PROTEINS: Structure, Function, and Genetics 27:567–575 (1997)

Molten Globule State of Equine β-Lactoglobulin

Masamichi Ikeguchi,* Shin-ichi Kato, Akio Shimizu, and Shintaro Sugai
Department of Bioengineering, Faculty of Engineering, Soka University, Hachioji, Tokyo, Japan

ABSTRACT **The acid-unfolded state of equine β-lactoglobulin was characterized by means of circular dichroism, nuclear magnetic resonance, analytical gel-filtration chromatography, and analytical centrifugation. The acid-unfolded state of equine β-lactoglobulin has a substantial secondary structure as shown by the far-ultraviolet circular dichroism spectrum but lacks persistent tertiary packing of the side chains as indicated by the near-ultraviolet circular dichroism and nuclear magnetic resonance spectra. It is nearly as compact as the native conformation as shown by the gel filtration and sedimentation experiments, and it has the exposed hydrophobic surface as indicated by its tendency to aggregate. All of these characteristics indicate that the acid-unfolded state of equine β-lactoglobulin is a molten globule state. The α helix content in the acid-unfolded state, which has been estimated from the circular dichroism spectrum, is larger than that in the native state, suggesting the presence of nonnative α helices in the molten globule state. This result suggests the generality of the intermediate with nonnative α helices during the folding of proteins having the β-clam fold.**
Proteins 27:567–575, 1997.

Key words: protein folding; lipocalin; calycin; α helix; β sheet; CD; NMR; ultracentrifugation

INTRODUCTION

Recently, it has been widely accepted that the molten globule state is a general intermediate populated during an early stage of protein folding.[1–6] The molten globule has a substantial secondary structure but lacks the persistent tertiary packing characteristic of native structures. It is nearly as compact as the native conformation but has an exposed hydrophobic surface that binds a hydrophobic dye or induces extensive aggregation. The detailed structures of the molten globule states have been revealed for more than a dozen proteins by combining various physical techniques, such as NMR and CD, with molecular biological techniques, such as isotope-labeling and site-directed mutagenesis.[7–10] The secondary structures in the molten globule that have been revealed so far are a part of the secondary structures in the native structure. There is little clear evidence indicating that nonnative structures formed during the folding process are essential to achieve the final native structure.

β-Lactoglobulin is a major component of mammalian milk. This protein is similar in both amino acid sequence[11] and three-dimensional structure[12] to serum retinol-binding protein (SRBP) and belongs to a superfamily of hydrophobic ligand-binding proteins[13] or calycins.[14] Although the biological function of β-lactoglobulin is not clear, it is thought to be a transporter of retinol for suckling animals because the protein has been known to strongly bind retinol.[15] The conformational properties of β-lactoglobulin, especially those of BLG, have been extensively investigated. From such studies, there are several interesting observations suggesting that some nonnative secondary structure is formed during an early stage of BLG folding. Kuwajima et al.[16] have investigated the folding kinetics of BLG by CD stopped-flow analysis. They have shown that the far-UV CD intensity larger than that of native BLG is regained within the dead time of the measurement. From the CD spectrum of the species formed within the dead time,[17] it seems that this folding intermediate assumes a considerable amount of α helix in spite of the predominant β sheet structure of the native BLG.[12,18] Goto and coworkers[19,20] have suggested the possibility of a "nonhierarchical" folding mechanism for BLG from the high helical propensity of BLG and its fragments in trifluoroethanol. Furthermore, it has been reported that the CRABP and SRBP, other members of the calycin superfamily, assume a partially folded conformation containing a substantially larger amount of α helix than the native conformation.[21–23] These results suggest that the proteins having the β-clam fold, which is a common folding

The present address of S. Kato is Department of Chemistry, Faculty of Science, Tokyo Metropolitan University.

Abbreviations: NMR, nuclear magnetic resonance; CD, circular dichroism; BLG, bovine β-lactoglobulin; ELG, equine β-lactoglobulin; CRABP, cellular retinoic acid binding protein; SRBP, serum retinol binding protein; GdnHCl, guanidine hydrochloride; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis; TSP, 2,2,3,3-tetradeutero-3-trimethylsilylpropionic acid

*Correspondence to: Dr. Masamichi Ikeguchi, Department of Bioengineering, Faculty of Engineering, Soka University, 1-236 Tangi-cho, Hachioji, Tokyo 192, Japan.

Received 12 August 1996; Accepted 24 October 1996

pattern of the calycin superfamily, generally fold via an α-helical intermediate.

To examine the generality of the α-helical intermediate in the folding of β-clam proteins, we have investigated the unfolding of ELG. It has been known that ELG contains two components (I and II). Component I consists of 162 amino acids with two disulfide bridges like BLG. However, the amino acid sequence of ELG is surprisingly different from that of BLG; 72 amino acids (44%) are different.[24] ELG is monomeric, whereas BLG shows various oligomeric states depending on pH.[25] ELG has no free cysteine residue, whereas BLG has a free cysteine residue and disulfide interchange readily occurs.[26,27] These properties of ELG are advantageous for simplifying the analysis of the folding reaction. In this report, we show that ELG assumes the molten globule state at an acidic pH. The CD spectrum has shown that the molten globule state of ELG has a larger amount of α helix than the native conformation. These results are very similar to the results reported for other proteins that belong to the calycin superfamily,[16,21–23] and suggest the importance of nonnative α helices in the folding of the β-clam proteins.

MATERIALS AND METHODS

Materials

ELG I was prepared as previously described for the preparation of equine α-lactalbumin.[28] Purity of the preparation was checked by SDS-PAGE, native PAGE at pH 8.0, and isoelectric focusing on a PhastGel IEF 4-6.5 using a PhastSystem (Pharmacia). Identification of ELG was made by amino acid analyses on a Waters Pico-Tag column (3.9 × 300 mm). The amino acid composition of our preparation agreed with the reported amino acid composition.[24]

Bovine α-lactalbumin was prepared as previously described.[28] Ribonuclease A, chymotrypsinogen A, ovalbumin, bovine serum albumin, and Blue Dextran 2000 were purchased from Pharmacia (Gel Filtration Calibration Kit). Cytochrome c, myoglobin, aldolase, catalase, and ferritin were from SERVA (Protein Molecular Weight Standard, Kit MS II). Bovine carbonic anhydrase and BLG were from Sigma, and bovine pancreatic trypsin inhibitor was from Boehringer Mannheim. Stable isotope compounds were purchased from Isotec, Merck, or Aldrich. The other chemicals were of analytical grade from Wako Pure Chemical Industries or Nacalai Tesque.

CD Measurements

CD spectra were measured with a Jasco J-720 spectropolarimeter at 25°C. Usually, 10 mM potassium citrate buffer, the ionic strength of which was adjusted to 0.1 with KCl, was used for the pH titration experiments. CD spectra at pH <2 were measured in 0.1 M HCl-KCl. In some cases, 10 mM phosphate buffer was also used. Typically, cuvettes with 1-mm and 10-mm path lengths were used for the measurements in the far- and near-UV regions, respectively. To obtain the CD spectrum of wavelengths below 200 nm, a cell with a 0.1-mm path length was used. The protein concentration used was 22–30 μM, and it was spectrophotometrically determined using the molar extinction coefficient, ϵ_{280} = 12,000 M^{-1} cm^{-1}, which was determined by the method of Gill and von Hippel.[29]

NMR Measurements

500 MHz ^{1}H-NMR spectra were recorded with a JEOL A-500 spectrometer at 25°C. For all measurements, 16 K data points were collected over spectral widths of 8 kHz, giving a digital resolution of 0.49 Hz. The residual HDO peak was suppressed by presaturation. Chemical shifts were measured relative to a TSP (0 ppm) or dioxane (3.75 ppm) internal standard. The reported pD values are uncorrected pH meter readings.

Analytical Gel Filtration Chromatography

The molecular dimensions and the extent of aggregation were determined by gel filtration chromatography on a Superdex 75 HR 10/30 column using a FPLC system (Pharmacia) at 4°C. The column was equilibrated with 20 mM phosphate buffer (pH 7.0) the ionic strength of which was adjusted to 0.2 with KCl (native condition), 0.1 M HCl-KCl (pH 1.5) (acidic condition) or 20 mM phosphate buffer containing 6 M GdnHCl (pH 7.0) (unfolding condition). The flow rate was 0.5 ml/min or 0.4 ml/min. The column was calibrated with the proteins described in the Materials section, and their Stokes radii were estimated from their intrinsic viscosities, diffusion coefficients or sedimentation coefficients given in the literature.[30–33]

Analytical Ultracentrifugation

Sedimentation equilibrium and velocity experiments were performed on a Beckmen XL-A analytical ultracentrifuge using an An-60 Ti rotor. Sedimentation equilibrium experiments were performed at pH 1.5, 3.3, 5.8, and 8.0. Buffers used were 0.1 M HCl-KCl (pH 1.5), 10 mM citrate buffer of which the ionic strength was adjusted to 0.1 with KCl (pH 3.3 and 5.8), and 10 mM Tris-HCl, 0.1 M KCl (pH 8.0). Samples at pH >3 were dialyzed against the respective buffers before the analyses. The protein concentration was 50–70 μM. Samples at pH 1.5 were passed through a Sephadex G-25 column equilibrated with 0.1 M HCl-KCl (pH 1.5) before the analyses. The protein concentration was 3 μM. Typically, a sample solution of 175 μl was placed in the sample sector of an aluminum-filled or charcoal-filled Epon centerpiece, and the reference sector was filled with 200 μl of the buffer used for the dialysis. The ultracentrifugation was done at a rotor speed of 18,000 rpm at 20°C. The concentration gradient formed as monitored by UV absorbance at 280 or 230

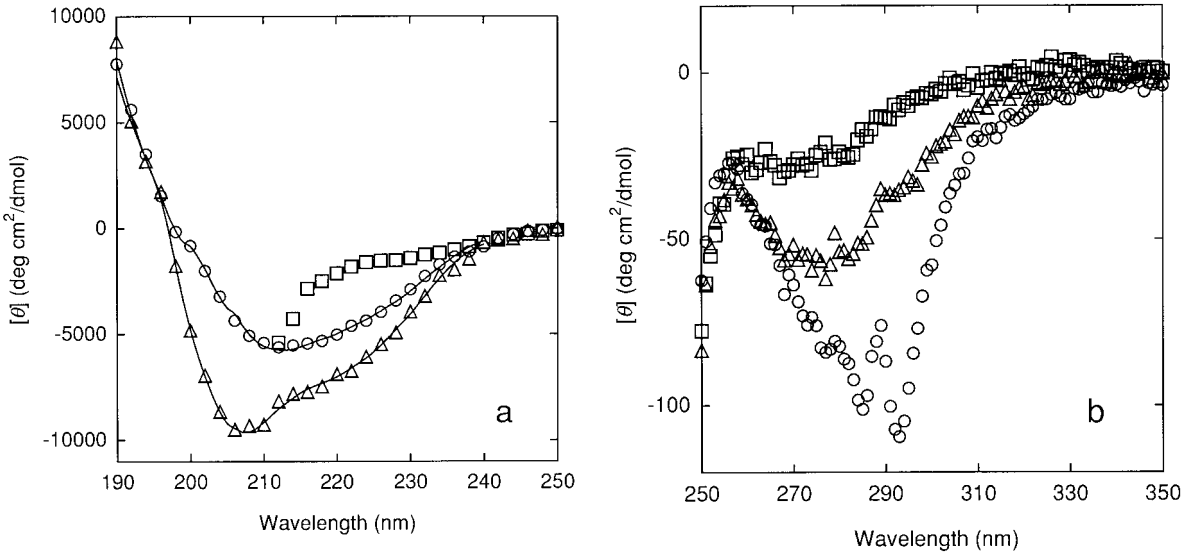

Fig. 1. Far- (**a**) and near- (**b**) UV CD spectra of ELG at pH 7 (circles), at pH 1.5 (triangles), and in 5.7 M GdnHCl (pH 5.6) (squares). To clarify the figure, data are shown only at 2-nm and 1-nm intervals in Figures a and b, respectively. Solid lines in Figure a are fitting results from the CONTIN program (see Table I).

TABLE I. Secondary Structure Contents of the N and A States of ELG*

	% α Helix	% β Sheet	% Remainder
N state†	9 (0.8)	41 (1.4)	50 (1.3)
A state‡	20 (0.9)	28 (1.6)	52 (1.5)
N state of BLG (x-ray)§	6	44	50

*The secondary structural contents of ELG were estimated from the CD spectrum using the CONTIN program.[37,38] Values in parentheses are standard errors.
†Conditions: 10 mM phosphate buffer (pH 7.1) at 25°C.
‡Conditions: 0.1 M HCl-KCl (pH 1.5) at 25°C.
§Taken from Monaco et al.[18]

nm. Sedimentation equilibrium was attained within 16 hours. The apparent molecular weight was determined using a program provided by the manufacturer, with the partial specific volume of 0.737 cm^3/g, which was calculated from the amino acid composition.[34]

Sedimentation and diffusion coefficients were simultaneously determined by the second moment method[35] and the boundary spreading method,[36] respectively. The samples were prepared like those used in the sedimentation equilibrium experiments. The rotor was accelerated by 80 rpm/s up to 40,000 rpm at 20°C. The boundary was formed using a synthetic boundary centerpiece and was detected by UV absorbance at a 3- or 5-minute interval.

RESULTS

pH Dependence of ELG Conformation

The pH dependence of the conformation of ELG was investigated by CD. In Figure 1, the CD spectra of ELG at neutral pH, at acidic pH, and in concentrated GdnHCl are shown. The CD spectrum of ELG at neutral pH is similar to that previously reported for BLG, suggesting that the native (N) conformation of ELG is similar to that of BLG. A CONTIN analysis[37,38] of the far-UV CD spectrum of the native ELG gives the following secondary structure contents; 9% α helix, 41% β sheet (Table I). This is consistent with the secondary structure content of BLG determined from x-ray crystallography.[12,18] The CD spectrum of ELG shows a significant change upon acidification. Figure 2 shows the pH dependence of the ellipticities at 293 and 222 nm. The acid-induced conformational transitions measured at two wavelengths coincide, indicating that the cooperative two-state transition occurs around pH 2.3. Because the CD spectrum at acidic pH shows the time-dependent change in the far-UV region that is associated with aggregation (see below), all data shown in Figures 1 and 2 were acquired within 2 hours after dissolution of the protein. The acid-induced conformational transition is fully reversible as long as the protein is not exposed to an acidic pH longer than 2 hours.

Aggregation of ELG

Before conformational characterization of the acid (A) state of ELG, the property of ELG to aggregate was investigated using analytical gel filtration chromatography and an analytical ultracentrifuge. Figure 3 shows the elution of ELG, which was incubated in 0.1 M HCl-KCl (pH 1.5) for various lengths of time at room temperature, from a Superdex 75 column equilibrated with the same solvent at 4°C. After a 5-minute incubation, ELG eluted as a single peak at

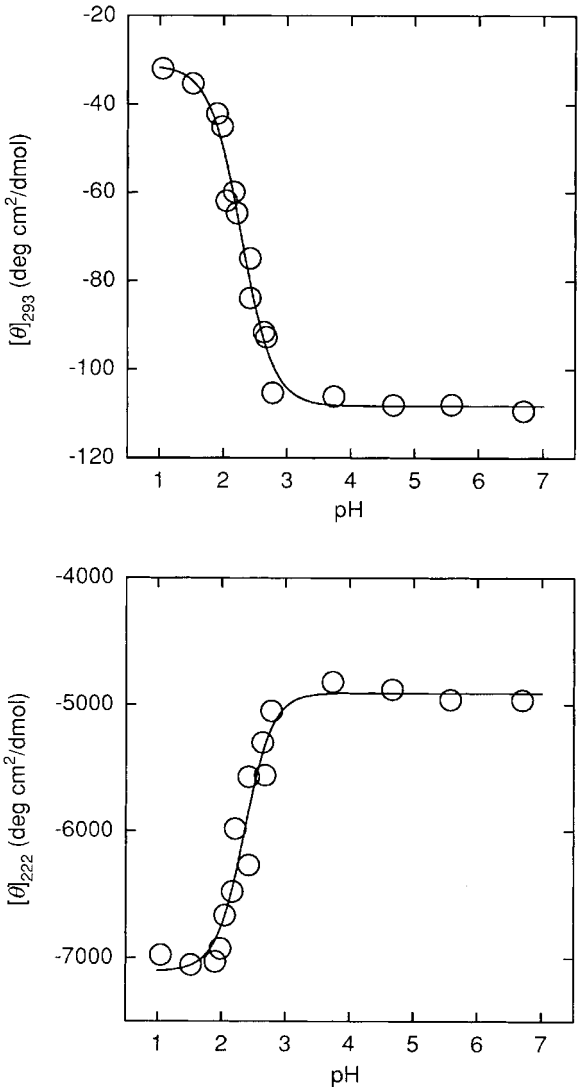

Fig. 2. Acid-unfolding transition of ELG monitored by CD at 293 and 222 nm. Ionic strength is constant (0.1) throughout the investigated pH range.

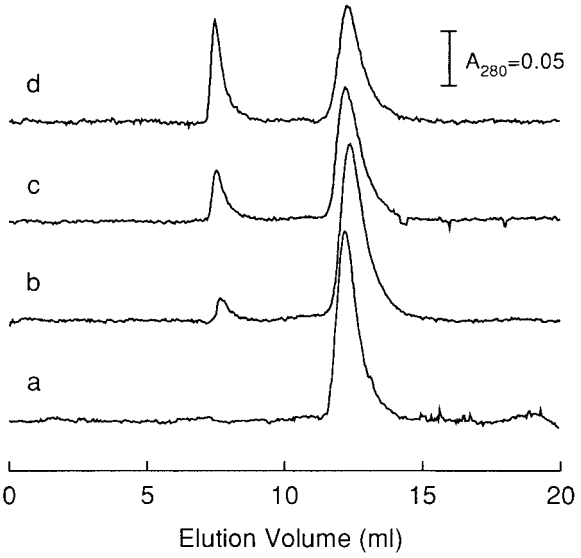

Fig. 3. Elution profile of ELG from a Superdex 75 column at 4°C. ELG (50 µM) was incubated in 0.1 M HCl-KCl (pH 1.5) at room temperature for 5 minutes (*a*), 120 minutes (*b*), 240 minutes (*c*), and 360 minutes (*d*) before chromatography. The protein was eluted with 0.1 M HCl-KCl (pH 1.5) at a flow rate of 0.5 ml/min. Elution was monitored by absorbance at 280 nm.

an elution volume of 12.4 ml. This elution volume is consistent with the elution of a protein having the molecular weight of 18.5 k Da, indicating that ELG is monomeric immediately after dissolving in 0.1 M HCl-KCl (pH 1.5) at a protein concentration of 50 µM. After a 120-minute incubation, however, two peaks were observed. The elution volume of the first peak (7.6 ml) corresponds to the void volume of the column, indicating that large aggregates (>100 kDa) are formed at acidic pH. The population of the aggregated ELG is approximately 5% for a 120-minute incubation. As the incubation time becomes longer, the population of aggregated species increases and reaches 30% after a 6-hour incubation at a protein concentration of 50 µM.

In order to confirm that ELG is monomeric at various pHs, sedimentation equilibrium experiments were performed at pH 1.5, 3.3, 5.8, and 8.0. At a pH between 3.3 and 8.0, where ELG assumes the native conformation as judged by the CD spectrum, the apparent molecular weight determined at a protein concentration of 50–70 µM is 16–17 K, which is consistent with the molecular weight calculated from the amino acid composition (18.5 K). Therefore, ELG is monomeric at a pH ranging from 3 to 8, in contrast to the pH-dependent oligomerization of BLG.[25] To prevent aggregation during ultracentrifugation, sedimentation equilibrium experiments at pH 1.5 were performed at a protein concentration of 3 µM. At this ELG concentration, no aggregated species was detected during ultracentrifugation, and the determined apparent molecular weight (17.4 K) shows a fair agreement with the calculated value (18.5 K).

Conformation of the Monomeric A State of ELG

The near- and far-UV CD spectra show that the tertiary packing of aromatic residues is largely lost and that some secondary structure is retained in the acid-unfolded ELG, respectively (Fig. 1). This result suggests that the ELG molecule assumes a molten globule state at acidic pH. At acidic pH, however, the broad CD band around at 275 nm remains, and the near-UV CD spectrum is different from that of the unfolded protein in concentrated GdnHCl (Fig. 1). To obtain further information on the side chain environment in the A state, the ^{1}H-NMR spectrum of ELG was measured at pD 5.8, at pD 1.2 or in 8 M urea (pD

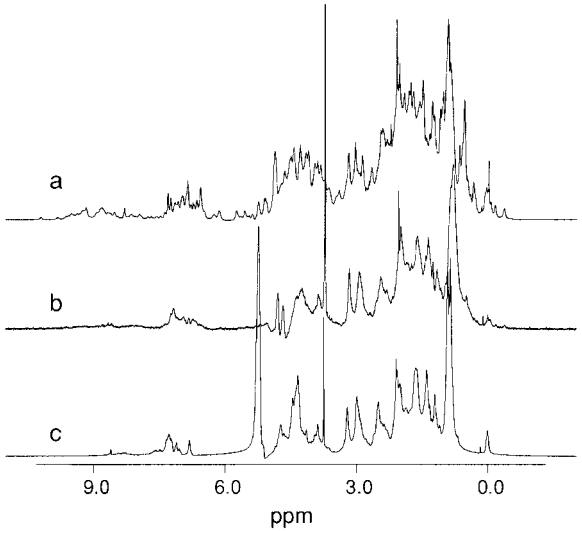

Fig. 4. 500 MHz ^{1}H-NMR spectra of native (*a*), acid-unfolded (*b*), and urea-unfolded ELG (*c*) at 25°C. The spectra of the native ELG (pD 6.0) and the unfolded ELG in 8 M urea (pD 1.6) were measured at the protein concentration of 1 mM by the accumulation of 32 scans. Urea was deuterated by repeated lyophilization from D_2O solution. The spectrum of the acid-unfolded ELG (pD 1.2) was measured at the protein concentrations of 50 μM by the accumulation of 1024 scans. All measurements were completed within 2 hours after sample preparation. A broad peak seen at 0 ppm in (*c*) is the resonance of TSP. The broadness of the TSP peak suggests that TSP interacts with ELG even in 8 M urea (see Ref. 53).

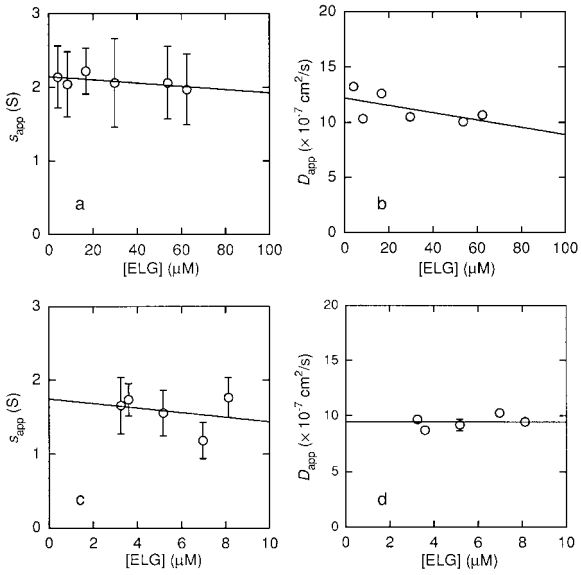

Fig. 5. Concentration dependence of the apparent sedimentation coefficient (s_{app}) and the apparent diffusion coefficient (D_{app}) of ELG under the native (*a* and *b*) and acid (*c* and *d*) conditions at 20°C. Buffers in the native and acid conditions were 10 mM Tris-HCl, 0.1 M KCl (pH 8.0) and 0.1 M HCl-KCl (pH 1.5), respectively.

TABLE II. Hydrodynamic Radii of N, A, and U States of ELG*

	N†	A‡	U§
Rs from s_{20} (Å)**	20.0 (0.5)	24.7 (5.0)	
Rs from D_{20} (Å)††	17.5 (1.2)	22.6 (0.6)	
Rs from gel filtration (Å)	22‡‡	24§§	37

*Values in parentheses are standard errors.
†Conditions: 10 mM Tris-HCl, 0.1 M KCl (pH 8.0) at 20°C.
‡Conditions: 0.1 M HCl-KCl (pH 1.5) at 20°C.
§Conditions: 20 mM phosphate, 6 M GdnHCl (pH 7.0) at 4°C.
**Stokes radius estimated from s_{20}, molecular weight (18.5 K), and partial specific volume (0.737 cm^3/g) by assuming a spherical shape of the ELG molecule.
††Stokes radius estimated from D_{20} by assuming a spherical shape of the ELG molecule.
‡‡Conditions: 20 mM phosphate (pH 7.0, I = 0.2) at 4°C.
§§Conditions: 0.1 M HCl-KCl (pH 1.5) at 4°C.

1.6) (Fig. 4). To prevent aggregation, the spectrum at acidic pD was acquired at a protein concentration of 50 μM within 2 hours after dissolving the protein. The spectrum at pD 5.8 is typical of globular proteins of this size. The wide dispersion of the chemical shifts reflects the highly specific interresidue interactions within the compact folded structure. These features are absent from the spectrum of the unfolded (U) state in 8 M urea. The spectrum of the U state shows a marginal deviation from the spectrum expected for a random coil conformation. The spectrum of the A state is closer to that of the U state rather than that of the N state. Several high-field shifted peaks are observed at around 0 ppm, indicating that some tertiary packing of the side chains remains in the A state. However, we emphasize again that the chemical shift dispersion is largely lost in the NMR spectrum of the A state and that the A state of ELG is similar to the molten globule state known so far.

We have also examined the compactness of ELG in the A state, which is another characteristic of the molten globule state. As described above, under acidic conditions, the monomeric ELG was eluted from the Superdex 75 column at 12.4 ml. This value is close to the elution volume of ELG under the native condition (11.9 ml), indicating that the A state of ELG is as compact as the N state, the Stokes radius of which is estimated to be approximately 22 Å. The elution volume of ELG under the unfolding condition is 8.3 ml, from which the Stokes radius of ELG in 6 M GdnHCl is estimated to be approximately 37 Å. To confirm the nativelike compactness of the A state of ELG, we have further measured the sedimentation coefficient (s_{20}) and the diffusion coefficient (D_{20}) of ELG under the native and acidic conditions. The results are summarized in Figure 5 and Table II. The Stokes radii of the N and A states were calculated from the sedimentation or diffusion coefficients, assuming a spherical shape of the molecule. The Stokes radius of ELG under the acidic condition is slightly larger than that under the

native condition but is much smaller than that of the unfolded protein. Therefore, it is concluded that the A state of ELG is nearly as compact as the N state.

It is clear from the far-UV CD spectrum that the A state of ELG has a considerable amount of secondary structure, although the spectrum of the A state is significantly different from that of the N state. To obtain more detailed information on the secondary structure in the A state, we have estimated the secondary structure content in the A state from the CD spectrum using the CONTIN program.[37,38] The results are shown in Table I. The α helix content in the A state is significantly larger than that in the N state, indicating that nonnative α helices are formed in the A state. Although the β sheet content in the A state decreases as the α helix content increases, this does not necessarily indicate that a part of the β sheet in the N state is converted to the α helix in the A state. Anyway, it is suggested by this analysis that the secondary structure in the A state is not native-like, although a part of the native secondary structures may be maintained in the A state.

DISCUSSION

The following characteristics of the acid-unfolded ELG elucidated in this study show that the A state of ELG is a molten globule state: (1) The A state of ELG has a substantial secondary structure as indicated by the far-UV CD spectrum, (2) it lacks the rigid tertiary packing of the side chains as shown by elimination of the near-UV CD intensity and of the wide dispersion of the chemical shifts, (3) it is nearly as compact as the N state as shown by the gel filtration and sedimentation experiments, and (4) it has an exposed hydrophobic surface as indicated by its tendency to aggregate. In the classical view of the molten globule, in which the molten globule is "a compact globule with nativelike secondary structure and with slowly fluctuating tertiary structure,"[39] the A state of ELG may not be regarded as a molten globule, because the A state contains nonnative secondary structures. Kim and Baldwin[40] proposed the term "collapsed form" to distinguish the experimentally observed intermediate from the original model of the molten globule. Ptitsyn[41] also proposed the terms "nativelike molten globule" and "disordered molten globule" for compact states with and without a native-like tertiary fold. As a result of finding various molten globule-like states, however, it has been revealed that the degree of structural organization in the molten globule-like state is considerably dependent on the proteins. Recently, the terms, "premolten globule,"[4] "molten coil,"[42] and "highly ordered molten globule,"[42–44] have been used. Because, at present, characterization of the A state of ELG is not sufficient to place the A state in a specified category, we conclude here that the A state of ELG is a molten globule, using the term in a broad sense.

An important feature of the molten globule state of ELG is that it contains nonnative α helices. Although BLG is stable at acidic pH and does not assume the molten globule state at acidic pH, an intermediate with non-native α-helices has been transiently observed during the folding from the GdnHCl-unfolded state.[16] The CD spectrum of the transient intermediate of BLG[17] is similar to those of the A state of ELG, suggesting that the molten globule state observed in this study generally appears during the folding reaction of β-lactoglobulin. The reason that the molten globule state is not observed for BLG is probably that the native state of BLG is more stable than that of ELG at acidic pH. Preliminary experiments have shown that the urea-induced unfolding transition of ELG occurs at a much lower urea concentration than that of BLG (unpublished results). A partially unfolded state showing the CD spectrum similar to that of the A state of ELG has been observed upon esterification of BLG.[45] A partially unfolded state very similar to the A state of ELG has also been observed for CRABP,[21] which is another member of the calycin superfamily. The secondary structure content in the A state of ELG (Table I) is very similar to that in the acid-unfolded CRABP, in which the α helix and β sheet contents have been reported to be 17–22% and 26–29%, respectively, suggesting the similarity of the secondary structure in the molten globule state between the proteins belonging to the calycin superfamily. SRBP, which is also homologous to β-lactoglobulin, has been reported to be in the molten globule state at acidic pH, although the secondary structure content in that state has not been reported.[22] Muccio et al.[23] have investigated the CD spectrum of thermally unfolded SRBP, which is similar to that of the acid unfolded SRBP, and have shown that both the α helix and β sheet contents are 20%. The similarity of these results suggests the generality of the α-helical intermediate during the folding of β-clam proteins.

It is of interest to ask whether nonnative secondary structures in the molten globule state play an important role in the folding of proteins. The fact that nonnative α helices are commonly observed for proteins of the calycin superfamily suggests the importance of the intermediate with nonnative α helices during the folding of β-clam proteins. The nonnative α helix in the A state of ELG, however, may be a result of local conformational preference, and it may not interact with other parts of the polypeptide chain. It may easily interconvert with other conformations and may not direct the succeeding folding process. In order to detect the local helical preference in the sequences of BLG, ELG, SRBP, and CRABP, we have analyzed the helix probability of these protein sequences with the HELIX program implementing the Lifson-Roig helix-coil theory with modifications to include N- and C-capping effects.[46,47] The helix probability of each residue is shown versus

```
Aligned Number      1        10        20        30        40        50        60

BLG                 1:-----LIVTQTMKGLDIQKVAGTWYSLAMAASDI---------------SLLDAQSAPLR 40
ELG                 1:-----TNIPQTMQDLDLQEVAGKWHSVAMAASDI---------------SLLDSEEAPLR 40
SRBP                1:ERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEG---------------LFLQDNIVAEF 45
CRABP               1:-----------------PNFAGTWKMRSSENFDELLKALGVNAMLRKVAVAAASKPHVEI 43

Aligned Number     61        70        80        90       100       110       120

BLG                41:VYVEELKPTPEGDLEILLQKWENGECAQKKIIAEKTKIPAVFKID---------------- 85
ELG                41:VYIEKLRPTPEDNLEIILREGENKGCAEKKIFAEKTESPAEFKIN---------------- 85
SRBP               46:SVDETGQMSATAKGRVRL-LNNWDVCADMVGTFTDTEDPAKFKMK-------------YW 91
CRABP              44:RQDGDQ-FYIKTSTT----------VRTTEINFKVGEGFEEETVDGRKCRSLPTWENENK 92

Aligned Number    121       130       140       150       160       170       180

BLG                86:----------ALNENKVLVLDTDYKKYLLFCM---ENSAEPEQSLACQCLVRTPEVDDEA 132
ELG                86:----------YLDEDTVFALDTDYKNYLFLCM---KNAATPGQSLVCQYLARTQMVDEEI 132
SRBP               92:GVAS----FLQKGNDDHWIVDTDYDTYAVQYSCRLLNLDGTCADSYSFVFSRDPNGLPPE 147
CRABP              93:IHCTQTLLEGDGPKTYWTREL--ANDELILTFGADD-------VVCTRIYVRE         136

Aligned Number    181       190       200       210       217

BLG               133:LEKFDKALKALPMHIRLSFNPTQLEEQCHI                              162
ELG               133:MEKFRRALQPLPGRVQIVPDLTRMAERCRI                              162
SRBP              148:AQKIVR-QRQEELCLARQYRLIVHNGYCDGRSERNLL                       183
```

a

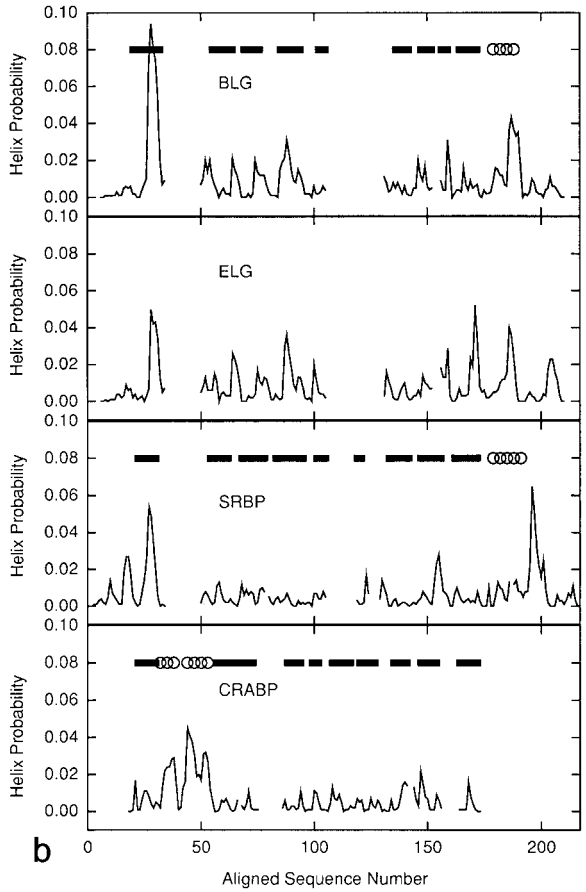

Fig. 6. (**a**) Sequence alignment of BLG, ELG, SRBP, and CRABP. The alignment was done according to Flower et al.[14] (**b**) Helix probabilities of BLG, ELG, SRBP, and CRABP sequences are plotted versus the aligned sequence number. The helix probability was calculated with the HELIX program.[46,47] The regions of α helix and β strand in the crystal structure[18,54,55] are shown as circles and bars, respectively.

the aligned sequence (Fig. 6b). The alignment is difficult especially between the extracellular hydrophobic ligand-binding proteins (lipocalins) such as BLG, ELG, and SRBP and the intracellular fatty acid-binding proteins (FABP) such as CRABP. The former possesses an eight-stranded antiparallel β barrel with an α helix at the C-terminal, whereas the latter has a ten-stranded antiparallel β barrel with two α helices between the first and second strands. Flower et al.[14] have established the equivalent regions between lipocalin (SRBP, bilin-binding protein, insecticyanin, and mouse major urinary protein) and FABP (myelin P2 protein and intestinal FABP) by comparing their amino acid sequences and three-dimensional structures. The sequence alignment shown in Figure 6a is based on the alignment of Flower et al.[14] The high helical probability has been commonly observed for BLG, ELG, and SRBP in the region of aligned number 26–32, which corresponds to the residue number 21–27 of the ELG sequence. This region corresponds to the β strand A in the native structure and is a possible candidate of the region to assume a nonnative α helix in the molten globule intermediate. This region is also adjacent to the consensus sequence Gly-X-Trp (aligned number 22–24) of the lipocalin superfamily, and the amino acid sequence of this region is relatively conserved at least for BLG, ELG, and SRBP. Although the functional role of the Gly-X-Trp motif has not yet been clarified,[48] the sequences that are conserved due to a functional requirement may happen to have a high helical propensity. An alternative possibility is that the conserved sequence in this region is a requirement for proper folding and/or for maintaining the native structure. Recent site-directed mutagenesis studies have shown that the substitution of Trp 19 of BLG does not affect the retinol-binding property but significantly destabilizes the native state.[49,50] The site-directed mutagenesis of the corresponding Trp of other members of a calycin superfamily has been reported not to affect the ligand binding property.[51,52] These results suggest the structural importance of the Gly-X-Trp motif rather than its functional importance.

In the case of CRABP, however, the region with high helical probability is located at the aligned numbers 33–53, which assumes α helices in the native structure, and no other region shows high helical probability. Therefore, the nonnative α helices in the acid-unfolded CRABP can not be explained by the high helical propensity in the region of aligned number 26–32. Identification of the region to assume the nonnative α helix in the molten globule intermediate is essential to clarify whether the nonnative secondary structures in the molten globule state play an important role in the folding of proteins. The stable molten globule state of ELG found in this study is useful to identify the region to assume nonnative α helices in the molten globule intermediate of β-lactoglobulin.

CONCLUSIONS

All of the characteristics of the acid-unfolded ELG obtained in this study show that ELG assumes the molten globule state at acidic pH. An important feature of the molten globule state of ELG is that it contains nonnative α helices. This result, together with the results obtained by other researchers, indicate that the molten globule intermediate with nonnative α helices generally accumulates during the folding process of proteins included in a calycin superfamily.

ACKNOWLEDGMENTS

We thank Prof. Shinji Takai of the Department of Veterinary, Kitasato University, who kindly supplied the equine milk; Dr. Stephen W. Provencher (Max-Plank Institut) for providing M.I. the CONTIN program; Dr. Avijit Chakrabartty and Dr. Carol Rohl (Stanford University) for providing M.I. the HELIX program; Dr. Lindsay Sawyer and Dr. J. Morais Cabral (Edinburgh University) for providing M.I. the coordinates of bovine β-lactoglobulin; Dr. K. Kuwajima (University of Tokyo) for his valuable comments to the manuscript; and Shuichiro Gohda for collaboration during the protein purification. This work was supported in part by a grant (RG-331/93M) from the International Human Frontier Science Program (HFSP) Organization.

REFERENCES

1. Kuwajima, K. The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. Proteins 6:87–103, 1989.
2. Matthews, C.R. Pathways of protein folding. Annu. Rev. Biochem. 62:653–683, 1993.
3. Baldwin, R.L. The nature of protein folding pathways: The classical versus the new view. J. Biomol. NMR 5:103–109, 1995.
4. Ptitsyn, O.B. Molten globule and protein folding. Adv. Protein Chem. 47:83–229, 1995.
5. Roder, H. Watching protein folding unfold. Nature Struct. Biol. 2:817–820, 1995.
6. Miranker, A.D., Dobson, C.M. Collapse and cooperativity in protein folding. Curr. Opin. Struct. Biol. 6:31–42, 1996.
7. Fersht, A. Protein folding and stability: The pathway of folding of barnase. FEBS Lett. 325:5–16, 1993.
8. Dobson, C.M. Protein folding: Solid evidence for molten globules. Curr. Biol. 4:636–640, 1994.
9. Peng, Z.-Y., Wu, L.C., Schulman, B.A., Kim, P.S. Does the molten globule have a native-like tertiary fold? Phil. Trans. R. Soc. Lond. B 348:43–47, 1995.
10. Shortle, D. Structural analysis of non-native states of proteins by NMR methods. Curr. Opin. Struct. Biol. 6:24–30, 1996.
11. Pervaiz, S., Brew, K. Homology of β-lactoglobulin, serum retinol-binding protein, and protein HC. Science 228:335–337, 1985.
12. Papiz, M.Z., Sawyer, L., Eliopoulos, E.E., North, A.C.T., Findlay, J.B.C., Sivaprasadarao, R., Jones, T.A., Newcomer, M.E., Kraulis, P.J. The structure of β-lactoglobulin and its similarity to plasma retinol-binding protein. Nature 324:383–385, 1986.
13. Banaszak, L., Winter, N., Xu, Z., Bernlohr, D.A., Cowan, S., Jones, T.A. Lipid-binding proteins: A family of fatty acid and retinoid transport proteins. Adv. Protein Chem. 45:89–151, 1994.
14. Flower, D.R., North, A.T.C., Attwood, T.K. Structure and

sequence relationships in the lipocalins and related proteins. Prot. Sci. 2:753–761, 1993.
15. Futterman, S., Heller, J. The enhancement of fluorescence and the decreased susceptability to enzymatic oxidation of retinol complexed with bovine serum albumin, β-lactoglobulin, and the retinol-binding protein of human plasma. J. Biol. Chem. 247:5168–5172, 1972.
16. Kuwajima, K., Yamaya, H., Miwa, S., Sugai, S., Nagamura, T. Rapid formation of secondary structure framework in protein folding studied by stopped-flow circular dichroism. FEBS Lett. 221:115–118, 1987.
17. Kuwajima, K. Circular dichroism. Methods Mol. Biol. 40:115–135, 1995.
18. Monaco, H.L., Zanotti, G., Spadon, P., Bolognesi, M., Sawyer, L., Eliopoulos, E.E. Crystal structure of the trigonal form of bovine β-lactoglobulin and its complex with retinol at 2.5 Å resolution. J. Mol. Biol. 197:695–706, 1987.
19. Shiraki, K., Nishikawa, K., Goto, Y. Trifluoroethanol-induced stabilization of the α-helical structure of β-lactoglobulin: Implication for non-hierarchical protein folding. J. Mol. Biol. 245:180–194, 1995.
20. Hamada, D., Kuroda, Y., Tanaka, T., Goto, Y. High helical propensity of the peptide fragments derived from β-lactoglobulin, a predominantly beta-sheet protein. J. Mol. Biol. 254:737–746, 1995.
21. Liu, Z.-P., Rizo, J., Gierasch, L. M. Equilibrium folding studies of cellular retinoic acid binding protein, a predominantly beta-sheet protein. Biochemistry 33:134–142, 1994.
22. Bychkova, V.E., Berni, R., Rossi, G.L., Kutyshenko, V.P., Ptitsyn, O.B. Retinol-binding protein is in the molten globule state at low pH . Biochemistry 31:7566–7571, 1992.
23. Muccio, D.D., Waterhous, D.V., Fish, F., Brouillette, C.G. Analysis of the thermal unfolding of serum retinol binding protein containing a single retinol ligand. Biochemistry 31:5560–5567, 1992.
24. Conti, A., Godovac-Zimmermann, J., Liberatori, J., Braunitzer, G. The primary structure of monomeric β-lactoglobulin I from horse colostrum (*Equus caballus*, Perissodactyla). Hoppe-Seyler's Z. Physiol. Chem. 365:1393–1401, 1984.
25. McKenzie, H.A. Milk proteins. Adv. Prot. Chem. 22:55–234, 1967.
26. McKenzie, H.A., Ralston, G.B., Shaw, D.C. Location of sulfhydryl and disulfide groups in bovine β-lactoglobulin and effects of urea. Biochemistry 11:4539–4547, 1972.
27. Phelan, P., Malthouse, J.P.G. C-13 n.m.r. of the cyanylated β-lactoglobulins: Evidence that cys-121 provides the thiol group of β-lactoglobulins A and B. Biochem. J. 302:511–516, 1994.
28. Gohda, S., Shimizu, A., Ikeguchi, M., Sugai, S. The superreactive disulfide bonds in α-lactalbumin and lysozyme. J. Prot. Chem. 14:731–737, 1995.
29. Gill, S.C., von Hippel, P.H. Calculation of protein extinction coefficients from amino acid sequence data. Anal. Biochem. 182:319–326, 1989.
30. Smith, M.H. Molecular weights of proteins and some other materials including sedimentation, diffusion and frictional coefficients and partial specific volume. In "Handbook of Biochemistry." Sober, H.A. (eds.). Cleveland: CRC Press, 1970:C3–C35.
31. Corbett, R.J.T., Roche, R.S. Use of high-speed size-exclusion chromatography for the study of protein folding and stability. Biochemistry 23:1888–1894, 1984.
32. Gallagher, W.H., Woodward, C.K. The concentration dependence of the diffusion coefficient for bovine pancreatic trypsin inhibitor: A dynamic light scattering study of a small protein. Biopolymers 28:2001–2024, 1989.
33. Uversky, V.N. Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule. Biochemistry 32: 13288–13298, 1993.
34. Cohen, E.J., Edsall, J.T. "Proteins, Amino Acids and Peptides." New York: Hafner, 1965.
35. Goldberg, R.J. Sedimentation in the ultracentrifuge. J. Phys. Chem. 57:194–202, 1953.
36. Muramatsu, N., Minton, A.P. An automated method for rapid determination of diffusion coefficients via measurements of boundary spreading. Anal. Biochem. 168:345–351, 1988.
37. Provencher, S.W., Glöckner, J. Estimation of globular protein secondary structure from circular dichroism. Biochemistry 20:33–37, 1981.
38. Provencher, S.W. CONTIN: A general purpose constrained regularization program for inventing noisy linear algebraic and integral equations. Comput. Phys. Commun. 27:229–242, 1982.
39. Dolgikh, D.A., Gilmanshin, R.I., Brazhnikov, E.V., Bychkova, V.E., Semisotnov, G.V., Venyaminov, S. Yu., Ptitsyn, O.B. α-Lactalbumin: Compact state with fluctuating tertiary structure? FEBS Lett. 136:311–315, 1981.
40. Kim, P.S., Baldwin, R.L. Intermediates in the folding reactions of small proteins. Annu. Rev. Biochem. 59:631–660, 1990.
41. Ptitsyn, O.B. The molten globule state. In "Protein Folding." Creighton, T.E. (eds.). New York: W.H. Freeman, 1992:243–300.
42. Redfield, C., Smith, R.A.G., Dobson, C.M. Structural characterization of a highly-ordered "molten globule" at low pH. Nature Struct. Biol. 1:23–29, 1994.
43. Ferrer, M., Barany, G., Woodward, C. Partially folded, molten globule and molten coil states of bovine pancreatic trypsin inhibitor. Nature Struct. Biol. 2:211–217, 1995.
44. Morozova, L.A., Haynie, D.T., Arico-Muendel, C., Van Dael, H., Dobson, C.M. Structural basis of the stability of a lysozyme molten globule. Nature Struct. Biol. 2:871–875, 1995.
45. Chobert, J.M., Briand, L., Grinberg, V., Haertle, T. Impact of esterification on the folding and the susceptibility to peptic proteolysis of β-lactoglobulin. Biochim. Biophys. Acta 1248:170–176, 1995.
46. Doig, A.J., Chakrabartty, A., Klingler, T.M., Baldwin, R.L. Determination of free energies of n-capping in α-helices by modification of the Lifson-Roig helix-coil theory to include N- and C-capping. Biochemistry 33:3396–3403, 1994.
47. Chakrabartty, A., Kortemme, T., Baldwin, R.L. Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. Prot. Sci. 3:843–852, 1994.
48. Godovac-Zimmermann, J. The structural motif of β-lactoglobulin and retinol-binding protein: a basic framework for binding and transport of small hydrophobic molecules? Trends Biochem. Sci. 13:64–66, 1988.
49. Katakura, Y., Totsuka, M., Ametani, A., Kaminogawa, S. Tryptophan-19 of β-lactoglobulin, the only residue completely conserved in the lipocalin superfamily, is not essential for binding retinol, but relevant to stabilizing bound retinol and maintaining its structure. Biochim. Biophys. Acta 1207:58–67, 1994.
50. Cho, Y., Batt, C.A., Sawyer, L. Probing the retinol-binding site of bovine β-lactoglobulin. J. Biol. Chem. 269:11102–11107, 1994.
51. Locke, B.C., MacInnis, J.M., Qian, S.J., Gordon, J.I., Li, E., Fleming, G.R., Yang, N.C. Fluorescence studies of rat cellular retinol binding protein-II produced in *E. coli*: An analysis of four tryptophan substitution mutants. Biochemistry 31:2376–2383, 1992.
52. Clark, P.L., Liu, Z.P., Zhang, J.H., Gierasch, L.M. Intrinsic tryptophans of CRABPI as probes of structure and folding. Prot. Sci. 5:1108–1117, 1996.
53. Shimizu, A., Ikeguchi, M., Sugai, S. Appropriateness of DSS and TSP as internal references for ^{1}H NMR studies of molten globule proteins in aqueous media. J. Biomol. NMR 4:859–862, 1994.
54. Cowan, S.W., Newcomer, M.E., Jones, T.A. Crystallographic refinement of human serum retinol binding protein at 2 Å resolution. Proteins 8:44–61, 1990.
55. Thompson, J.R., Bratt, J.M., Banaszak, L.J. Crystal structure of cellular retinoic acid binding protein I shows increased access to the binding cavity due to formation of an intermolecular β-sheet. J. Mol. Biol. 252:433–446, 1995.