Anal. Chem. **2002,** *74,* 6031−6039

Detection and Resolution of Intermediate Species in Protein Folding Processes Using Fluorescence and Circular Dichroism Spectroscopies and Multivariate Curve Resolution

Susana Navea, Anna de Juan, and Romà Tauler*

Department of Analytical Chemistry, University of Barcelona, Diagonal 647, Barcelona 08028, Spain

Thermally induced protein unfolding/folding processes have been studied on α-lactalbumin and α-apolactalbumin. Experiments monitored by fluorescence and circular dichroism spectroscopic techniques on α-apolactalbumin showed the formation of an intermediate species, whereas in the case of α-lactalbumin, this intermediate species was not detected. The presence and resolution of this intermediate species, its spectrum, and the evolution of all conformations during protein unfolding/folding processes were estimated using the multivariate curve resolution-alternating least-squares method. Elucidation of the nature and contribution of the different secondary structure motifs in each of the resolved protein conformations, including the intermediate, was also carried out. Multivariate resolution has shown to be an excellent tool for the complete characterization of all protein conformations involved in folding processes, including intermediate species that cannot be isolated by physical or chemical means. Indeed, it is in the determination and modeling of these intermediates that this chemometric approach outperforms in power and reliability previous methodologies based on simpler measurements and data treatments and fills the void linked to the elucidation and interpretation of complex mechanisms in protein folding processes.

Proteins participate in essential processes and structures of organisms. To be biologically active, proteins must adopt specific folded three-dimensional structures; hence, the relevance of protein folding studies.[1,2] The structure of the proteins is organized in four different levels,[3] namely: (1) *primary structure*, sequence of amino acids in the polypeptide chains; (2) *secondary structure*, regular spatial arrangements of the backbone of the polypeptide chain stabilized by hydrogen bonds that give rise to helical or flat sheet arrangements; (3) *tertiary structure*, spatial arrangement of the secondary structure motifs within a polypeptide chain, responsible for the globular or fibrillar nature of proteins; and (4) *quaternary structure*, union of several polypeptide chains by weak forces or disulfide bridges. All biologically active proteins present their secondary and tertiary structures ordered. When a protein is completely unfolded, these two structural levels are unordered.

Unfolding transitions of single proteins are often one-step processes, i.e., only the folded or native (N) and the unfolded (U) states are detected. However, a partially folded intermediate, called molten globule state,[4] may be detected as a third protein conformation in some proteins under certain experimental conditions. The "*molten globule*" state has been shown to exhibit organized secondary structure motifs and no nativelike tertiary structure.[5,6] Thus, the process of protein folding or unfolding can follow two mechanisms: (a) one-step process N ⇔ U; or (b) two-step process, N ⇔ intermediate ("molten globule") (I) ⇔ U. The attempt to characterize protein folding pathways has been largely a search for folding intermediates.[2] However, the detection and characterization of this type of species is not easy because either the lifetime of these transient intermediates is frequently too short to be detected or it is not possible to separate and isolate them from other protein conformations simultaneously present.

The mechanism and the identity of protein conformations involved in a protein folding process can be studied by monitoring spectrometrically changes on the tertiary and secondary structure. In this work, fluorescence and far- and near-UV circular dichroism have been used to study protein folding processes. Circular dichroism measurements are specially suitable because they are very sensitive to protein conformation changes and show two specific wavelength regions (far- and near-UV) to monitor changes in the secondary and in the tertiary structure,[7] respectively. Fluorescence is sensitive to changes in the environment around the tryptophan aromatic system, which is more solvent-exposed in unfolded states[8] and mainly reveals changes in the tertiary structure.

* To whom correspondence should be addressed. E-mail: roma@apolo.qui.ub.es.

(1) Anfinsen, C. B. *Science* **1973**, *181*, 223−230.
(2) Pain, R. H., Ed. *Mechanisms of Protein Folding*; Oxford University Press: New York, 1994.
(3) Schulz, G. E.; Schirmer, R. H. *Principles of Protein Structure*; Springer-Verlag Inc: New York, 1979.
(4) Creigthon, T. E. *Proteins: Structures and Molecular Properties*, 2nd ed.; W. H. Freeman and Co.: New York, 1997.
(5) Dolgikh, D. A.; Gilmanshin, R. I.; Brazhnikov, E. V.; Bychkova, V. E.; Semisotnov, G. V.; Venyaminov, S. Y.; Ptitsyn, O. B. *FEBS Lett.* **1981**, *136*, 311−315.
(6) Kuwajima, K. *Proteins* **1989**, *6* (2), 87−103.
(7) Berova, N.; Nakanishi, K.; Woody, R. W. *Circular Dichroism. Principles and Applications*; Wiley: New York, 2000.
(8) Mills, O. E. *Biochim. Biophys. Acta* **1976**, *434*, 324−332.

10.1021/ac025914d CCC: $22.00
Published on Web 10/25/2002

Protein folding processes have been classically studied through examination of melting curves, i.e., monitoring the evolution of one-wavelength response values (e.g., absorbances or ellipticities) as a function of the variable controlling the folding process (temperature, pH, or concentration of denaturant agent[9,10]). In this work, multiwavelength spectra are recorded at each stage of the protein folding process and all the information collected is analyzed using multivariate data analysis.

Multivariate curve resolution-alternating least squares (MCR-ALS) is a powerful chemometric method allowing mathematical resolution of concentration and spectra profiles of pure chemical species in mixtures and in complex chemical evolving processes.[11−14] MCR-ALS is especially appropriate and powerful in the study of processes where physical or chemical isolation of all species is not possible, as it is the case in the detection and resolution of protein folding intermediate species. The resolved concentration profiles and spectra in protein folding do provide information about the mechanism of the process and about the structural identity of the protein conformations involved.

Thermal-induced (melting) protein folding processes of α-lactalbumin and α-apolactalbumin have been monitored spectroscopically using fluorescence and circular dichroism. Both compounds are globular proteins with the same amino acid sequence formed by 123 residues and a very similar secondary structure, predominantly α-helix with an antiparallel β-sheet.[15] Both proteins differ in that α-lactalbumin is a metalloprotein with a Ca^{2+} atom, whereas α-apolactalbumin lacks this metal ion in its structure. The combined use of full spectral information and MCR-ALS has allowed for the detection and complete modeling of an intermediate species in the α-apolactalbumin protein folding process. The results obtained have been useful to address other chemical issues, such as the influence on the evolution of protein folding processes due to the presence of Ca^{2+} in their structure and due to the effect of variations in the salt concentration of the medium.

EXPERIMENTAL SECTION

Chemicals. α-Lactalbumin and α-apolactalbumin from bovine milk were purchased from Sigma and were used without further purification. Sodium chloride (A.R.) was purchased from Merck. The protein was dissolved in bidistilled water and adjusted to the suitable salt concentration when needed. Protein solutions of 0.1 mg/mL were prepared to record fluorescence and far-UV circular dichroism spectra. For near-UV circular dichroism spectra, protein solutions of 1 mg/mL were used. All solutions were 0.15 M in NaCl (physiological conditions), except those used to estimate the salt effect (see below).

Instrumentation. CD spectra were obtained using a Jasco spectropolarimeter (model J-720). Far- and near-UV CD spectra covered 195−250 and 250−330 nm, respectively, and ellipticity readings were recorded every 1 nm. A 10-mm-path length closed quartz cell was used. Fluorescence emission spectra were recorded with an Aminco Bowman AB-2 fluorometer (λ_{ex} = 285 nm). Emission spectra were measured between 300 and 430 nm, and fluorescence intensities were recorded every 1 nm. A 10-mm-path length closed quartz cell was used. Both instruments were equipped with a cell holder thermostated by a water circulation bath for controlling temperature of protein solutions. Data acquisition and storage were carried out using the specific software of each instrument.

Experimental Procedure. The temperature was modified between 5 and 85 °C, increasing in protein unfolding experiments and decreasing in protein folding experiments. Temperature intervals among consecutive spectra ranged from 2 to 5 °C, depending on the spectral variation. A 2-min waiting period was used before recording each spectrum. This period of time is established to stabilize the protein solution and is fixed to avoid the presence of uncontrolled time-dependent effects on the evolution of the monitored processes. In this way, protein conformations at equilibrium can be detected.[2] No information about transient kinetic intermediates is expected to be obtained because they usually evolve in a too fast (millisecond) time scale.[16]

Data Treatment. Protein folding has been studied by recording a complete spectrum at each stage of the monitored process. The spectra recorded in a thermal-dependent protein folding process are organized in a data matrix **D**, whose rows are the spectra recorded at each temperature and whose columns represent the melting curves (absorbance versus temperature profiles) for each wavelength (see Figure 1).

Mixture analysis and resolution of complex chemical evolving processes are often tackled using multivariate resolution techniques.[17] In this work, the MCR-ALS has been selected because of the versatility to adapt to very diverse chemical data sets, the flexibility in the application of any constraint, and the possibility to handle either one data matrix (two-way data sets) or several data matrices together (three-way data sets).[11,12] MCR-ALS decomposes a data matrix **D** into the product of two small matrices, **C** and $\mathbf{S}^T$.

$$\mathbf{D} = \mathbf{C}\mathbf{S}^T + \mathbf{E} \qquad (1)$$

The columns in matrix **C** are the concentration profiles of the pure species detected during the experiment and the rows in $\mathbf{S}^T$ are their related pure spectra. Matrix **E** describes the experimental error.

The general steps of MCR-ALS are enumerated below: (1) determination of the number of species independently contributing to the measured signal in matrix **D** (using singular value decomposition (SVD)[18]); (2) building initial estimates of the **C** matrix (e.g., using evolving factor analysis, EFA[19]) or of the $\mathbf{S}^T$ matrix (e.g., using methods to select the purest spectra in **D**[20]); (3) given **D** an **C**, constrained least-squares calculation (see below) of $\mathbf{S}^T$; (4) given **D** an $\mathbf{S}^T$, constrained least-squares calculation (see below) of **C**; (5) go to step 3 until convergence is achieved.

(9) Manderson, G. A.; Creamer, L. K.; Hardman, M. J. *J. Agric. Food Chem.* **1999**, *47*, 4557−4567.
(10) Creamer, L. K. *Biochemistry* **1995**, *34*, 7170−7176.
(11) Tauler, R.; Smilde, A. K.; Kowalski, B. R. *J. Chemom.* **1995**, *9*, 31−58.
(12) Tauler, R. *Chemom. Intell. Lab. Syst.* **1995**, *30*, 133−146.
(13) Mendieta, J.; Díaz-Cruz, M. S.; Esteban, M.; Tauler, R. *Biophys. J.* **1998**, *74*, 2876−2888.
(14) Navea, S.; de Juan, A.; Tauler, R. *Anal. Chim. Acta* **2001**, *446*, 187−197.
(15) Chrysina, E. D.; Brew, K.; Acharya, K. R. *J. Biol. Chem.* **2000**, *275* (47), 37021−37029.
(16) Nölting, B. *Protein Folding Kinetics*; Springer-Verlag: Berlin, 1999.
(17) de Juan, A.; Casassas, E.; Tauler, R. In *Encyclopedia of Analytical Chemistry: Applications, Theory, and Instrumentation*; Meyers, R. A., Ed.; John Wiley & Sons Ltd.: Chichester, U.K., 2000; pp 9800−9837.
(18) Golub, G. H.; Reinsch, C. *Numer. Math.* **1970**, *14*, 403−420.
(19) Maeder, M. *Anal. Chem.* **1987**, *59*, 527−530.
(20) Windig, W.; Guilment, J. *Anal. Chem.* **1991**, *63*, 1425−1432.

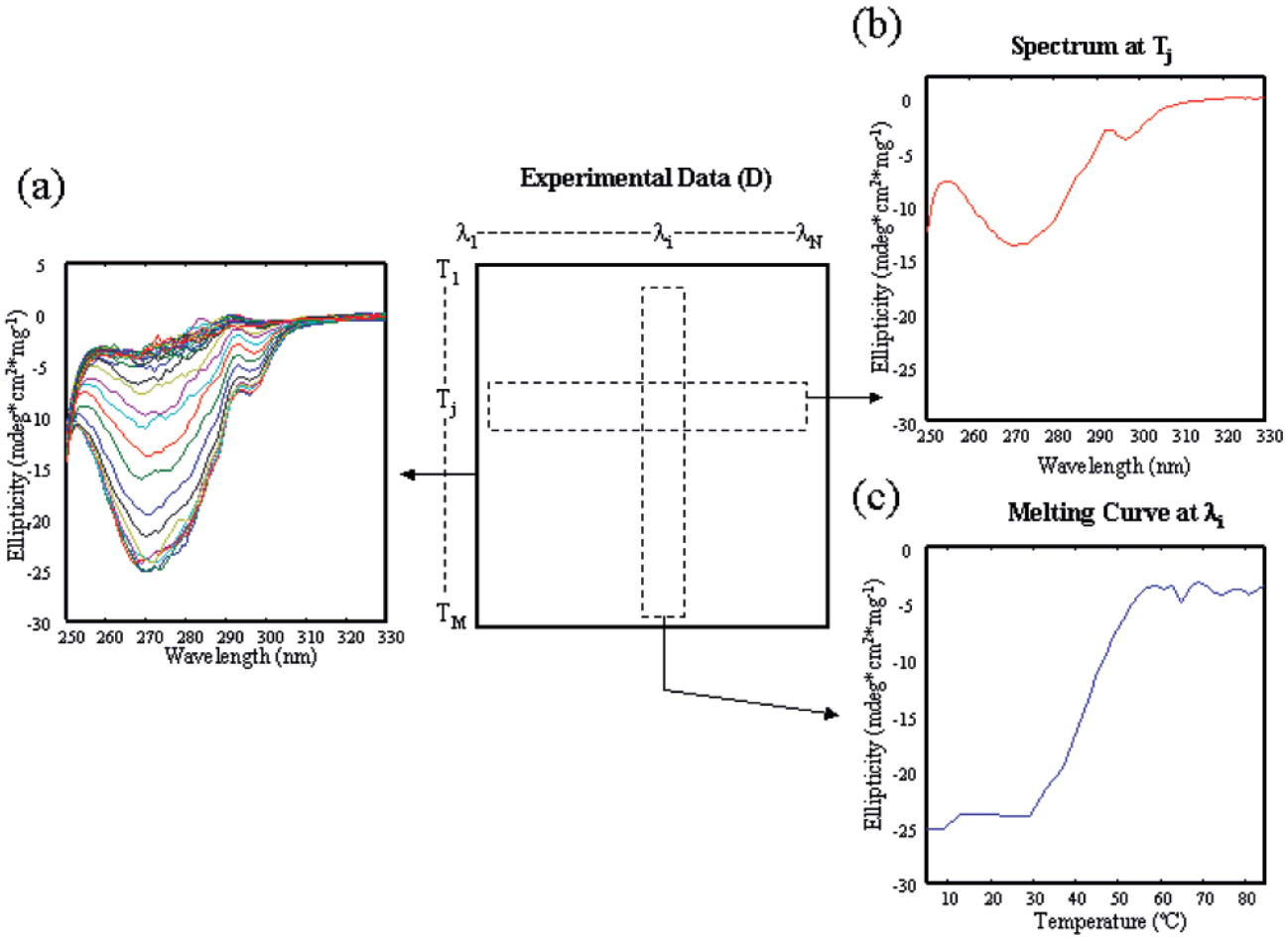

Figure 1. Measurements used in protein folding monitoring. Thick line: multivariate approach. (a) One spectrum per process stage. Dashed line: classical univariate approaches. (b) Spectrum at one single temperature. (c) Melting curve at one single wavelength (λ).

Note that this sequence is valid when a **C**-type initial estimate is used; working with $\mathbf{S}^T$-type initial estimates, the order of steps 3 and 4 is reversed. Constraining concentration profiles in **C**, the spectra in $\mathbf{S}^T$, or both is crucial to drive the iterative process to a satisfactory solution. A constraint can be defined as any systematic feature fulfilled by one or more profiles in the data set. Among the most commonly used, nonnegativity, unimodality, and closure are related to chemical properties of the data set, whereas the detection of selective zones and regions of the absence/presence of species often requires the use of complementary mathematical tools to be set.[11,12,21−23]

In protein folding, MCR-ALS can be applied to the process monitored using only one spectroscopic technique or to a row-wise augmented matrix, derived from appending spectroscopic measurements from different techniques monitoring the same process (see Figure 2). The treatment of several spectroscopic measurements simultaneously improves the quality of the obtained solution because more information is included in the analysis. The ambiguities related to the use of multivariate resolution methods are drecreased.[11]

All the chemometric methods employed in this work are found in MATLAB[24] programs developed by our investigation group, available at the WEB page: http:/www.ub.es/gesq/eq1_eng.htm.

To display more clearly the trends of the resolved concentration profiles, plots of both the actual ALS-resolved elements in the profiles and the best nonlinear least-squares fitted lines using LSQNONLIN, LSQCURVEFIT MATLAB, or both functions are included in Figures 3−6 of the Results section (see Figures 3−6).

RESULTS AND DISCUSSION

The application of the MCR-ALS method to diverse experimental spectroscopic data arrangements has answered different questions related to protein folding and will be commented on in detail in the next sections (see Table 1).

Resolution of Changes in the Protein Secondary Structure. The evolution of the secondary structure of α-lactalbumin and α-apolactalbumin was studied by far-UV (190−250 nm) circular dichroism (see Figure 2a for data arrangement). MCR-ALS-resolved concentration profiles in **C** were constrained to be nonnegative and unimodal (i.e., the evolution of each protein conformation can be appropriately represented by an emergence-decay profile having a single peak maximum). Since the total

(21) de Juan, A.; Vander Heyden, Y.; Tauler, R.; Massart, D. L. *Anal. Chim. Acta* **1997**, *346*, 307−318.
(22) Bro, R.; de Jong, S. *J. Chemom.* **1997**, *11*, 393−401.
(23) Bro, R.; Sidiropoulos, N. D. *J. Chemom.* **1998**, *12*, 223−247.
(24) The Mathworks Inc, Natick, MA, version 6, http://www.mathworks.com.

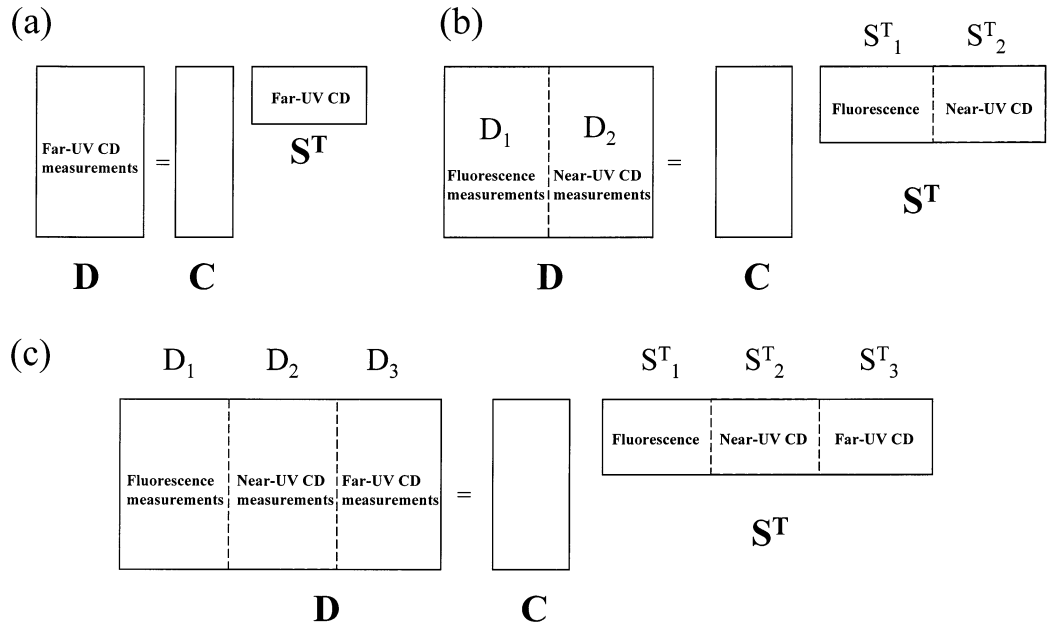

Figure 2. Different forms to organize the information from the multivariate monitoring of protein folding. (a) Single matrix for the study of one process with one technique (one experiment). (b) Row-wise augmented matrix for the study of one process with several techniques (several experiments).

Table 1. Description of the Experiments Related to Protein Folding Processes and Results Obtained by MCR-ALS

			α-lactalbumin		α-apolactalbumin	
information sought	techniques	process	T_{sec}[a]	T_{tert}[b]	T_{sec}[a]	T_{tert}[b]
resolution of changes in the protein secondary structure	far-UV CD	unfolding	60.10		59.70	
		folding	63.41		57.05	
resolution of changes in the protein tertiary structure	near-UV CD and fluorescence	unfolding		62.47		42.73
		folding		58.78		39.30
detection of intermediates	near- and far-UV CD and fluorescence	unfolding	61.80		58.80	44.00
		folding	58.72		58.10	42.80
salt concn effect	near- and far-UV CD	unfolding (C_{salt} = 0 M in NaCl)			64.80	30.15
		unfolding (C_{salt} = 0.15 M in NaCl)			63.70	43.94

[a] Temperature (°C) at which 50% of secondary structure is formed (in folding process) or disordered (in unfolding process). [b] Temperature (°C) at which 50% of tertiary structure is formed (in folding process) or disordered (in unfolding process).

concentration of protein remains constant along the whole unfolding/folding process, the condition of a closed system was also applied as an additional constraint in the MCR-ALS optimization. Selectivity constraints were applied at the lowest temperatures, where only native conformation was supposed to be present. In contrast, spectra in $\mathbf{S^T}$ were not forced to obey any particular constraint. Resolved concentration and spectra profiles for α-lactalbumin and α-apolactalbumin are shown respectively in Figure 3a and b. Analogous results were obtained for the folding process. In both cases, only two contributions were resolved from the thermal unfolding or folding process, the native and the unfolded species.

The resolved dichroism spectra associated with both native and unfolded conformations in both proteins have very similar shapes. The native conformations show the typical spectral features associated with a major contribution of the α-helix motif in their secondary structure, i.e., an intense negative band with two shoulders located around 220 and 210 nm[7,25] The resolved spectra for the unfolded species show the typical spectral features linked to a random coil motif (a sharper negative band at short wavelengths and weaker features around 220 nm).[7,25]

The MCR-ALS-resolved concentration profiles show the evolution of the concentration of the different protein conformations in the process. The crossing point in these concentration profiles plot gives the temperature at which 50% of the native protein has lost its initial secondary structure (T_{sec}) (see Table 1). As can be seen, the evolution of the secondary structure in α-lactalbumin and α-apolactalbumin along protein folding is very similar. The evolution of the concentration profiles and the temperature crossing point between folded and unfolded conformations are rather similar. These facts seem to indicate that the chemical differences between both proteins do not affect significantly the formation or destruction (in folding and unfolding processes, respectively) of the secondary structure. The differences observed

(25) Greenfield, N. J. *Anal. Biochem.* **1996**, *235*, 1−10.

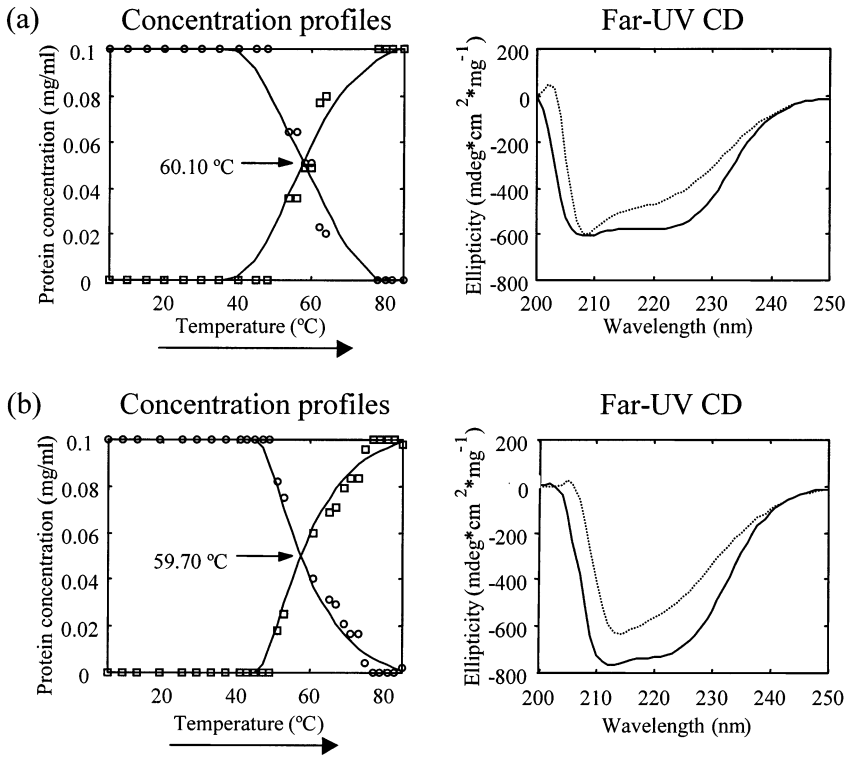

Figure 3. Pure concentration profiles and spectra obtained by MCR-ALS on data sets monitoring changes on the protein secondary structure. (a) Results for α-lactalbumin. (b) Results for α-apolactalbumin. In (a) and (b), pure spectra (solid line, native conformation; dashed line, unfolded conformation); experimental points in concentration profiles (○ native conformation and □ unfolded conformation), solid lines (nonlinear fitted profiles obtained from the ALS-resolved estimation points).

in Figure 3 between experimental points and the fitted concentration profiles are due to the low signal-to-noise ratios associated with experimental far-UV CD measurements.

Resolution of Changes in the Protein Tertiary Structure. Changes in tertiary structure along protein folding have been studied using fluorescence and near-UV (250−330 nm) circular dichroism. Analysis of the row-wise augmented matrix formed by the two data sets coming from each of these two techniques has been carried out (see data arrangement in Figure 2b). Fluorescence spectra have been normalized to suppress random intensity changes unrelated to the monitored process. Data matrices within the row-wise augmented data set have been conveniently scaled, i.e., multiplied by a certain factor, to balance the different scale units of the combined techniques. Concentration profiles in **C** have been constrained as in the study of changes in the secondary structure because the monitored process is the same. In contrast, fluorescence spectra profiles in the augmented $\mathbf{S}^{\mathbf{T}}$ matrix have been constrained to be nonnegative, whereas near-UV CD spectra are left unconstrained. Figure 4 shows the MCR-ALS-resolved concentration profiles and spectra related to the different tertiary structures present during the unfolding of α-lactalbumin and α-apolactalbumin. The resolved spectra have been rescaled by taking into account the concentrations of the protein solutions used for the different spectroscopic techniques and the scaling factors applied to append each matrix in the augmented data set.

There is a big similarity in the spectra associated with the folded and the unfolded conformations of both proteins; i.e., despite the chemical difference between both proteins, the native tertiary structure acquired in both cases is practically identical. The near-UV dichroism spectra of the folded protein conformations show a very intense negative band,[7] and the fluorescence spectra present a band with a maximum intensity at 340 nm due to the presence of tryptophan amino acid residues.[8] The spectrum of the denatured/unfolded species shows a substantial decrease in the intensity of the negative band showing a nearly flat signal in the near-UV dichroism spectrum. A red shift of the maximum in the fluorescence band can be also noticed.

In contrast, the evolution of the resolved concentration profiles related to the protein folding of the two proteins is clearly different (Figure 4). Now, the crossing point in the concentration profile plot represents the temperature at which 50% of the native protein has lost its initial tertiary structure (T_{tert}). Whereas the crossing point of concentration profiles takes place at T_{tert} = 62.47 °C for α-lactalbumin, this temperature drops to 42.73 °C for α-apolactalbumin, ∼20 °C difference. Bearing in mind that the experiments performed in both proteins have been carried out at exactly the same experimental conditions, the calcium atom present in α-lactalbumin appears to produce a large stabilizing role of the folded tertiary structure of the protein.[26] Indeed, it is in the α-apolactalbumin, where this metal atom is absent, that the native tertiary structure starts to disappear at much lower temperatures. Within the same protein, results obtained in the study of folding and unfolding processes did not differ significantly (see Table 1).

Detection of Intermediates. A preliminary check to detect the presence of an intermediate in a protein folding process can be carried out by comparing the temperatures at which the secondary structure (T_{sec}) and the tertiary structure (T_{tert}) of the folded conformation are half-formed. If both coincide, then the

(26) Griko, Y. V.; Remeta, D. P. *Protein Sci.* **1999**, *8*, 554−561.

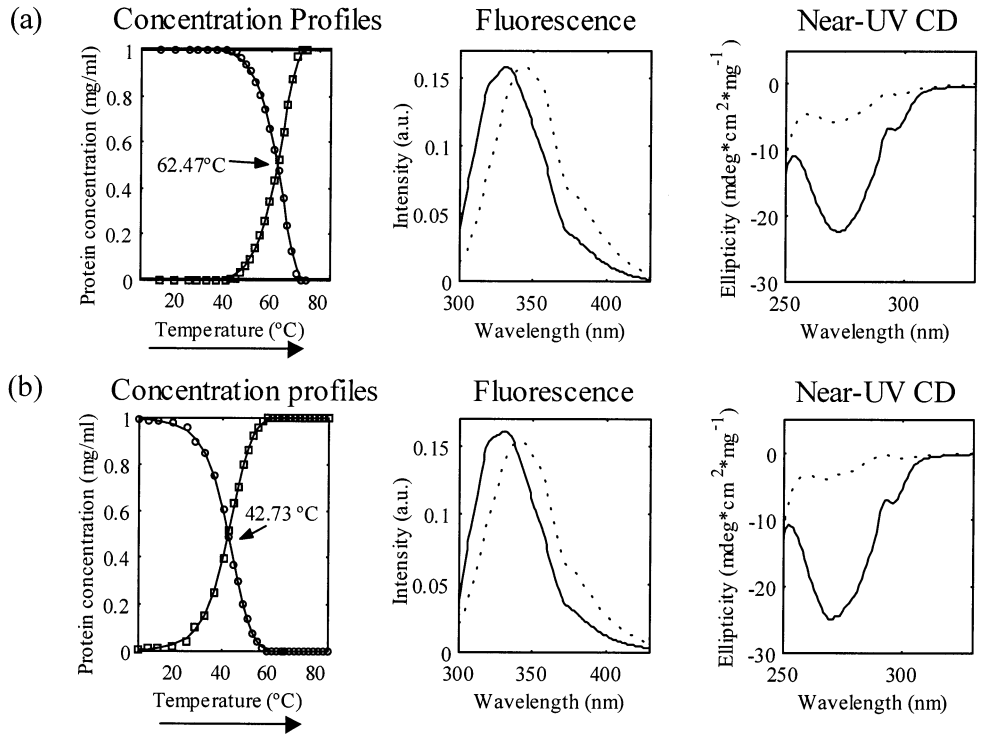

Figure 4. Pure concentration profiles and spectra obtained by MCR-ALS on data sets monitoring changes on the protein tertiary structure. (a) Results for α-lactalbumin. (b) Results for α-apolactalbumin. In (a) and (b), pure spectra (solid line, native conformation; dashed line, unfolded conformation); experimental points in concentration profiles (○ native conformation and □ unfolded conformation), solid lines (nonlinear fitted profiles obtained from the ALS-resolved estimation points).

protein loses the tertiary and the secondary structures at the same time and only a native conformation (with secondary and tertiary structures ordered) and an unfolded conformation (with both structural levels unordered) are required to describe the process. Instead, if significant differences are observed in the crossing points of concentration profiles, a new third species (an intermediate with the secondary structure ordered and the tertiary unordered) may be needed to explain the different shifts in the loss of the tertiary and secondary structures. As noted in Table 1, the shift between T_{sec} and T_{tert} is much larger in the α-apolactalbumin (∼20 °C) than in the α-lactalbumin (∼5 °C). This fact would suggest the presence of an intermediate species in the α-apolactalbumin folding and a one-step mechanism in the folding of the α-lactalbumin.

In this work, we propose the application of the MCR-ALS method as a more accurate way to confirm these conclusions and to model the evolution of the intermediate species. For this purpose, the measurements related to the tertiary and secondary structures must be analyzed together. To do so, a row-wise augmented data matrix formed by the fluorescence, near- and far-UV CD measurements was used (see Figure 2c).The scaling whithin the augmented matrix and the constraints applied to concentration profiles and spectra are applied as in previous sections.

Figure 5 shows the concentration profiles and spectra resolved for the protein conformations present during the α-apolactalbumin and α-lactalbumin unfolding processes. Whereas only one folded and one unfolded protein conformation were necessary to describe the process for the α-lactalbumin (Figure 5a), one additional intermediate conformation was necessary to explain the protein folding process of α-apolactalbumin (Figure 5b). From the concentration profiles in Figure 5b, the temperatures T_{sec} (crossing point between the concentration profiles of native and intermediate conformations) and T_{tert} (crossing point between the concentration profiles between intermediate and unfolded conformations), indicators for the half-formation of secondary and tertiary protein structures, in two-step protein folding processes, can be now accurately determined. This is a remarkable and novel contribution because it allows the description of intermediate conformations, impossible to be physically or chemically isolated.

The rescaled resolved spectra obtained for the folded and unfolded conformations of α-apolactalbumin and α-lactalbumin resemble analogous spectra shown in Figures 3 and 4. The spectra obtained for the α-apolactalbumin intermediate presents an ordered secondary structure similar to the native folded protein and an unordered tertiary structure similar to the unfolded protein at high temperatures. This fact is especially noticeable in the near- and far-UV circular dichroism spectra. The near-UV circular dichroism spectrum has a practically flat signal like the spectrum related to the unfolded conformation. On the other hand, the far-UV circular dichroism spectrum keeps the typical spectral features present in the native protein conformation. These spectral features help to confirm the intermediate detected as a molten globule state because of the unordered tertiary structure and the almost completely ordered secondary structure.

The molten globule state detected in the α-apolactalbumin confirms the active role of the calcium atom in the protein folding process.[26,27] In the absence of this metal ion, the tertiary structure of the protein is much less stabilized. As a consequence, the whole mechanism of protein folding is modified in such a way that the

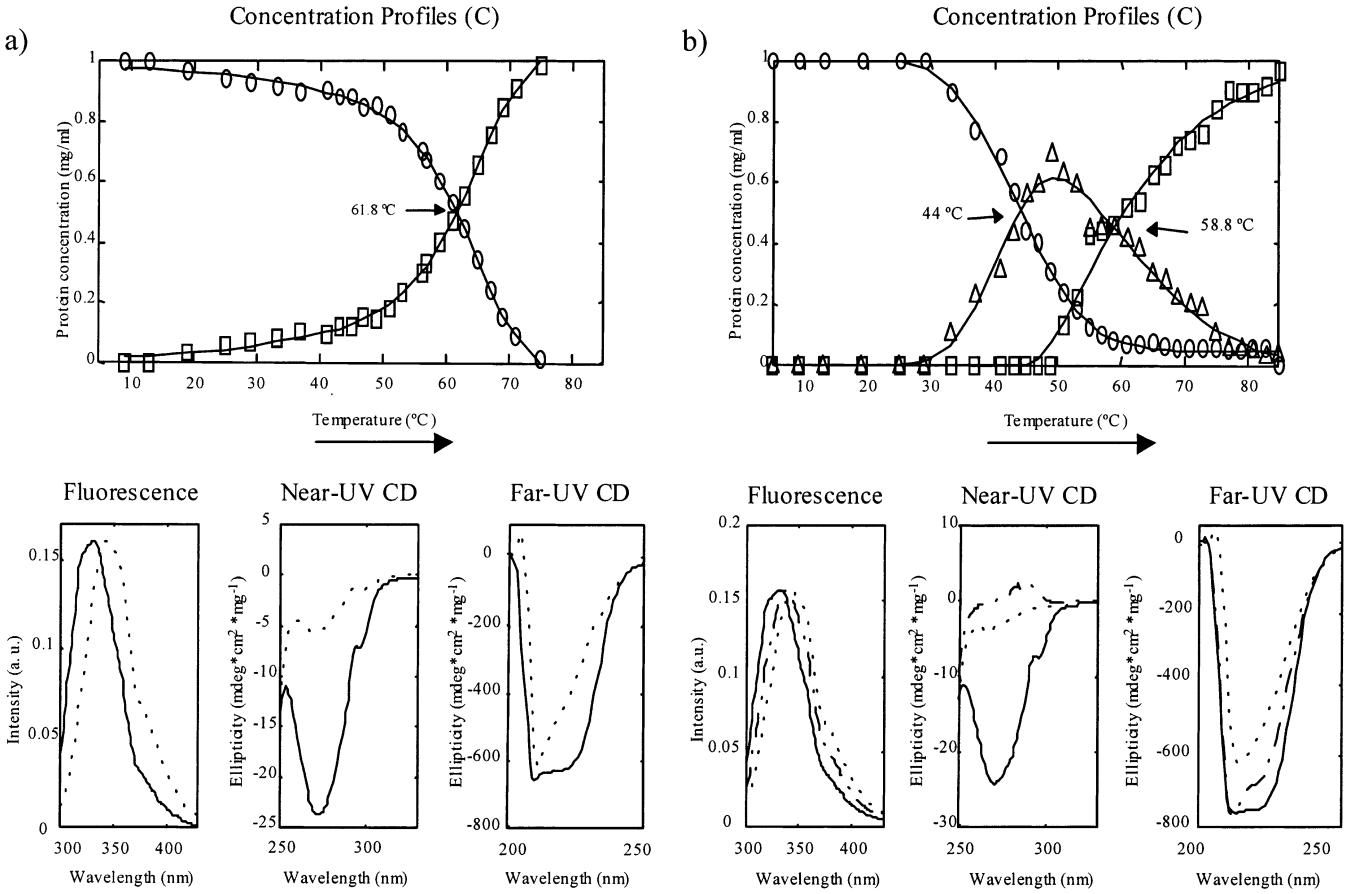

Figure 5. Pure concentration profiles and spectra obtained by MCR-ALS on data sets monitoring changes on both secondary and tertiary structure. (a) Results for α-lactalbumin. (b) Results for α-apolactalbumin. In (a) and (b), pure spectra (solid line, native conformation; dash−dot line, intermediate conformation; dashed line, unfolded conformation); experimental points in concentration profiles (○ native conformation, △ intermediate conformation, and □ unfolded conformation), solid lines (nonlinear fitted profiles obtained from the ALS-resolved estimation points).

presence of a new intermediate species is necessary in the transition between the completely ordered and unordered protein conformations.

Salt Concentration Effect. Given the proven influence of the calcium atom in the protein folding, salt concentration effects have also been studied, in particular the influence of sodium ion in the medium. For this purpose, experiments identical to those previously described in the absence of any additional salt concentration have been carried out with α-apolactalbumin. Since circular dichroism has been confirmed to be the most sensitive and discriminant technique to monitor the protein unfolding/folding processes, only the results obtained with this technique are used in this section.

In the absence of any NaCl salt concentration, the protein folding mechanism of α-apolactalbumin also takes place through an intermediate. The CD spectra obtained in all performed experiments are quite similar in shape (not shown). In contrast, concentration profiles showed noticeable differences in the temperatures indicated by the crossing points between the different protein conformations (see Figure 6). Whereas the formation of the protein secondary structure did not seem to be much affected by the concentration of sodium ions (T_{sec} does not vary significantly with the inclusion of Na^+ ions, 63.7 °C without and 64.8 °C with, respectively), the stabilization of the protein tertiary structure was clearly affected.[26] Indeed, T_{tert} was higher in the presence of Na^+ ions than in its absence (43.9 °C compared with 30.15 °C). Therefore, higher temperatures were required to break the tertiary structure of α-apolactalbumin in the presence of a high concentration of Na^+ ions.

It has been postulated that monovalent cations, such as Na^+, may be bound to the vacant calcium binding sites in the α-apolactalbumin.[28] However, the different charge and size of the sodium ion result in a lower binding affinity and the protein structure cannot achieve the same stability as the natural metalloprotein α-lactalbumin.[26,28−30] This would be the reason an intermediate species is needed to describe the protein folding process of α-apolactalbumin, no matter whether metal ions such as Na^+ may be present in the medium. In any case, the effect of Na^+ mimics that of Ca^{2+}, although in a lesser extent. Indeed, the drop in the temperature needed to break the tertiary structure and the wider range of existence of the intermediate protein conformation in solutions with no presence of Na^+ would confirm the role played also by this ion in the stabilization of the protein tertiary structure.

(27) Kuwajima, K.; Harushima, Y.; Sugai, S. *Int. J. Pept. Protein Res.* **1986**, *27* (1), 18−27.
(28) Hiraoka, Y.; Sugai, S. *Int. J. Pept. Protein Res.* **1985**, *26*, 252−261.
(29) Permyakov, E. A.; Morozova, L. A.; Burstein, E. A. *Biophys. Chem.* **1985**, *21*, 21−31.
(30) Kronman, M. *J. Crit. Rev. Biochem. Mol. Biol.* **1989**, *24*, 565−667.

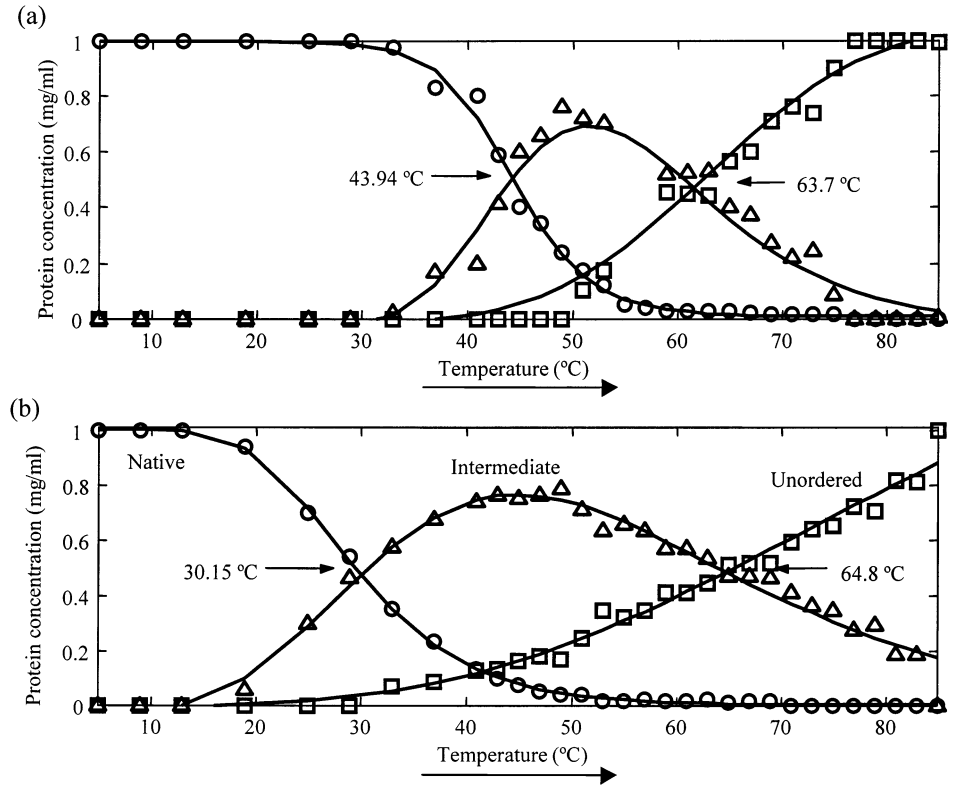

Figure 6. Effect of salt concentration in the protein unfolding of α-apolactalbumin. Representation of the pure concentration profiles (a) in the presence of C_{salt}= 0.15 M of NaCl and (b) in the absence of salt (NaCl). Experimental points in concentration profiles (○ native conformation, △ intermediate conformation, and □ unfolded conformation), solid lines (nonlinear fitted profiles obtained from the ALS-resolved estimation points).

Elucidation of the Secondary Structure of Protein Conformations. An alternative way to describe the structural changes in protein folding is displaying the evolution of the formation or disappearance of the different secondary structural motifs (helices, random coil, etc.) along the process. Classically, this information has been obtained from far-UV CD spectra measured at different stages of the process (e.g., temperatures).[9,10] Each of these raw spectra ($\mathbf{s}_i$) can be defined as the weighted sum of the spectral contributions related to the pure secondary structure motifs:

$$\mathbf{s}_i = \sum_j \mathbf{f}_j \mathbf{p}_j$$

where $\mathbf{p}_j$ and $\mathbf{f}_j$ are the spectrum and the fraction related to the jth pure structural motif, respectively, i.e., ($\sum_j \mathbf{f}_j = 1$).

There are many established methods to elucidate the secondary structure associated with far-UV CD protein spectra.[31,32] Although very diverse in mathematical background, all of them relate the measured protein spectrum to a reference database formed by the far-UV CD spectra and the associated secondary structure composition of a group of representative proteins. As a result, the secondary structure composition (the j values of $\mathbf{f}$) linked to the measured spectrum is obtained. In a thermal-induced protein folding, the plot of the $\mathbf{f}$ values calculated for each measured spectrum as a function of the temperature displays the evolution of the different secondary structure motifs along the process.

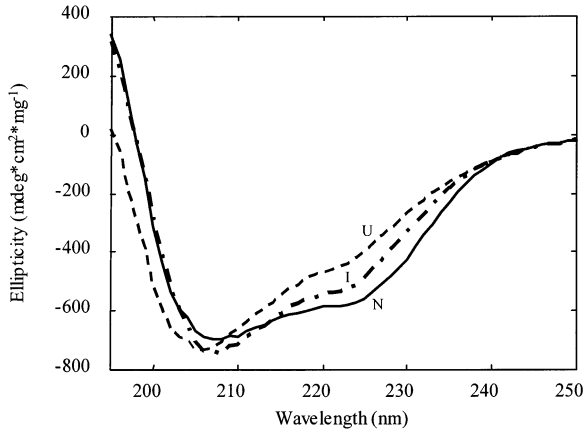

Figure 7. Resolved MCR-ALS far-UV CD spectra in the protein unfolding of α-apolactalbumin in the absence of salt (NaCl). Solid line, native conformation; dash−dot line, intermediate conformation; dashed line, unfolded conformation.

Using the results obtained by MCR-ALS, this information can also be improved with respect to classical analyses. Now, not only the description of the folding process as a function of the evolution of the structural motifs is obtained but, most importantly, the secondary structure of each of the resolved protein conformations involved becomes available. Note that this is essential in the case of the presence of an intermediate, where the usual elucidation from X-ray diffraction of pure crystallized structures is not an option.

Below, we present the information obtained for the thermal-induced protein unfolding of α-apolactalbumin in the absence of

(31) Sreerama, N.; Woody, R. W. *Anal. Biochem.* **2000**, *287*, 252−260.
(32) Johnson, W. C. *Proteins: Struct., Funct., Genet.* **1999**, *35*, 307−312.

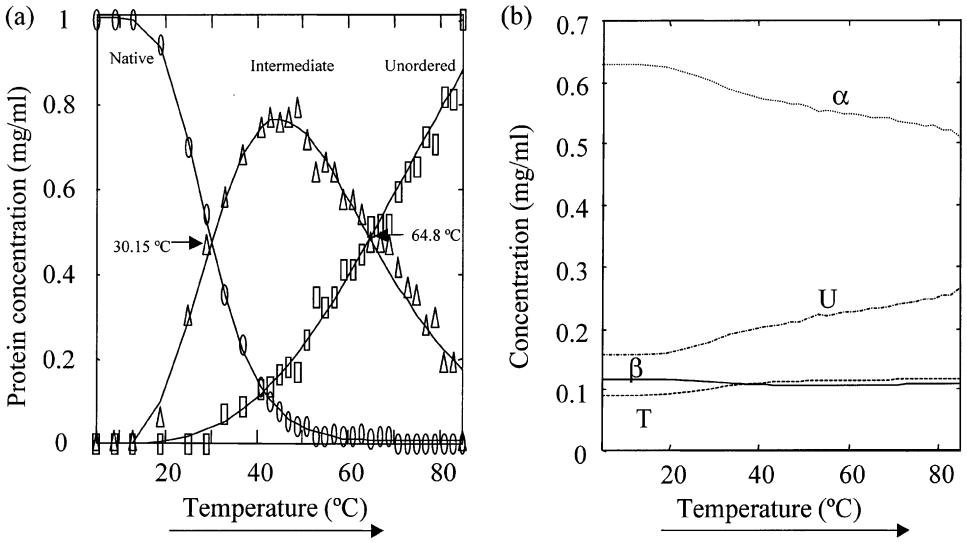

Figure 8. Description of the thermal-induced protein unfolding process of α-apolactalbumin in absence of sodium ion. (a) Evolution of the pure protein conformations involved; (b) evolution of the secondary structural elements.

Table 2. Estimation of the Fraction of the Different Secondary Structure Motifs in Protein Conformations Present in the Unfolding Process of α-Apolactalbumin in the Absence of Sodium Ion

protein conformation	secondary structure motifs			
	α (helix)	β (sheets)	turns	random coil
native	0.630	0.118	0.090	0.158
intermediate	0.574	0.106	0.114	0.201
unordered	0.510	0.109	0.118	0.264

sodium ions. No results are shown for experiments performed in the presence of Na^+ ions because the protein spectral features linked to wavelengths lower than 200 nm are masked by the too intense signal of the metal ion. The elucidation method used, proposed by Johnson, is CDSSTR,[32] which averages the best results coming from calculations performed on a large number of protein subset combinations from the original huge reference data set. The application of this method to each of the resolved far-UV CD spectra (profiles in the matrix **S** obtained by MCR-ALS; see Figure 7) provides the secondary structure related to the protein conformations linked to the unfolding process of α-apolactalbumin, i.e., native, intermediate, and unfolded (see Table 2).

Keeping the structure and information of Table 2, we may define a composition matrix, **F**, the rows of which contain the ratio of the structural elements related to each protein conformation. Now, the evolution of the secondary structure with the temperature is straightforwardly obtained. Thus, using the composition matrix, **F**, the resolved concentration profiles obtained for each protein conformation (matrix **C** from MCR-ALS) (Figure 8a) are transformed into the profiles related to the evolution of each of the secondary structure motifs, $\mathbf{C_f}$. This transformation process can be expressed by the simple matrix equation below:

$$\mathbf{C_f} = \mathbf{CF}$$

The columns in matrix $\mathbf{C_f}$ give the evolution profile of each of the secondary structure motifs along the process studied (see Figure 8b).

The secondary structure obtained for the resolved protein conformations (see Table 2) and the derived evolution of the structural motifs along the unfolding process of α-apolactalbumin (see Figure 8b) are consistent with the pathway expected in this thermal-induced process. Thus, a decrease in the α-helix proportion, most abundant in the native conformation, and an increase in random coil and turns, elements typically associated with unfolded states, are observed as the conformations evolve in the sequence native → intermediate → unfolded.

CONCLUSIONS

Complex protein folding processes involving the presence of intermediate conformations can be successfully described by combining multispectroscopic monitoring and multivariate curve resolution. The amount and quality of the information acquired and the efficient data analysis allow for the detection and modeling of intermediate species that cannot be isolated either by physical or by chemical means. The fate of the intermediate along the process, i.e., when it is present and in which amount, is unraveled from the original raw measurements. The pure spectrum resolved for this species is the essential starting point, nonobtainable otherwise, to get the secondary structure linked to this conformation using the appropriate deconvolution approaches.

ACKNOWLEDGMENT

This work has been financially supported by the Spanish and Catalan governments through projects BQU2000-0788 and 2001SGR 00056, respectively. S.N. also acknowledges the Universitat de Barcelona for a Ph.D. scholarship.

Received for review July 3, 2002. Accepted September 20, 2002.

AC025914D