Structure and Thermodynamics of the Extraordinarily Stable Molten Globule State of Canine Milk Lysozyme[†,‡]

Takumi Koshiba,[§] Min Yao,[§] Yoshihiro Kobashigawa,[§] Makoto Demura,[§] Atsushi Nakagawa,[§,||,⊥] Isao Tanaka,[§,⊥] Kunihiro Kuwajima,[#] and Katsutoshi Nitta*[,§]

Division of Biological Sciences, Graduate School of Science, Hokkaido University, Kita-ku, Sapporo 060-0810, and Department of Physics, School of Science, The University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan

Received July 2, 1999; Revised Manuscript Received November 30, 1999

ABSTRACT: Here, we show that an unfolded intermediate of canine milk lysozyme is extraordinarily stable compared with that of the other members of the lysozyme-α-lactalbumin superfamily, which has been studied previously. The stability of the intermediate of this protein was investigated using calorimetry, CD spectroscopy, and NMR spectroscopy, and the results were interpreted in terms of the structure revealed by X-ray crystallography at a resolution of 1.85 Å to an *R*-factor of 17.8%. On the basis of the results of the thermal unfolding, this protein unfolds in two clear cooperative stages, and the melting temperature from the intermediate to the unfolded states is about 20 °C higher than that of equine milk lysozyme. Furthermore, the ^{1}H NMR spectra of canine milk lysozyme at 60 °C, essentially 100% of which exists in the intermediate, showed that small resonance peaks that arise from ring-current shifts of aliphatic protons are still present in the upfield region from 0 to −1 ppm. The protein at this temperature (60 °C) and pH 4.5 has been found to bind 1-anilino-naphthalene-8-sulfonate (ANS) with enhancement of the fluorescence intensity compared with that of native and thermally unfolded states. We interpret that the extraordinarily stable intermediate is a molten globule state, and the extraordinary stabilization of the molten globule state comes from stronger protection around the C- and D-helix of the aromatic cluster region due to the His-21 residue. The conclusion helps to explain how the molten globule state acquires its structure and stability.

To understand the protein folding mechanism, the concept of a folding intermediate, such as a molten globule state, is extremely important. The molten globule state has been thought to be an equilibrium and kinetic intermediate between the native and unfolded states of a number of globular proteins (*1*−*6*). For some time, the molten globule state of a protein has been assumed to be a compact intermediate that possesses a stable secondary structure without native tertiary contacts, and α-lactalbumin, cytochrome *c*, apomyoglobin, and several other proteins have often been utilized as a useful model protein for studying the folding mechanisms (*4*−*11*). However, with the extensive progress in studying the molten globule state, new evidence has been accumulated that shows specific tertiary interactions or nativelike topology within the molten globule state of such proteins as α-lactalbumin (*12*−*19*), apomyoglobin (*20*−*23*), and equine milk lysozyme (EML)[1] (*24*−*28*).

Canine milk lysozyme (CML) belongs to the family of calcium-binding lysozymes (*29*), and its physicochemical properties are known to be rather different from those of conventional, non-calcium-binding lysozymes. In particular, a notable difference has been observed in their unfolding profiles (*30*). A stable equilibrium intermediate has been shown to accumulate during the guanidine hydrochloride-induced unfolding of CML, although conventional lysozymes usually unfold in a cooperative two-state manner without the accumulation of an intermediate. Thus, the unfolding behaviors of this protein and EML are rather similar to those of α-lactalbumin, which has been shown to exhibit a well characterized molten globule state in its equilibrium unfolding (*4*, *7*, *9*).

Recently, the expression system for CML has been established in *Escherichia coli* (*E. coli*), and we have found that the intermediate of this protein shows extraordinary stabilization (*31*). In this paper, we show that CML is an excellent model for elucidating the molecular mechanisms by which the intermediate is stabilized. To investigate the stability of the intermediate of CML, its unfolding behavior was studied by differential scanning calorimetry (DSC),

[†] This study was supported by Grants-in-aid from the Ministry of Agriculture, Forestry and Fisheries of Japan and also supported in part by Tukuba Advanced Research Alliance (TARA) Sakabe project.

[‡] The atomic coordinates and observed structure factors of canine milk lysozyme (apo-type) have been submitted to the Protein Data Bank, the Research Collaboratory for Structural Bioinformatics (RCSB) (entry code 1QQY)

[§] Graduate School of Science, Hokkaido University.

[||] Institute for Protein Research, Osaka University.

[⊥] Members of the TARA project of Tsukuba University, Japan.

[#] Graduate School of Science, The University of Tokyo.

* To whom correspondence should be addressed. Telephone: +81-11-706-2773. Fax: +81-11-706-4992. E-mail: nitta@sci.hokudai.ac.jp.

[1] Abbreviations: ANS, 1-anilino-naphthalene-8-sulphonate; CML, canine milk lysozyme; CD, circular dichroism; ΔC_p, heat capacity change; ΔH, enthalpy change; DSC, differential scanning calorimetry; EDTA, ethylenediaminetetraacetic acid; EML, equine milk lysozyme; ΔG, Gibbs free-energy change; GPLA, guinea-pig α-lactalbumin; MG, molten globule; NMR, nuclear magnetic resonance; PEG, poly(ethylene glycol); ΔS, entropy change; T_m, melting temperature; UV, ultraviolet.

10.1021/bi991525a CCC: $19.00
Published on Web 03/01/2000

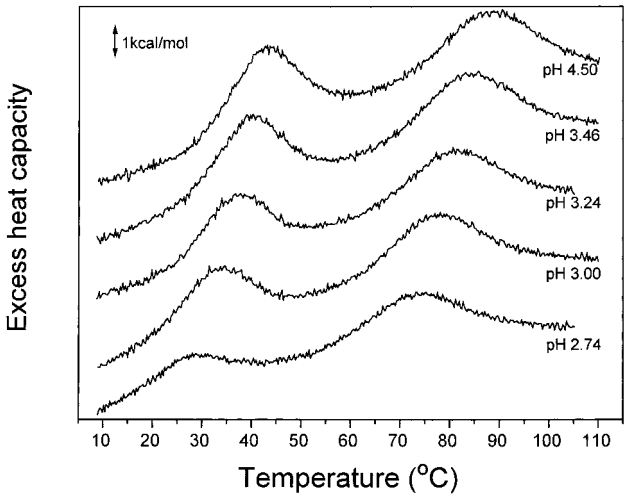

FIGURE 1: The thermal unfolding transition of CML in the absence of calcium ions at various pHs (2.74, 3.00, 3.24, 3.48, and 4.50). The unfolding transition is expressed by excess heat capacity as a function of temperature.

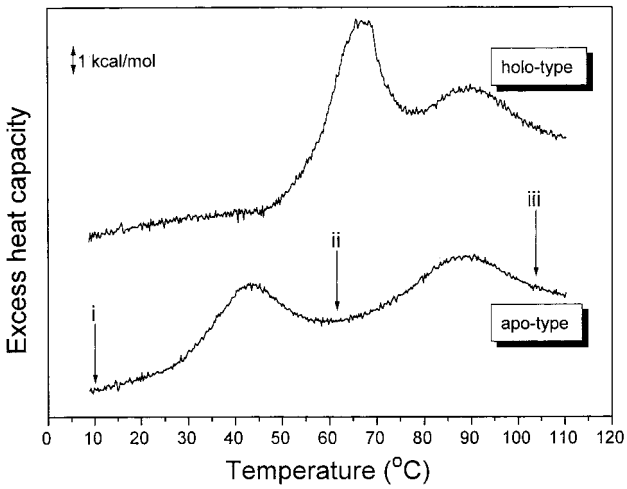

FIGURE 2: The thermal unfolding transition of CML in the absence (apo-type) and presence of 10 mM $CaCl_2$ (holo-type) at pH 4.5. The unfolding transition is expressed by excess heat capacity as a function of temperature.

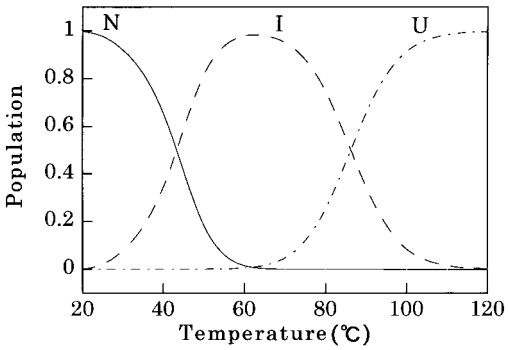

FIGURE 3: Temperature dependence of the populations of the three states, the native (N), folding intermediate (I), and unfolded (U) states in the absence (apo-type) of $CaCl_2$ at pH 4.5.

UV−CD spectroscopy, and NMR spectroscopy, and the results were interpreted in terms of the structure revealed by X-ray crystallography. Moreover, the 1-anilino-naphthalene-8-sulfonate (ANS)-binding experiment was also performed for the purpose of detecting the formation of the loose hydrophobic core and investigating the hydrophobic surface accessible to solvent of the intermediate of this protein. The X-ray crystal structure of this protein was determined by a molecular replacement method and refined at a resolution of 1.85 Å to an *R*-factor of 17.8% (free *R*-factor = 21.5%).

From the result of ANS-binding experiment, it appears that the intermediate has been found to bind ANS with enhancement of the fluorescent intensity compared with that of the native and the thermal unfolded states. We interpret that the extraordinarily stable intermediate of this protein is a molten globule state, and the extraordinarily stabilization come from its specific packing interactions in the α-domain. In support of this interpretation, the ^{1}H NMR spectra show that the ring-current effect is certainly present in the intermediate of CML. The relevance of the present results to the physicochemical properties of CML and the structural origins of the stability of the molten globule state will be discussed.

MATERIALS AND METHODS

Materials. According to the procedures of Koshiba et al. (1999), expression and purification of CML were carried out using the expression system in *E. coli* (*31*). EML has been prepared previously (*32*). All other reagents were of biochemical research grade.

Preparation of Apo- and Holoproteins. Apo- and holoproteins were prepared as described previously (*31*, *33*).

Differential Scanning Calorimetry (DSC). Differential scanning calorimetry (DSC) measurements of CML were carried out with an MC-2 microcalorimeter (MicroCal, Inc., Northampton, MA) at a scan rate of 1.0 K/min. Sample solutions for DSC measurements between pH 4.0 and 4.5 were prepared by dissolving the lysozymes in 50 mM sodium acetate buffer, and solutions between pH 2.5 and 4.0 were prepared using 50 mM glycine hydrochloride. The lysozyme concentrations used were 1.2−1.8 mg/mL. The pH of the sample solution was confirmed before and after each measurement. Calorimetric enthalpies (ΔH_{cal}) were obtained using Origin computer software (Microcal, Inc.).

Measurements of CD Spectra. CD spectra of CML and EML were measured with a Jasco J-500 spectropolarimeter (Tokyo, Japan). The protein concentrations were 0.5 mg/mL for all measurements. Far- and near-UV−CD spectra were measured using a cell with 0.1 and 1 cm optical path lengths, respectively, at pH 4.5 and three different temperatures (9, 63, and 98 °C).

Measurements of ANS-binding. Fluorescence spectra of 50 μM ANS in the absence and presence of 25 μM apo-CML were collected using a Hitachi spectrofluorimeter 650−40 (Tokyo, Japan), equipped with a thermostatically controlled cell-holder to keep the temperature. ANS fluorescence was measured using an excitation wavelength of 380 nm. The emission spectrum was recorded over the range of 400−640 nm. An ANS stock solution was prepared in methanol, and 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 (*49*). Fluorescence measurements were carried out using a cell with a 1 cm optical path length at pH 4.5 and three different temperatures (9, 63, and 98 °C).

One-dimensional ^{1}H NMR Spectra. One-dimensional ^{1}H NMR spectra were obtained on a JEOL α-400 spectrometer

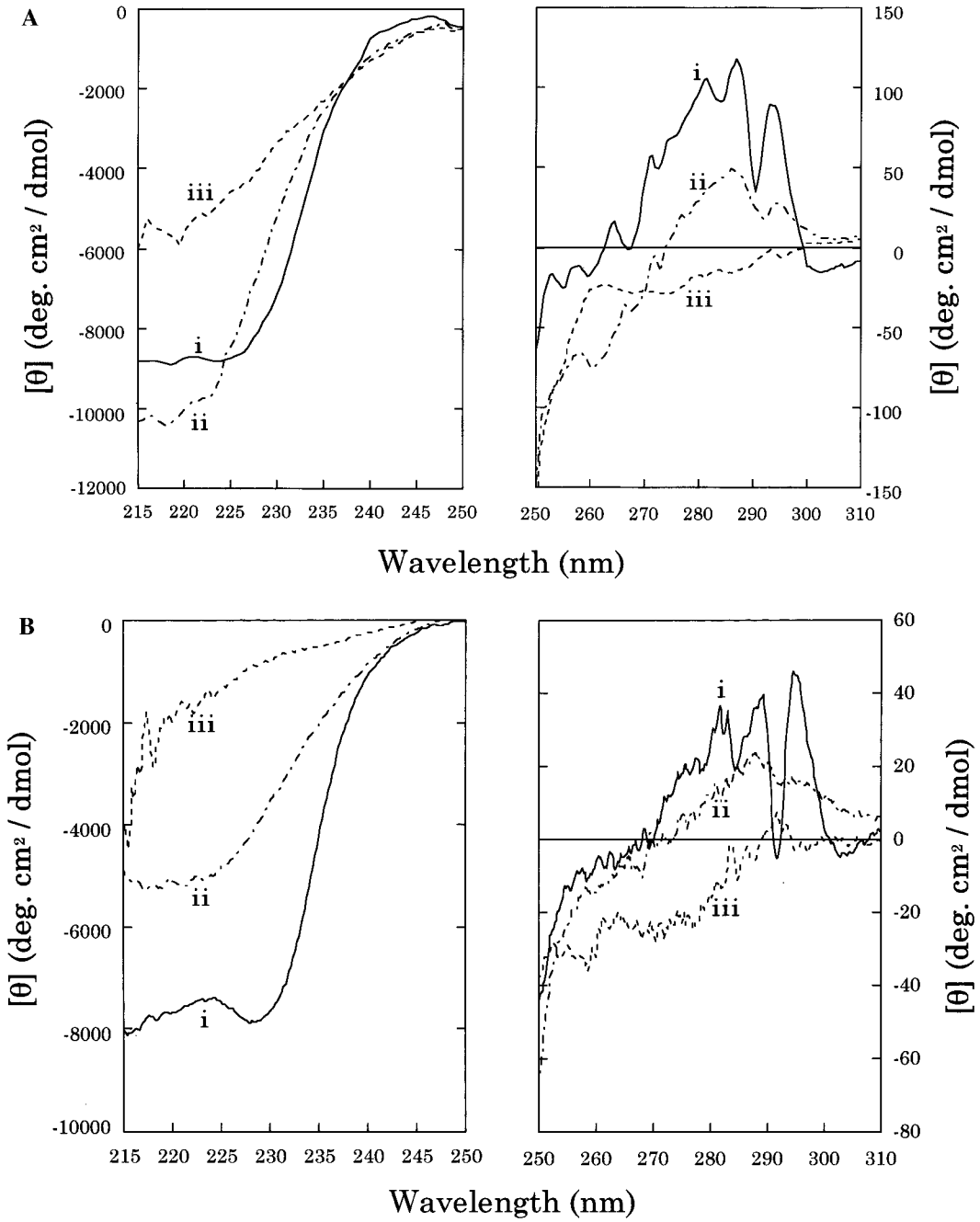

FIGURE 4: CD spectra of (A) CML and (B) EML in the far-UV (left) and near-UV (right) regions at pH 4.5. Three spectra (i, ii, and iii) were measured at the different temperatures indicated by arrows in Figure 2.

operated at 400 MHz and several temperatures (30, 40, 50, 60, 65, and 70 °C). The sample solution for the ^{1}H NMR measurement was prepared by dissolving lyophilized lysozyme (of apo-type) in D_2O, and the pH was adjusted with small quantities of NaOD or DCl solution to pH 4.5. In all experiments, the sweep width was 15.0 ppm, and the digital resolution was 0.36 Hz/point. The water signal was suppressed by a DANTE pulse-sequence (*34*).

Crystallization, Data Collection, and Structure Refinement. A diffraction-quality crystal of apo-CML was obtained by vapor diffusion of hanging drops using 0.1 M trisodium citrate buffer (pH 5.6) containing 10 mg/mL protein, 0.2 M ammonium acetate, 30% (w/v) PEG-4000, and 1 mM EDTA. After 7−10 days, a crystal with the size of 0.8 × 0.2 × 0.2 mm^3 was obtained. The space group of the crystal was $P6_1$-22 with cell dimensions $a = b = 68.6$ Å, $c = 107.1$ Å. There is one monomeric molecule in an asymmetric unit. The X-ray diffraction data of the apo-CML crystals were collected from a single crystal at room temperature using a Fuji imaging plate detector with a Rigaku IPR-4080 imaging scanner and a Weissenberg camera for macromolecular crystallography (*35*) on BL-18B at the Photon Factory, KEK, Japan. Integrated intensities were obtained with DENZO (*36*) and were reduced using a CCP4 package (*37*). The initial model of the apo-type CML was obtained by molecular replacement using the program AMoRe (*38*). A search model was constructed from the atomic coordinates of EML (*39*). The structural refinements were performed using the program X-PLOR (*40*).

Estimation of Protein Concentration. The protein concentrations were estimated spectrophotometrically using the

following extinction coefficients at 280 nm: $E^{1\%}_{1cm} = 23.2$ for CML (*30*) and $E^{1\%}_{1cm} = 23.5$ for EML (*41*).

RESULTS

Differential Scanning Calorimetry (DSC) of CML. To determine the thermodynamic parameters of thermal unfolding, DSC measurements of the CML were carried out at acidic pHs between 2.7 and 4.5 in the absence of calcium ions. Figure 1 shows that the temperature dependence of the partial heat capacity of CML in acidic solutions without calcium ions. Under these conditions, CML is in apo-type. In this pH region, the unfolding of CML is reversible, and the protein exhibits a cooperative thermal unfolding with two remarkably positive heat absorption peaks, as previously reported in EML (*25*, *26*). It is indicating that the first peak represents the unfolding transition from the native state to an unfolded intermediate, and the second peak represents the unfolding transition from the intermediate to the thermally unfolded state. The melting temperatures (T_m), at which both heat absorption peaks reach their maximum, decreases at lower pH values in almost the same manner.

The temperature dependence of the partial heat capacity of CML in 50 mM sodium acetate buffer in the absence (apo-type) and presence (holo-type) of 10 mM $CaCl_2$ at pH 4.5 is shown in Figure 2. In this pH, the thermal stability and enthalpy of unfolding of CML are insensitive to variations in ionic strength (sodium acetate buffer; from 10 to 100 mM) (data not shown). The temperature dependence of the fractional population of apo-CML at pH 4.5, which was estimated from Figure 2, is shown in Figure 3. From this figure, we can easily see that the intermediate is extraordinarily stable, and essentially 100% of the protein is in this state at 60 °C, compared with 70% of EML (*26*). Under the presence of 10 mM $CaCl_2$ conditions, α-lactalbumin and EML are known to show a single heat absorption peak caused by a cooperative transition from the native to the thermally unfolded states (*25*, *42*−*45*). However, unlike α-lactalbumin and EML, CML shows the second heat absorption peak centered at 85 °C, indicating that there is a stable intermediate even in the thermal unfolding in the presence of 10 mM $CaCl_2$. Thus, the comparison of the DSC curves in the apo- and holo-types of CML also indicates that the extraordinary stabilization of the intermediate of this protein has led to the appearance of the second heat absorption peak in the holoprotein observed above (Figure 2).

Measurements of CD Spectra. The far- (215−250 nm) and near-UV (250−310 nm) CD spectra of apo-CML and apo-EML at pH 4.5 and the temperature indicated by arrows (i−iii) in Figure 2 are depicted in Figure 4a and b. The far-UV−CD spectrum of apo-CML at 63 °C still shows the presence of the secondary structure, much as in the native state (i); moreover, the near-UV−CD spectrum at this temperature shows the presence of elements of a tertiary-like structure (Figure 4a). In Figure 4b, however, for EML, we can see that CD intensity around 220 nm at 63 °C is half as much as that of native state (i). Thus, the intermediate of EML has already started to disorder at this temperature. Therefore, it appears that the intermediate of CML still possesses a partially nativelike structure and is highly populated at elevated temperatures (70−100 °C) both in the absence and the presence of Ca^{2+} (Figures 2, 3, and 4a).

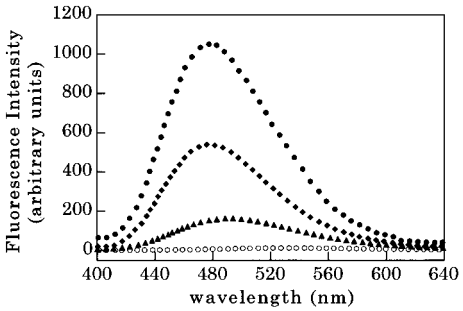

FIGURE 5: Fluorescence spectra of 50 μM ANS in the absence (open symbols) and presence (filled symbols) of 25 μM apo-CML at pH 4.5. The spectra of the complex of ANS with apo-CML were obtained at 9 (◆), 63 (●), and 98 °C (▲). The spectrum of ANS without lysozyme were obtained at 25 °C (○). These three temperatures are indicated in Figure 2 with arrows.

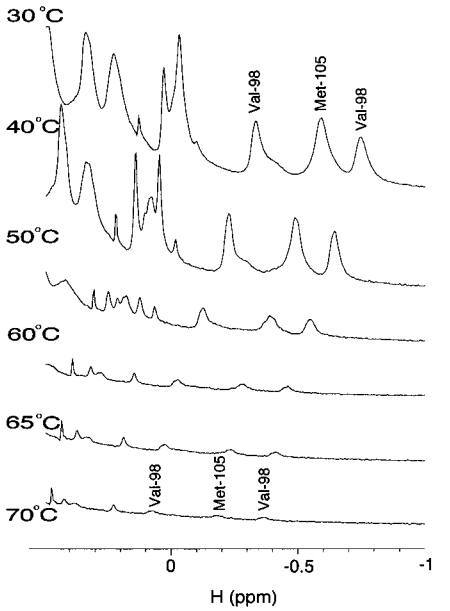

FIGURE 6: One-dimensional ^{1}H NMR spectra (400 MHz) of apo-CML as a function of temperature at pH 4.5.

ANS Binding. The fluorescent hydrophobic molecule, 1-anilino-naphthalene-8-sulfonate (ANS), is commonly used as a probe of the hydrophobic surface exposed to solvent (*47*, *48*). The fluorescence spectra of apo-CML at pH 4.5 and the temperatures indicated by arrows (i−iii) in Figure 2 are depicted in Figure 5. The fluorescence intensity of 50 μM ANS in the presence of 25 μM apo-CML at 63 °C is greater than that of the temperatures at 9 and 98 °C. This is the characteristics of the molten globule state, which has the hydrophobic core exposed to solvent (*19*, *28*, *48*, *49*). The flexible nature of the molten globule state permits some internal nonpolar groups to become exposed to solvent, thus making the surface of this state more hydrophobic than that of the native or unfolded states (*48*, *50*). Therefore, the result of ANS-binding measurement indicates that the intermediate of this protein has a characteristic of the molten globule state.

One-dimensional ^{1}H NMR Spectra. Figure 6 shows the thermal unfolding transitions of apo-CML, which were monitored with ^{1}H NMR. Heating of the apo-CML induced a major change in the NMR spectra between 40 and 50 °C; this temperature was comparable to the temperature (30−55 °C) at which DSC and CD spectra show their first thermal transition (Figures 2, 3, and 4a). The highly dispersed

Table 1: Thermodynamic Parameters for the Unfolding of CML and EML in the Absence of Ca^{2+} at pH 4.5

lysozyme	type of transition[a]	T_m (°C)	ΔH^b (kcal/mol)	$T\Delta S^b$ (kcal/mol)	ΔG^b (kcal/mol)	ΔC_p (kcal/mol/K)
			at 41.5 °C			
EML[c]	N−I	41.5	36.6	36.6	0.0	1.20
CML	N−I	39.0	44.6	44.9	−0.3	1.03[d]
			at 66.4 °C			
EML[c]	I−U	66.4	29.5	29.5	0.0	0.61
CML	I−U	86.9	30.9	28.8	2.1	0.70[d]

[a] N, I, and U mean native, intermediate, and unfolded states, respectively. [b] These values are extrapolated to 41.5 or 66.4 °C. [c] From Griko et al. (*26*). [d] From Koshiba et al. (*46*).

Table 2: Crystallographic Data, X-ray Processing Statistics, and Refinement Statistics of the apo-type CML

crystal data collection	
space group	$P\,6_122$
cell constants (Å)	
a	68.58
b	68.58
c	107.06
resolution (Å)	100−1.85
no. of measured reflections	199512
no. of independent reflections	13034
R merge (%)[a]	6.00
refinement	
resolution used (Å)	8.0−1.85
no. of used reflections ($F > 2\sigma$)	12679
data completeness (%)	95.5
Δ bond length (Å)	0.007
Δ bond andle (°)	1.414
average *B*-factor (Å^2)	
main-chain	23.8
side-chain	28.3
R-factor (%)[b]	17.8
free *R*-factor (%)[c]	21.5

[a] $R_{merge} = \sum|I - \{I\}|/\sum|\{I\}| \times 100\%$. [b] *R*-factor $= \sum|F_o| - |F_c|/\sum|F_o| \times 100\%$. [c] Free *R*-factor is the same as *R*-factor, except for a 10% subset of all reflections.

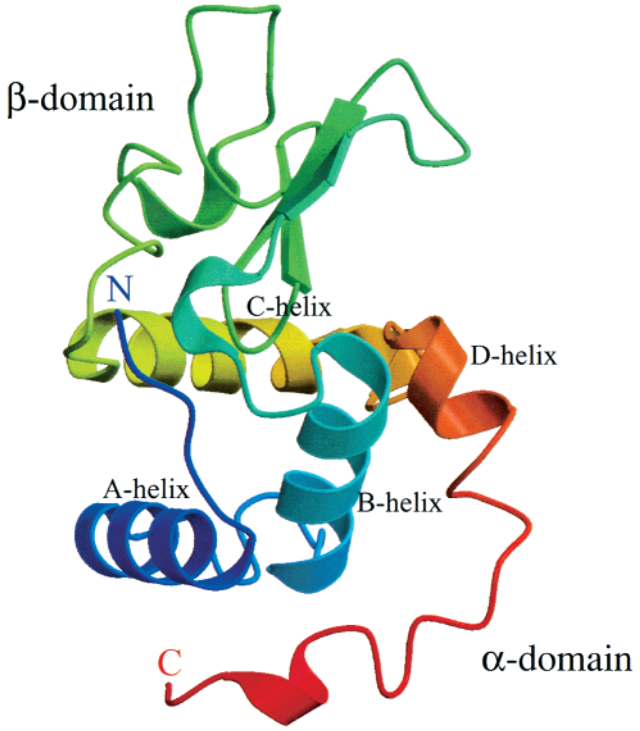

FIGURE 7: Ribbon model of CML in which α-helices are sequentially labeled from A to D. The structure shows a rainbow color (red to blue) from the N-terminus to the C-terminus (figure produced using MOLSCRIPT (*74*)).

resonances in the chemical shifts from 0 to −1 ppm corresponding to the native conformation show a decrease in intensity concurrent with the downfield chemical shift (*31*). Similar decreasing in intensity and chemical shift changes as a function of temperature have been observed in α-lactalbumin and EML. However, unlike the ^{1}H NMR spectra of α-lactalbumin and EML, which show no resonances from 0 to −1 ppm above 60 °C (*25*, *28*, *51*−*53*), the spectra of apo-CML show small resonance peaks still present in the upfield region from 0 to −1 ppm above 60 °C (Figure 6). Thus, the structure of the intermediate of apo-CML is more order than that of the other members of the lysozyme-α-lactalbumin superfamily.

X-ray Crystal Structure of apo-CML. Apo-CML was crystallized in space group of $P6_122$, and the three-dimensional structure was determined by a molecular replacement method. The 1.85 Å structure was refined to an *R*-factor of 17.8% ($R_{free} = 21.5\%$). The crystallographic parameters and refinement statistics are summarized in Table 2.

Figure 7 shows the ribbon diagram of apo-CML. A deep cleft containing the active site divides the CML molecule into a large domain (α-domain) and a small domain (*β*-domain), both of which are similar to those previously reported for the lysozyme-α-lactalbumin superfamily (*39*, *54*−*59*). The α-domain comprises residues −1−42 and 86−129, and it is composed of four major α helices (A-helix, 5−16; B-helix, 25−36; C-helix, 88−100; D-helix, 109−113) and two short helices (104−107; 123−126). The *β*-domain is composed of a series of loops and a small three-stranded antiparallel *β*-pleated sheet (S1: 43−45, S2: 51−55, and S3: 58−61).

The overall structure of the determined apo-CML is shown in Figure 8, together with those of apo-EML (*39*) and guinea-pig α-lactalbumin (GPLA) (*58*). The three proteins compared here are very similar to each other, reflecting their high similarity of their amino acid sequences (Figure 9). The root-mean-square deviations between CML and EML and between CML and GPLA for the main-chain Cα atoms were 1.294 and 1.950 Å, respectively. Differences in main-chain conformation are observed in several distinct regions of these structures, the largest of which being in the regions between residues 43−55. There are also considerable differences in the regions between residues 64−75. One of the features of the crystal structure of apo-CML is that the loop region of the *β*-domain (residues 43−55) is much closer to the α-domain than in EML and GPLA. Therefore, according to the structure of native CML, it is likely that the *β*-domain

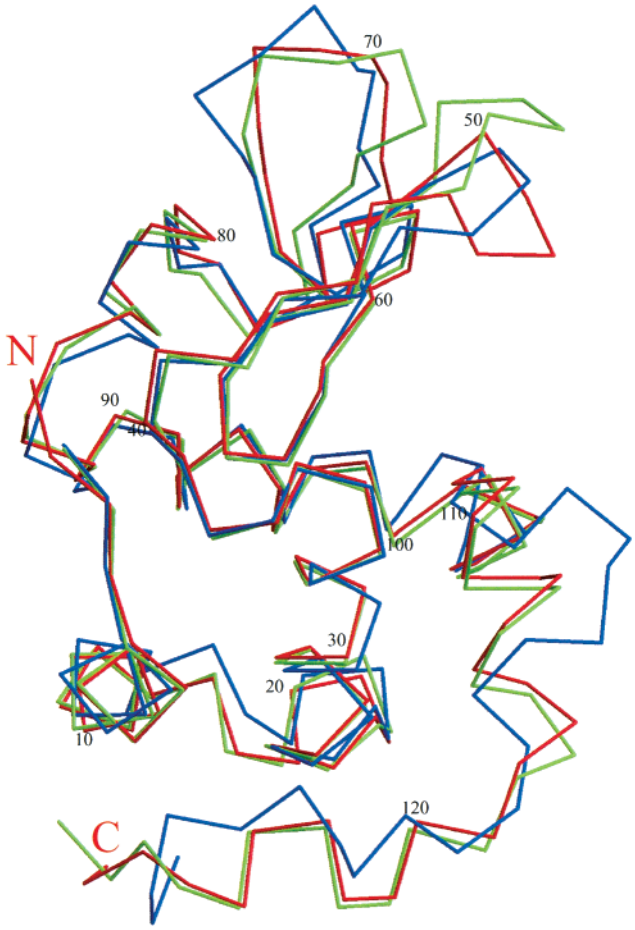

FIGURE 8: Superimposition of the structures of CML, EML, and GPLA. A structure of the Cα traces of CML (red), EML (green), and GPLA (blue) are shown after the least-squares superposition. Every tenth residue of CML is labeled (figure produced using MOLSCRIPT (*74*)).

structure of CML may be more stable than those of EML and GPLA.

DISCUSSION

The purpose of the present study is to elucidate the relationship between the structure and the thermodynamics of the intermediate of CML and compare the intermediate with those of the other members of the lysozyme-α-lactalbumin superfamily, which have been studied previously. To fully understand the effect on the conformational stability of a protein, it is necessary to elucidate the thermodynamic parameters of the unfolding. The thermodynamic parameters for the unfolding [$\Delta H(T)$, $T\Delta S(T)$, and $\Delta G(T)$] of the CML and EML at 41.5 °C (transition from the native state to the intermediate; N−I transition) and 66.4 °C (transition from the intermediate to the thermally unfolded state; I−U transition) were obtained according to the following equations using the data from the DSC measurements in this study and Griko et al. (*26*):

$$\Delta H(T) = \Delta H(T_m) + \int_{T_m}^{T} \Delta C_p(T)\mathrm{d}T \quad (1)$$

$$\Delta S(T) = \Delta S(T_m) + \int_{T_m}^{T} \Delta C_p(T)\mathrm{d}\ln T \quad (2)$$

$$\Delta G(T) = \Delta H(T) - T\Delta S(T) \quad (3)$$

where T_m is the melting temperature and $\Delta S(T_m)$ equals to $\Delta H(T_m)/T_m$. The $\Delta C_p(T)$ values of CML and EML at these temperatures were obtained using the data of Koshiba et al. (*46*) and Griko et al. (*26*), respectively. The parameters obtained are listed in Table 1. The ΔG value of the I−U transition for CML is 2.1 kcal/mol more stable than that of EML at 66.4 °C, although the ΔG values of the N−I transition between two proteins are quite similar (−0.3 kcal/mol) to each other at 41.5 °C. It is indicating that the thermal stability of the intermediates between CML and EML is very different despite the similarity of the N−I transition. In the I−U transition at 66.4 °C, the ΔH value for the unfolding of the intermediate of CML (30.9 kcal/mol) is larger than that of EML (29.5 kcal/mol). In addition, the entropy term ($T\Delta S$) in the unfolding of the intermediate of CML (28.8 kcal/mol) is smaller than that of EML (29.5 kcal/mol). These evaluations indicate that both the enthalpy and entropy terms seem to be the major terms for stabilization of the intermediate of CML.

From the results of this study, the intermediate of CML has been shown its extraordinary stabilization, and we propose that the intermediate is a molten globule state. There are at least three lines of strong evidence for the identification of the intermediate of CML as the molten globule state. First, the intermediate of CML is stably populated at pH 2.0, the ^{1}H NMR spectrum is more similar to the spectrum in the unfolded state, and the characteristic chemical shift dispersion has mostly disappeared (*31*). The spectral width is broader than those in the native and the fully unfolded states, and these are characteristics of the molten globule state (*25*, *53*, *80*). Second, CML shows a three-state unfolding transition in guanidine hydrochloride (*30*) with the unfolded intermediate state very similar to the intermediate of the thermal unfolding (this study), and these intermediates are structurally very similar to the acid state (*31*). Such unfolding behavior and the properties of the intermediate or the acid states are essentially identical to those observed previously in α-lactalbumin and EML (*4*, *7*, *9*, *26*, *27*, *60*, *61*), for which the unfolding intermediate state has been well-characterized and identified as the molten globule state. Finally, the results of the ANS-binding experiment presented in this study show that the hydrophobic fluorescence probe (ANS) has a different affinity to various conformations of CML (Figure 5). This probe is not practically bound to fully unfolded proteins as well as to a number of native protein (*48*). Nevertheless, the results of this study show a very good correlation of the highest protein affinity to ANS with the conformation of the intermediate. This is a characteristic of the molten globule state, which has the hydrophobic core exposed to solvent (*19*, *28*, *48*, *49*).

Recently, with the extensive progress in the study of the molten globule state, new evidence has been accumulated that shows the presence of nativelike tertiary fold within the molten globule state of α-lactalbumin (*12*−*19*). Peng and Kim (1994) have shown that the helical domain (α-domain) of human α-lactalbumin, in isolation, forms a molten globule state with the same overall tertiary fold as that found in intact α-lactalbumin (*12*). The same group has also reported that the α-helical domain forms a helical structure with a nativelike tertiary fold, while the β-sheet domain is largely unstructured (*14*). The studies indicate that the molten globule state of α-lactalbumin has a nativelike overall fold,

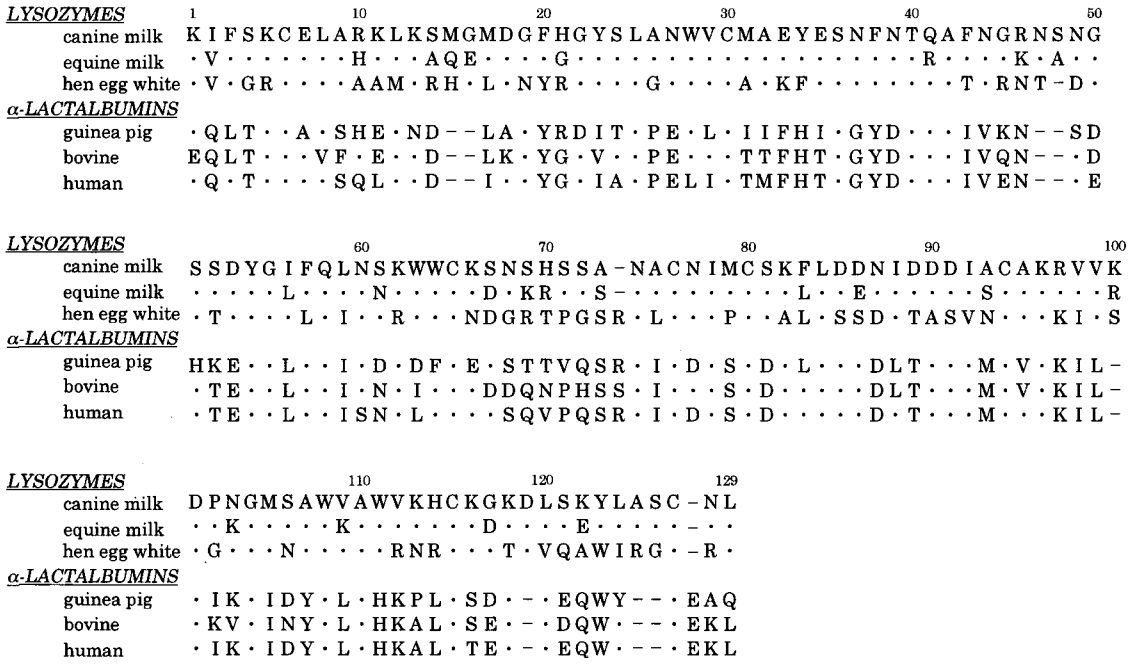

FIGURE 9: Amino acid sequences of selected lysozymes and α-lactalbumins. The sequence of CML (*29*) appears in the single letter code. The residues present in the rows of other lysozymes (*75*,*76*) and α-lactalbumins (*77*−*79*) are shown only where are different from the CML sequence, with identity being indicated by a dot. Dashes indicate deletions relative to the other sequences.

even though it lacks the extensive fixed tertiary interactions that are characteristic of native proteins (*12*, *62*).

Furthermore, Dobson and co-workers have revealed that EML, unlike other lysozymes, can be transformed into the equilibrium molten globule state (*24*, *25*, *27*, *28*) and that this state is more similar to the native state than the equilibrium state of apo-type α-lactalbumin, which is usually regarded as a "classical" molten globule state (*62*, *63*). Moreover, it has been reported that the molten globule state of EML is stabilized, at least in part, by nativelike specific interactions and is more stable than that of α-lactalbumin (*60*). The α-helices A, B, and D of EML are especially strongly protected in the molten globule state (*27*, *28*).

Thus, the intermediate of CML can be identify as the molten globule state, and it is thought to be similar to the molten globule state of EML, which has been distinguished from a "classical" molten globule state. However, the ^{1}H NMR spectrum of CML at 60 °C (Figure 6), the temperature at which 100% of the protein exists in the stable molten globule state (Figure 3), has shown interesting characteristics as compared with that of the other lysozyme-α-lactalbumin superfamily. Small resonance peaks are still present in the upfield region from 0 to −1 ppm. These peaks arise from ring-current shifts of methyl protons and are due to specific packing interactions of the aliphatic side chains around the aromatic rings; moreover, this characteristic has previously been observed in the acidic-folding intermediate (pH 2.0) (*31*). Recently, we have sequentially assigned the chemical shifts of CML using ^{15}N-labeled protein (unpublished results), and these peaks appear to come from the methyl protons of Val-98 and βH of Met-105, as shown in Figure 6. Therefore, our NMR studies suggest that the extraordinary stabilization of the molten globule state of CML is due to the specific packing interactions of C-helix (Val-98) and Met-105 around the aromatic rings.

However, why can CML fold to the extraordinarily stable molten globule state as compared with that of EML? In the following, we address this question and show possible reasons that interpret such difference in the molten globule stability between the proteins.

The Role of His-21 for the Stabilization. The hydrophobic cluster of CML (Figure 10a), which originates from the aromatic residues (Tyr-23, Trp-28, −63, −64, −108, and −111) around Val-98 and Met-105, is compared with that of EML (Figure 10b). In the crystal structure of CML, His-21 is located at the turn region between the A- and B-helices and is close to the C-helix; however, this position in EML is replaced by glycine (Figure 9). The distance between the $N^{\epsilon 2}$ atom of His-21 and O atom (C═O) of Val-99 is 3.13 Å, and a hydrogen bond is formed between these two atoms, as shown in Figure 11. As a result, the A- and B-helix become much closer to the C-helix, and hydrophobicity around this region is increased. Thus, the substitution of Gly-21 with His-21 can be suggested to have two effects on the stabilization of the molten globule state. First, by forming a hydrogen bond between the side chain of His-21 and the main chain of Val-99, the α-helices A and B are connected to the C-helix, and these chains become less flexible. Second, the involvement of the His-21 residue on the hydrophobic cluster makes the structure more stable, and the molten globule state of CML is further stabilized. Indeed, substitutions of some residues in the α-domain of α-lactalbumin with more hydrophobic amino acids increase the stability of the molten globule state of α-lactalbumins (*16*, *18*, *61*)

Stability of D-Helix. Residue 109 is located in the N-terminal of the D-helix; these residues are lysine and valine in EML and CML, respectively (Figures 9, 10a, and 10b). It is well-known that a protein is stabilized by an electrostatic interaction between a negatively charged side chain in a helical N-terminal and a positively charged side chain in a helical C-terminal due to the α-helix dipole (*64*−*67*). In the case of EML, these two residues, which are located in the helical N- and C-terminals of the D-helix, are both lysine (Lys-109 and 113). From the results of the X-ray structure

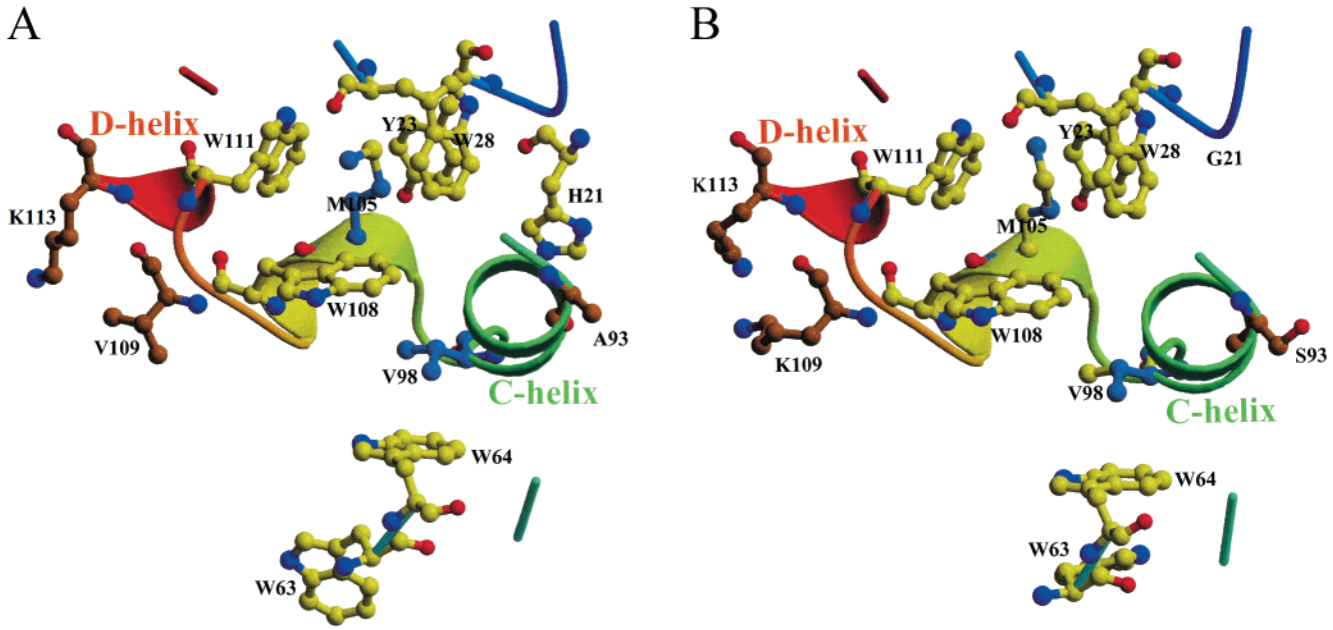

FIGURE 10: Conformation of the aromatic cluster around the residues Val-98 and Met-105 within 10 Å in distances from their Cα atoms. The local structures are shown schematically for (A) CML and (B) EML. Two molecules are viewed in the same orientation (figure produced using MOLSCRIPT (*74*)).

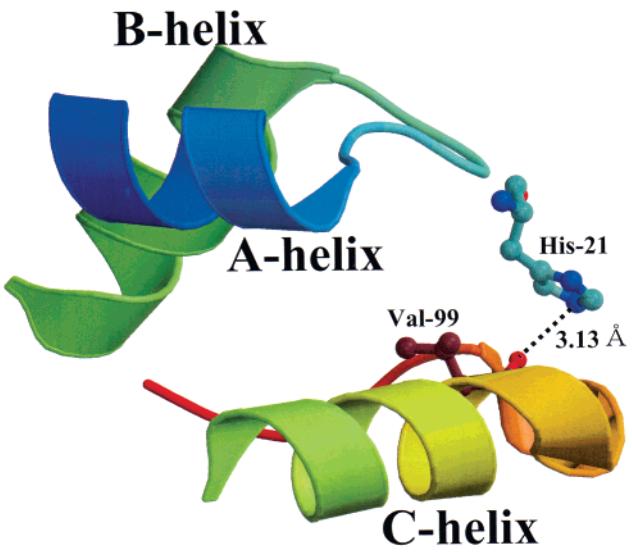

FIGURE 11: Partial structure of CML around the region of residue His-21. The broken line indicates hydrogen bond (figure produced using MOLSCRIPT (*74*)).

of EML, the distance between the N^{ζ} atom of Lys-109 and the N^{ζ} atom of Lys-113 appears to be 3.08 Å. The pK_a value of lysine residue at 25 °C is 10.5 (*68*), so both lysine residues (Lys-109 and 113) are protonated under this experimental condition (acidic pH of 4.5) of DSC (*26*), UV−CD spectra (Figure 4b), and ^{1}H NMR spectra (*25*). Thus, these two positively charged side chains are thought to destabilize the D-helix due to electrostatic repulsion. However, this helical N-terminal residue of CML is valine, so that this α-helix dipole is thought to give more stability to the D-helix compared than that in EML. In addition, the distance between the C^{β} of Val-109 and the N^{ζ} atom of Lys-113 is 5.58 Å, which might relieve the steric compression (Figure 10a).

Difference of Amino Acid Sequences. Moreover, we have previously noted the differences between CML and EML in their amino acid sequences (Figure 9), and these differences should be responsible for the differences in the stability of the molten globule state (*16*, *18*, *61*, *69*−*72*). CML has 23 amino acid substitutions in comparison with EML. Among these substitutions, that at residue 93 seems to be important in relation to difference in the stability of the molten globule state for the following reason. Residue 93 is located in the middle position of the C-helix (Figure 10a and b). Residue 93 is serine in EML, and alanine in CML. Alanine is thought to be more favorable for forming an α-helix than serine when the alanine residue is located in the central position of an α-helix (*73*).

Thus, it seems likely that such substitutions (from Gly-21 to His-21; from Lys-109 to Val-109; from Ser-93 to Ala-93) in the aromatic cluster around the C-helix and D-helix increase the stability of the molten globule state of CML. Therefore, the molten globule state of CML is probably strongly protected by the aromatic cluster around the C-helix and D-helix as compared with EML. This is one of the main reasons why CML has an extraordinarily stable molten globule state as compared with other proteins in the lysozyme-α-lactalbumin superfamily.

In conclusion, the thermal unfolding equilibrium of CML is well-represented by the three-state mechanism in which only the native, the intermediate, and unfolding states are populated. Both CD and NMR data indicate that the intermediate of this protein has secondary structure and elements of tertiary structure. Moreover, the results of ANS-binding experiment show the highest affinity of ANS to the intermediate. From these results, we interpret that the extraordinarily stable intermediate is a molten globule state. The molten globule state of this protein is extraordinarily stable compared with those of the other members of the lysozyme-α-lactalbumin superfamily, and this stabilization comes from the stronger protection around the C- and D-helix of the aromatic cluster region. This molten globule state may be more appropriate to define as nativelike tertiary folded

state. Our structural and thermodynamically information will facilitate the study of protein folding.

ACKNOWLEDGMENT

We gratefully acknowledge Prof. Yue Ke of Normal University of Inner Mongolia, China, for the generous gift of equine milk and are grateful to Dr. M. Arai (The University of Tokyo) for invaluable advice. Prof. N. Sakabe and Dr. N. Watanabe, at the Photon Factory for help with high-resolution data collection, are also gratefully acknowledged. The experiments using synchrotron radiation were performed under the approval of the Photon Factory admission committee, KEK, Japan (PAC No. 98G369).

REFERENCES

1. Aune, K. C., Salahuddin, A., Zarlengo, M. H., and Tanford, C. (1967) *J. Biol. Chem. 242*, 4486-4489.
2. Tanford, C., Aune, K. C., and Ikai, I. (1973) *J. Mol. Biol. 73*, 185−197.
3. Ikai, A., Fish, W. W., and Tanford, C. (1973) *J. Mol. Biol. 73*, 165−184.
4. Kuwajima, K. (1989) *Proteins: Struct. Funct. Genet. 6*, 87−103.
5. Ptitsyn, O. B. (1992) *Protein Folding* (Creighton, T. E., Ed.) pp 243−300, Freeman, New York.
6. Ptitsyn, O. B. (1995) *Adv. Protein Chem. 47*, 83−229.
7. Kuwajima, K., Nitta, K., Yonayama, M., and Sugai, S. (1976) *J. Mol. Biol. 106*, 359−373.
8. Griko, Y. V., Privalov, P. L., Venyaminov, S. Y., and Kutyshenko, V. P. (1988) *J. Mol. Biol. 202*, 127−138.
9. Kuwajima, K. (1996) *FASEB J. 10*, 102−109.
10. Dabora, J. M., Pelton, J. G., and Marqusee, S. (1996) *Biochemistry 35*, 11951−11958.
11. Tsui, V., Gracia, C., Cavagnero, S., Siuzdak, G., Dyson, H. J., and Wright, P.E (1999) *Protein Sci. 8*, 45−49.
12. Peng, Z-yu, and Kim, P. S. (1994) *Biochemistry 33*, 2136−2141.
13. Peng, Z-yu, Wu, L. C., and Kim, P. S. (1995) *Biochemistry 34*, 3248−3252.
14. Wu, L. C., Peng, Z-yu and Kim, P. S. (1995) *Nat. Struct. Biol. 2*, 281−286.
15. Schulman, B. A., and Kim, P. S. (1996) *Nat. Struct. Biol. 3*, 682−687.
16. Wu, L. C., and Kim, P. S. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 14314−14319.
17. Song, J., Bai, P., Luo, L., and Peng, Z-yu (1998) *J. Mol. Biol. 280*, 167−174.
18. Wu, L. C., and Kim, P. S. (1998) *J. Mol. Biol. 280*, 175−182.
19. Mizuguchi, M., Masaki, K., and Nitta, K. (1999) *J. Mol. Biol. 280*, 175−182.
20. Kay, M. S., and Baldwin, R. L. (1996) *Nat. Struct. Biol. 3*, 439−445.
21. Eliezer, D., Yao, J., Dyson, H. J., and Wright, P. E. (1998) *Nat. Struct. Biol. 5*, 148−155.
22. Fink, A. L., Oberg, K. A., and Seshadri, S. (1998) *Folding Des. 3*, 19−25.
23. Kay, M. S., and Ramos, C. H., and Baldwin, R. L. (1999) *Proc. Natl. Acad. Sci. 96*, 2007−2012.
24. Morozova, L. A., Haezebrouck, P., and Van Cauwelaert, F. (1991) *Biophys. Chem. 41*, 185−191.
25. Van Dael, H., Haezebrouck, P., Morozova, L. A., Arico-Muendel, C., and Dobson, C. M. (1993) *Biochemistry 32*, 11886−11894.
26. Griko, Y. V., Freire, E., Privalov, G., Van Dael, H., and Privalov, P. L. (1995) *J. Mol. Biol. 252*, 447−459.
27. Morozova, L. A., Haynie, D. T., Arico-Muendel, C., Van Dael, H., and Dobson, C. M. (1995) *Nat. Struct. Biol. 2*, 871−875.
28. Morozova, L. A., Arico-Muendel, C., Haynie, D. T., Emelyanenko, V. I., Van Dael, H., and Dobson, C. M. (1997) *J. Mol. Biol. 268*, 903−921.
29. Grobler, J. A., Rao, K. R., Pervaiz, S., and Brew, K. (1994) *Arch. Biochem. Biophys. 313*, 360−366.
30. Kikuchi, M., Kawano, K., and Nitta, K. (1998) *Protein Sci. 7*, 2150−2155.
31. Koshiba, T., Hayashi, T., Ishido, M., Kumagai, I., Ikura, T., Kawano, K., Nitta, K., and Kuwajima, K. (1999) *Protein Eng. 12*, 429−435.
32. Nitta, K., Tsuge, H., Sugai, S., and Shimazaki, K. (1987) *FEBS Lett. 223*, 405−408.
33. Koshiba, T., Tsumoto, K., Masaki, K., Kawano, K., Nitta, K., and Kumagai, I. (1998) *Protein Eng. 11*, 683−690.
34. Morris, G. A., and Freeman, R. (1978) *J. Magn. Reson. 29*, 433−462.
35. Sakabe, N., Ikemizu, S., Sakabe, K., Higashi, T., Nakagawa, A., Watanabe, N., Adachi, S., and Sasaki, K. (1995) *Rev. Sci. Instrum. 66*, 1276−1281.
36. Otwinowski, Z. (1993) pp 80−86, SERC Daresbury Laboratory, Warrington, UK.
37. Collaborative Computational Project, Number 4. (1994) *Acta Crystallogr. D50*, 760−763.
38. Navaza, J. (1994) *Acta Crystallogr. A50*, 157−163.
39. Tsuge, H., Ago, H., Noma, M., Nitta, K., Sugai, S., and Miyano, M. (1992) *J. Biochem. 111*, 141−143.
40. Brünger, A. T. (1993) X-PLOR, Ver. 3.1. A system for X-ray Crystallography and NMR, Yale University Press, New Haven, CT.
41. Desmet, J., Van Dael, H., Van Cauwelaert, F., Nitta, K., and Sugai, S. (1989) *J. Inorg. Biochem. 37*, 185−191.
42. Yutani, K., Ogasahara, K., and Kuwajima, K. (1992) *J. Mol. Biol. 228*, 347−350.
43. Griko, Y., Freire, E., and Privalov, P. L. (1994) *Biochemistry 33*, 1889−1899.
44. Pfeil, W. (1998) *Proteins: Struct. Funct. Genet. 30*, 43−48.
45. Griko, Y., and Remeta, D. P. (1999) *Protein Sci. 8*, 554−561.
46. Koshiba, T., Kuwajima, K., Kobashigawa, Y., and Nitta, K. (2000) to be published.
47. Stryer, L. (1965) *J. Mol. Biol. 13*, 482−495.
48. Semisotnov, G. V., Rodionova, N. A., Razgulyaev, O. I., Uversky, V. N., Gripas, A. F., and Gilmanshin, R. I. (1991) *Biopolymers 31*, 119−128.
49. Demarest, S. J., Fairman, R., and Raleigh, D. P. (1998) *J. Mol. Biol. 283*, 279−291.
50. Semisotnov, G. V., Rodionova, N. A., Kutyshenko, V. P., Ebert, B., Blanck, J., and Ptitsyn, O. B. (1987) *FEBS Lett. 224*, 9−13.
51. Kuwajima, K., Harushima, Y., and Sugai, S. (1986) *Int. J. Peptide Protein Res. 27*, 18−27.
52. Baum, J., Dobson, C. M., Evans, P. A., and Hanley, C. (1989) *Biochemistry 28*, 7−13.
53. Nitta, K., Tsuge, H., and Iwamoto, H. (1993) *Int. J. Peptide Protein Res. 41*, 118−123.
54. Artymiuk, P. J., and Blake, C. C. F. (1981) *J. Mol. Biol. 152*, 737−762.
55. Stuart, D. I., Acharya, K. R., Walker, N. P. C., Smith, S. G., Lewis, M., and Phillips, D. C. (1986) *Nature 324*, 84−87.
56. Acharya, K. R., Stuart, D. I., Walker, N. P. C., Lewis, M., and Phillips, D. C. (1989) *J. Mol. Biol. 208*, 99−127.
57. Acharya, K. R., Ren, J., Stuart, D. I., Phillips, D. C., and Fenna, R. E. (1991) *J. Mol. Biol. 221*, 571−581.
58. Pike, A. C. W., Brew, K., and Acharya, K. R. (1996) *Structure 4*, 691−703.
59. Guss, J. M., Messer, M., Costello, M., Hardy, K., and Kumar, V. (1997) *Acta Crystallogr. D53*, 355−363.
60. Mizuguchi, M., Arai, M., Yue, K., Nitta, K., and Kuwajima, K. (1998) *J. Mol. Biol. 283*, 262−277.
61. Uchiyama, H., Perez-Prat, E. M., Watanabe, K., Kumagai, I., and Kuwajima, K. (1995) *Protein Eng. 8*, 1153−1161.
62. Alexandrescu, A. T., Evans, P. A., Pitkeathly, M., Baum, J., and Dobson, C. M. (1993) *Biochemistry 32*, 1707−1718.
63. Dobson, C. M. (1995) *Nat. Struct. Biol. 2*, 513−517.
64. Blagdon, D. E., and Goodman, M. (1975) *Biopolymers 14*, 241−245.
65. Wada, A. (1976) *Adv. Biophys. 9*, 1−63.

66. Hol, W. G. J., van Duijnen, P. T., and Berendeson, H. J. C. (1978) *Nature 273*, 443−446.
67. Nicholson, N., Becktel, W. J., and Matthews, B. W. (1988) *Nature 336*, 651−656.
68. Tanford, C. (1962) *Adv. Protein. Chem. 17*, 69−165.
69. Hughson, F. M., Barrick, D., and Baldwin, R. L. (1991) *Biochemistry 30*, 4113−4118.
70. Kiefhaber, T., and Baldwin, R. L. (1995) *J. Mol. Biol. 252*, 122−132.
71. Marmorino, J. L., and Pielak, G. J. (1995) *Biochemistry 34*, 3140−3143.
72. Marmorino, J. L., Lehti, M., and Pielak, G. J. (1998) *J. Mol. Biol. 275*, 379−388.
73. Merutka, G., Lipton, W., Shalongo, W., Park, S. H., and Stellwagen, E. (1990) *Biochemistry 29*, 7511−7515.
74. Kraulis, P. J. (1991) *J. Appl. Cryst. 24*, 946−950.
75. McKenzie, H. A., and Shaw, D. C. (1985) *Biochem. Int. 10*, 23−31.
76. Brew, K., Vanaman, T. C., and Hill, R. L. (1967) *J. Biol. Chem. 242*, 3747−3749.
77. Brew, K. (1972) *Eur. J. Biochem. 27*, 341−353.
78. Shewale, J. G., Sinha, S. K., and Brew, K. (1984) *J. Biol. Chem. 259*, 4947−4956.
79. Hall, L., Craig, R., Edbrooke, M. R., and Campbell, P. N. (1982) *Nucl. Acids Res. 10*, 3503−3515.
80. Ikeguchi, M., Kuwajima, K., Mitani, M., and Sugai, S. (1986) *Biochemistry 25*, 6965−6972.

BI991525A