and

Abbreviations used: ANS, 1-anilinonaphthalene-8-sulphonate; CD, circular dichroism; DQF-COSY, double quantum filtered correlated spectroscopy; E.COSY, exclusive correlation spectroscopy; FMOC, 9-fluorenylmethyloxycarbonyl; PAL, 5-(4′-Fmoc-aminomethyl-3′,5′-dimethoxyphenoxy) valeric acid; ppb, parts per billion; ppm, parts per million; ROE, rotating frame nuclear Overhauser effect; ROESY, rotating frame nuclear Overhauser effect spectroscopy; TBTU, 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium tetrafluoroborate; TFA, trifluoroacetic acid; TFE, trifluoroethanol; TOCSY, total correlated spectroscopy; TSP, 3-(trimethylsilyl) propionate.

E-mail address of the corresponding author: draleigh@ccmail.sunysb.edu

0022–2836/98/410279–13 $30.00/0

interactions that stabilize molten globule states, is difficult to obtain, because their NMR spectra are poorly dispersed with broadened lines (Dobson, 1992).

The A-state of α-lactalbumin is populated under acidic conditions and is one of the best characterized of molten globule states. The molten globule state of α-lactalbumin can be populated at neutral pH by the addition of moderate concentrations of denaturants, by the removal of Ca^{2+} and other ionic salts, or by thermal denaturation (Kuwajima *et al.*, 1976; Maruyama *et al.*, 1977; Hiraoka & Sugai, 1984; Haynie & Friere, 1993). α-Lactalbumin is a 14 kDa protein whose structure is virtually identical with its evolutionary partner lysozyme (Acharya *et al.*, 1989; Phillips, 1974). It contains two subdomains: one is a primarily helical domain comprised of residues 1 to 39 and 81 to 123 (the α-domain) and the other is a predominantly sheet and coil domain that encompasses the remainder of the molecule (the β-domain). All of the available evidence indicates that the α-domain is structured and the β-domain is unstructured in the molten globule state (Baum *et al.*, 1989; Peng & Kim, 1994; de Laureto *et al.*, 1995; Schulman & Kim, 1996). Kim and co-workers have shown that a peptide model of the α-domain consisting of residues 1 to 39 and 81 to 123 connected by a three-residue glycine linker can fold into a molten globule-like state. This construct is highly helical and shows a strong preference for native-like disulfide bond formation (Peng & Kim, 1994; Wu *et al.*, 1995). The α-domain includes four α-helices, denoted A through D, and a short 3_{10} helix. It is not known if smaller constructs are capable of forming a molten globule-like structure or if all of the helices are required (Schulman & Kim, 1996; Kuhlman *et al.*, 1997).

Interactions between the B and D helices are likely to be important in stabilizing the molten globule state. In the native state, the B-helix is crosslinked *via* the 28 to 111 disulfide to the region that includes the D-helix and the 3_{10} helix. Studies of mutants of α-lactalbumin containing only a single disulfide demonstrate that the 28 to 111 disulfide has the highest propensity to form in the molten globule state and hence indicate a high propensity to form structure in this region (Peng *et al.*, 1995). Computational studies have also suggested that association of these regions of the protein is likely to be an important event in the folding process. Algorithms that are designed to predict potential folding pathways by considering the amount of hydrophobic surface area buried upon the folding and packing together of segments of the polypeptide chain predict that the association of the regions of the chain corresponding to the B and D helices is an important event in the folding process of the closely related protein lysozyme (Chelvanayagam *et al.*, 1992; Moult & Unger, 1991).

Here, we report the results of a study of a set of four peptides designed to directly probe the role of local interactions in the regions of the B helix and the D helix as well as the role of potential long-range interactions between these regions in stabilizing the molten globule. We report the results of detailed CD and NMR studies of the conformational tendencies of these peptides at both low and neutral pH and in the presence and absence of organic cosolvents.

Molten globule states are often considered to be stabilized by relatively non-specific interactions; nevertheless, it is clearly important to establish whether there are specific interactions, local or long-range that stabilize these states. If so, are these interactions identical when molten globule states are isolated under different conditions? This question is a particularly important issue for the case of α-lactalbumin, since the molten globule state can be stabilized at both neutral and low pH, where it is known that the structure of the native state differs in the region of the D helix (Acharya *et al.*, 1991; Harata & Muraki, 1992; Pike *et al.*, 1996). Interestingly, in the low pH form of the molten globule state, this region of the protein chain appears to adopt an entirely different conformation involving non-native side-chain interactions (Alexandrescu *et al.*, 1993). Our peptides are well suited to study the role that local interactions play in defining these conformations.

Results

The ribbon diagram of the structure of α-lactalbumin shown in Figure 1 highlights the regions probed by our set of peptides. The first peptide, designated αlac:20-36 consists of residues 20 through 36 of human α-lactalbumin and encompasses the entire B helix as well as several additional residues at the N and C termini. The second peptide, αlac:101-120, contains both the D and 3_{10} helices. A third peptide, designated the BD pair, was prepared and consists of the 20 to 36 fragment crosslinked *via* the native disulfide to the

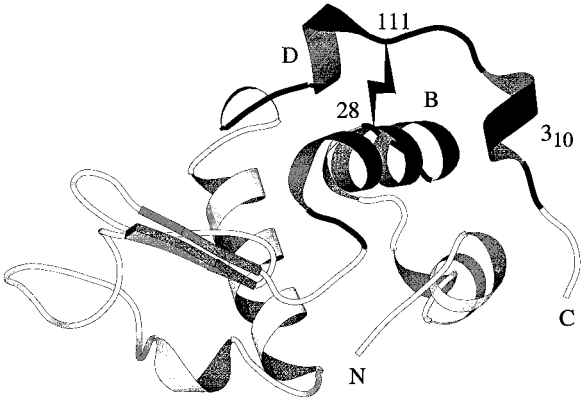

Figure 1. Ribbon diagram of the high-pH crystal structure of human α-lactalbumin (Acharya *et al.*, 1991). The shaded regions correspond to residues 20 to 36 and 101 to 120. The N and C termini are labeled and the 28 to 111 disulfide is indicated. The individual elements of secondary structure are labeled. The diagram was created using the program MOLSCRIPT (Kraulis, 1991).

101 to 120 fragment. Finally, a fourth peptide comprised of residues 101 to 111, denoted αlac:101-111, was synthesized. This peptide was designed to help elucidate the local conformational propensities of the D-helix region. The remainder of this section is organized as follows. We first describe experiments designed to probe the structural preferences of αlac:20-36. Experiments are then described that examine the conformation of αlac:101-111 at low pH. We compare the behavior of the longer αlac:101-120 peptide to that of αlac:101-111, and examine the effect of varying pH on both of these fragments. The conformation of both of these peptides is then analyzed in 30% (v/v) TFE. We conclude by examining the crosslinked BD pair. Comparison of the structural preferences of the BD pair to the conformational tendencies of the two individual peptides provides a sensitive test for any mutually stabilizing interactions.

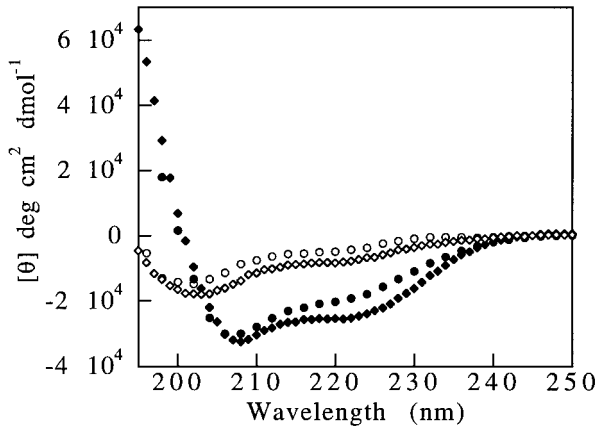

Figure 2. CD spectra of αlac:101-111 (circles) and αlac:101-120 (diamonds) at pH 2.8, 25°C in the presence and in the absence of TFE. Open symbols indicate that the solvent is water. Filled symbols indicate that the solvent is 30% (v/v) TFE.

CD experiments show that αlac:20-36 is unstructured

The CD spectrum of αlac:20-36 recorded at pH 2.8, 25°C indicates that the peptide is largely unstructured. The mean residue ellipticity at 222 nm is only −1000 deg cm^2 dmol^{-1}. This peptide encompasses the B helix and the random-coil nature of the CD spectrum demonstrates that long-range interactions are required to induce structure in this region of α-lactalbumin in the molten globule state. There is no significant change in the CD spectrum over the pH range of 2 to 11 (D. Moriarty, S.J.D., & D.P.R., unpublished results). Unfortunately, the peptide is not soluble above approximately 50 μM, precluding NMR studies.

NMR and CD experiments provide evidence for the formation of non-random structure in αlac:101-111 in aqueous solution at low pH

Concentration-dependent CD and NMR experiments show that αlac:101-111 is monomeric up to at least 6 mM peptide concentration at pH 2.8. CD experiments clearly demonstrate that this peptide forms non-random structure in aqueous solution. The CD spectrum of αlac:101-111 recorded at pH 2.8 and 25°C in aqueous solution shows a substantial signal at 222 nm (−6000 deg cm^2 dmol^{-1}; Figure 2). Due to the small size of the peptide and possible contributions from the aromatic residues to the intensity observed at 222 nm, interpretation of the signal in terms of fraction helix is complicated (Chakrabartty *et al.*, 1993). Nonetheless, non-random structure in the peptide is evident from the shoulder at 222 nm in the spectrum.

NMR experiments provide additional evidence for non-random structure in this peptide. The spectra are well resolved and can be easily assigned using standard methods. The fingerprint region of the TOCSY spectrum is shown in Figure 3A and the deviation of the C$^\alpha$H chemical shifts from random coil values is included in Figure 4(A) (Wishart *et al.*, 1992, 1995). The first eight C$^\alpha$H resonances all show an upfield shift from random coil values. The C$^\alpha$H resonances of residues D102 to K108 are all shifted by at least 0.14 ppm and several of these residues have significantly larger deviations, which presumably reflects a propensity to preferentially sample the helical region of ϕ,ψ space. Several amide protons are shifted well upfield of random coil values (Wüthrich, 1986). The amide proton chemical shift values of W104 (7.80 ppm), A106 (7.77 ppm) and L105 (7.58 ppm) at 25°C are particularly noteworthy. Upfield shifts of this nature are often the result of interactions with nearby aro-

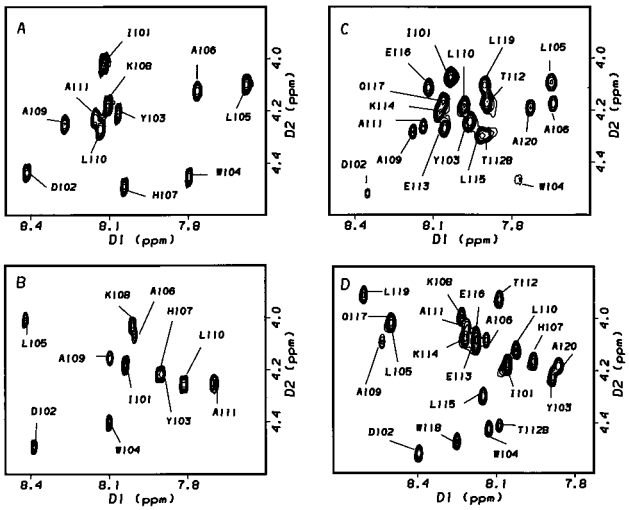

Figure 3. The fingerprint region of the TOCSY spectra of αlac:101-111 and αlac:101-120. A, αlac:101-111 at 25°C, pH 2.8 in water. B, αlac: 101-111 at 25°C, pH 2.8, 30% (v/v) TFE. C, αlac:101-120 at 45°C in water, pH 2.8 (the K108 crosspeak is unassigned and the H107 and W118 crosspeaks are saturated). D, αlac:101-120 at 25°C, pH 2.8, 30% (v/v) TFE.

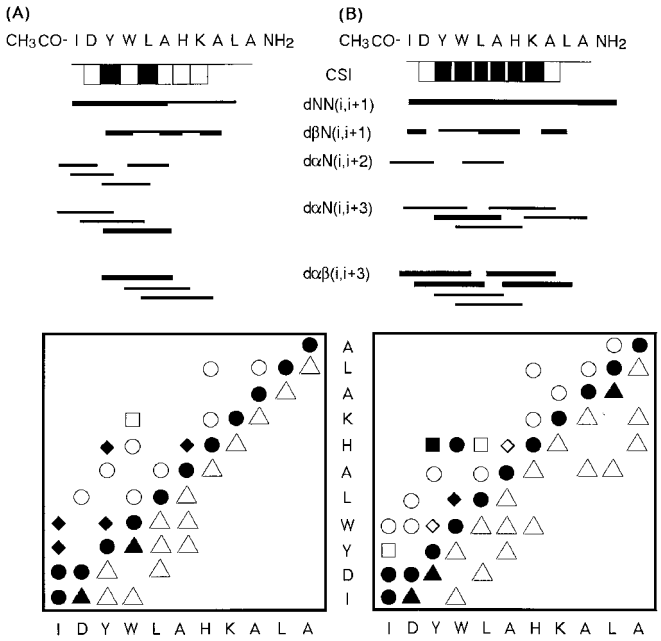

Figure 4. Deviations of the $C^{\alpha}H$ chemical shifts from random coil values, main-chain to main-chain ROEs, and ROESY contact map of (A) αlac:101-111 in aqueous solution at pH 2.8, 5°C and (B) αlac:101-111 in 30% TFE at pH 2.8, 25°C. The deviations of the $C^{\alpha}H$ chemical shifts are displayed with downward open rectangles indicating shifts of at least −0.1 ppm from random-coil values and filled rectangles indicating $C^{\alpha}H$ chemical shifts that are shifted by at least −0.2 ppm from random-coil values. In the contact maps, triangles below the diagonal indicate main-chain to main-chain ROEs. The symbols above the diagonal indicate side-chain to side-chain (squares), side-chain to main-chain (circles), or both side-chain to side-chain and side-chain to main-chain (diamonds) ROEs. Open symbols (both above and below the diagonal) indicate one or more weak ROEs and filled symbols indicate at least two or more ROEs of at least medium intensity.

matic residues (Kemmink *et al.*, 1993; Kemmink & Creighton, 1995).

The observed pattern of ROEs provides additional evidence for deviations from random coil behavior in this peptide, particularly for residues 101 through 107. A total of 166 inter- and intraresidue ROEs could be assigned in the ROESY spectrum of this peptide, 130 of which involve residues located in the first seven positions of the sequence. All of the sequential amide to amide ROEs are observed, except for that between the C-terminal two residues. Weak $(i, i+2)_{\alpha N}$, $(i, i+3)_{\alpha N}$ and $(i, i+3)_{\alpha\beta}$ ROEs also indicate that there is some propensity for non-random backbone structure in the N-terminal seven residues. The C-terminal four residues appear to be less ordered. A plot of the observed main-chain to main-chain ROEs is included in Figure 4(A).

A hydrophobic cluster involving residues I101, Y103 and W104 was identified on the basis of a set of ROEs between the side-chains of these residues. A total of 18 interresidue side-chain to side-chain (ten) and side-chain to main-chain (eight) ROEs were observed between I101, Y103 and W104. An additional seven ROE peaks from these hydrophobic residues could not be unambiguously assigned. Two of the seven are between the side-chain of W104 and the side-chain of Y103. The other five are from the indole ring of W104 to methyl resonances of either I101 or L105 and to β and γ resonances. An interaction between the side-chains of Y103 and H107 was identified by ROEs between these two residues. A contact map depicting the pattern of the backbone to side-chain and side-chain to side-chain ROEs is shown in Figure 4(A).

The amide proton chemical shift temperature coefficients offer more evidence of non-random structure in the N-terminal region of αlac:101-111 at pH 2.8. Residues 104 through 108 have small amide temperature coefficients, possibly due to solvent inaccessibility resulting from shielding by the side-chains in the hydrophobic cluster and/or residual hydrogen bonding (Deslauriers & Smith, 1980). Interactions with nearby aromatic rings also may contribute (Kemmink *et al.*, 1993). The remaining six residues have values between −6.8 and −7.3 ppb/K.

The only $^{3}J_{NH\alpha}$ coupling constant that deviates significantly from random-coil values is that for Y103, which has a value of 4.9 Hz. Dobson and coworkers have shown that fluctuations about helical φ,ψ values can, in isolated helical peptides, increase the $^{3}J_{NH\alpha}$ coupling constants above the values expected for a static helix (Bolin *et al.*, 1996). The $^{3}J_{\alpha\beta}$ coupling constants of the AMX spin systems were determined from E.COSY spectra. All of the residues except H107 have $^{3}J_{\alpha\beta}$ coupling constants indicative of conformational averaging. The $^{3}J_{\alpha\beta}$ coupling constants of H107 are 9.9 and 5.6 Hz. Assignment of a specific χ^{1} rotamer was not possible, due to the inconclusive pattern of NOEs (Wagner *et al.*, 1987).

Side-chain to side-chain interactions in αlac:101-111 are not disrupted by the addition of urea

CD and NMR experiments indicate that significant structure remains even in 6 M urea. θ_{222} decreased from −6000 deg cm^{2} dmol^{-1} in the absence of urea to −3800 deg cm^{2} dmol^{-1} in 6 M urea. The CD spectrum of the peptide in 6 M urea does not correspond to what is expected for an unstructured peptide, indicating either residual structure or a significant contribution to the far-UV signal from the aromatic residues. A ROESY experiment in $^{2}H_{2}O$ in 6 M urea provides unambiguous evidence that the hydrophobic cluster involving residues I101, Y103 and W104 as well as the interaction between Y103 and H107 are still intact. The ten side-chain to side-chain ROEs that were observed in water are still present in 6 M urea and are of comparable intensity. Two of the eight side-

chain to main-chain ROEs that were observed in water were to amide protons and hence would not be detectable in ^2H_2O. Of the remaining six ROEs, two are detectable. A set of seven additional side-chain to side-chain ROEs was observed in this region of the peptide in water that could not be unambiguously assigned to particular proton pairs (although some could be assigned to particular residues). Six of these seven ROEs are still present in 6 M urea, including peaks from the indole ring of W104 to the side-chain of Y103. Because the hydrophobic cluster is so robust, it must dictate, in some fashion, the allowable conformations that the peptide backbone can adopt even under denaturing conditions.

Conformational analysis of αlac:101-120 in H_2O provides evidence for non-random structure in the region of residues 115 through 119

Extending the C terminus of the peptide does not appear to significantly perturb the structure formed in αlac:101-111. The addition of nine residues to the C terminus increases the helical content, as judged by CD. The mean residue ellipticity of αlac:101-120 at 222 nm is -9500 deg cm^2 dmol^{-1} (Figure 2). Substitution of Ala for the wild-type Cys at position 111 has no detectable effect on the CD spectra, and all NMR experiments were performed with the Ala mutant to eliminate any potential complications due to disulfide formation.

αlac:101-120 associates at low temperatures at concentrations above 500 μM; however, the peptide is monomeric at NMR concentrations at 45°C and above. At this temperature, the peptide appears to be slightly less structured, as judged by CD, θ_{222} decreases to -7400 deg cm^2 dmol^{-1}. The pattern of the fingerprint peaks of the first 11 residues of αlac:101-120 at 45°C is similar to the pattern found in the fingerprint region of αlac:101-111 at 25°C (Figure 3C). Most of the ROEs between residues 101 through 108 that were observed in the shorter peptide are visible in the longer peptide. Interestingly, the pattern of C$^\alpha$H chemical shifts observed in this peptide are similar to those observed in a peptide corresponding to the same region in hen lysozyme, even though the primary sequences are very different (Yang *et al.*, 1996).

There is evidence for non-random interactions involving residues in the C-terminal region of αlac:101-120. The C$^\alpha$H chemical shifts of four of the five C-terminal residues are shifted upfield of random coil values by at least 0.15 ppm. These shifts are even larger at 25°C; however, the spectra at low temperature are complicated by some self-association. Non-native side-chain to side-chain ROEs are observed between the indole ring of W118 and methyl groups of both L115 and L119, suggestive of a second hydrophobic cluster near the C-terminal end of the peptide. A weak $(i, i+3)_{\alpha N}$ ROE was observed between L115 and W118.

The effect of pH on the structure of αlac:101-111 and αlac:101-120

Unfortunately, αlac:101-120 associates at neutral pH and could not be studied under these conditions. αlac:101-111 is monomeric up to at least 3 mM at pH 7.0 as judged by concentration-dependent CD and NMR measurements. The CD spectrum of αlac:101-111 recorded at pH 7.0 is very similar to the spectrum recorded at pH 2.8, although the value of θ_{222} is slightly more negative, -6600 deg cm^2 dmol^{-1} compared to -6000. The C$^\alpha$H chemical shifts are also very similar; however, there are noticeable shifts in the amide proton resonances and the dispersion in the amide region is considerably reduced. Residues 104 through 106 all experience downfield shifts, and the amide chemical shift temperature coefficients of W104 and L105 are both more negative (i.e. closer to random-coil values) at pH 7.0 than at pH 2.8. These observations suggest that there is some change in the conformational preferences of this region of the peptide.

The effect of varying the pH on the Y103 chemical shifts provides additional indirect evidence for a pH-dependent conformational change. The apparent pK_a values of the ionizable residues were determined by monitoring their chemical shifts from pH 1.4 to pH 11.0. The pK_a values of D102 (3.6) and H107 (6.4) are relatively unperturbed compared to random-coil values. The chemical shift of the 2,6 protons of Y103 display a sigmoidal transition as the pH is varied through the pK_a of the D102 side-chain, shifting approximately 0.12 ppm upfield between pH 2 and 5. As the pH is increased through the pK_a of the H107 side-chain, this resonance shifts back downfield. The H107 imidazole protons also display a minor perturbation as D102 is protonated. Although it is clearly hazardous to associate pH-dependent chemical shift perturbations with specific structural changes, the data do suggest that there is a pH-dependent interaction between Y103 and H107 that may be modulated, in some fashion, by D102. Taken together, the pH-dependent perturbations of the Y103 ring protons, the perturbations of the amide chemical shifts and of the amide chemical shift temperature coefficients indicate that there must be some pH-dependent change in the structure of this region of the peptide.

pH-dependent changes in the H107 $^3J_{\alpha\beta}$ coupling constants are consistent with a weaker Y103-H107 interaction at pH 7.0, 5°C. The difference between the two $^3J_{\alpha\beta}$ coupling constants of H107 is noticeably smaller at pH 7.0 (1.9 Hz) than at pH 2.8 (4.3 Hz), suggesting that rotation about the H107 α-β bond is less hindered at neutral pH. ROESY experiments provide direct evidence that the interaction between Y103 and H107 is weaker at neutral pH (Figure 5). No ROE is observed between these two side-chains at 25°C, pH 7, while a set of four ROEs was observed at pH 2.8. The interactions involving the hydrophobic side-chains also appear

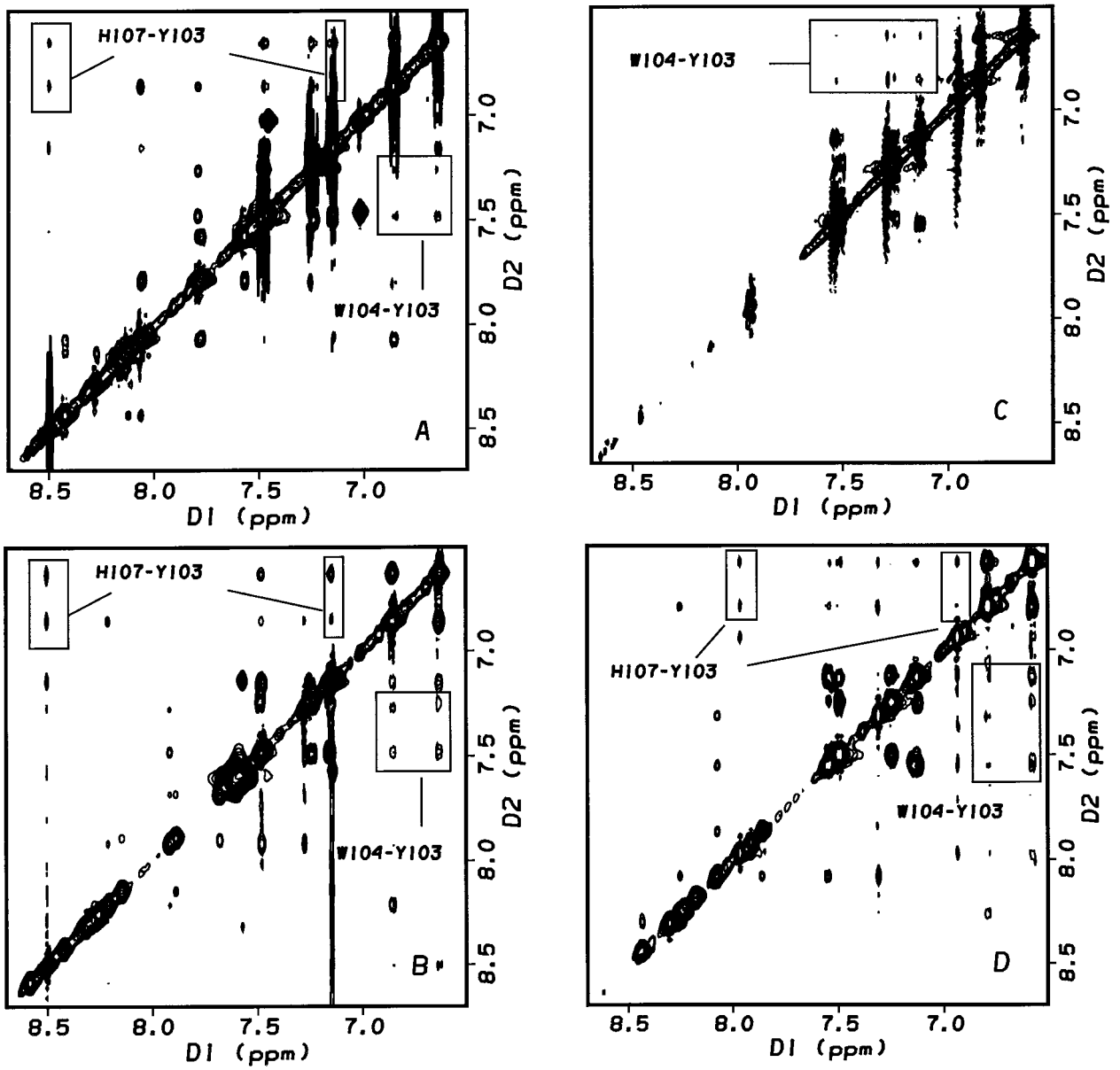

Figure 5. Sections of the ROESY spectrum of αlac:101-111. A, pH 2.8, 25°C; B, pH 2.8, 5°C; C, pH 7.0, 25°C; D, pH 7.0, 5°C. Crosspeaks from the Y103 ring to the H107 side-chain are indicated.

to be less extensive at pH 7.0, 25°C than at pH 2.8, 25°C. Only three of the eight main-chain to side-chain ROEs that were observed at pH 2.8 are present at pH 7.0, and seven of the ten side-chain to side-chain ROEs that were detected at pH 2.8 are present at pH 7.0. Reducing the temperature to 5°C stabilizes the structure, and the pattern of crosspeaks detected in the ROESY spectra of αlac:101-111 recorded at pH 7.0, 5°C is very similar to the pattern of crosspeaks observed in the pH 2.8 spectrum measured at 25°C. Three of the four ROEs involving Y103 and H107 are detectable and the fourth would be obscured by the diagonal. Nine of the ten side-chain to side-chain and seven of the eight side-chain to main-chain ROEs between residues in the hydrophobic cluster are present at pH 7.0, 5°C. ROEs between Y103 and A106 and between the acetylated N-terminal methyl group and Y103 are present at pH 7.0, 5°C.

It is important to point out that the His107 side-chain will not be fully deprotonated at pH 7.0. Based on the experimentally determined pK_a value, approximately 20% of the molecules will be protonated at pH 7.0. It would clearly have been desirable to obtain two-dimensional NMR spectra at higher pH values; unfortunately, the samples undergo slow, time-dependent aggregation at higher pH values and it was not possible to obtain two-dimensional spectra. Nonetheless, it is clear that the Y103-H107 interaction is weaker at neutral pH.

The effects of TFE on αlac:101-111 and αlac:101-120

TFE is often used to induce structure in peptides and there are several studies of the effects of TFE on the structure of intact proteins. Of relevance to this work are the studies by Dobson and co-workers of α-lactalbumin in TFE. They have used two-dimensional NMR and CD spectroscopy to probe the structure of α-lactalbumin in 50% TFE (Alexandrescu *et al.*, 1994) and it is of interest to examine the effect of TFE on the local structural preferences of αlac:101-111 and αlac:101-120. Dobson and co-workers have determined the structure of a slightly shorter peptide corresponding to residues 101 to 110 of human α-lactalbumin in 95% (v/v) TFE (Smith *et al.*, 1995) at low pH. The thermodynamic properties of TFE/water solvent mixtures are sensitive to the solvent composition and solutions of low concentrations of TFE in water behave very differently from solutions of low concentrations of water in TFE (Blandamer *et al.*, 1990; Cammers-Goodwin *et al.*, 1996).

TFE increases the apparent helical content of αlac:101-111

The addition of TFE to αlac:101-111 at pH 2.8 increased the apparent helicity, as judged by the intensity of the CD signal at 222 nm. The mean residue ellipticity reaches its maximum intensity by 30% (v/v) TFE. In 30% TFE, the measured value of θ_{222} is −22,100 deg cm^2 dmol^{-1} (Figure 2).

NMR spectra of αlac:101-111 recorded in 30% TFE are well resolved and could be assigned in a straightforward manner. The fingerprint region of the TOCSY spectrum is shown in Figure 3B. The C$^\alpha$H resonances of residues W104 through L109 are all shifted upfield by at least 0.1 ppm relative to their values in water at pH 2.8 (Figure 4(B)). The C$^\alpha$H resonance of H107 experiences the largest upfield shift, 0.36 ppm. The amide proton chemical shift of L105 also undergoes a dramatic change. In water, this resonance is the most upfield amide and is found at 7.56 ppm. In 30% TFE, this peak shifts by almost 0.9 ppm to 8.42 ppm and is the most downfield amide resonance. The large chemical shift change indicates that there must be a change in the local environment of the L105 amide. A likely explanation is a change in the distance and/or relative orientation of an aromatic side-chain.

In contrast to the results in water, αlac:101-111 displays a complete set of sequential amide to amide ROEs in 30% TFE. No long-range amide to amide ROE is detected. The main-chain to main-chain ROEs are summarized in Figure 4(B). All of the $^3J_{NH\alpha}$ coupling constants decrease in 30% TFE to below 5.4 Hz, except for those of L110 and A111. It is interesting to note that the coupling constants are lower than those observed for the C-helix peptide of hen lysozyme in 50% TFE (Bolin *et al.*, 1996). That peptide forms a helix in 50% TFE

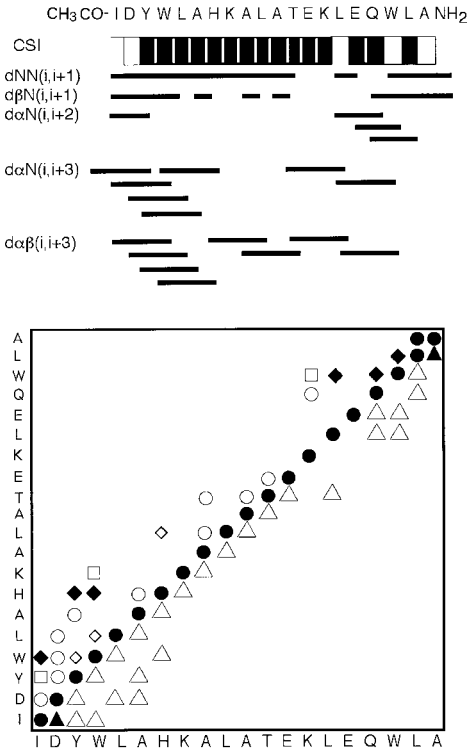

Figure 6. Deviations of the C$^\alpha$H chemical shifts from random-coil values, main-chain to main-chain ROEs, and overall ROESY contact map of αlac:101-120 in 30% TFE at pH 2.8, 25°C. Deviations of the C$^\alpha$H chemcial shifts are displayed with downward open rectangles indicating shifts of at least −0.1 ppm from random-coil values and filled rectangles indicating C$^\alpha$H chemical shifts that are shifted by at least −0.2 ppm from random-coil values. In the contact maps, triangles below the diagonal indicate main-chain to main-chain ROEs. The symbols above the diagonal indicate side-chain to side-chain (squares), side-chain to main-chain (circles), or both side-chain to side-chain and side-chain to main-chain (diamonds) ROEs. Open symbols (both above and below the diagonal) indicate one or more weak ROEs and filled symbols indicate at least two or more ROEs of at least medium intensity.

and the majority of the $^3J_{NH\alpha}$ coupling constants are between 5.0 and 6.5 Hz. The smaller $^3J_{NH\alpha}$ coupling constants, the shifts of the C$^\alpha$H resonances, and the observed pattern of ROEs all indicate that αlac:101-111 has a pronounced tendency to adopt helical or helical-like turn conformations in 30% TFE. The lack of amide to amide (i, $i+2$) and α to amide (i, $i+4$) ROEs demonstrates, however, that the peptide does not form a well-ordered α-helix.

TFE modulates the non-native side-chain interactions present in αlac:101-111

Many of the ROEs between hydrophobic side-chains that were observed in water are missing or weaker in 30% TFE. Fewer ROEs are detected between the I101, Y103 and W104 side-chains, and

the intensity of the remaining ROEs is diminished compared to what is observed in aqueous solution. The ROEs between the Y103 and H107 side-chains are, however, still present and are of comparable intensity. A contact map of the backbone to side-chain and side-chain to side-chain ROEs is shown in Figure 4(B).

The effect of TFE on the structure of αlac:101-120

Addition of TFE increases the apparent helicity of αlac:101-120. The maximum value of θ_{222}, −26,500 deg cm^2 dmol^{-1}, is reached by 28% (v/v) TFE (Figure 2). The peptide is more soluble in 30% TFE than in water; therefore, NMR spectra could be recorded at 25°C. This allows direct comparison with the studies of αlac:101-111. The fingerprint region of the TOCSY spectrum is shown in Figure 3D. The $C^{\alpha}H$ chemical shifts of all of the residues of αlac:101-120 suggest an increased propensity to adopt values of ϕ,ψ in the helical region of the Ramachandran map. The $C^{\alpha}H$ chemical shifts of the N-terminal five residues and the C-terminal five residues change the least in TFE, suggesting that structure already present in these areas is not disrupted by the addition of TFE. The deviations of the $C^{\alpha}H$ chemical shifts in TFE from random-coil values are shown in Figure 6. The largest perturbations in the $C^{\alpha}H$ chemical shifts compared to their values in water arise from residues H107 to K114, all of which experience significant upfield shifts with the addition of TFE. H107 has the largest upfield shift of −0.47 ppm. This resonance displays the largest shift upon addition of TFE to the shorter peptide. This is suggestive additional evidence that the local structure in this region of αlac:101-111 is preserved in αlac:101-120.

ROESY spectra also indicate an increased propensity to populate helical-like conformations for αlac:101-120 in the presence of 30% TFE. Sixteen of the 20 possible sequential amide to amide ROEs are observed. Of the four missing crosspeaks, three would lie too close to the diagonal to detect. Seven $(i,\ i+3)_{\alpha N}$ and eight $(i,\ i+3)_{\alpha\beta}$ medium to weak ROEs are also detected. Five of the seven $(i,\ i+3)_{\alpha N}$ and four of the eight $(i,\ i+3)_{\alpha\beta}$ ROEs are within the N-terminal seven residues. Three medium intensity $(i,\ i+2)_{\alpha N}$ ROE peaks are present in the C-terminal region. The main-chain to main-chain ROEs are summarized in Figure 6. With the exception of H107 and A120, the measured $^3J_{NH\alpha}$ coupling constants are all below 5.5 Hz and all except D102, T112, E116 and W118 are below 5.0 Hz.

The pattern of side-chain to side-chain ROEs between the residues that make up the N-terminal hydrophobic cluster is very similar to that observed for the shorter peptide in 30% TFE. Interestingly, the C-terminal cluster identified in water is more structured in 30% TFE. A large number of interresidue side-chain to side-chain (17) and main-chain to side-chain (eight) ROEs were identified in the C-terminal region. Particularly noteworthy are ROEs of medium strength from W118 to both L115 and L119, and a set of weaker ROEs from W118 to Q117 and to K114. A contact plot of the side-chain to side-chain and side-chain to main-chain ROEs is shown in Figure 6. Several ROE crosspeaks that were observed in the ROESY spectrum of the shorter peptide are not resolved in the spectra of αlac:101-120, due to spectral overlap. The $^3J_{\alpha\beta}$ coupling constants of residues Y103, H107 and W118 deviate from the values expected for conformational averaging, indicating restriction in the mobility about the α-β bonds. The differences between $^3J_{\alpha\beta2}$ and $^3J_{\alpha\beta3}$ are 2.2 Hz for Y103, 3.0 Hz for H107 and 3.6 Hz for W118. It is interesting that the side-chain interactions involved in the N-terminal cluster are weaker in 30% TFE, while the interactions that lead to the C-terminal cluster appear to be favored in 30% TFE. This may reflect the fact that more residues are involved in the N-terminal clus-

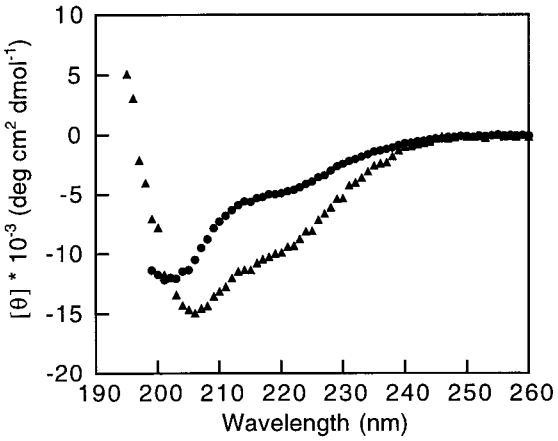

Figure 7. CD spectrum of the BD pair (triangles) compared to the sum of the spectra of the reduced peptides (circles). The spectra were recorded at pH 2.8, 25°C.

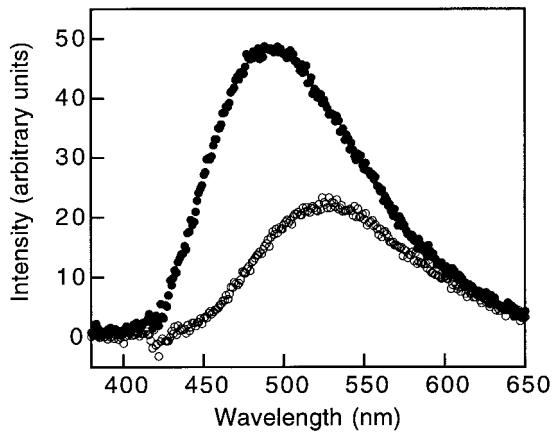

Figure 8. Fluorescence emission spectra of ANS in the presence (filled circles) and in the absence (open circles) of the BD pair. Spectra were recorded at pH 2.8, 25°C.

ter, and that this cluster appears to impose more structural constraints upon the peptide backbone.

Crosslinking αlac:20-36 to αlac:101-120 by the native 28 to 111 disulfide results in a significant increase in structure

Sedimentation experiments indicate that the BD pair is monomeric in aqueous solution at a concentration of 5 μM in 10 mM NaCl at pH 2.8, 25°C. The apparent molecular mass derived from a global fit to the centrifugation data is 4590 (+/ − 260) Da, which is within 5% of the known actual mass, 4366 Da. The data fit well using a single-species model and the residuals are random. The BD pair associates at concentrations above 25 μM; therefore, all CD experiments were performed at 5 μM.

The CD spectrum of the BD pair recorded at pH 2.8 shows a dramatic increase in the value of θ_{222} compared to the value calculated for the sum of the spectra of the two reduced peptides (Figure 7). The molar ellipticity of the BD pair at 222 nm is −9700 deg cm^2 $dmol^{-1}$, which corresponds to approximately 30% helicity, assuming no significant contribution from either the disulfide or the aromatics. The sum of the molar ellipticities at 222 nm of the reduced peptides is −5600 deg cm^2 $dmol^{-1}$, corresponding to roughly 17% helicity.

The BD pair contains two of the four tyrosine residues found in human α-lactalbumin and two of the three tryptophan residues. The fluorescence emission maximum of native α-lactalbumin, measured at 25°C, pH 7.0, 100 μM Ca^{2+} is 330 nm. The emission maximum of the molten globule state at pH 2.8, 25°C, 1 mM EDTA is 342 nm. The emission maximum of the BD pair is 352 nm at pH 2.8. This value of the emission maximum is red shifted from the value measured for the native state and from the value observed for the molten globule state. This indicates that the tryptophan residues are likely to be more accessible to solvent in the BD pair than they are in the molten globule state of intact α-lactalbumin. This is not at all unreasonable, given the relatively small size of the peptide.

The fluorescent hydrophobic molecule 1-anilinonaphthalene-8-sulphonate (ANS) is commonly used as a probe of molten globule state formation (Semisotnov *et al.*, 1991). The BD pair binds ANS. The ANS emission maximum is blue shifted to 492 nm in the presence of an equimolar amount of BD pair (2 μM ANS, 2 μM peptide) and there is a 2.2-fold increase in fluorescence intensity (Figure 8). Binding is abolished by the addition of urea. The BD pair associates at neutral pH, precluding ANS binding studies. A smaller blue shift of the ANS emission maximum (511 nm) and weaker (1.6 fold) enhancement of fluorescence intensity is observed in the presence of αlac:101-120 at low pH, indicating that this peptide weakly binds ANS and hence, presumably, contains a hydrophobic surface. This result is consistent with the observation of hydrophobic cluster formation by NMR. Binding is abolished by the addition of 6 M urea. ANS binding is not observed for this peptide at pH 7.0, indicating either a pH-dependent change in structure or some electrostatic contributions to the weak bindng observed at low pH. αlac:20-36 and αlac:101-111 do not bind ANS.

The position of the tryptophan emission maximum of the construct and its ability to bind ANS are consistent with a loose packing of the hydrophobic side-chains, reminiscent of the molten globule state.

Conclusions

Our data demonstrate that the N-terminal region of αlac:101-111 and αlac:101-120 has a high propensity for non-random structure in water. It is clear that a hydrophobic cluster can be formed *via* local interactions in these peptides without the aid of stabilizing interactions from more distant regions of the peptide chain. This strongly argues that these interactions should persist in the denatured state under conditions that favor folding. The N-terminal cluster persists in 6 M urea, indicating that it may also be populated under conditions that favor the denatured state. The Y103-H107 interaction is clearly non-native. In all crystal structures, the phenol ring of Y103 is at least 10 Å from the imidazole side-chain of H107. The interaction is weaker at neutral pH, but is still present. The pH-dependence of this interaction demonstrates that it could preferentially stabilize the acid-induced molten globule state. Because this interaction is non-native, it will need to be disrupted during the transition from the molten globule to the native state. This interaction may be an example of a non-native interaction that contributes to the roughness of the folding free-energy landscape. It is likely to play a less significant role at neutral pH, since it is weaker under these conditions.

We have found evidence for a second hydrophobic cluster localized in residues 115 to 119. The local hydrophobic clusters characterized in this work may provide a manifold for other regions of the protein to pack against during the formation of the molten globule state. In a recent study of ^{15}N-labeled human α-lactalbumin, Kim, Dobson and co-workers found that most of the ^{1}H-^{15}N amide peaks of the α-domain could not be observed, due to broadening caused by slow conformational fluctuations (Schulman *et al.*, 1997). Upon unfolding of the molten globule by temperature and urea, the last ^{1}H-^{15}N amide resonances to appear were those from the D helix and from residues 115 to 119. This was interpreted as indirect evidence that this region of the molten globule is the most resistant to denaturation. Our demonstration that the local interactions present in αlac:101-111 are not eliminated in urea is completely consistent with the results reported by Schulman *et al.* (1997). Both clusters may be key determinants in the formation

of the molten globule state during the initial events of protein folding.

Both αlac:101-111 and αlac:101-120 are noticeably more structured in TFE, and both appear to have an increased propensity to populate the helical region of the ϕ,ψ map. This may be relevant to studies of intact α-lactalbumin in TFE, which have demonstrated a dramatic increase in the intensity of the CD signal at 222 nm (Alexandrescu *et al*., 1994). It is interesting to speculate on the origins of the changes in the structure of the αlac:101-111 peptide in the two different solvents. In water, the structure appears to be stabilized by side-chain to side-chain interactions involving Y103, H107 and several apolar residues. These interactions are weaker in 30% TFE. The weakening of the side-chain interactions may be the result of backbone-driven structure formation in TFE/water mixtures *versus* structure formation driven by hydrophobic clustering and the Y103-H107 interaction in water.

The existence of long-range stabilizing interactions between the B helix and the C-terminal region of the protein have been clearly demonstrated by our studies of the BD model peptide and indicate that interactions between these regions of the protein are sufficient to drive formation of secondary structure. The measured helicity of the BD pair is approximately twofold greater than that of the reduced peptides. The fluorescence emission spectrum of the BD pair suggests that the tryptophan side-chains are exposed to solvent, and its ability to bind ANS indicates a loose packing of the side-chains at the interface of the two peptides. These results are fully consistent with the notion that relatively non-specific hydrophobic interactions between segments of the peptide chain that have some intrinsic propensity to form local structure can play an important role in stabilizing the molten globule state.

It is interesting to compare the results reported here with studies of peptide fragments derived from other proteins that form well-characterized molten globule states. Mutagenesis studies of the A-state of cytochrome c have shown that interactions between the N and C-terminal helices stabilize the A-state (Marmorino & Pielak, 1995). Peptide fragments derived from the N and C-terminal regions of cytochrome *c* have little tendency to adopt structure in isolation. However, Roder and co-workers have shown that a heme-containing N-terminal fragment associates with a peptide corresponding to the C-terminal helix to form a highly helical, non-covalent complex (Wu *et al*., 1993). Wright, Dyson and co-workers have performed extensive studies of peptide fragments of myoglobin and have shown that a peptide encompassing the G helix is largely unstructured, while a peptide corresponding to the H helix is approximately 30% helical (Waltho *et al*., 1993). These helices interact in the native state and in the molten globule state (Eliezer *et al*., 1998; Kay & Baldwin, 1996). A peptide model consisting of the G and H helices crosslinked by a disulfide bond is highly helical in water, demonstrating that intermolecular association also stabilizes secondary structure in this system (Shin *et al*., 1993).

Our crosslinked BD peptide is considerably more helical than the individual reduced peptides, but the observed helicity cannot account for the known helical content of the molten globule state. The mean residue ellipticity of the low-pH form of the molten globule state at 25°C, is −12,000 deg cm^2 $dmol^{-1}$ at 222 nm. A mean residue ellipticity of this magnitude could be generated by assuming that 42 of the 123 residues in the full-length protein adopt a helical conformation. The BD pair would have to be fully helical to account for this amount of helical structure. The experimentally measured ellipticity of the BD pair can, however, be generated by only an 11 residue helix. Although clearly a gross oversimplification, this crude calculation demonstrates that the BD pair does not account for the total helicity observed in the molten globule state. Therefore, it will be interesting to explore what other elements of structure are necessary to arrive at a more adequate model of the α-lactalbumin molten globule state. Studies of larger peptides that include both the A and B helices crosslinked to residues 101 to 120 are underway.

Materials and Methods

Peptide synthesis and purification

αlac:101-111 and αlac:101-120 were designed with the native Cys residues at positions 111 and 120 replaced by Ala. A second version of the 101-120 peptide was prepared with the wild-type Cys at position 111. Peptide synthesis reagents were purchased from PerSeptive Biosystems Inc. or Advanced ChemTech. Solvents were purchased from Fisher Scientific. αlac:101-111 was synthesized using a Rainin Model PS3 automated peptide synthesizer. αlac:20-36 and αlac:101-120 were prepared on a 0.2 mmol scale using a Millipore 9050 automated peptide synthesizer. Fmoc-L-amino acids were coupled *via* *O*-(7-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium tetrafluoroborate (TBTU) activation. The polystyrene support with a peptide amide linker (PAL resin) provided amidated carboxy termini. The amino termini were acetylated upon completion of the syntheses. Peptides were cleaved from the resin using a solution of 90% TFA, 3.3% ethanedithiol, 3.3% anisole, 3.3% thioanisole. The peptides were purified using reverse-phase HPLC with water/CH_3CN gradients containing 0.1% TFA using a Vydac C18 column. The BD pair was formed by oxidation of a mixture of αlac:20-36 and αlac:101-120 at pH 8.4 in aqueous solution. The BD pair was purified from the other oxidized species by reverse-phase HPLC. αlac:101-111 and the BD pair were characterized by MALDI mass spectrometry: αlac:101-111, expected 1341.5 Da, determined 1341.0 Da; BD pair, expected 4366.0 Da, determined 4364.8 Da. αLac:20-36, αlac:101-120 (Ala) and αlac:101-120 (Cys) were characterized by FAB mass spectrometry at the University of Illinois Mass Spectrometry Facility: αlac:20-36, expected 1898.0 Da, determined 1898.6 Da; αlac:101-120 (Ala), expected 2441.0 Da, determined 2441.0 Da; αlac:101-120 (Cys), expected 2473.9 Da, determined 2473.4 Da.

CD

CD measurements were performed on an Aviv model 62A circular dichroism spectrometer. Wavelength scans were performed with a minimum of three repeats and an averaging time of three seconds. A 2 mM phosphate, 2 mM borate, 2 mM citrate, 10 mM NaCl buffer was used for all CD experiments. Concentration-dependence studies were performed by dilution from a concentrated peptide stock solution into the buffer solution. For measurements of 100 μM peptide and below, a separate stock solution of 100 μM peptide was prepared and diluted by weight measurement into buffer, assuming a 1:1 density ratio. The concentrations of the stock solutions were determined by absorbance measurements at 280 nm using extinction coefficients calculated *via* the method of Pace *et al*. (1995). Fraction helix f_{hlx} was calculated from the measured molar ellipticity at 222 nm using the method of Rohl & Baldwin (1997):

$$f_{hlx} = \frac{(\theta_{222} - \theta_C)}{(\theta_H - \theta_C)}$$

$$\theta_C = 2220 - (53t)$$

$$\theta_H = (-44{,}000 + 250t)\left(1 - \frac{3}{N}\right)$$

θ_C and θ_H are the molar ellipticities for pure coil and pure helix respectively, t is temperature in °C, and N is the number of amino acids in the peptide.

^{1}H NMR

^{1}H NMR experiments were carried out on Varian Unity INOVA 500 and 600 MHz spectrometers. Standard presaturation techniques were used for solvent suppression. All TOCSY and ROESY experiments were performed with either 256 or 512 t_1 increments and 2048 data points in t_2. TOCSY experiments utilized a 75 ms mixing time and ROESY experiments used 250 to 300 ms mixing times. The intensities of the amide to amide and alpha to amide ROE crosspeaks were corrected for offset effects (Griesinger & Ernst, 1987). DQF-COSY and E.COSY experiments were performed with 512 t_1 increments and 8 K and 4 K data points in t_2, respectively. All experiments used the TPPI method of data collection (Marion & Wüthrich, 1983), except the E.COSY experiments, which used the method of States *et al*. (1982). All data sets were processed and analyzed using Felix 95.0 (Biosym). 70 to 90° shifted sinebell window functions were applied to all 2D data sets. The data sets were zero-filled once in the first dimension and twice in the second dimension before Fourier transformation. Sequence-specific assignments were determined by assigning spin systems from the TOCSY spectra and using sequential alpha to amide ROE peaks (Wüthrich, 1986). $^3J_{NH\alpha}$ coupling constants were estimated by measurement of υ_a and υ_d from DQF-COSY experiments (Kim & Prestegard, 1989). $^3J_{\alpha\beta}$ coupling constants were measured as passive couplings from E.COSY experiments (Griesinger *et al*., 1987).

Apparent pK_a values for αlac:101-111 were determined by fitting chemical shift changes with pH to a modified form of the Henderson-Hasselbalch equation:

$$\delta(\mathrm{pH}) = \frac{[\delta_{base} + \delta_{acid}(1 + 10^{(\mathrm{p}K_a - \mathrm{pH})})]}{(1 + 10^{(\mathrm{p}K_a - \mathrm{pH})})}$$

δ_{base} and δ_{acid} are the chemical shifts of the ionizable residues in their fully deprotonated and fully protonated forms, respectively, and δ(pH) is the measured chemical shift. The change in the chemical shift of the amide proton of D102 and the average of the chemical shift changes of the two βCH_2 protons of D102 were used to determine the apparent pK_a of the D102 side-chain. The chemcial shift changes of the 2H and 4H protons of the imidazole ring of H107 were fitted in order to determine pK_a values for this residue. The data were fit using KaleidaGraph™.

Sedimentation equilibrium

A solution of the BD pair was dialyzed against a buffer containing 2 mM phosphate, 2 mM borate, 2 mM citrate and 10 mM NaCl. Experiments were performed at 25°C with a Beckman XL-A analytical ultracentrifuge, using rotor speeds of 30,000, 40,000 and 50,000 rpm. The concentration of the BD pair was 5 μM. Experiments were carried out using 12 mm pathlength, six-channel, charcoal-filled, Epon cells with quartz windows. Data were collected at 230 nm. Partial specific volumes were calculated from the weighted average of the partial specific volumes of the individual amino acids (Cohn & Edsall, 1943). The data were globally fit with a single-species model with the molecular mass treated as a fitting parameter. The HID program from the Analytical Ultracentrifugation Facility at the University of Connecticut was used for the fitting analysis.

Fluorescence measurements

All fluorescence measurements were performed using an ISA Fluorolog spectrometer. ANS fluorescence was measured using an excitation wavelength of 370 nm. The emission spectrum was recorded over the range of 380 to 650 nm. An ANS stock solution was prepared in methanol. The concentrations of the ANS stock solutions were determined using an extinction coefficient of 8.0×10^3 M^{-1} cm^{-1} at 372 nm in methanol (provided by Molecular Probes). The stock solution was diluted into 3 ml of buffer at pH 2.8 to give a final ANS concentration of 2 μM. Peptide was titrated into this solution.

Acknowledgments

This work was supported by NIH grant GM544233 to D.P.R., who is a Pew Scholar in the Biomedical Sciences. S.D. is supported by a GAANN Fellowship from the Department of Education. The NMR facility at SUNY Stony Brook is supported by a grant from the NSF CHE9413510. FAB mass spectrometry was performed at the University of Illinois Mass Spectrometry Center. The fluorescence spectrometer was purchased with funds from NSF grant CHE9709164. The analytical ultracentrifuge was purchased with funds from a grant to R.F. from the Zimmer Corporation. We thank Preston Hensley for the Igor based non-linear fitting algorithm used to generate the sedimentation equilibrium figure and Donna Luisi for proofreading the manuscript.

References

Acharya, K. R., Stuart, D. I., Walker, N. P. C., Lewis, M. & Phillips, D. C. (1989). Refined structure of baboon

α-lactalbumin at 1.7 Å resolution. *J. Mol. Biol.* **208**, 99–127.

Acharya, K. R., Ren, J., Stuart, D. I., Phillips, D. C. & Fenna, R. E. (1991). Crystal structure of human α-lactalbumin at 1.7 Å resolution. *J. Mol. Biol.* **221**, 571–581.

Alexandrescu, A. T., Evans, P. A., Pitkeathly, M., Baum, J. & Dobson, C. M. (1993). Structure and dynamics of the acid-denatured molten globule state of α-lactalbumin: a two-dimensional NMR study. *Biochemistry,* **32**, 1707–1718.

Alexandrescu, A. T., Ng, Y.-L. & Dobson, C. M. (1994). Characterization of a trifluoroethanol-induced partially folded state of α-lactalbumin. *J. Mol. Biol.* **235**, 587–599.

Baum, J., Dobson, C. M., Evans, P. A. & Hanley, C. (1989). Characterization of a partly folded protein by NMR methods: studies on the molten globule state of guinea pig α-lactalbumin. *Biochemistry,* **28**, 7–13.

Blandamer, M. J., Burgess, J., Cooney, A., Cowles, H. J., Horn, I. H., Martin, K. J., Morcom, K. W. & Warrick, P. (1990). Excess molar Gibbs energies of mixing of water and 1,1,1,3,3,3–hexafluoropropan-2–ol mixtures at 298.15 K. *J. Chem. Soc. Faraday Trans.* **86**, 2209–2213.

Bolin, K. A., Pitkeathly, M., Miranker, A., Smith, L. J. & Dobson, C. M. (1996). Insight into a random coil conformation and an isolated helix: structural and dynamical characterization of the C-helix peptide from hen lysozyme. *J. Mol. Biol.* **261**, 443–453.

Cammers-Goodwin, A., Allen, T. J., Oslick, S. L., McClure, K. F., Lee, J. H. & Kemp, D. S. (1996). Mechanism of stabilization of helical conformations of polypeptides by water containing trifluoroethanol. *J. Am. Chem. Soc.* **118**, 3082–3090.

Chakrabartty, A., Kortemme, T., Padmanabhan, S. & Baldwin, R. L. (1993). Aromatic side-chain contribution to far-ultraviolet circular dichroism of helical peptides and its effect on measurement of helix propensities. *Biochemistry,* **32**, 5560–5565.

Chelvanayagam, G., Reich, Z., Bringas, R. & Argos, P. (1992). Prediction of protein folding pathways. *J. Mol. Biol,* **227**, 901–916.

Cohn, E. J. & Edsall, J. T. (1943). *Proteins, Amino Acids and Peptides as Ions and Dipolar Ions*, Reinhold Publishing Corporation, New York.

Creighton, T. E. (1997). How important is the molten globule for correct folding? *Trends Biochem. Sci.* **22**, 6–10.

de Laureto, P. P., De Filippis, V., Di Bello, M., Zambonin, M. & Fontana, A. (1995). Probing the molten globule state of α-lactalbumin by limited proteolysis. *Biochemistry,* **34**, 12596–12604.

Deslauriers, R. & Smith, I. C. P. (1980). In *Biological Magnetic Resonance* (Berliner, L. J. & Reuben, J., eds), vol. 2, pp. 243–344, Plenum Press, New York.

Dobson, C. M. (1992). Unfolded proteins, compact states and molten globules. *Curr. Opin. Struct. Biol.* **2**, 6–12.

Dolgikh, D. A., Gilmanshin, R. I., Brazhnikov, E. V., Bychkova, V. E., Semisotnov, G. V., Venyaminov, S. Y. & Ptitsyn, O. B. (1981). α-Lactalbumin: compact state with fluctuating tertiary structure? *FEBS Letters,* **136**, 311–315.

Eliezer, D., Yao, J., Dyson, J. & Wright, P. E. (1998). Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding. *Nature Struct. Biol.* **5**, 148–155.

Griesinger, C. & Ernst, R. R. (1987). Frequency offset effects and their elimination in NMR rotating-frame cross-relaxation spectroscopy. *J. Magn. Reson.* **75**, 261–271.

Griesinger, C., Sorensen, O. W. & Ernst, R. R. (1987). Practical aspects of the E.COSY technique. Measurement of scalar spin-spin coupling constants in peptides. *J. Magn. Reson.* **75**, 474–492.

Harata, K. & Muraki, M. (1992). X-ray structural evidence for a local helix-loop transition in α-lactalbumin. *J. Biol. Chem.* **267**, 1419–1421.

Haynie, D. T. & Friere, E. (1993). Structural energetics of the molten globule state. *Proteins: Struct. Funct. Genet.* **16**, 115–140.

Hiraoka, Y. & Sugai, S. (1984). Thermodynamics of thermal unfolding of bovine apo-α-lactalbumin. *Int. J. Pept. Protein Res.* **23**, 535–542.

Ikeguchi, M., Kuwajima, K., Mitani, M. & Sugai, S. (1986). Evidence for identity between the equilibrium unfolding intermediate and a transient folding intermediate: a comparative study of the folding reactions of α-lactalbumin and lysozyme. *Biochemistry,* **25**, 6965–6972.

Kay, M. S. & Baldwin, R. L. (1996). Packing interactions in the apomyoglobin folding intermediate. *Nature Struct. Biol.* **3**, 439–445.

Kemmink, J. & Creighton, T. E. (1995). The physical properties of local interactions of tyrosine residues in peptides and unfolded proteins. *J. Mol. Biol.* **245**, 251–260.

Kemmink, J., van Mierlo, C. P. M., Scheek, R. M. & Creighton, T. E. (1993). Local structure due to an aromatic-amide interaction observed by ^{1}H-NMR spectroscopy in peptides related to the N terminus of BPTI. *J. Mol. Biol.* **230**, 312–322.

Kim, Y. & Prestegard, J. H. (1989). Measurement of vicinal couplings from cross-peaks in COSY spectra. *J. Magn. Reson.* **84**, 9–13.

Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.

Kuhlman, B., Boice, J. A., Wu, W., Fairman, R. & Raleigh, D. P. (1997). Calcium binding peptides from α-lactalbumin: implications for protein folding and stability. *Biochemistry,* **36**, 4607–4615.

Kuwajima, K. (1989). The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. *Proteins: Struct. Funct. Genet.* **6**, 87–103.

Kuwajima, K. (1996). The molten globule state of α-lactalbumin. *FASEB J.* **10**, 102–108.

Kuwajima, K., Nitta, K., Yoneyama, M. & Sugai, S. (1976). Three-state denaturation of α-lactalbumin by guanidine hydrochloride. *J. Mol. Biol.* **106**, 359–373.

Kuwajima, K., Mitani, M. & Sugai, S. (1989). Characterization of the critical state in protein folding. *J. Mol. Biol.* **206**, 547–561.

Marion, D. & Wüthrich, K. (1983). Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of ^{1}H-^{1}H spin-spin coupling constants in proteins. *Biochem. Biophys. Res. Commun.* **113**, 967–974.

Marmorino, J. L. & Pielak, G. J. (1995). A native tertiary interaction stabilizes the A state of cytochrome *c*. *Biochemistry,* **34**, 3140–3143.

Maruyama, S., Kuwajima, K., Nitta, K. & Sugai, S. (1977). Thermodynamic characterization of partially denatured states in the denaturation process of

bovine α-lactalbumin by inorganic denaturants. *Biochim. Biophys. Acta,* **494**, 343–353.
Moult, J. & Unger, R. (1991). An analysis of protein folding pathways. *Biochemistry,* **30**, 3816–3824.
Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. & Gray, T. (1995). How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* **4**, 2411–2423.
Peng, Z.-y. & Kim, P. S. (1994). A protein dissection study of a molten globule. *Biochemisty,* **30**, 2136–2141.
Peng, Z.-y., Wu, L. C. & Kim, P. S. (1995). Local structural preferences in the α-lactalbumin molten globule. *Biochemistry,* **34**, 3248–3252.
Phillips, D. C. (1974). Crystallographic studies of lysozyme and its interactions with inhibitors and substrates. In *Lysozyme* (Osserman, E. F., Canfield, R. E. & Beychock, S., eds), pp. 9–30, Academic Press, New York.
Pike, A. C., Brew, K. & Acharya, K. R. (1996). Crystal structures of guinea-pig, goat and bovine α-lactalbumin highlight the enhanced conformational flexibility of regions that are significant for its action in lactose synthase. *Structure,* **4**, 691–703.
Privalov, P. L. (1996). Intermediate states in protein folding. *J. Mol. Biol.* **258**, 707–725.
Ptitsyn, O. B. (1995). Molten globule and protein folding. *Advan. Protein Chem.* **47**, 83–222.
Rohl, C. A. & Baldwin, R. L. (1997). Comparison of NH exchange and circular dichroism as techniques for measuring the parameters of helix-coil transition in peptides. *Biochemistry,* **36**, 8435–8442.
Schulman, B. A. & Kim, P. S. (1996). Proline scanning mutagenesis of a molten globule reveals non-cooperative formation of a protein's overall topology. *Nature Struct. Biol.* **3**, 682–687.
Schulman, B. A., Kim, P. S., Dobson, C. M. & Redfield, C. (1997). A residue-specific NMR view of the non-cooperative unfolding of a molten globule. *Nature Struct. Biol.* **4**, 630–634.
Semisotnov, G. V., Rodionova, N. A., Razgulyaev, O. I., Uversky, V. N., Gripas, A. F. & Gilmanshin, R. I. (1991). Study of the ''molten globule'' intermediate state in protein folding by a hydrophobic fluorescent probe. *Biopolymers,* **31**, 119–128.
Shin, H-C., Merutka, G., Waltho, J. P., Tennant, L. L., Dyson, H. J. & Wright, P. E. (1993). Peptide models of protein folding initiation sites. 3. The G-H helical hairpin of myoglobin. *Biochemistry,* **32**, 6356–6364.
Smith, L. J., Alexandrescu, A. T., Pitkeathly, M. & Dobson, C. M. (1995). Solution structure of a peptide fragment of human α-lactalbumin in trifluoroethanol: a model for local structure in the molten globule. *Structure,* **2**, 703–712.
States, D. J., Haberkorn, R. A. & Ruben, D. J. (1982). A two-dimensional nuclear Overhauser experiment with pure absorption phase in four quadrants. *J. Magn. Reson.* **48**, 286–292.
Wagner, G., Braun, W., Havel, T. F., Schaumann, T., Go, N. & Wüthrich, K. (1987). Protein structures in solution by NMR and distance geometry: the polypeptide fold of BPTI determined using two different algorithms, DISGEO and DISMAN. *J. Mol. Biol.* **196**, 611–639.
Waltho, J. P., Feher, V. A., Merutka, G., Dyson, H. J. & Wright, P. E. (1993). Peptide models of protein folding initiation sites. 1. Secondary structure formation by peptides corresponding to the G- and H-helices of myoglobin. *Biochemistry,* **32**, 6337–6347.
Wishart, D. S., Sykes, B. D. & Richards, F. M. (1992). The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry,* **31**, 1647–1651.
Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S. & Sykes, B. D. (1995). ^{1}H, ^{13}C, ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J. Biomol. NMR,* **5**, 67–81.
Wu, L. C., Laub, P. B., Elöve, G. A., Carey, J. & Roder, H. (1993). A noncovalent peptide complex as a model for an early folding intermediate of cytochrome *c*. *Biochemistry,* **32**, 10271–10276.
Wu, L. C., Peng, Z.-y. & Kim, P. S. (1995). Bipartite structure of the α-lactalbumin molten globule. *Nature Struct. Biol.* **2**, 281–285.
Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, Inc., New York.
Yang, J. J., van den Berg, B., Pitkeathly, M., Smith, L. J., Bolin, K. A., Keiderling, T. A., Redfield, C., Dobson, C. M. & Radford, S. E. (1996). Native-like secondary structure in a peptide from the α-domain of hen lysozyme. *Folding Des.* **1**, 473–484.

Edited by P. E. Wright

(Received 28 April 1998; received in revised form 17 July 1998; accepted 22 July 1998)

http://www.hbuk.co.uk/jmb

Supplementary material comprising five Tables and one Figure is available from JMB Online