Localized Nature of the Transition-state Structure in Goat α-Lactalbumin Folding

ARTICLE *in* JOURNAL OF MOLECULAR BIOLOGY · SEPTEMBER 2004

Impact Factor: 4.33 · DOI: 10.1016/j.jmb.2004.06.010 · Source: PubMed

doi:10.1016/j.jmb.2004.06.010 *J. Mol. Biol.* (2004) **341**, 589–604

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Localized Nature of the Transition-state Structure in Goat α-Lactalbumin Folding

Kimiko Saeki†, Munehito Arai†, Takao Yoda, Masaharu Nakao and Kunihiro Kuwajima*

Department of Physics Graduate School of Science University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

**Corresponding author*

To investigate whether the structure partially formed in the molten globule folding intermediate of goat α-lactalbumin is further organized in the transition state of folding, we constructed a number of mutant proteins and performed Φ-value analysis on them. For this purpose, we measured the equilibrium unfolding transitions and kinetic refolding and unfolding reactions of the mutants using equilibrium and stopped-flow kinetic circular dichroism techniques. The results show that the mutants with mutations located in the A-helix (V8A, L12A), the B-helix (V27A), the β-domain (L52A, W60A), the C-helix (K93A, L96A), the C–D loop (Y103F), the D-helix (L105A, L110A), and the C-terminal 3_{10}-helix (W118F), have low Φ-values, less than 0.2. On the other hand, D87N, which is located on the Ca^{2+}-binding site, has a high Φ-value, 0.91, indicating that tight packing of the side-chain around Asp87 occurs in the transition state. One β-domain mutant (I55V) and three C-helix mutants (I89V, V90A, and I95V) demonstrated intermediate Φ-values, between 0.4 and 0.7. These results indicate that the folding nucleus in the transition state of goat α-LA is not extensively distributed over the α-domain of the protein, but very localized in a region that contains the Ca^{2+}-binding site and the interface between the C-helix and the β-domain. This is apparently in contrast with the fact that the molten globule state of α-lactalbumin has a partially formed structure inside the α-domain. It is concluded that the specific docking of the α and β-domains at a domain interface is necessary for this protein to organize its native structure from the molten globule intermediate.

Keywords: protein folding; α-lactalbumin; the transition state; the molten globule state; Φ-value analysis

Introduction

Elucidation of protein folding mechanisms is one of the most important problems in modern biophysics and molecular biology.[1] The accumulation of molten globule-like intermediates in protein folding is common to many globular proteins with more than 100 amino acid residues; the folding mechanism of the proteins is considered to be as shown in Scheme 1, where U, I, and N are the unfolded, the intermediate and the native states, respectively.[2–4] The process from the unfolded state to the intermediate is rapid, usually occurring within the dead time of stopped-flow measurement (1–30 ms), while the rate-limiting step in the folding reaction is the process from the intermediate to the native state.

$$U \rightleftharpoons I \rightarrow N$$

Scheme 1.

† K.S. and M.A. contributed equally to this work.

Present addresses: M. Arai, Institute for Biological Resources and Functions, National Institute of Advanced Industrial Science and Technology (AIST), 1-1-1 Higashi, Tsukuba, Ibaraki 305-8566, Japan; T. Yoda, School of Bioscience, Nagahama Institute of Bio-Science and Technology, 1266 Tamura, Nagahama, Shiga 526-0829, Japan.

Abbreviations used: α-LA, α-lactalbumin; CD, circular dichroism; GdnHCl, guanidine hydrochloride; I, the intermediate; N, the native state; NMR, nuclear magnetic resonance; U, the unfolded state; UV, ultra violet.

E-mail address of the corresponding author: kuwajima@phys.s.u-tokyo.ac.jp

α-Lactalbumin (α-LA) is a monomeric globular protein with a molecular mass of 14,200 (Figure 1),[5,6] and is one of the best-understood proteins in protein folding studies.[2,4] α-LA has two domains, an α-domain comprising the residues 1–39 and 81–123 and a β-domain comprising the residues 40–80. The structure is stabilized by four disulfide bridges (6–120, 61–77, 73–91, and 28–111). The two domains are held together by the cystine bridge between residues 73 and 91, forming a loop, which resides between these domains and binds Ca^{2+}. The Ca^{2+}-binding loop contains two helices, a portion of the C-helix (residues 86–98) as well as a 3_{10}-helix (residues 77–80), which are analogous to, but smaller than, the helix-loop-helix motif of the classical EF-hand. The Ca^{2+} binding is coordinated with three aspartic acid side-chains (Asp82, Asp87, and Asp88), two backbone carbonyl oxygen atoms (Lys79 and Asp84), and two water molecules, which contribute the last two oxygen ligands to a distorted pentagonal bipyramidal structure.

It has been known that α-LA forms a typical molten globule state at equilibrium, equivalent to the kinetic folding intermediate formed at early stages of refolding, and that the folding reaction of α-LA can be described by Scheme 1.[2,4,7] The molten globule state typically exhibits native-like compactness and the native tertiary fold of the molecule, but does not have specific side-chain packing interactions.[4,7] The structure of the molten globule state of bovine and human α-LA has been characterized in great detail by various techniques including circular dichroism (CD), fluorescence, NMR, hydrogen-exchange labeling[8–14] and protein engineering techniques.[15–17] These studies have shown that the α-LA molten globule has a bipartite structure in which the α-domain is weakly formed and the β-domain is disordered. It has also been shown that the molten globule state has a loosely packed hydrophobic core and is stabilized by non-specific hydrophobic interactions.[2,4,18]

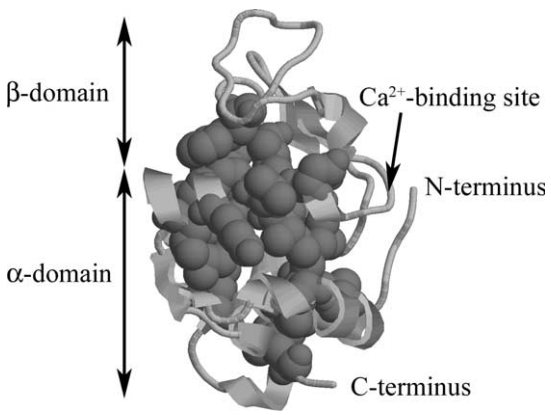

Figure 1. A schematic representation of goat α-LA (PDB code: 1HMK). The side-chains of the residues where mutations have been introduced are shown by space filling.

Although the molten globule state of goat α-LA has not been studied in as much detail as the molten globules of the bovine and human proteins, the amino acid sequence of goat α-LA has 94% identity with that of the bovine protein (only 7 of 123 residues in the sequence are different between the two proteins), so that the molten globule of the goat protein must be very similar to that of the bovine protein. Although the α-domain is more organized in the molten globule, as commonly found in members of the α-LA/lysozyme family, the helices in the domain are not highly stabilized, and their protection factors against hydrogen exchange do not exceed a factor of 10 in bovine α-LA.[14,19] Therefore, the molten globule of bovine and probably also goat α-LA is a highly fluctuating ensemble of the compact intermediate formed at an early stage of folding.

Compared with the well-characterized structure of the molten globule state of α-LA, the structure of the transition state of the folding from the molten globule to the native state of the protein has been less clear. Elucidation of the structural relationship between the molten globule and the transition state of folding must be important for understanding the mechanism of folding of proteins. It is thus certainly important to address a question such as whether the fluctuating structure organized in the molten globule state of goat α-LA is tightened and further organized in the transition state of folding.

Here, we have thus studied the equilibrium unfolding transitions and the kinetic refolding and unfolding reactions of various mutants of goat α-LA using equilibrium and stopped-flow kinetic CD techniques.[20] For this purpose, we constructed 16 mutant proteins that had mutation sites on the A-helix (V8A, and L12A) and the B-helix (V27A), within the β-domain (L52A, I55V, and W60A), within the Ca^{2+}-binding site (D87N), on the C-helix (I89V, V90A, K93A, I95V, and L96A), within the C–D loop (Y103F), on the D-helix (L105A and L110A), and on the C-terminal 3_{10}-helix (W118F). We chose the D87N mutant instead of D87A, because the D87A displayed a total loss of rigid tertiary structure, a dramatic loss in secondary structure and negligible Ca^{2+} affinity;[21] D87N α-LA retains each of these structural and functional features.[22] Including mutants (T29I and the N terminus) previously studied,[6,20] the mutation sites are distributed over both the α and β-domains; hence, this set of mutants will be useful for investigating the transition-state structure of folding in goat α-LA.

We carried out Φ-value analysis of the transition state of folding for these mutant proteins,[23] and the results rather surprisingly show that the transition-state structure is not further organized in the α-domain hydrophobic core, composed of the A, B, D, and the 3_{10}-helices. A high Φ-value was observed for D87N, and fractional Φ-values were observed for I55V, I89V, V90A, and I95V. The results indicate that the folding nucleus in the

transition state of goat α-LA is not distributed extensively over the protein, but very localized in the Ca^{2+}-binding site and at a part of the interface between the α and the β-domains. These results will be compared with the transition-state structures of other proteins characterized by a molten globule-like folding intermediate and also our previous results from the unfolding simulations of goat α-LA by molecular dynamics.[20]

Results

CD spectra

Figure 2 shows the far and near-ultra violet (UV) CD spectra of wild-type and mutant goat α-LA in the native state and the spectra of the wild-type protein in the unfolded state in 6 M guanidine hydrochloride (GdnHCl). The far and near-UV spectra except for those of W60A, Y103F, and W118F are coincident with those of the wild-type protein, suggesting that all the proteins except W60A, Y103F, and W118F proteins form the same secondary and tertiary structures in the native state as the wild-type protein (Figure 2(c) and (d)). The spectra in the unfolded state for the mutant proteins were essentially the same as those for the wild-type protein.

The mutants in which substitutions were made for aromatic residues, W60A, Y103F, and W118F, showed far and near-UV CD spectra markedly different from those of the wild-type protein (Figure 2(a) and (b)). Y103F showed decreases in the CD intensity around 225 and 280 nm, and these decreases most likely arose from the disappearance of the L_a (225 nm) and L_b (280 nm) tyrosyl bands caused by the mutation.[24,25] W118F showed a decrease in CD intensity over a broad range in the near-UV region from 260 to 300 nm; this decrease was probably due to disappearance of the L_a (269 nm) and L_b (282 nm) tryptophan bands caused by the mutation.[24,26] W60A, however, showed more complex behavior. The CD intensity of W60A increased around 280 nm, decreased around 230 nm, and again increased around 215 nm. This behavior of the CD spectral change of W60A may be interpreted in terms of the

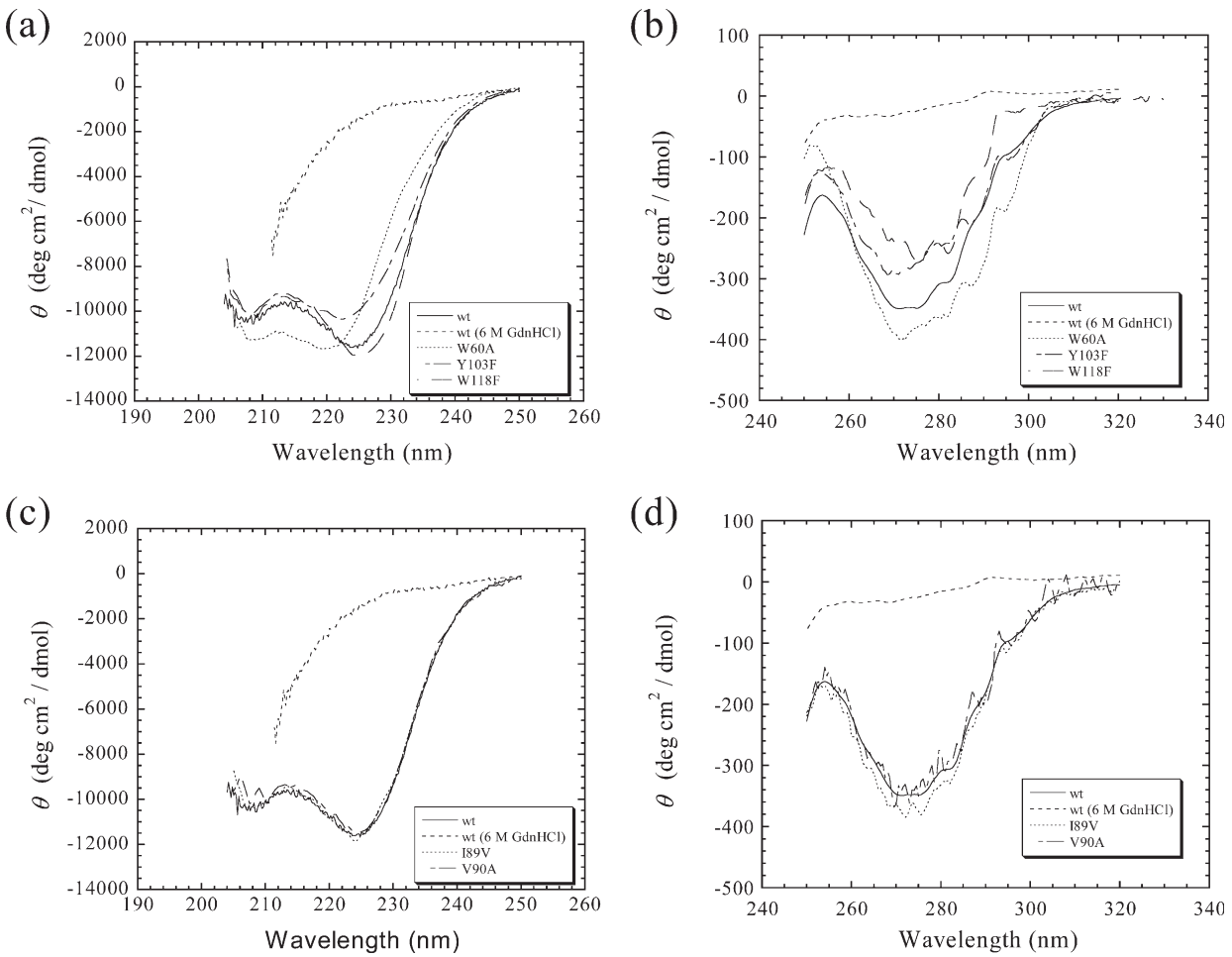

Figure 2. (a) and (c) Far and (b) and (d) near-UV CD spectra of the wild-type (—) and mutants of goat α-LA: (a) and (b) W60A (···), Y103F (– – –), and W118F (- –). (c) and (d) I89V (···) and V90A (- –). The spectra of the wild-type protein in the unfolded state at 6 M GdnHCl (- - -) are also shown.

disappearance of aromatic–aromatic interactions, which are known to exhibit large coupled-oscillator contributions in the CD spectra.[24,27] In the native structure of goat α-LA, the side-chain of Trp60 is adjacent to those of Tyr103 and Trp104, forming an aromatic cluster. The W60A mutation thus eliminated aromatic–aromatic interactions in this cluster, leading to the complex CD change in the mutant. These results suggest that the CD spectral changes of these mutants can be interpreted in terms of the changes of the aromatic side-chains with no essential changes occurring in the native conformation.

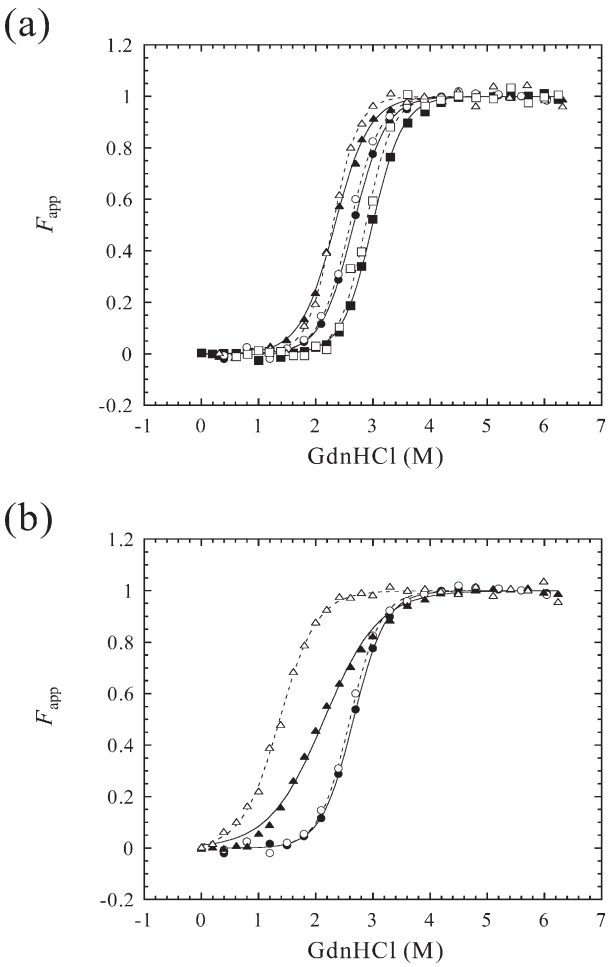

Figure 3. (a) GdnHCl-induced equilibrium unfolding transition curves of the wild-type (circles), I89V (triangles) and V90A (squares) of goat α-LA monitored by far-UV CD (filled symbols) and near-UV CD (open symbols). The apparent fractional extents of unfolding (F_{app}) are plotted against the molar concentration of GdnHCl. Similar coincident transition curves were observed for V8A, I95V, L96A, L110A, and W118F. (b) GdnHCl-induced equilibrium unfolding transition curves of the wild-type (circles) and Y103F (triangles) of goat α-LA monitored by far-UV CD (filled symbols) and near-UV CD (open symbols). Non-coincident transition curves, similar to those observed for Y103F, were also observed for L12A, V27A, L52A, I55V, W60A, D87N, K93A, and L105A. The data for the wild-type protein were taken from Yoda *et al*.[20]

Equilibrium unfolding transitions

Figure 3 shows transition curves of the GdnHCl-induced equilibrium unfolding of wild-type and mutant α-LA as monitored by the CD ellipticity at 225 nm and 270 nm. The transition curves are represented by the apparent fractional extents of unfolding, which were calculated from the observed ellipticity values and the values linearly extrapolated from pre and post-transition regions (i.e. baselines for the native and the unfolded states, respectively).[28] The transition curve monitored by the far-UV CD (225 nm) indicates disruption of the secondary structure of the protein by addition of GdnHCl, while the transition curve monitored by the near-UV CD (270 nm) indicates disruption of the specific side-chain packing of the protein.[19] For the wild-type and seven mutants (V8A, I89V, V90A, I95V, L96A, L110A, and W118F), the transition curves measured at the two different wavelengths were essentially coincident with each other.[28] Thus, we analyzed the transition curves assuming a two-state transition between the native and the unfolded states for these proteins (see below).

For the other nine mutants (L12A, V27A, L52A, I55V, W60A, D87N, K93A, Y103F, and L105A), however, the transition measured by the far-UV CD occurred at a significantly higher concentration of GdnHCl than the transition measured by the near-UV CD (Figure 3(b)). This disagreement of the transition curves indicates the accumulation of an unfolding intermediate (the molten globule state), in which the secondary structure is retained, but the specific side-chain packing is disrupted, at intermediate concentrations of GdnHCl. To analyze the transition curves, we assumed a three-state model for the unfolding transitions of these mutants (see below).

Kinetic refolding and unfolding reactions

Kinetic refolding and unfolding reactions of the mutants of goat α-LA were measured by the stopped-flow CD technique at 225 nm. Figure 4 shows the representative refolding and unfolding curves of the I89V mutant. The kinetic unfolding curve was well fitted to a single-exponential function, and this behavior was observed in other mutants as well. On the other hand, the refolding curve at 0.5 M GdnHCl was fitted to a two-exponential function, and additionally a burst phase occurring within the dead time of the stopped-flow measurement (25 ms) was also observed in refolding, indicating the accumulation of an intermediate within the burst phase. The same behavior was also observed in other mutant proteins although the kinetics were two or three-exponential depending on the mutant proteins. The exception was the D87N mutant, which showed a

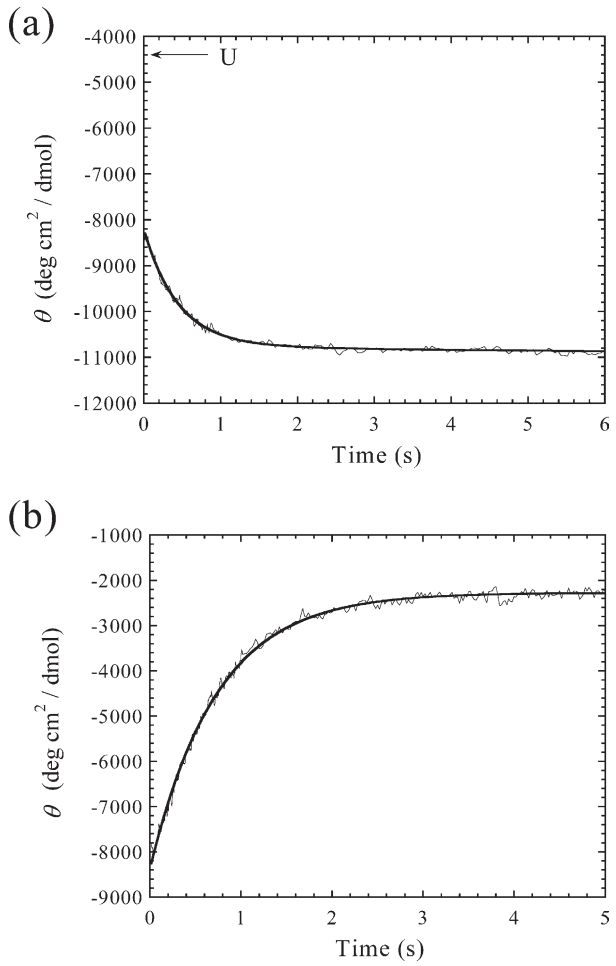

Figure 4. (a) Kinetic refolding and (b) unfolding reaction curves of the I89V mutant at 0.5 M and 4.0 M GdnHCl, respectively. Thick continuous lines are theoretical kinetic progress curves assuming a double and single-exponential function for refolding and unfolding, respectively (equation (6)). (a) The refolding rate constants of the fast and slow phases of the double-exponential fitting were 2.3 s^{-1} and 0.2 s^{-1}, respectively. The amplitude of the fast phase was about 89% of the total amplitude of the two phases. U denotes the CD value of I89V in the unfolded state obtained by extrapolation of the baseline of the unfolded state to 0.5 M GdnHCl. (b) The unfolding rate constant of the single-exponential fitting was 1.4 s^{-1}.

single-exponential kinetics and a burst phase occurring within the dead time of the measurement in the refolding reaction. Therefore, except for D87N, there are three or four phases (the burst phase, the fast phase, the middle phase, and the slow phase) in the refolding of these proteins at low GdnHCl concentration. A similar four-phases reaction has also been observed for the wild-type protein.[20] In the case of D87N, the rate of refolding after the burst phase was 27-fold slower than the corresponding refolding rate of the wild-type protein (see Table 2 and Figure 8(c), below); this may have made the apparent kinetics of refolding single-exponential.

Analysis of the unfolding transition of the burst phase intermediate

One of the advantages of the use of the stopped-flow CD technique in measuring the folding of α-LA is its large burst phase amplitude, which enables us to measure the unfolding transition curve of the burst phase intermediate. Such measurement is not possible when we use a stopped-flow fluorescence technique, because the fluorescence change within the burst phase is very small in the refolding of α-LA.[19] The ellipticity value extrapolated to zero time, θ_{zero}, along the kinetic refolding curve at 225 nm can be taken as a measure of the concentration of the species formed within the burst phase; hence, the GdnHCl concentration dependence of θ_{zero} gives an apparent unfolding transition curve of the burst phase intermediate.[19] Figure 5 shows typical unfolding transition curves of the burst phase folding intermediate thus obtained for the wild-type and three mutant proteins. Although the equilibrium unfolding transition curves of the wild-type and different mutant proteins are significantly different from each other, the transition curves of the burst phase intermediate of all these proteins are fairly similar.

For the sake of simplicity, we assume a two-state transition for the unfolding of the burst phase intermediate. Hence, θ_{zero} is given by:

$$\theta_{zero} = \frac{\theta_I + \theta_U \exp\{(-\Delta G^0_{IU} + m_{IU}c)/RT\}}{1 + \exp\{(-\Delta G^0_{IU} + m_{IU}c)/RT\}} \quad (1)$$

where ΔG^0_{IU} is the Gibbs free energy difference between the burst phase intermediate and the unfolded state at 0 M GdnHCl, and m_{IU} is the

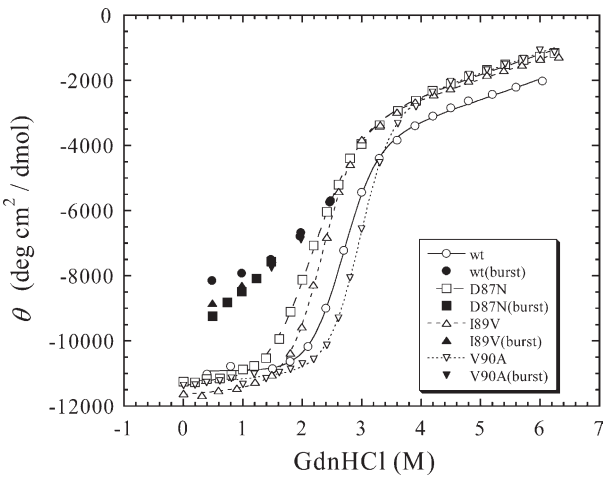

Figure 5. The GdnHCl-induced unfolding transition curves of the burst phase intermediate (filled symbols) and the equilibrium unfolding transition curves (open symbols), measured by the CD ellipticity at 225 nm of wild-type (- ○ - ○ -, ●), D87N (- □ - □ -, ■), I89V (- - △ - - △ - -, ▲) and V90A (· · · ▽ · · · ▽ · · ·, ▼). Each value for the burst phase intermediate was obtained by the extrapolation to zero time of the kinetic refolding curve measured under the corresponding GdnHCl concentration.

cooperativity index of the unfolding transition from the burst phase intermediate. θ_I and θ_U are the ellipticity values of the pure burst phase intermediate and the pure unfolded state, respectively, and are assumed to be identical to the θ_N and θ_U obtained in the analysis of the equilibrium unfolding transition from the native state (see below).

For the proteins that showed two-state equilibrium unfolding, ΔG^0_{IU} and m_{IU} values were obtained from the unfolding transition curves of the burst phase intermediate by the non-linear least-squares method using equation (1). For the proteins that show the three-state equilibrium unfolding, we used global fitting analysis, using the equilibrium unfolding data at 225 nm and 270 nm as well as the unfolding data of the burst phase intermediate at 225 nm, to calculate the ΔG^0_{IU} and m_{IU} (see below). The parameter values of ΔG^0_{IU} and m_{IU} thus obtained are summarized in Table 1.

Analysis of the equilibrium unfolding

Two-state transition

We used the following equation for the observed ellipticity, θ_{obs}, as a function of the GdnHCl concentration, c, to analyze the equilibrium transition curves for the two-state proteins (wild-type, V8A, I89V, V90A, I95V, L96A, L110A, and W118F):

$$\theta_{obs} = \frac{\theta_N + \theta_U \exp\{(-\Delta G^0_{NU} + mc)/RT\}}{1 + \exp\{(-\Delta G^0_{NU} + mc)/RT\}} \quad (2)$$

Here, we assumed that θ_N and θ_U have linear

Table 1. Thermodynamic parameters of the wild-type and 16 mutants of goat α-LA

Protein	Mechanism[a]	c_{NU} (M)	m_{NU} (kcal/mol M)	ΔG^0_{NU} (kcal/mol)	c_{NI} (M)	m_{NI} (kcal/mol M)	ΔG^0_{NI} (kcal/mol)	c_{IU} (M)	m_{IU} (kcal/mol M)	ΔG^0_{IU} (kcal/mol)	$\Delta\Delta G$[b] (kcal/mol)
Wild-type[c]	2	2.66 (0.01)	2.08 (0.05)	5.54 (0.13)	3.19 (0.18)	1.59 (0.08)	5.06 (0.16)	0.97 (0.06)	0.49 (0.06)	0.48 (0.08)	
Wild-type (au)[d]	2	3.15 (0.01)	2.6 (0.1)	8.2 (0.4)							
Wild-type (r)[d]	2	2.67 (0.01)	2.3 (0.1)	6.2 (0.1)							
V8A	2	2.26 (0.01)	2.10 (0.08)	4.74 (0.19)	2.76 (0.25)	1.48 (0.11)	4.10 (0.21)	1.04 (0.07)	0.62 (0.07)	0.64 (0.09)	−0.85 (0.05)
L12A	3	1.21 (0.06)	2.60 (0.26)	3.17 (0.36)	1.37 (0.06)	2.07 (0.32)	2.84 (0.46)	0.62 (1.20)	0.53 (0.41)	0.33 (0.58)	−3.75 (0.41)
V27A	3	1.81 (0.01)	2.70 (0.11)	4.88 (0.20)	2.17 (0.04)	1.90 (0.12)	4.13 (0.27)	0.94 (0.46)	0.80 (0.16)	0.75 (0.34)	−2.31 (0.10)
L52A	3	1.15 (0.02)	2.03 (0.12)	2.34 (0.14)	1.16 (0.03)	1.52 (0.12)	1.76 (0.15)	1.13 (0.55)	0.51 (0.17)	0.58 (0.21)	−3.06 (0.19)
I55V	3	1.65 (0.09)	1.37 (0.24)	2.25 (0.42)	2.38 (0.84)	0.62 (0.37)	1.48 (1.02)	1.04 (1.60)	0.75 (0.44)	0.78 (1.10)	−1.38 (0.28)
W60A	3	1.59 (0.06)	2.40 (0.25)	3.82 (0.43)	1.58 (0.05)	1.79 (0.28)	2.82 (0.46)	1.65 (1.45)	0.61 (0.38)	1.00 (0.62)	−2.55 (0.30)
D87N	3	2.07 (0.01)	2.18 (0.08)	4.51 (0.16)	2.36 (0.04)	1.46 (0.09)	3.43 (0.21)	1.49 (0.44)	0.72 (0.12)	1.08 (0.27)	−1.29 (0.06)
I89V	2	2.33 (0.01)	1.95 (0.03)	4.54 (0.08)	2.95 (0.27)	1.30 (0.11)	3.83 (0.15)	1.09 (0.10)	0.65 (0.10)	0.71 (0.13)	−0.64 (0.02)
V90A	2	2.99 (0.01)	2.22 (0.05)	6.63 (0.16)	3.50 (0.20)	1.64 (0.08)	5.74 (0.19)	1.54 (0.08)	0.58 (0.06)	0.89 (0.10)	0.73 (0.03)
K93A	3	1.83 (0.04)	2.12 (0.15)	3.88 (0.29)	2.07 (0.04)	1.51 (0.13)	3.13 (0.27)	1.24 (0.77)	0.61 (0.20)	0.76 (0.40)	−1.71 (0.15)
I95V	2	1.92 (0.01)	1.92 (0.05)	3.69 (0.09)	2.46 (0.17)	1.21 (0.07)	2.98 (0.12)	1.00 (0.06)	0.71 (0.05)	0.71 (0.06)	−1.42 (0.04)
L96A	2	2.00 (0.01)	1.72 (0.05)	3.44 (0.10)	2.68 (0.23)	1.09 (0.08)	2.91 (0.12)	0.83 (0.06)	0.63 (0.06)	0.53 (0.06)	−1.14 (0.04)
Y103F	3	1.76 (0.05)	1.82 (0.11)	3.20 (0.22)	1.64 (0.04)	1.30 (0.13)	2.13 (0.22)	2.07 (0.91)	0.52 (0.17)	1.08 (0.31)	−1.63 (0.14)
L105A	3	2.18 (0.02)	2.75 (0.22)	6.00 (0.47)	2.55 (0.09)	2.00 (0.25)	5.10 (0.66)	1.22 (1.21)	0.74 (0.33)	0.90 (0.81)	−1.31 (0.12)
L110A	2	2.38 (0.01)	2.29 (0.08)	5.45 (0.19)	2.99 (0.20)	1.69 (0.09)	5.04 (0.19)	0.69 (0.05)	0.60 (0.05)	0.41 (0.05)	−0.63 (0.04)
W118F	2	2.49 (0.01)	2.12 (0.04)	5.27 (0.10)	2.97 (0.25)	1.64 (0.12)	4.86 (0.18)	0.85 (0.24)	0.48 (0.11)	0.41 (0.15)	−0.36 (0.02)

[a] For the two-state mutants, thermodynamic parameters (c_{NU}, m_{NU}, and ΔG^0_{NU}) were obtained from the equilibrium unfolding transition curves. The values of c_{IU}, m_{IU}, and ΔG^0_{IU} were obtained from analysis of the unfolding transition of the burst phase intermediate. For the three-state mutants, the thermodynamic parameters were calculated by global fitting.

[b] $\Delta\Delta G$ is the difference in the ΔG between the mutant and the wild-type at the transition midpoint of the wild-type protein.

[c] The data for the recombinant wild-type protein were taken from Yoda *et al*.[20]

[d] The data for the authentic (au) and recombinant (r) wild-type protein were taken from Chaudhuri *et al*.[53] (Note that the parameter values by Chaudhuri[53] were corrected.)[54] Standard errors are given in parentheses.

dependence on c, as $\theta_N = a_1 + a_2c$ and $\theta_U = a_3 + a_4c$, and the values of a_1, a_2, a_3, and a_4 were obtained from the dependence of the ellipticity on c in the pre and post-transition regions. ΔG^0_{NU} is the Gibbs free energy difference between the native and the unfolded states at 0 M GdnHCl, and m is the cooperativity index of the equilibrium unfolding transition; R and T are the gas constant and absolute temperature, respectively.[29] The values of ΔG^0_{NU} and m were obtained by the non-linear least-squares method using equation (2), and these values thus obtained, as well as the transition midpoint, c_M ($= \Delta G^0_{NU}/m$), are listed in Table 1. The results show that all of the mutants except V90A are less stable than the wild-type protein. This destabilization of the mutants may be due to the removal of the side-chains, which are tightly packed in the native state of the wild-type protein. Because alanine stabilizes a helix more than valine,[23] the valine-to-alanine substitution of V90A may stabilize the C-helix on which the mutation site is located, hence making this mutant more stable than the wild-type.

Three-state transition

We used the following equation for the observed ellipticity, θ_{obs}, as a function of the GdnHCl concentration, c, to analyze the transition curves for the three-state proteins (L12A, V27A, L52A, I55V, W60A, D87N, K93A, Y103F, and L105A) (equation (3)).[30] Here, we assumed that θ_N, θ_I, and θ_U have linear dependence on c, as $\theta_N = a_1 + a_2c$, $\theta_I = a_3 + a_4c$, and $\theta_U = a_5 + a_6c$. ΔG^0_{NU} and ΔG^0_{NI} are the Gibbs free energy differences at 0 M GdnHCl between the native and the unfolded states and between the native and the intermediate states, respectively, and m_{NU} and m_{NI} are the respective cooperativity indexes. As shown above for the two-state proteins, the values of a_1, a_2, a_5, and a_6 were obtained from the dependence of the ellipticity on c in the pre and post-transition regions. We further assumed that $\theta_I = \theta_N$ for the far-UV ellipticity (225 nm) and that $\theta_I = \theta_U$ for the near-UV ellipticity (270 nm).

We analyzed the equilibrium unfolding data measured at 225 nm and 270 nm for the three-state proteins and the unfolding data of the burst phase refolding intermediate measured at 225 nm for the same proteins by global fitting using equations (1) and (3) (Figure 6). We assumed that the equilibrium unfolding intermediate in these three-state proteins is identical to the burst phase refolding intermediate. The fitting variables were ΔG^0_{NI}, ΔG^0_{NU}, m_{NI} and m_{NU} for each protein. The parameter values thus obtained are listed in Table 1; they produce theoretical transition curves in good agreement with the experimental data, supporting our assumption that the equilibrium intermediate is identical to the burst phase intermediate in refolding.

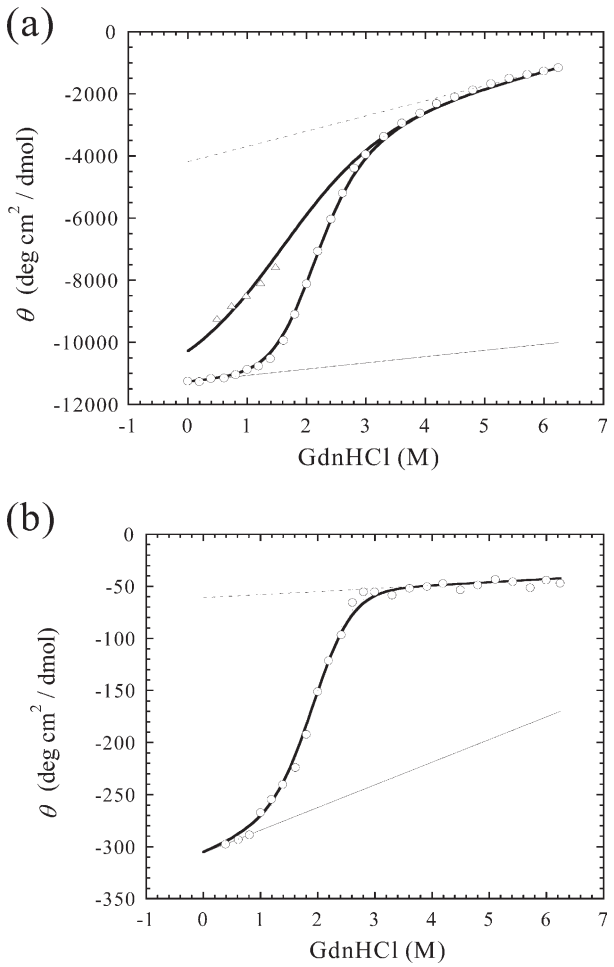

Figure 6. (a) The GdnHCl-induced equilibrium unfolding transition curve (circles) and the unfolding transition curve of the burst phase intermediate (triangles) of the D87N mutant measured by the CD at 225 nm. (b) The GdnHCl-induced equilibrium unfolding transition curve (circles) of the D87N mutant measured by the CD at 270 nm. Continuous and broken lines show the baselines of the N and U states, respectively.

The ΔG^0_{NU} values of the above mutants were all lower than that of the wild-type, indicating that the native state is less stable than that of the wild-type protein for the three-state unfolding proteins. In other words, the destabilization of the native state may change the unfolding behavior from the two-state type to the three-state type because the relative population of the intermediate is increased by the destabilization. The destabilization may be due to the removal of side-chain atoms, which are tightly packed in the native state of the wild-type protein.

$$\theta_{obs} = \frac{\theta_N + \theta_I \exp\{(-\Delta G^0_{NI} + m_{NI}c)/RT\} + \theta_U \exp\{(-\Delta G^0_{NU} + m_{NU}c)/RT\}}{1 + \exp\{(-\Delta G^0_{NI} + m_{NI}c)/RT\} + \exp\{(-\Delta G^0_{NU} + m_{NU}c)/RT\}} \quad (3)$$

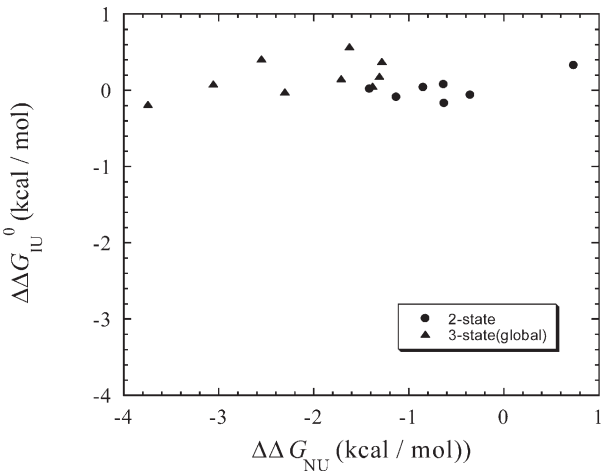

Figure 7. The plot of $\Delta\Delta G^0_{IU}$ *versus* $\Delta\Delta G_{NU}$ for the unfolding of goat α-LA. Filled circles indicate the two-state mutants (V8A, I89V, V90A, I95V, L96A, L110A, and W118F), and filled triangles indicate the three-state mutants (L12A, V27A, L52A, I55V, W60A, D87N, K93A, Y103F, and L105A).

Comparison between the unfolding free energies of the intermediate and the native state

Figure 7 shows the values of $\Delta\Delta G^0_{IU}$, which is the difference in ΔG^0_{IU} between the mutant and the wild-type proteins, as compared with $\Delta\Delta G_{NU}$, which is the difference in the equilibrium unfolding free energy between the mutant and the wild-type proteins at 2.66 M GdnHCl, the midpoint of the unfolding transition of the wild-type protein. The $\Delta\Delta G^0_{IU}$ values distributed in a narrow region around zero, whereas the $\Delta\Delta G_{NU}$ value changed from −4 to 1 kcal/mol. Thus, the stabilities of the intermediate state of the different proteins are more similar to each other than to the native-state stabilities. Figure 7 also indicates that there is no significant correlation between $\Delta\Delta G^0_{IU}$ and $\Delta\Delta G_{NU}$.

Φ-Value analysis

Figure 8 shows the GdnHCl concentration dependence of the rate constants of the refolding and the unfolding reactions (i.e. chevron plots) of the proteins measured by the stopped-flow CD. As described above, the burst phase and the subsequent three phases were observed in the kinetic refolding reaction of goat α-LA. The slow phase may arise from proline isomerization from *cis* to *trans* configuration,[31] because the rate constant of this phase of the wild-type protein has little dependence on GdnHCl concentration.[20] The middle phase was slightly dependent on GdnHCl concentration, and the amplitude of this phase was less than 20%. The fast phase had the largest amplitude among the phases (more than 80% of the total amplitude), and its rate constant remarkably decreased with increasing GdnHCl concentration, indicating that this phase reflected the real folding reaction. Therefore, we consider only the fast phase in the following analysis.

For the mutants with mutations located on the A-helix (V8A, L12A), the B-helix (V27A) (Figure 8(a)), the D-helix (L105A, L110A), and the C-terminal 3_{10}-helix (W118F) (Figure 8(e)), the unfolding rate was significantly faster than that of the wild-type, although the refolding rates were similar to that of the wild-type protein. Similar results were also obtained for the mutants with mutations on the β-domain (L52A, W60A), the C-helix (K93A, L96A), and the C–D loop (Y103F) (Figure 8(b) and (d)). On the other hand, compared with the wild-type protein, the D87N mutant, which had the mutation on the Ca^{2+}-binding site, showed a remarkably lower folding (27-fold), but a comparable unfolding rate (Figure 8(c)). In the β-domain mutant (I55V) and the C-helix mutants (I89V, V90A, and I95V), mutationally induced change in the folding rate is similar to the corresponding change in the unfolding rate (Figure 8(d)).

Figure 8 shows that the chevron plots of wild-type and mutant goat α-LAs do not have a "rollover" at low GdnHCl concentrations. Usually, the absence of the rollover is interpreted as an indication of a two-state folding reaction without accumulation of intermediates.[32,33] However, the results of the stopped-flow CD measurement of goat α-LA refolding clearly show the accumulation of the molten globule intermediate with sufficient structure and stability (Figure 5). Such apparent contradiction may be due to the high stability of the molten globule state[30] and/or the small cooperativity of the unfolding transition from the molten globule. To obtain kinetic parameters of unfolding and refolding, we analyzed the chevron plots of the wild-type and mutant proteins, based on the following equation, which represent the observed rate constant, k_{obs}, as a function of c:

$$\ln k_{obs} = \ln(k^0_{unf}\exp(m_{unf}c/RT) + k^0_{ref}\exp(-m_{ref}c/RT)) \quad (4)$$

where m_{unf}/RT and m_{ref}/RT are the respective derivatives of the natural logarithms of the unfolding and refolding rate constants, k_{unf} and k_{ref}, with respect to GdnHCl concentration; k^0_{unf} and k^0_{ref} are the k_{unf} and k_{ref} values at zero concentration of GdnHCl.[23] Here we assumed the apparent two-state model, in which k_{ref} and k_{unf}, respectively, are the rate constants of the refolding and unfolding reactions between the native state and the denatured state ensemble that involves both the molten globule and more extensively unfolded states. Although k_{ref} is affected by the fast equilibrium between the molten globule and the fully unfolded states, k_{unf} is independent of this equilibrium. Kinetic parameters thus obtained are listed in Table 2.

The Φ-values of folding for the mutants were

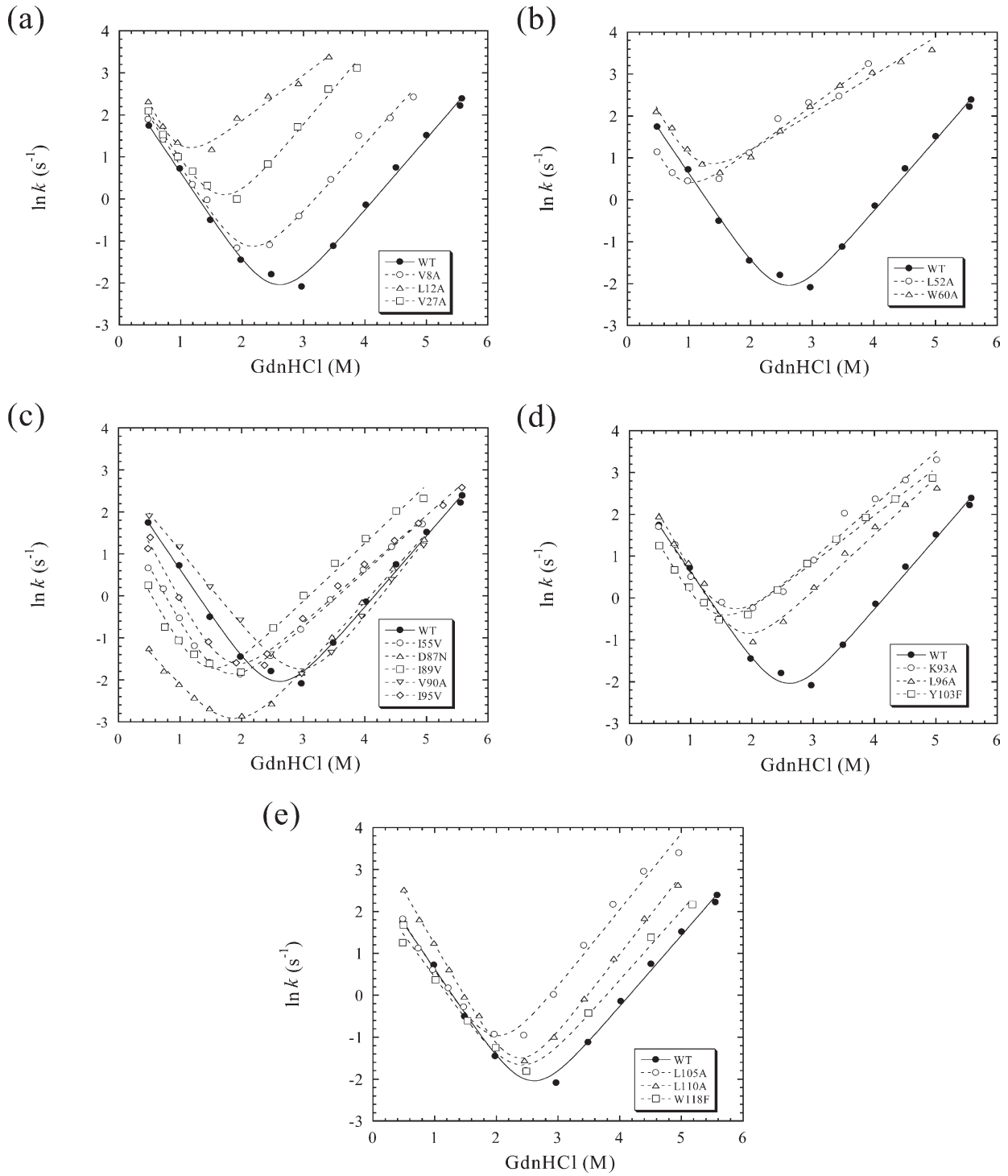

Figure 8. Chevron plots of the wild-type (– ● –) and mutants of goat α-LA. Only the rate constant of the fast phase is shown. (a) V8A (- -○- -), L12A (- -△- -), and V27A (- -□- -). (b) L52A (- -○- -) and W60A (- -△- -). (c) I55V (- -○- -), D87N (- -△- -), I89V (- -□- -), V90A (- -▽- -), and I95V (- -◇- -). (d) K93A (- -○- -), L96A (- -△- -), and Y103F (- -□- -). (e) L105A (- -○- -), L110A (- -△- -), and W118F (- -□- -). Theoretical curves fitted to equation (2) are shown for the wild-type (thick continuous lines) and mutants (broken lines). The data for the wild-type protein were taken from Yoda *et al*.[20]

calculated using the following equation:

$$\Phi = 1 - \frac{RT \ln(k_{unf}^{WT}(c_M^{WT})/k_{unf}^{mutant}(c_M^{WT}))}{\Delta\Delta G_{NU}(c_M^{WT})} \quad (5)$$

where k_{unf}^{WT} and k_{unf}^{mutant} are the unfolding rate constants of the wild-type and a mutant protein, respectively.[23] The Φ-value thus obtained does not depend on whether the chevron plots are analyzed based on the two-state or three-state folding

Table 2. Kinetic parameters of the wild-type and 17 mutants of goat α-LA obtained from the chevron plots

Protein	k^0_{ref} (s^{-1})	m_{ref} (kcal/mol per M)	k^0_{unf} (s^{-1})	m_{unf} (kcal/mol per M)	Φ
Wild-type[a]	16.3 (4.3)	1.28 (0.11)	0.0009 (0.0004)	1.00 (0.06)	
V8A	22.9 (7.4)	1.40 (0.18)	0.0056 (0.0033)	0.96 (0.09)	−0.16 (0.06)
L12A	48.9 (37.4)	2.09 (0.77)	0.73 (0.32)	0.65 (0.10)	0.19 (0.09)
V27A	23.5 (6.8)	1.35 (0.19)	0.033 (0.016)	1.03 (0.09)	0.05 (0.04)
T29I[a]	32.1 (6.3)	1.26 (0.05)	4.7×10^{-8} (9.2×10^{-8})	1.64 (0.23)	−0.18 (0.05)
L52A	14.2 (17.5)	2.12 (1.31)	0.38 (0.17)	0.63 (0.09)	0.14 (0.05)
I55V	7.71 (1.16)	1.61 (0.10)	0.0092 (0.0016)	0.78 (0.04)	0.43 (0.12)
W60A	34.2 (23.4)	1.77 (0.61)	0.52 (0.20)	0.54 (0.06)	0.004 (0.12)
D87N	0.60 (0.08)	0.97 (0.09)	0.0013 (0.0003)	0.97 (0.03)	0.91 (0.004)
I89V	7.0 (1.04)	1.33 (0.08)	0.0015 (0.0004)	1.01 (0.04)	0.46 (0.02)
V90A	17.4 (2.5)	1.06 (0.05)	0.00034 (0.00018)	1.11 (0.07)	0.62 (0.02)
K93A	18.7 (10.2)	1.45 (0.40)	0.056 (0.030)	0.76 (0.08)	−0.05 (0.09)
I95V	13.9 (3.8)	1.63 (0.18)	0.011 (0.003)	0.76 (0.04)	0.45 (0.02)
L96A	27.8 (11.0)	1.60 (0.24)	0.020 (0.010)	0.80 (0.08)	−0.17 (0.04)
Y103F	12.9 (4.4)	1.58 (0.26)	0.080 (0.022)	0.66 (0.04)	0.09 (0.10)
L105A	17.9 (8.1)	1.36 (0.26)	0.0053 (0.0037)	1.08 (0.11)	0.04 (0.09)
L110A	43.6 (5.8)	1.52 (0.07)	0.0014 (0.0004)	1.11 (0.04)	0.07 (0.05)
W118F	11.7 (3.4)	1.20 (0.15)	0.0019 (0.0017)	0.98 (0.12)	−0.12 (0.07)

[a] The data for the wild-type protein and T29I were taken from Yoda *et al.*[20] Standard errors are in parentheses.

mechanism, because the unfolding rate constants calculated by assuming the different folding mechanisms should be essentially identical to each other. Here, we calculated the Φ-values at the midpoint of the equilibrium unfolding transition of the wild-type protein, c_M^{WT} ($= 2.66$ M), because long extrapolation of thermodynamic and kinetic parameters to 0 M GdnHCl could have caused large errors in estimating the parameters.[23] The Φ-value is one if the structure around the mutation site is fully formed in the transition state of folding, and zero if the structure around the mutation site is not yet formed in the transition state.[23]

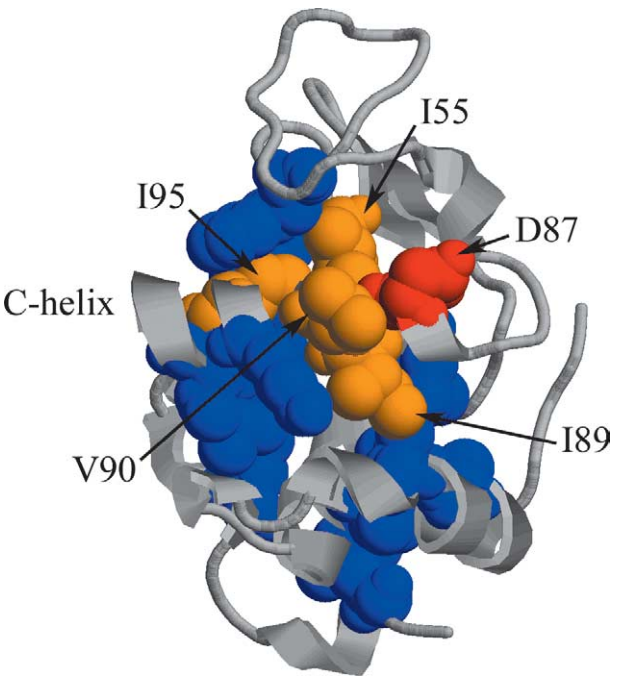

Figure 9. A schematic representation of the structure of goat α-LA in the transition state in folding. Residues with Φ-values between 0.8 and 1.0 are colored red; between 0.3 and 0.8, orange; lower than 0.4, blue.

We obtained the Φ-values of all the mutants from equation (5), and mapped out the transition-state structure of the folding from the molten globule state to the native state (Figure 9). The residue with a high Φ-value (0.8–1.0) is colored in red, the residues with intermediate Φ-values (0.3–0.8) are colored orange, and those with low Φ-values (<0.3) are colored blue. These results thus indicate that the structure around Asp87 is fully formed in the transition state of folding of goat α-LA and that the structures around Ile55, Ile89, Val90 and Ile95 are partially formed, while the structures around the other mutation sites are not yet organized.

Discussion

To characterize the structure in the transition state of folding/unfolding of goat α-LA, we constructed 16 mutants of the protein and studied their equilibrium unfolding transitions and kinetic refolding reactions. Including the two mutation sites (Thr29 and the N-terminal Met) previously studied, these 18 mutation sites are scattered around the whole molecule of α-LA. Any changes in the CD spectra in either the far or near-UV region of the mutant proteins were interpreted in terms of the amino acid replacement of an aromatic residue, suggesting that all the mutants had essentially the same native three-dimensional structure as the wild-type protein. We measured the stabilities of the intermediate and native states, and carried out Φ-value analysis of the transition states for these mutants of goat α-LA to map out the structure in the transition state.

The present results suggesting that the transition-state structure of goat α-LA is not

extensively distributed over the protein, but is localized around the Ca^{2+}-binding site and at the interface between the C-helix and the β-domain, are fully consistent with our previous studies. Kuwajima *et al*.[34] studied the Ca^{2+} concentration dependence of the refolding kinetics of bovine α-LA, which is almost exactly similar to the goat protein (with only seven differences in their amino acid sequences), and have shown that the structure around the Ca^{2+} binding site (Asp82, Asp87, and Asp88) is already formed in the transition state. In previous studies, Yoda *et al*. performed ϕ-value analysis of the T29I mutant of goat α-LA,[20] and Chaudhuri *et al*. measured the refolding and unfolding reactions of the wild-type protein in the authentic and recombinant forms.[6] Thr29 is on the B-helix, and buried within the hydrophobic core of the molecule at the center of the α-domain; an additional methionine residue attached to the N terminus in the recombinant form is located at a peripheral portion of the α-domain.[6] The native structures in these mutation sites were shown to be not yet organized in the transition state of folding.

In the following, we further discuss the unfolding free energies of the intermediate and the native state of the α-LA/lysozyme family, characterize the structure of the transition-state of folding of goat α-LA, compare its transition-state structure with those of other proteins that show molten globule-like folding intermediates, and check the present results against those of our previous molecular dynamics simulation studies of goat α-LA unfolding.[20]

Comparison between the unfolding free energies of the intermediate and the native state

We evaluated the free energy changes, ΔG_{NU} and ΔG^0_{IU}, for the wild-type and 16 mutant proteins, and calculated the differences, $\Delta\Delta G_{NU}$ and $\Delta\Delta G^0_{IU}$, of these parameter values between each mutant protein and the wild-type protein. The values of $\Delta\Delta G_{NU}$ of all the mutants except V90A were less than zero, indicating that these mutant proteins are less stable than the wild-type protein in the native state. Values for $\Delta\Delta G_{NU}$ varied from -3.7 to 0.8 kcal/mol, but $\Delta\Delta G^0_{IU}$ varied only in a narrow region around zero, and there was no significant correlation between $\Delta\Delta G^0_{IU}$ and $\Delta\Delta G_{NU}$ (Figure 7).

The narrow distribution of $\Delta\Delta G^0_{IU}$ around zero and the absence of a significant correlation between $\Delta\Delta G^0_{IU}$ and $\Delta\Delta G_{NU}$ may both arise from the molten globule characteristics of the intermediate state of goat α-LA. The molten globule state has a native-like secondary structure, and retains the compact shape of the protein molecule, but has lost most of the specific side-chain packing interactions.[2,4] Because the specific packing interactions are what contribute most to the mutationally induced free energy differences ($\Delta\Delta G_{NU}$ and $\Delta\Delta G^0_{IU}$), the absence of these interactions in the intermediates explains why the $\Delta\Delta G^0_{IU}$ values were close to zero.

The molten globule intermediate has been well characterized for the α-LA/lysozyme family, which includes goat, bovine, human and guinea pig α-LA, and equine and canine milk lysozyme.[2,4,7] All of these molten globule intermediates have similar overall structures in which the α-domain is more ordered than the β-domain. However, stabilities of the molten globule states are different among the proteins. This difference in the stabilities of the molten globules may be explained by the difference in packing density between secondary structural elements.[35,36]

Among the molten globule states of the α-LA/lysozyme family, the molten globule of goat α-LA is the least stable.[18] The molten globule of this protein seems to be an ideal molten globule in which the specific side-chain packing interactions are almost fully lost. The molten globules of the other proteins in this family are more organized and thus more stable. In fact, the molten globule states of some proteins in this family, especially equine and canine milk lysozyme, are highly structured and partially preserve the part of native tertiary packing interactions.[30,37] While there is no significant correlation between $\Delta\Delta G^0_{IU}$ and $\Delta\Delta G_{NU}$ in goat α-LA, these proteins with a structured molten globule would show a correlation between the two parameters. Hence the molten globule of goat α-LA, in which the specific side-chain packing interactions are almost fully lost, could be considered the closest to an ideal molten globule state.

Structure of the transition state of folding in goat α-LA

The degree of structural organization

The transition state of folding is located between the intermediate and the fully native state, and the ratio of the kinetic and equilibrium *m*-values ($\alpha = (1 - m_{unf})/m_{NU}$) represents the degree of structural organization in the transition state in terms of solvent accessibility. The mutants that unfold very rapidly have low values of m_{unf}. In particular, mutants L12A, L52A, and W60A, show a decrease in m_{unf} compared with the wild-type. In these mutants, the increase in α on mutation was thus observed, reflecting movement of the transition state closer to the native state. This is known as the Hammond effect.[23]

The folding nucleus

The structured region in the transition state is called the folding nucleus, and the folding nucleus of goat α-LA is very localized around the Ca^{2+}-binding site and at the interface between the C-helix and the β-domain. The structure around one of the Ca^{2+}-binding ligands (Asp87) is very highly organized in the transition state as indicated by the Φ-value (0.91) for the D87N mutant, and the

structures around the residues (Ile89 and Val90) near the Ca^{2+}-binding site and around the residues (Ile55 and Ile95) located at the interface between the α and β-domains are partially organized as indicated by the fractional Φ-values (0.43–0.62) for the I55V, I89V, V90A and I95V mutants. However, the structures around all the other mutation sites investigated have little or no organization as indicated by the Φ-values close to zero for those mutants. Therefore, the majority of the α-domain, including the A-helix, the B-helix, the α-domain side of the C-terminal half of the C-helix, and the D-helix, is not further organized during the folding process from the intermediate to the native state.

The presence of the folding nucleus at the interface between the α and the β-domains indicates that the specific docking of the α and the β-domains at this interface is necessary for α-LA to correctly organize the specific native structure from the molten globule intermediate. In accordance with this result, Peng & Kim have reported that a single-chain recombinant model of the α-domain of human α-LA, consisting of residues 1–39 and 81–123 of the human protein, forms a molten globule with the same overall tertiary fold as that found in intact α-LA, but does not form the native structure with specific side-chain interactions.[15] The α-domain itself is thus not sufficient to form the native state, although it is sufficient to form the molten globule state. Therefore, not only the α-domain but also the β-domain is necessary for the folding from the molten globule to the native state of goat and human α-LA.

The present results on the folding nucleus of goat α-LA may be compared with the folding nucleus predicted by a computational method based on Ting & Jernigan's sequence conservation cluster analysis.[38] Although Ting & Jernigan have proposed that the predicted folding nucleus may correspond to contacting residues in the early folding intermediate (the molten globule) of α-LA, it is also likely that it may correspond to the most critical residues to be structured in the transition state of folding. The folding nucleus predicted by Ting & Jernigan is formed by a large cluster of 19 residues, the side-chains of which are mostly hydrophobic and point toward the core of the protein to form the cluster in the native structure, although the residues are distributed over different structural elements. Among the 19 residues, there are only four (Val8, Leu12, Leu52 and Ile55) for which Φ-values have been experimentally determined, and among these four residues, only Ile55 has a significantly large Φ-value (0.43). Furthermore, there are four other residues exhibiting significantly large Φ-values (0.45–0.91) that are not among the 19 residues predicted by Ting & Jernigan. These latter four residues are all located within the C-helix; three of them (Asp87, Ile89 and Val90) are at or near the Ca^{2+}-binding site, and the last one (Ile95) is at the interface between the C-helix and the β-domain. Therefore, the present folding nucleus determined by Φ-value analysis is not identical to the folding nucleus predicted by Ting & Jernigan. However, it should be noted that both the nuclei involve the same portion of the interface between the α and β-domains, suggesting that the specific docking of the two domains may be an important step in the folding of α-LA.

The transition state and the intermediate state

There are many studies on the structure of the molten globule state of α-LA, and they are consistent in suggesting that the α-domain is more organized than the β-domain in this intermediate state of the protein. Hydrogen-exchange protection profiles of the molten globules of bovine and human α-LA show that the α-domain is more highly protected than the β-domain.[12,14] From limited proteolysis studies, the β-domain appears to be highly mobile or unfolded in the bovine α-LA molten globule, while the rest of the protein chain maintains sufficient structure and rigidity to prevent extensive proteolysis.[39] Peptide studies have shown that a peptide segment, containing the A and B-helices linked by a disulfide bond to the D and C-terminal 3_{10}-helices, is important for the stability of the α-domain structure in the molten globule state of human α-LA.[40] A study of urea-induced unfolding in the human α-LA molten globule has shown that the A, B, D and the C-terminal 3_{10}-helices are involved in a stable core that is resistant to a high concentration of denaturant.[12] These results thus indicate that the α-domain is more organized in the molten globule state of α-LA, and suggest a stable hydrophobic core composed of the A, B, D and the C-terminal 3_{10}-helices in human α-LA.

This presence of a stable core composed of the A, B, D and the 3_{10}-helices without extensive involvement of the C-helix in the molten globule state makes an apparent contrast with the transition-state structure of goat α-LA, in which the C-helix is much more organized than the A, B, and the D-helices. However, this difference is probably due to the difference in the α-LA species rather than the difference in stages, because a stable core in the molten globule state of α-LA is not found in many other species besides humans, and the comparable state in goat α-LA is expected to be different.

Although there are insufficient residue-specific studies on the goat α-LA molten globule, there are such studies on bovine α-LA, which has an amino acid sequence that is 94% identical with the sequence of the goat protein, but only 75% identical with that of the human protein.[5] Though the molten globule states of the bovine and human proteins exhibit similar behavior in many respects, a hydrogen exchange study of the bovine α-LA molten globule has shown that this state is substantially less protected than that of human α-LA and that the pattern of protected residues is also different between the two proteins. The hydrogen-exchange protection profile of the molten globule

state of bovine α-LA shows the highest, albeit small, protection factor (5–11) in the C-helix; this is in contrast with the molten-globule protection profile of the human protein, which shows the B-helix as most protected.[11,14] The A and B-helices of human α-LA are known to be significantly more hydrophobic and more stable against the urea-induced unfolding transition than those of bovine α-LA, showing that the unfolding patterns of individual helices in the two proteins are clearly different. In fact, the residue-specific NMR study of the urea-induced unfolding of the bovine and human α-LA molten globule shows that all four α-helices and the C-terminal 3_{10} helix unfold over a relatively narrow range of urea concentration in bovine α-LA, whereas the C-helix is found to be far less stable against unfolding than the other helices in human α-LA.[41]

Altogether, the above results strongly suggest that the C-helix is much more important for the folding in bovine α-LA, and this may also be true in goat α-LA as well, being consistent with the organization of the C-helix in the transition state of the goat protein revealed here. It is thus possible that the initiation site of the α-LA folding might be different for different α-LA species. Further studies are warranted for elucidating the species dependence, if any, of the folding initiation site in α-LA.

Comparison with other proteins

Here we compare the present results with those of other proteins that show the molten globule-like folding intermediate. Structures of the intermediate and the transition state of barnase have been well characterized by Φ-value analysis.[23] The transition state of barnase shows a heterogeneous structure in which some regions have Φ-values of zero and other regions have Φ-values close to 1. A core region of ribonuclease H has Φ-values close to 1 both in the intermediate and the transition state, while the mutations outside the core region do not affect the stability of the intermediate but show fractional Φ-values.[42] Tumor suppressor p16 consists of four ankyrin repeats. C-terminal repeats of p16 have native structure in the rate-determining transition state and their Φ-values are close to 1, whereas the N-terminal repeats are largely unstructured and the Φ-values are close to zero.[43] These proteins provide examples of transition states in which Φ-values are a mixture of close to 0 and 1.

On the other hand, there are proteins that have transition states with mainly fractional Φ-values, such as CheY,[44] CD2.d1,[45] FNIII,[46] and Im7.[47] The transition state of CD2.d1 has Φ-values of 0–0.5; the residues with non-zero Φ-values are located in close spatial proximity to the central core tryptophan residue, suggesting that the transition state is attained by formation of a tightly localized hydrophobic nucleus.[45] The transition-state structure of goat α-LA is thus more similar to this type of transition state in that it has zero or fractional Φ-values and a highly localized structure.

Comparison with the results of molecular dynamics simulations

Molecular dynamics simulation is a powerful tool to characterize protein (un)folding reactions with the resolution of atoms.[20,48] We have previously performed molecular dynamics simulation of the unfolding of goat α-LA and compared those results with experimental data.[20] In the analysis, the structure of the transition state was considered to be similar to the structure at the initial stages of the unfolding observed in the simulation trajectories. We showed that the simulation results were consistent with the experimental data in that the structures around Thr29 and the N terminus are not yet folded in the transition state of folding.

The present results show that Ile55 and Ile95 are partially folded in the transition state but that Val8, Leu12, Val27, Leu52, Trp60, Tyr103, Leu105 and Trp118 are not; these are also consistent with the results of molecular dynamics simulations performed by Yoda *et al*.[20] The simulations suggested that the side-chain–side-chain contacts between Ile55 and Ile95 (aromatic cluster II) are formed in the transition state (not disrupted at the initial stages of the unfolding). On the other hand, the non-local contacts of Leu105 with the B-helix, the non-local contacts between the A and B-helices (Val8, Leu12, and Val27), and the non-local contacts of Trp118 in aromatic cluster I were all partially disrupted at an initial stage of the unfolding simulation.[20] Leu52 and Trp60 are located near the inter-domain boundary. Leu52, which had non-local inter-residue contacts with amino acid residues in the loop between the A and B-helices in the native state showed an increase of the side-chain accessible surface area at an early stage of the unfolding simulation. Some of the native contacts formed by Trp60 with the C-helix were disrupted at an early stage of the unfolding simulation. Although most of the native inter-residue contacts in aromatic cluster II were stable, some of the native contacts around Tyr103 were found to be unstable. These simulation results are thus consistent with the present experimental results. Therefore, we have shown that molecular dynamics simulations can be used to characterize the detailed structure of the transition state of protein folding with a resolution of atoms.

Materials and Methods

Materials

Goat α-LA expression plasmid pSCREEN-LA was constructed by inserting a XbaI-XhoI fragment of pUT7-LA, which was constructed by Uchiyama *et al*.[18] into the XbaI-XhoI site of pSCREEN-1b(+) (Novagen). A plasmid encoding an α-LA mutant was constructed from the

plasmid pSCREEN-LA by the Kunkel method with a MUTA-GENE M13 *in vitro* mutagenesis kit (Bio-Rad) or QuikChange™ site-directed mutagenesis kit (Stratagene).

All mutants of goat α-LA were purified as described by Yoda *et al*.[20] In brief, all proteins were expressed in *Escherichia coli* as inclusion bodies, which were solubilized by 7 M urea and applied to a DEAE Sepharose FF column equilibrated with 20 mM Tris–HCl (pH 8.5), 1 mM EDTA, 7 M urea, and 5 mM DTT, and were eluted by a linear gradient of NaCl concentration from 0 to 0.3 M. Oxidative refolding of the unfolded proteins was performed by dialysis against a refolding buffer containing 50 mM NaCl, 1 mM $CaCl_2$, 2.5 mM reduced glutathione, and 0.5 mM oxidized glutathione at pH 8.5 and 4 °C. The refolded proteins were again applied to the DEAE Sepharose FF column equilibrated with 20 mM Tris–HCl (pH 8.5) and 1 mM $CaCl_2$, and were eluted by a linear gradient of NaCl concentration from 0 to 0.3 M. Purity of the mutant proteins was checked by native poly-acrylamide gel electrophoresis under non-denaturing conditions. The purified proteins were dialyzed against water and lyophilized.

GdnHCl was of a specially prepared reagent grade for biochemical use from Nacalai Tesque, Inc. (Kyoto, Japan). Other chemicals were of guaranteed reagent grade. The concentration of α-LA except W60A, Y103F and W118F was determined spectrophotometrically using an extinction coefficient, $E^{1\%}_{1\,cm} = 20.1$, at 280 nm.[49] The extinction coefficients for the W60A, Y103F and W118F mutants were calculated using the method of Pace *et al*.[50] The concentration of GdnHCl was determined from the refractive index at 589 nm.[29] Solutions were filtered through membrane filters (pore size 0.45 μm) before spectroscopic measurements. All solutions contained 50 mM sodium cacodylate (pH 7.0), 50 mM NaCl, and 1.0 mM $CaCl_2$. The protein concentrations were typically 0.5 and 0.15 mg/ml for the equilibrium and kinetic measurements, respectively.

CD and fluorescence measurements

Equilibrium and kinetic CD measurements were carried out in a Jasco J-720 spectropolarimeter calibrated with ammonium *d*-10-camphorsulfonate. The temperature of the cuvette for equilibrium and kinetic CD and fluorescence measurements was controlled by circulating water. All experiments were done at 25 °C. Path lengths of the cuvette were 1.00 nm or 2.00 mm for the far-UV CD measurements and 10.0 mm for the near-UV CD measurements.

The stopped-flow apparatus attached to the spectropolarimeter was the same as that described.[19,28,51,52] The path length was 3.9 mm and the dead time of measurement was 25 ms. The refolding reaction was initiated by mixing the unfolded protein in 5.5 M GdnHCl with the refolding buffer containing an appropriate amount of GdnHCl to give the final folding conditions. The unfolding reaction was initiated by mixing the native protein in 0 or 0.5 M GdnHCl with the unfolding buffer containing an appropriate amount of GdnHCl to give the final unfolding conditions.

Data analysis

The time-dependent change in the far-UV CD at 225 nm was fitted by the method of non-linear least-squares to the following equation:

$$A(t) = A(\infty) + \sum_i \Delta A_i \exp(-k_i t) \qquad (6)$$

where $A(t)$ and $A(\infty)$ are the CD intensities at time t and at infinite time, respectively, ΔA_i is the change in the amplitude of the ith phase, and k_i is the rate constant of the reaction at the ith phase.

Acknowledgements

We thank Ms Mihoko Takahashi, Mr Kouji Kawase, Mr Hirotaka Ishii, and Mr Tomonori Nozawa for technical assistance. This study was supported by Grants-in-Aid for Scientific Research from Ministry of Education, Science and Culture of Japan.

References

1. Kim, P. S. & Baldwin, R. L. (1990). Intermediates in the folding reactions of small proteins. *Annu. Rev. Biochem.* **59**, 631–660.
2. 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.
3. Ptitsyn, O. B. (1995). Molten globule and protein folding. *Advan. Protein Chem.* **47**, 83–229.
4. Arai, M. & Kuwajima, K. (2000). Role of the molten globule state in protein folding. *Advan. Protein Chem.* **53**, 209–282.
5. Pike, A. C., Brew, K. & Acharya, K. (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.
6. Chaudhuri, T. K., Horii, K., Yoda, T., Arai, M., Nagata, S., Terada, T. P. *et al.* (1999). Effect of the extra N-terminal methionine residue on the stability and folding of recombinant α-lactalbumin expressed in *Escherichia coli*. *J. Mol. Biol.* **285**, 1179–1194.
7. Arai, M., Ito, K., Inobe, T., Nakao, M., Maki, K., Kamagata, K. *et al.* (2002). Fast compaction of α-lactalbumin during folding studied by stopped-flow X-ray scattering. *J. Mol. Biol.* **321**, 121–132.
8. 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.
9. 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.
10. Chyan, C. L., Wormald, C., Dobson, C. M., Evans, P. A. & Baum, J. (1993). Structure and stability of the molten globule state of guinea-pig α-lactalbumin: A hydrogen exchange study. *Biochemistry*, **32**, 5681–5691.
11. Schulman, B. A., Redfield, C., Peng, Z. Y., Dobson, C. M. & Kim, P. S. (1995). Different subdomains are most protected from hydrogen exchange in the

molten globule and native states of human α-lactalbumin. *J. Mol. Biol.* **253**, 651–657.
12. 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.
13. Balbach, J., Forge, V., Lau, W. S., Jones, J. A., Van Nuland, N. A. J. & Dobson, C. M. (1997). Detection of residue contacts in a protein folding intermediate. *Proc. Natl Acad. Sci. USA*, **94**, 7182–7185.
14. Forge, V., Wijesinha, R. T., Balbach, J., Brew, K., Robinson, C. V., Redfield, C. & Dobson, C. M. (1999). Rapid collapse and slow structural reorganisation during the refolding of bovine α-lactalbumin. *J. Mol. Biol.* **288**, 673–688.
15. Peng, Z. & Kim, P. S. (1994). A protein dissection study of a molten globule. *Biochemistry*, **33**, 2136–2141.
16. Wu, L. C., Peng, Z. & Kim, P. S. (1995). Bipartite structure of the α-lactalbumin molten globule. *Nature Struct. Biol.* **2**, 281–286.
17. 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.
18. Uchiyama, H., Perez-Prat, E. M., Watanabe, K., Kumagai, I. & Kuwajima, K. (1995). Effects of amino acid substitutions in the hydrophobic core of α-lactalbumin on the stability of the molten globule state. *Protein Eng.* **8**, 1153–1161.
19. Arai, M. & Kuwajima, K. (1996). Rapid formation of a molten globule intermediate in refolding of α-lactalbumin. *Fold. Des.* **1**, 275–287.
20. Yoda, T., Saito, M., Arai, M., Horii, K., Tsumoto, K., Matsushima, M. *et al.* (2001). Folding-unfolding of goat α-lactalbumin studied by stopped-flow circular dichroism and molecular dynamics simulations. *Proteins: Struct. Funct. Genet.* **42**, 49–65.
21. Anderson, P. J., Brooks, C. L. & Berliner, L. J. (1997). Functional identification of calcium binding residues in bovine α-lactalbumin. *Biochemistry*, **36**, 11648–11654.
22. Permyakov, S. E., Uversky, V. N., Veprintsev, D. B., Cherskaya, A. M., Brooks, C. L., Permyakov, E. A. & Berliner, L. J. (2001). Mutating aspartate in the calcium-binding site of α-lactalbumin: effects on the protein stability and cation binding. *Protein Eng.* **14**, 785–789.
23. Fersht, A. R. (1999). *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, W.H. Freeman and Company, New York.
24. Cantor, C. R. & Schimmel, P. R. (1980). *Biophysical Chemistry Part II: Techniques for the Study of Biological Structure and Function*, W.H. Freeman and Company, New York.
25. Kurapkat, G., Kruger, P., Wollmer, A., Fleischhauer, J., Kramer, B., Zobel, E. *et al.* (1997). Calculation of the CD spectrum of bovine pancreatic ribonuclease. *Biopolymers*, **41**, 267–287.
26. Woody, R. W. (1994). Contributions of tryptophan side chains to the far-ultraviolet circular dichroism of proteins. *Eur. Biophys. J.* **23**, 253–262.
27. Grishina, I. B. & Woody, R. W. (1994). Contributions of tryptophan side chains to the circular dichroism of globular proteins: exciton couplets and coupled oscillators. *Faraday Discuss.* **99**, 245–262.
28. Arai, M., Hamel, P., Kanaya, E., Inaka, K., Miki, K., Kikuchi, M. & Kuwajima, K. (2000). Effect of an alternative disulfide bond on the structure, stability, and folding of human lysozyme. *Biochemistry*, **39**, 3472–3479.
29. Pace, C. N. (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* **131**, 266–280.
30. Mizuguchi, M., Arai, M., Ke, Y., Nitta, K. & Kuwajima, K. (1998). Equilibrium and kinetics of the folding of equine lysozyme studied by circular dichroism spectroscopy. *J. Mol. Biol.* **283**, 265–277.
31. Balbach, J. & Schmid, F. X. (2000). Proline isomerization and its catalysis in protein folding. In *Mechanisms of Protein Folding* (Pain, R. H., ed.), 2nd edit., pp. 212–249, Oxford University Press, New York.
32. Jackson, S. E. & Fersht, A. R. (1991). Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry*, **30**, 10428–10435.
33. Baldwin, R. L. (1996). On-pathway versus off-pathway folding intermediates. *Fold. Des.* **1**, R1–R8.
34. Kuwajima, K., Mitani, M. & Sugai, S. (1989). Characterization of the critical state in protein folding. Effects of guanidine hydrochloride and specific Ca^{2+} binding on the folding kinetics of α-lactalbumin. *J. Mol. Biol.* **206**, 547–561.
35. Griko, Y. V., Freire, E., Privalov, G., van Dael, H. & Privalov, P. L. (1995). The unfolding thermodynamics of c-type lysozymes: a calorimetric study of the heat denaturation of equine lysozyme. *J. Mol. Biol.* **252**, 447–459.
36. Bai, P. & Peng, Z. (2001). Cooperative folding of the isolated α-helical domain of hen egg-white lysozyme. *J. Mol. Biol.* **314**, 321–329.
37. Koshiba, T., Yao, M., Kobashigawa, Y., Demura, M., Nakagawa, A., Tanaka, I. *et al.* (2000). Structure and thermodynamics of the extraordinarily stable molten globule state of canine milk lysozyme. *Biochemistry*, **39**, 3248–3257.
38. Ting, K. L. & Jernigan, R. L. (2002). Identifying a folding nucleus for the lysozyme/α-lactalbumin family from sequence conservation clusters. *J. Mol. Evol.* **54**, 425–436.
39. Polverino de Laureto, P., Scaramella, E., Frigo, M., Gefter-Wondrich, F., De Filippis, V., Zambonin, M. & Fontana, A. (1999). Limited proteolysis of bovine α-lactalbumin: Isolation and characterization of protein domains. *Protein Sci.* **8**, 2290–2303.
40. Demarest, S. J., Boice, J. A., Fairman, R. & Raleigh, D. P. (1999). Defining the core structure of the α-lactalbumin molten globule state. *J. Mol. Biol.* **294**, 213–221.
41. Wijesinha-Bettoni, R., Dobson, C. M. & Redfield, C. (2001). Comparison of the denaturant-induced unfolding of the bovine and human α-lactalbumin molten globules. *J. Mol. Biol.* **312**, 261–273.
42. Raschke, T. M., Kho, J. & Marqusee, S. (1999). Confirmation of the hierarchical folding of RNase H: a protein engineering study. *Nature Struct. Biol.* **6**, 825–831.
43. Tang, K. S., Fersht, A. R. & Itzhaki, L. S. (2003). Sequential unfolding of ankyrin repeats in tumor suppressor p16. *Structure*, **11**, 67–73.
44. Lopez-Hernandez, E. & Serrano, L. (1996). Structure of the transition state for folding of the 129 aa protein CheY resembles that of a smaller protein, CI-2. *Fold. Des.* **1**, 43–55.
45. Lorch, M., Maso, J. M., Clarke, A. R. & Parker, M. J. (1999). Effects of core mutations on the folding of a β-sheet protein: implications for backbone organization in the I-state. *Biochemistry*, **38**, 1377–1385.

46. Cota, E., Steward, A., Fowler, S. B. & Clarke, J. (2001). The folding nucleus of a fibronectin type III domain is composed of core residues of the immunoglobulin-like fold. *J. Mol. Biol.* **305**, 1185–1194.
47. Capaldi, A. P., Kleanthous, C. & Radford, S. E. (2002). Im7 folding mechanism: misfolding on a path to the native state. *Nature Struct. Biol.* **9**, 209–216.
48. Daggett, V. (2002). Molecular dynamics simulations of the protein unfolding/folding reaction. *Accts Chem. Res.* **35**, 422–429.
49. Kuwajima, K., Hiraoka, Y., Ikeguchi, M. & Sugai, S. (1985). Comparison of the transient folding intermediates in lysozyme and α-lactalbumin. *Biochemistry*, **24**, 874–881.
50. 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.
51. Kuwajima, K. (1996). Stopped-flow circular dichroism. In *Circular Dichroism and the Conformational Analysis of Biomolecules* (Fasman, G. D., ed.), pp. 159–182, Plenum, New York.
52. Arai, M., Ikura, T., Semisotnov, G. V., Kihara, H., Amemiya, Y. & Kuwajima, K. (1998). Kinetic refolding of β-lactoglobulin. Studies by synchrotron X-ray scattering, and circular dichroism, absorption and fluorescence spectroscopy. *J. Mol. Biol.* **275**, 149–162.
53. Chaudhuri, T. K., Arai, M., Terada, T. P., Ikura, T. & Kuwajima, K. (2000). Equilibrium and kinetic studies on folding of the authentic and recombinant forms of human α-lactalbumin by circular dichroism spectroscopy. *Biochemistry*, **39**, 15643–15651.
54. Chaudhuri, T. K., Horii, K., Yoda, T., Arai, M., Nagata, S., Terada, T. P. *et al.* (2004). Erratum to: effect of the extra N-terminal methionine residue on the stability and folding of recombinant α-lactalbumin expressed in *Escherichia coli*. (*J. Mol. Biol.* (1999) **285**, 1179–1194). *J. Mol. Biol.* **336**, 825.

Edited by F. Schmid

(Received 22 March 2004; received in revised form 3 June 2004; accepted 4 June 2004)