Anal. Chem. **2003,** *75,* 5592−5601

Modeling Temperature-Dependent Protein Structural Transitions by Combined Near-IR and Mid-IR Spectroscopies and Multivariate Curve Resolution

Susana Navea, Anna de Juan,* and Romà Tauler

Chemometrics Group, Department of Analytical Chemistry, Diagonal, 647, 08028 Barcelona, Spain

The combination of near- and midinfrared spectroscopies (NIR and MIR) is proposed to monitor temperature-dependent transitions of proteins. These techniques offer a high discriminating power to distinguish among protein structural conformations but, in temperature-dependent processes, present the drawback associated with the intense and evolving absorption of the deuterium oxide, used as a solvent in the protein solutions. Multivariate curve resolution−alternating least squares (MCR−ALS) is chosen as the data analysis technique able to unravel the contributions of the pure protein and deuterium oxide species from the mixed raw experimental measurements. To do so, MCR−ALS works by analyzing simultaneously experiments from MIR and NIR on pure deuterium oxide solutions and protein solutions in D_2O. This strategy has proven to be effective for modeling the protein process in the presence of D_2O and, therefore, for avoiding the inclusion of artifacts in the data stemming from inadequate baseline corrections. The use of MIR and NIR and MCR−ALS has been tested in the study of the temperature-dependent evolution of β-lactoglobulin. Only the combined use of these two infrared techniques has allowed for the distinction of the three pure conformations involved in the process in the working thermal range: native, R-type state, and molten globule.

Infrared spectroscopy has been widely used for the characterization of protein secondary structure and the identification of protein components involved in events such as ligand binding and electron-transfer reactions.[1] Whereas midinfrared (MIR) spectroscopy has been used extensively in protein structural studies, near-infrared (NIR) spectroscopy has focused mainly on agricultural and bioindustrial applications for process control or quantitative purposes. Although the practical applications of NIR spectroscopy have been extensive, the potential of this spectroscopic technique for protein structural analysis has been hardly explored.[2]

Several spectroscopic techniques are currently used for the conformational analysis of proteins, namely, nuclear magnetic resonance (NMR), circular dichroism (CD), and X-ray crystallography. Two-dimensional NMR, recently introduced in the study of proteins in solution, was originally limited to the study of low molecular weight proteins[3] but now offers a reasonable alternative to X-ray crystallography, limited by the growth of fine protein crystals.[4] ESR and optical spectroscopy (visible, UV, fluorescence, CD, etc.) also have their specific limitations as techniques for studying protein conformations, particularly those related to light scattering, which restrict their application to clear protein solutions.

Fourier transform infrared spectroscopy has been shown to be useful for studying protein structures[5] and to be able to cope with light-scattering problems, which allows for the analysis of turbid solutions, aggregates, etc., when an attenuated total reflectance (ATR) accessory is employed. In midinfrared, the two dominant features in any protein spectrum are the amide I and amide II absorption bands located at around 1650 and 1550 cm^{-1}, respectively. These bands are sensitive to hydrogen-bonding interactions and, therefore, to differences in protein secondary structure. The differences in amide bond geometric orientations of the α-helix, β-sheet, turn, and random coil structural motifs result in differences in vibrational frequencies and distinct IR signals.

The basic research on the protein structures by NIR spectroscopy has concerned mainly proteins in solid state,[6,7] and the structural investigations of proteins in aqueous solutions have been far behind. Proteins in their physiological state are most often in aqueous solution, and it is obvious that structure−function relationships cannot be established in "dry" samples. However, it is well-known that NIR spectra are also rich in molecular structural information and their spectral range can also monitor changes in hydration and hydrogen bonding in protein samples.

NIR spectroscopy offers some advantages over midinfrared spectroscopy, such as the less intense overlap between the water and the protein absorption bands in some spectral ranges and an easier experimental manipulation due to the sample size and the

* Corresponding author. E-mail: annaj@apolo.qui.ub.es.

(1) Jackson, M.; Mantsch, H. H. *Crit. Rev. Biochem. Mol. Biol.* **1995**, *30*, 95−120.
(2) Robert, P.; Devaux, M. F.; Mouhous, N.; Dufour, E. *Appl. Spectrosc.* **1999**, *53* (2), 226−232.
(3) Wüthrich, K. *Science* **1989**, *243*, 45−50.
(4) Fiaux, J.; Bertelsen, E. B.; Horwich, A. L.; Wüthrich, K. *Nature* **2002**, *418*, 207−211.
(5) Byler, D. M.; Susi, H. *Biopolymers* **1986**, *25* (3), 469−487.
(6) Liu, Y.; Cho, R. K.; Sakurai, K.; Miura, T.; Ozaki, Y. *Appl. Spectrosc.* **1994**, *48* (10), 1249−1253.
(7) Cho, R. K.; Lee, J. H.; Ozaki, Y.; Iwamoto, M. *J. Near Infrared Spectrosc.* **1995**, *3*, 73−79.

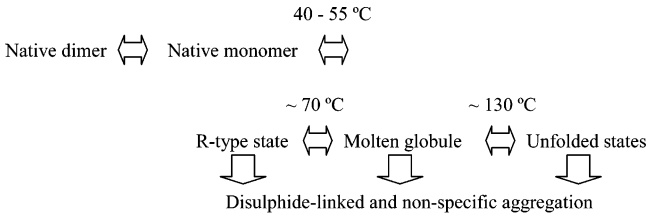
10.1021/ac0343883 CCC: $25.00
Published on Web 09/03/2003

instrumental setup required. On the other hand, the interpretation of the structural and conformational information included in MIR protein spectra is much better established. Therefore, the combined use of the mid- and near-infrared spectroscopies can be a new and improved way to study structural changes in protein processes.

In both MIR and NIR spectroscopies, the protein absorption bands in aqueous solutions are much weaker than those due to water. For this reason, protein solution spectra are recorded using deuterium oxide as solvent. Raw spectra obtained in this way are often subject to spectral manipulations, such as spectral subtraction of the D_2O signal or baseline correction. When a temperature-dependent protein-folding process is monitored, a series of protein spectra are collected as a function of the temperature. In this case, the solvent subtraction and subsequent baseline correction is not advisable because of the high intensity of deuterium oxide bands, the intense band overlap between protein and deuterium oxide, and the evolving nature of the deuterium oxide signal with the temperature. The use of these classical signal processing methodologies can easily introduce artifacts into the data collected.

Temperature-dependent protein processes have been classically studied through examination of melting curves, that is, by monitoring the evolution of one-wavelength response values (e.g., absorbances or ellipticities) as a function of the temperature[8,9] or performing deconvolution of individual spectra collected at few particular temperatures. In the present work, a multiwavelength spectrum is recorded at each stage of the protein process, covering exhaustively the whole thermal range, and all the information collected is organized in a data table (matrix) and analyzed using multivariate data analysis. Multivariate curve resolution-alternating least squares (MCR−ALS) is a chemometric method that allows for the mathematical resolution of concentration and spectra profiles of pure chemical species from raw data sets recorded in multicomponent mixtures or chemical evolving processes.[10−12] MCR−ALS has been selected because it can be applied simultaneously to several matrices coming from the same process monitored with several spectroscopic techniques or to matrices collected in several experiments and has been proven to be suitable for the modeling of other protein processes.[13−15] In comparison with previous works devoted to protein folding, the protein data sets obtained with MIR and NIR spectroscopies have an additional problem: the rank-deficiency.[16−18] This problem occurs when two or more independent closed reaction systems evolve and are analyzed simultaneously. In this context, a closed system is a reaction in which the concentrations of all species involved sum up to a constant value along the different stages of the process, that is, a dynamic chemical system in which the mass balance is preserved. In our experiments, two independent closed systems evolve as a function of the temperature: the deuterium oxide and the protein system. As a consequence, the number of pure contributions detected is smaller than the number of real chemical contributions (species) in the data set, and the whole system cannot be resolved properly. Since MCR−ALS can work with several experiments, correct resolution is obtained by analyzing a full-rank augmented data set formed by the protein experiments and analogous experiments performed on solutions of pure deuterium oxide.

The thermal denaturation of β-lactoglobulin (BLG) using mid- and near-infrared spectroscopy is studied as a real model process to explore the combined use of these spectroscopic techniques. BLG is found in the whey fraction of the milk of many mammals. This protein has 162 residues and contains two disulfide bonds (C106−C119 and C66−C160). It is well-known that, at room temperature and physiological pH, BLG exists mainly as a dimer in which the monomers are noncovalently linked. The thermal denaturation of BLG at neutral and alkaline pH value is supposed to proceed by a multistep mechanism,[19,20] according to the scheme below:



It has been postulated that the dissociation of the dimer is coupled with a conformational transition to an R-type state, although this change has not always been detected. At physiological concentrations, this seems to take place at 40−55 °C. The R-type state, in which R− originally stood for "reversible",[21] differs from the native conformation in the environment of a few side chains and the accessibility of a free thiol to intermolecular reactions, but there are no major differences of the protein secondary structure,[21] as has been confirmed by solution CD spectroscopy.[22] Other authors[23] report the first denaturation step around 70 °C and postulate a molten globule state at these high temperatures, with a somewhat lesser compact secondary structure than the native structure, and a second denaturation transition at ∼130°C. The results obtained from this study are oriented to provide more information on this unresolved problem.

EXPERIMENTAL PROCEDURES

Chemicals. β-Lactoglobulin from bovine milk and deuterium oxide were purchased from Sigma and Aldrich, respectively.

(8) Manderson, G. A.; Creamer, L. K.; Hardman, M. J. *J. Agric. Food Chem.* **1999**, *47*, 4557−4567.
(9) Creamer, L. K. *Biochemistry* **1995**, *34*, 7170−7176.
(10) Tauler, R.; Smilde, A. K.; Kowalski, B. R. *J. Chemom.* **1995**, *9*, 34−58.
(11) Tauler, R. *Chemom. Intell. Lab. Syst.* **1995**, *30*, 133−146.
(12) de Juan, A.; Casassas, E.; Tauler, R. Soft modelling of analytical data. In *Encyclopedia of Analytical Chemistry: Applications, Theory, and Instrumentation*; Meyers, R. A., Ed.; John Wiley & Sons Ltd.: Chichester, U.K., 2000; pp 9800−9837.
(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) Navea, S.; de Juan, A.; Tauler, R. *Anal. Chem.* **2002**, *74*, 6031−6039.
(16) Amrhein, M.; Srinivasan, B.; Bonvin, D.; Schumacher, M. M. *Chemom. Intell. Lab. Sys.* **1996**, *33* (1), 17−33.
(17) Izquierdo-Ridorsa, A.; Saurina, J.; Hernández-Cassou, S.; Tauler, R. *Chemom. Intell. Lab. Sys.* **1997**, *38* (2), 183−196.
(18) Diewok, J.; de Juan, A.; Tauler, R.; Lendl, B. *Appl. Spec.* **2002**, *56* (1), 40−50.
(19) Qi, X. L.; Brownlow, S.; Holt, C.; Sellers, P. *Biochim. Biophys. Acta* **1995**, *1248*, 43−49.
(20) Iametti, S.; De Gregori, B.; Vecchio, G.; Bonomi, F. *Eur. J. Biochem.* **1996**, *237*, 106−112.
(21) Tanford, C.; Taggart, V. G. *J. Am. Chem. Soc.* **1961**, *83*, 1634−1638.
(22) Matsuura, J. E.; Manning, M. C. *J. Agric. Food Chem.* **1994**, *42*, 1650−1656.
(23) Qi, X. L.; Holt, C.; McNulty, D.; Clarke, D. T.; Brownlow, S.; Jones, G. R. *Biochem. J.* **1997**, *324*, 341−346.

β-Lactoglobulin was used without further purification. The protein was dissolved in deuterium oxide, and 50 mg/mL protein solutions were used to record midinfrared (MIR) and near-infrared (NIR) spectra.

Instrumentation. MIR data were collected with a Bomem Fourier transform infrared spectrometer (FTIR) equipped with a Peltier temperature controller (precision ± 0.1°C), using calcium fluoride windows with 50-μm Teflon spacers. FT-IR spectra were measured from 4000 to 720 cm^{-1} at 4 cm^{-1} resolution. Each spectrum is the averaged result of 50 scans. The background spectrum was recorded after a 10-min purge with nitrogen, and sample spectra were recorded under nitrogen atmosphere.

NIR spectra were recorded with a Perkin-Elmer λ-19 spectrometer equipped with a Peltier-type thermostated cell holder. Absorbance spectra were measured in the spectral range 900−2500 nm, and the wavelength step was 1 nm. A 1-mm path length closed quartz cell was used to avoid sample evaporation during the experiment.

Experimental Procedure. The proper analysis of a temperature-dependent protein process monitored with MIR or NIR spectroscopies needed the measurement of two different series of spectra: one on the protein solution and another one on pure deuterium oxide. Both series of spectra were recorded under the same experimental conditions and using the same spectroscopic technique. The temperature was modified between 14 and 85 °C. Temperature intervals among consecutive spectra ranged from 2 to 4 °C, depending on the spectral change detected. A 2-min waiting period was used before recording each spectrum to avoid possible uncontrolled time-dependent effects on the evolution of the monitored process.

Data Sets and Data Treatment. The spectra recorded during the monitoring of a temperature-dependent protein folding process are organized in a data matrix $\mathbf{D_i}$, the rows of which are the spectra recorded at each temperature and the columns of which represent the melting curves (absorbance versus temperature profiles) at each wavelength or wavenumber. Using NIR or IR as monitoring spectroscopic techniques, the data set $\mathbf{D_i}$ contains two closed systems that absorb and evolve with the temperature: the protein system and the deuterium oxide system. This situation provides a rank-deficient data set, in which the number of distinguishable components is lower than the total number of chemical contributions to the data set.[16−18] Therefore, the straightforward resolution of this kind of system is not feasible.

Breaking rank deficiencies can be accomplished using matrix augmentation, for instance, through appending full-rank appropriate information to the rank-deficient data set.[16−18] In the temperature-dependent study of a protein solution with mid- or near-infrared spectroscopy (or both), the protein rank-deficient data set can be augmented with a full-rank matrix coming from the temperature-dependent monitoring of a pure deuterium oxide solution performed in the same experimental conditions using the same spectroscopic technique. The column-wise augmented data set **D** formed by the deuterium oxide data matrix and the protein data matrix (see Figure 1) has full rank and can be satisfactorily analyzed with three-way multivariate resolution methods. All of the experiments performed and their related data sets are described and coded in Table 1. These codes will be adopted throughout the text.

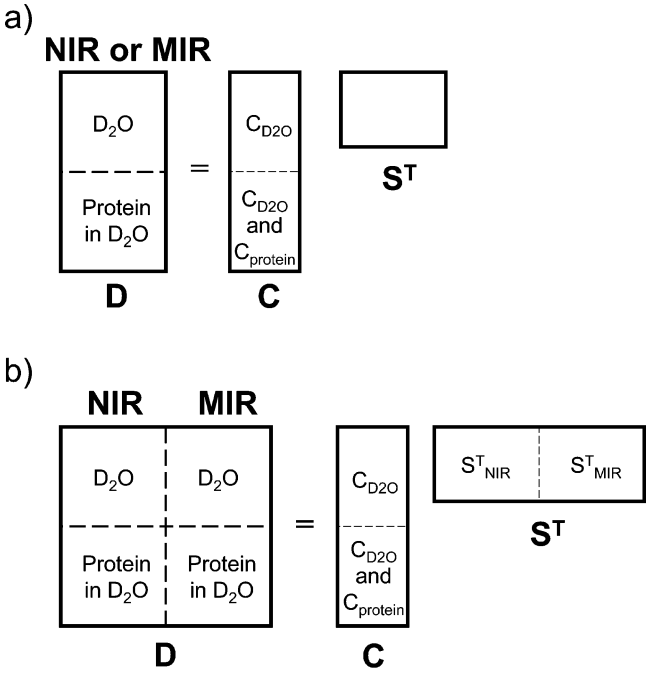


Figure 1. MCR−ALS decomposition of (a) a full-rank column-wise augmented matrix used to model temperature-dependent protein changes monitored by NIR or MIR, (b) a full-rank row- and column-wise augmented matrix used to model temperature-dependent protein changes monitored by NIR and MIR.

Table 1. Experiment Coding Used in This Work

data matrix	system studied	spectrometric technique
$\mathbf{D_1}$	deuterium oxide (D_2O)	near-infrared spectroscopy
$\mathbf{D_2}$	β-lactoglobulin in D_2O	near-infrared spectroscopy
$\mathbf{D_3}$	deuterium oxide (D_2O)	midinfrared spectroscopy
$\mathbf{D_4}$	β-lactoglobulin in D_2O	midinfrared spectroscopy

Multivariate curve resolution-alternating least squares was selected for the data analysis of the full rank augmented data sets described above.[10−12] As any other resolution method, MCR−ALS decomposes the original data set, **D**, into the product of two small matrices, **C** and $\mathbf{S^T}$, containing the pure profiles of all the chemical contributions present in the raw mixed experimental measurements.

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

C contains the pure column concentration profiles, and $\mathbf{S^T}$ contains the pure row signal profiles. Matrix **E** describes the experimental error, that is, the residual variation of the data set that is not attributed to any chemical contribution. In the context of protein folding, the column-wise augmented matrix **C** is formed by two $\mathbf{C_i}$ submatrices, one with the temperature-dependent concentration profiles of the species in the deuterium oxide experiment and the other related to the thermally induced protein experiment and formed by the concentration profiles of the deuterium oxide species and of all protein conformations as a function of the temperature. The rows in matrix $\mathbf{S^T}$ are the pure spectra related to the components in matrix **C**.

MCR−ALS works according to the following steps:

1. Determination of the number of contributions. The number of species that independently contribute to the measured signal in the column-wise augmented matrix **D** is determined by singular value decomposition (SVD)[24] or other methods based on factor analysis.

2. Building initial estimates. Generally speaking, estimates of the **C** matrix (e.g., usually obtained by evolving factor analysis, EFA[25,26]) or of the $\mathbf{S}^T$ matrix (e.g., using methods to select the purest spectra in **D**[27]) can be used to start the iterative process. In this work, we have not used spectral estimates obtained from the protein experimental measurements because of the unavoidable overlap between protein and deuterium oxide signals in the measured protein spectra. Classical EFA applied to the pure deuterium oxide $\mathbf{D_i}$ submatrix has been used to obtain initial estimates of the deuterium contributions in matrix **C**.[28,29] A modified algorithm based on evolving factor analysis, adapted to work with rank-deficient systems, has been applied to the column-wise augmented matrix **D** to obtain the initial concentration profiles for the protein system.[30] Two main advantages are linked to this algorithm: the capacity to detect all chemical contributions in the rank-deficient matrix (protein experiment) because it is adapted to work with full-rank augmented matrices and the ability to derive initial concentration profiles and local rank information exclusively related to the evolution of the protein contributions.

3. Given **D** and **C**, constrained least-squares calculation of $\mathbf{S}^T$.

4. Given **D** and $\mathbf{S}^T$, constrained least-squares calculation of **C**. Considering the protein folding problem, all concentration profiles in the different $\mathbf{C_i}$ submatrices have been forced to be nonnegative and unimodal[14] (i.e., presence of one maximum per profile). The profiles related to the protein conformations have also been constrained to form a closed system, and the local rank information obtained previously has been introduced in the iterative process. It must be noted that the local rank information is not limited to the detection of selective zones (none for the protein system since there is always overlap with the deuterium oxide species) but to the detection of thermal ranges with some protein conformations absent. This does not mean that the concentration windows for each protein conformation are completely fixed. Instead, the protein conformations are only set to have null concentration in the thermal ranges where this information can be unambiguously concluded. Pure spectra in $\mathbf{S}^T$ have been constrained to be nonnegative according to the properties of mid- and near-infrared spectroscopy.[10,12,31−33]

5. Iterate until convergence is achieved. The convergence criterion in the MCR−ALS optimization is based on the comparison of the lack of fit obtained in two consecutive iterations. When the relative difference in fit is below a threshold value, the optimization is finished. The lack of fit is calculated according to the expression

$$\text{lack of fit} = 100\sqrt{\frac{\sum_{i,j} r_{ij}^2}{\sum_{i,j} d_{ij}^2}} \qquad (2)$$

where d_{ij} designs an element of the raw data matrix **D** and r_{ij}, its related residual. A maximum number of iterative cycles may also be used as a stopping criterion.

All the chemometric methods employed in this work are in MATLAB[34] programs developed by our research group, available at the WEB page http://www.ub.es/gesq/eq1_eng.htm.

RESULTS AND DISCUSSION

Resolution of Changes in the Protein Structure Using Near-Infrared Spectroscopy. The evolution of the protein structure of β-lactoglobulin was studied by NIR. As mentioned above, when a thermally induced protein denaturation is monitored using infrared spectroscopy, two closed absorbing systems evolve simultaneously: the protein and the D_2O system, and a rank-deficient data set is obtained. To achieve the complete resolution of the protein system, the MCR−ALS analysis must be performed on the full-rank three-way data set $[\mathbf{D_1};\mathbf{D_2}]$ (the semicolon between matrices indicates a column-wise augmented matrix according to MATLAB notation). The results obtained describe the changes in the protein structure of β-lactoglobulin in D_2O.

As shown in Figure 1a, MCR−ALS analysis of a column-wise augmented matrix provides a column-wise augmented **C** matrix, formed by $\mathbf{C_i}$ submatrices that contain nonnull concentration profiles for the species present in the related $\mathbf{D_i}$ submatrix and a matrix $\mathbf{S}^T$, common to all matrices treated together. To start the resolution process, it is necessary to obtain initial concentration estimates of each submatrix $\mathbf{D_i}$. $\mathbf{D_1}$ (series of pure deuterium oxide spectra) is a full-rank data matrix, and $\mathbf{D_2}$ (series of β-lactoglobulin spectra in D_2O) is a rank-deficient data matrix.

Classical evolving factor analysis is used to construct the concentration estimates for the deuterium species present in the submatrix $\mathbf{D_1}$, symbolized by d_1 and d_2 in Figure 2b. Since the thermal variation in the series of pure deuterium oxide and protein solution are identical, the concentration estimates for the deuterium oxide system obtained from $\mathbf{D_1}$ are valid for $\mathbf{D_2}$. These EFA results agree with other studies performed to model the temperature-dependence of D_2O that preponderantly support the description of this variation with a noncontinuous model involving two water species differing in their H bonding properties. This conclusion has been extracted using different spectroscopic techniques (e.g., NIR, IR, and Raman, fundamentally) and working with thermal and spectral ranges larger than the ones used in this study.[35−43]

(24) Golub, G. H.; Reinsch, C. *Numer. Math.* **1970**, *14*, 403−420.
(25) Gampp, H.; Maeder, M.; Meyer, Ch. J.; Zuberbühler, A. D. *Talanta* **1985**, *32* (12), 1133−1139.
(26) Maeder, M. *Anal. Chem.* **1987**, *59*, 527−530.
(27) Winding, W.; Guilment, J. *Anal. Chem.* **1991**, *63*, 1425−1432.
(28) Gampp, H.; Maeder, M.; Meyer Ch. J.; Zuberbühler, A. *Talanta* **1986**, *33* (12), 943−951.
(29) Tauler, R.; Casassas, E. J. *Chemometrics* **1988**, *3* (Suppl. A), 151−161.
(30) de Juan, A.; Navea, S.; Diewok, J.; Tauler, R. *Chemom. Intell. Lab. Syst.*, In press.
(31) de Juan, A.; Vander Heyden, Y.; Tauler, R.; Massart, D. L. *Anal. Chim. Acta* **1997**, *346* (3), 307−318.
(32) Bro, R.; de Jong, S. *J. Chemom.* **1997**, *11* (5), 393−401.
(33) Bro, R.; Sidiropoulos, N. D. *J. Chemom.* **1998**, *12* (4), 223−247.
(34) Version 6; The Mathworks Inc.: Natick, MA; http://www.mathworks.com.
(35) Libnau, F. O.; Toft, J.; Christy, A. A.; Kvalheim, O. M. *J. Am. Chem. Soc.* **1994**, *116*, 8311−8316.
(36) Segtnan, V. H.; Sasic, S.; Isaksson, T.; Ozaki, Y. *Anal. Chem.* **2001**, *73*, 3153−3161.
(37) Benson, S, W.; Siebert, E. D. *J. Am. Chem. Soc.* **1992**, *114*, 4269−4271.
(38) Hare, D. E.; Sorensen, C. M. *J. Am. Chem. Soc.* **1990**, *93*, 25−33.
(39) Hare, D. E.; Sorensen, C. M. *J. Am. Chem. Soc.* **1990**, *93*, 6954−6961.

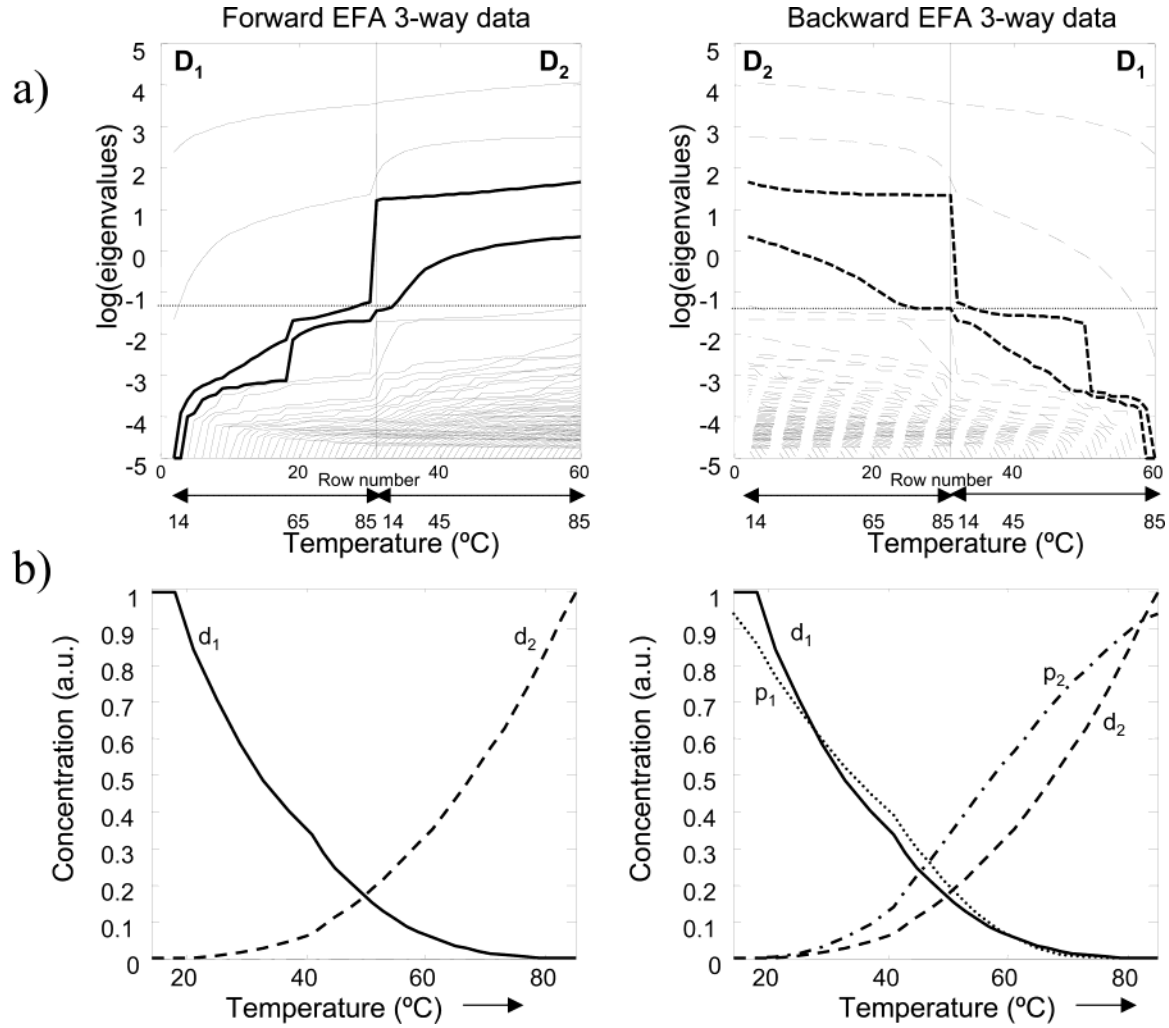


Figure 2. NIR data exploratory analysis. (a) Results from adapted EFA algorithm on the full-rank column-wise augmented data set [$\mathbf{D_1}$;$\mathbf{D_2}$]. Left plot: forward EFA. Right plot: backward EFA. The dotted horizontal lines in the plots mark the noise level; the vertical lines separate the information related to the deuterium oxide experiment ($\mathbf{D_1}$) and the protein solution experiment ($\mathbf{D_2}$). Bold lines mark protein significant contributions, only present in the protein experiment. (b) Initial concentration estimates derived from EFA analysis. Left plot: concentration estimates related to $\mathbf{D_1}$. Right plot: concentration estimates related to $\mathbf{D_2}$. d_1 and d_2 design deuterium oxide species, p_1 and p_2 protein species.

To get initial estimates specific for the protein species, an adaptation of the EFA algorithm for rank-deficient systems has been used.[30] Working analogously to the parent method on the full-rank augmented data set formed by $\mathbf{D_1}$ and $\mathbf{D_2}$, this new algorithm depicts the evolution of the eigenvalues obtained in subsequent principal component analyses performed on gradually increasing submatrices in the process direction, enlarged by adding one new row (spectrum) at a time. This procedure is performed on the augmented matrix [$\mathbf{D_1}$;$\mathbf{D_2}$] from top to bottom of the data set (forward EFA) and on the augmented matrix [$\mathbf{D_2}$;$\mathbf{D_1}$] from bottom to top of the data set (backward EFA) to investigate the emergence and the decay of the process contributions in the protein system, respectively. As soon as the PCA results start to be obtained on submatrices including the D_2O matrix and spectra from the protein experiment, that is, as soon as the eigenvalue lines cross the vertical separation in Figure 2a to go from $\mathbf{D_1}$ to $\mathbf{D_2}$, new contributions absent in the D_2O full-rank matrix used for augmentation are detected, that is, those exclusively related to the protein system (bold lines in forward and backward EFA plots in Figure 2a). The EFA plots obtained suggest that only two protein species are needed to describe the protein changes detected by NIR spectroscopy. The somewhat unclear division among noise-related and significant eigenvalues is due to the structured noise of the signal. The maximum value of the third eigenvalue in the deuterium oxide matrix ($\mathbf{D_1}$), where only two contributions are needed to describe the system, has been used to mark the noise level threshold (It should be noted that the logarithmic scale used in Figure 2a enhances in a misleading way the importance of some of the noise-related eigenvalues. Thus, the third eigenvalue in the deuterium oxide matrix accounts for a variance percent lower than 0.1). The concentration estimates of the protein contributions (hereafter p_1 and p_2) are derived, as in classical EFA, from the combination of the protein-related eigenvalues in the $\mathbf{D_2}$ part of the EFA plots in Figure 2a. In Figure 2b, the EFA-derived input initial estimates to be optimized by MCR−ALS are shown.

(40) Hare, D. E.; Sorensen, C. M. *J. Am. Chem. Soc.* **1992**, *96*, 13−22.
(41) Walrafen, G. E. In *Water, a Comprehensive Treasite*; Franks, F., Ed.; Plenum Press: New York, 1972; Vol. 1, p 208.
(42) Maréchal, Y. *Chem. Phys.* **1991**, *95* (8), 5565−5573.
(43) Giguére, P. A. J. *Raman Spectrosc.* **1984**, *15* (5), 354−359.

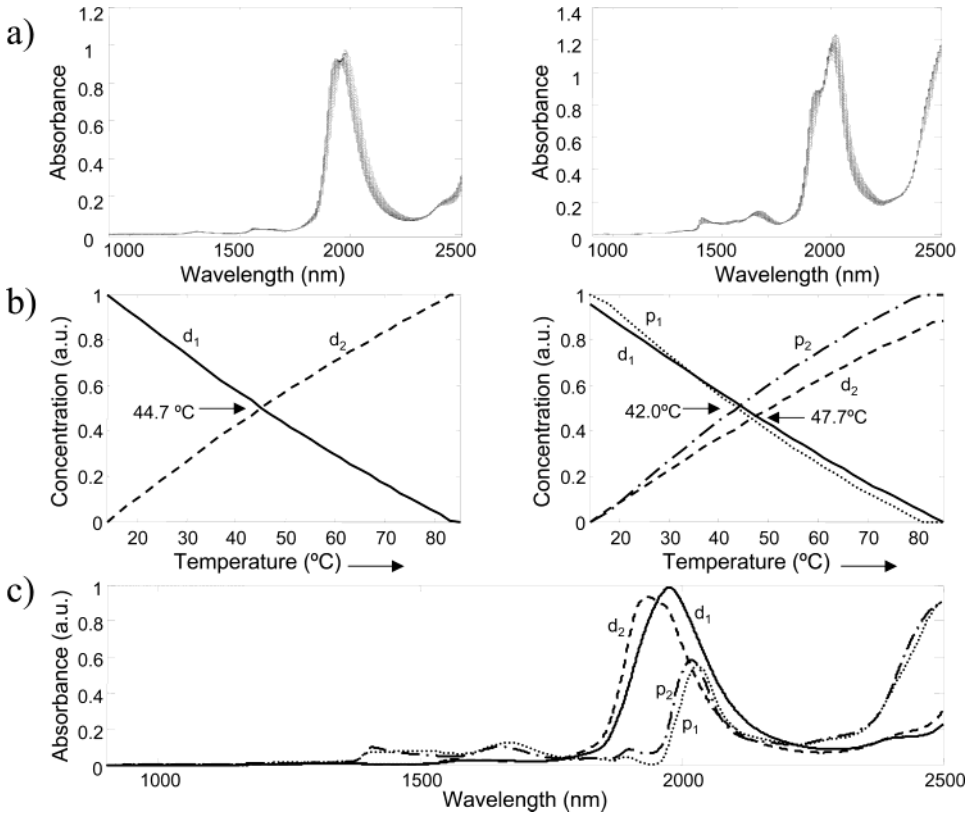


Figure 3. Raw data [$\mathbf{D_1}$;$\mathbf{D_2}$] and MCR−ALS results from temperature-dependent monitoring of β-lactoglobulin with NIR. (a) Raw data. Left plot: series of deuterium oxide spectra ($\mathbf{D_1}$). Right plot: Series of protein spectra in D_2O ($\mathbf{D_2}$). (b) Resolved concentration profiles. Left plot: related to $\mathbf{D_1}$. Right plot: related to $\mathbf{D_2}$. d_1, d_2, p_1, and p_2 as in Figure 2.

MCR−ALS is applied constraining all the concentration profiles in the $\mathbf{C_i}$ submatrices to be nonnegative and unimodal (i.e., the temperature-dependent evolution of each protein and deuterium oxide species can be appropriately represented by an emergence−decay profile having a single peak maximum). The condition of a closed system was also applied to the protein concentration profiles because the total concentration of protein remains constant during the whole temperature-dependent process. Local rank constraints for the protein system, that is, setting the absence of a certain protein conformation at some temperatures, were applied according to the information on the EFA plot.[10,44] Spectra in $\mathbf{S^T}$ were forced to be nonnegative according to the properties of NIR spectroscopy. The constraint of correspondence of species has been applied by setting null protein concentration profiles in the pure deuterium experiment[10,11,45].

The resolved concentration profiles and spectra are shown in Figure 3b and c, respectively. The MCR−ALS resolved concentration profiles show the evolution of the concentration of the different species present in the process. The temperature-dependent evolution of D_2O can be modeled with two species, d_1 and d_2; they represent the predominant deuterium oxide species at low and high temperatures, respectively. As a confirmation of this two-component model, resolution trying to include a third D_2O species did not improve the overall description (fit) of the variation of the data set. MCR−ALS can describe separately the protein and the solvent contributions from the series of protein spectra in D_2O. In the related $\mathbf{C}_i$ submatrix, the deuterium oxide species modeled are d_1 and d_2, and the protein species modeled are the p_1 and p_2 conformations. The resolved concentration profiles of the deuterium oxide species in both $\mathbf{D_1}$ and $\mathbf{D_2}$ experiments are very similar, as was expected. The crossing temperature between the protein conformations p_1 and p_2 is 42.0 °C. Comparing this temperature with the proposed mechanism in the scheme of the introduction, NIR spectroscopy has been shown to be able to detect the transition from the native protein to the R-type state. The individual analysis of the spectroscopic changes provided by this technique does not seem to allow for the detection of additional conformational changes.

MCR−ALS resolves the pure spectra of both the deuterium and protein systems. As can be observed from the results, the protein bands of conformations p_1 and p_2, which are mostly embedded in the D_2O bands and show similar shapes, can be successfully resolved. These pure spectral shapes agree with the tendency shown by the raw spectra collected (see Figure 3a). The protein bands present various maximums located at 2037, 1670, and 1410 nm for the p_1 conformation. These maximums of protein bands are displaced to lower wavelengths for the p_2 conformation. The increase in the temperature causes a shift of the D_2O band to a lower wavelength, as well. Therefore, the increase in the temperature produces a blue shift in both D_2O and β-lactoglobulin bands. Protein band assignments in NIR spectroscopy have not been well-established and remain a subject of research. Sadler et al.,[46] using photoacoustic spectroscopy, observed the overtones and combination bands in the near-infrared region. The deeper insight into the 1400−1600-nm region made it possible to assign

(44) de Juan, A.; Maeder, M.; Martínez, M.; Tauler, R. *Anal. Chim. Acta* **2001**, *442* (2), 337−350.
(45) Tauler, R.; Barceló, D. *Trends Anal. Chem.* **1993**, *12* (8), 319−327.

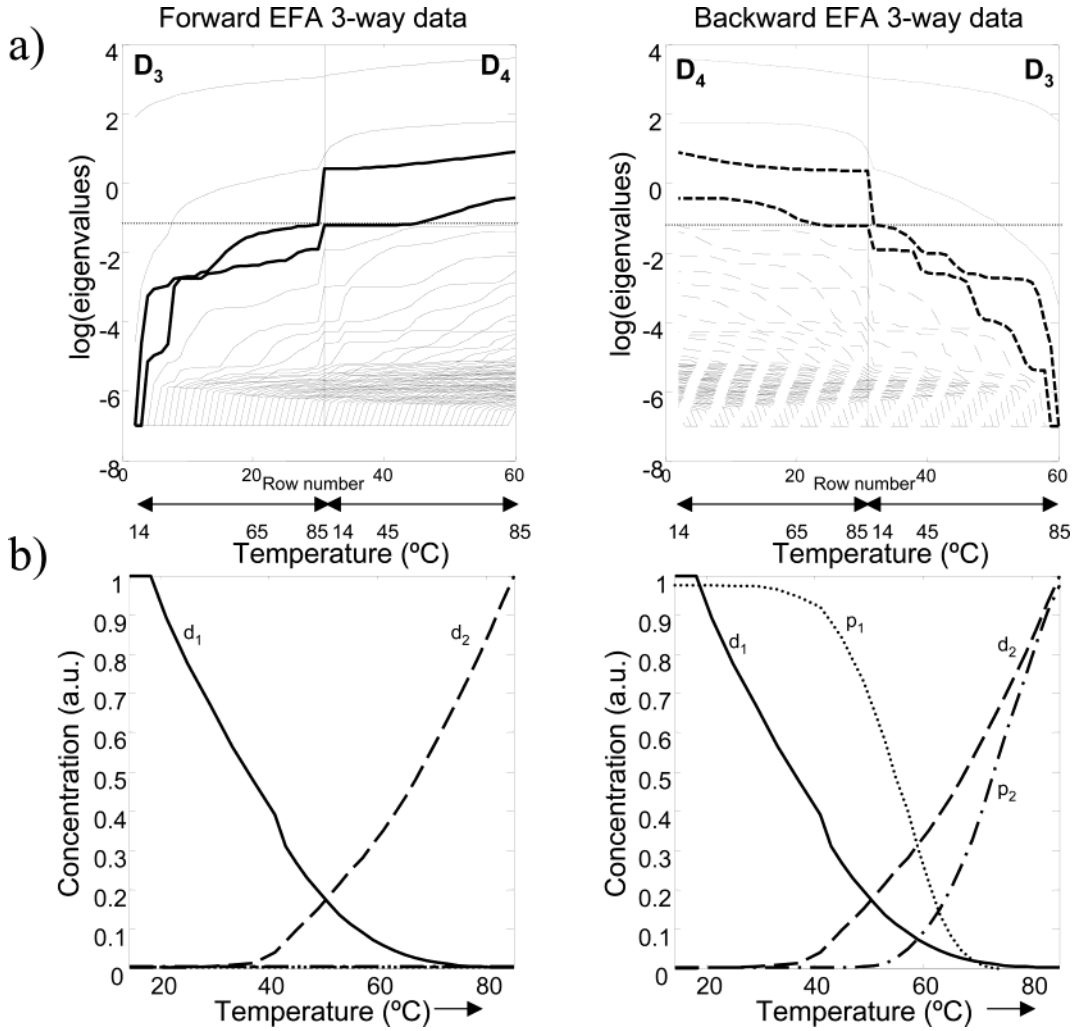


Figure 4. MIR data exploratory analysis. (a) Results from adapted EFA algorithm on the full-rank column-wise augmented data set [$\mathbf{D_3}$;$\mathbf{D_4}$]. Left plot: forward EFA. Right plot: Backward EFA. The vertical lines separate the information related to the deuterium oxide experiment ($\mathbf{D_3}$) and the protein solution experiment ($\mathbf{D_4}$). Unexplained lines and linestyles as in Figure 2. (b) Initial concentration estimates derived from EFA analysis. Left plot: concentration estimates related to $\mathbf{D_2}$. Right plot: concentration estimates related to $\mathbf{D_4}$. d_1, d_2, p_1 and p_2 as in Figure 2.

1490 nm to the first overtone of free NH stretching, 1523 nm to the first overtone of amide A, and 1600 nm to the combination of free NH stretching with the first overtone of amide II. The peak at 2050 nm was thought to describe a combination band of amide A and amide II. According to that mentioned above in this section, these spectral features could also be assigned to the pure protein spectra recovered.

Resolution of Changes in the Protein Structure Using Midinfrared Spectroscopy. Midinfrared spectroscopy has been used extensively in protein structural studies. The temperature-dependence of β-lactoglobulin protein structure was also studied from the MIR measurements recorded in the spectral region 1950−1350 cm^{-1}, which contains the amide I′, amide II′, and amide III′ protein bands. MCR−ALS analysis of the three-way data set [$\mathbf{D_3}$;$\mathbf{D_4}$] is performed to describe the changes in the protein structure of β-lactoglobulin in D_2O. The initial concentration estimates of this data matrix are obtained as in the previous NIR section. As in NIR, only two protein conformations, p_1 and p_2, are also needed to explain the protein structure changes increasing the temperature using MIR measurements (see Figure 4). Analogous constraints to the NIR analysis are applied to the $\mathbf{C}$ and $\mathbf{S}^T$ matrices, because the general properties of the monitored process and the spectroscopic techniques used are similar.

Resolved concentration and spectra profiles are shown in Figure 5b and c, respectively. The MCR−ALS-resolved concentration profiles show the evolution of the concentration of the different species present in the process. The temperature-dependent evolution of D_2O can be also modeled with two species, d_1 and d_2, and is rather similar to the evolution obtained with NIR analysis (see Figure 3). The comments in the NIR section about the inadequacy of a model with more than two water species apply also to this section. The inclusion of a third species does not make sense in view of the variance percent explained by the third eigenvalue, <0.1%, and the lack of improvement in the resolution results. In the $\mathbf{C_i}$ submatrix related to the series of β-lactoglobulin spectra in D_2O, the species from the protein (p_1 and p_2) and from the deuterium oxide can be successfully modeled. The crossing temperature between the conformations p_1 and p_2 is 63.9 °C. Comparing this crossing temperature with the proposed mechanism in the Introduction scheme, this conformational transition

(46) Sadler, A. J.; Horsch, J. G.; Lawson, E. Q.; Harmatz, D.; Brandau, D. T.; Middaugh, C. R. *Anal. Biochem.* **1984**, *138* (1), 44−51.

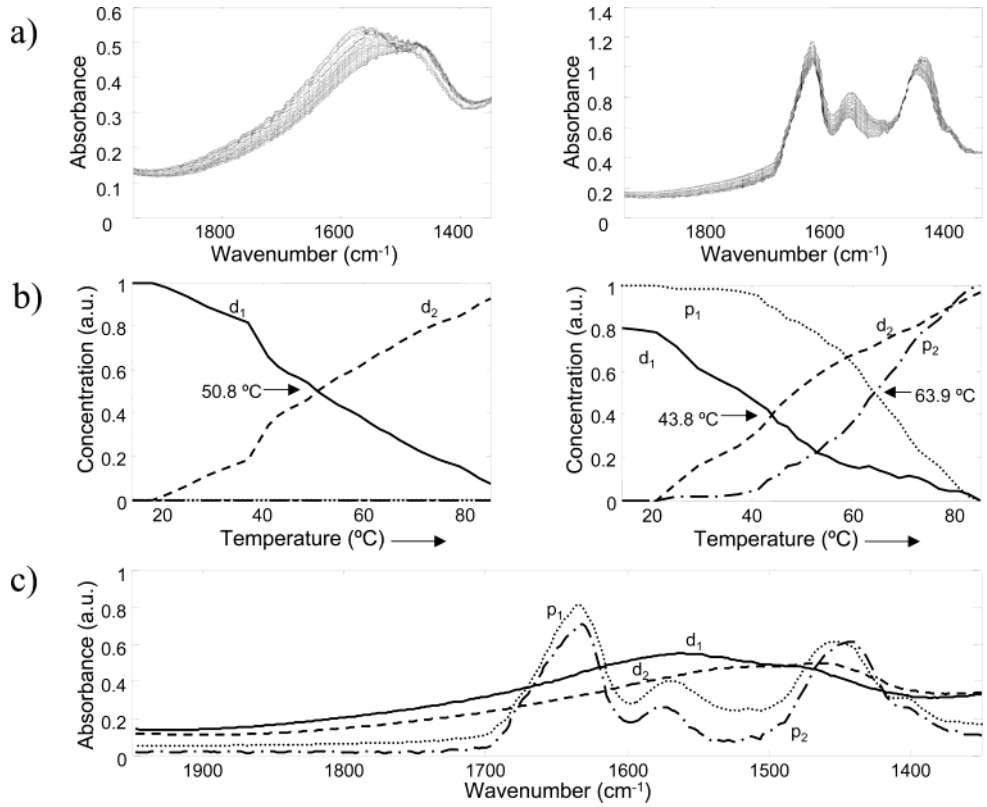


Figure 5. Raw data [$\mathbf{D_3}$;$\mathbf{D_4}$] and MCR−ALS results from temperature-dependent monitoring of β-lactoglobulin with MIR. (a) Raw data. Left plot: series of deuterium oxide spectra ($\mathbf{D_3}$). Right plot: series of protein spectra in D_2O ($\mathbf{D_4}$). (b) Resolved concentration profiles. Left plot: related to $\mathbf{D_3}$. Right plot: related to $\mathbf{D_4}$. d_1, d_2, p_1, and p_2 as in Figure 2.

may be identified as the step from the R-type state to the molten globule. The crossing protein temperature obtained with MIR measurements is different from that obtained with NIR measurements. Therefore, each region of the infrared spectrum seems to detect a different temperature-dependent conformational change of the β-lactoglobulin.

The resolved spectra confirm the capability of MCR−ALS to model separately the protein and the solvent contributions from the series of β-lactoglobulin spectra in D_2O. Indeed, this fact becomes evident through the consistency among the shapes of the pure spectra recovered and the different shape of the raw spectra collected in the deuterium oxide and the protein experiments. The increase in the temperature causes a shift of the D_2O band to a lower wavenumber. Spectra obtained for the p_1 and p_2 protein conformations present the amide I′ maximum around 1635 cm^{-1}, characteristic of β-proteins. The differences between the p_1 and p_2 resolved spectra are principally observed in the region of the amide II′, where the absorbance in the amide II′ band decreases with the temperature. The absorbance in the pure protein spectra recovered is null in the spectral region of 1950−1700 cm^{-1}, in accordance with MIR protein measurements reported in the literature.[47] In this case, the pure protein spectra obtained by MCR−ALS circumvent the dangerous approach of baseline correction that often ends up in introduction of artifacts in raw protein spectra, particularly when the background contribution is intense and does not remain constant during the process.[47,48]

Resolution of Changes in the Protein Structure Using Near- and Midinfrared Spectroscopy. MCR−ALS analysis of the augmented data set [$\mathbf{D_1}$ $\mathbf{D_3}$;$\mathbf{D_2}$ $\mathbf{D_4}$] is performed to describe changes in the protein structure of β-lactoglobulin in D_2O as monitored by NIR and MIR spectroscopies (the space between matrices indicates a row-wise augmented matrix according to the MATLAB notation). As shown in Figure 1b, the MCR−ALS analysis of this column- and row-wise augmented matrix provides a column-wise augmented matrix **C** formed by two submatrices, one related to the row-wise appended MIR and NIR experiments performed with pure deuterium oxide [$\mathbf{D_1}$ $\mathbf{D_3}$] and the other related to the analogous experiments performed with protein solution in D_2O [$\mathbf{D_2}$ $\mathbf{D_4}$], and a row-wise augmented $\mathbf{S^T}$ matrix formed by two $\mathbf{S_i^T}$ submatrices, one with the pure spectra related to the column-wise appended experiments performed with NIR [$\mathbf{D_1}$;$\mathbf{D_2}$] and the other related to the analogous experiments performed with MIR [$\mathbf{D_3}$;$\mathbf{D_4}$].

Now, the initial concentration estimates of the D_2O system are made using the classical EFA approach on the row-wise data matrix [$\mathbf{D_1}$ $\mathbf{D_3}$] that monitors the temperature-dependent evolution of deuterium oxide by NIR and MIR. Again, a model with two water species is sufficient to describe the variation measured by MIR and NIR spectra. This agrees with previous results and confirms that the temperature-dependence of D_2O is equally described by NIR and MIR spectroscopies. The variance percent linked to the third eigenvalue is lower than 0.1, as well.

The EFA adapted to rank-deficient systems is now applied to the full-rank augmented matrix [$\mathbf{D_1}$ $\mathbf{D_3}$;$\mathbf{D_2}$ $\mathbf{D_4}$] to obtain the initial concentration estimates for the protein. In this occasion, three protein conformations (p_1, p_2, and p_3) seem to be detected (see

(47) Rahmelow, K.; Hübner, W. *Appl. Spectrosc.* **1997**, *51* (2), 160−170.

(48) Dong, A.; Huang, P.; Caughey, W. S. *Biochem.* **1990**, *29* (13), 3003−3008.

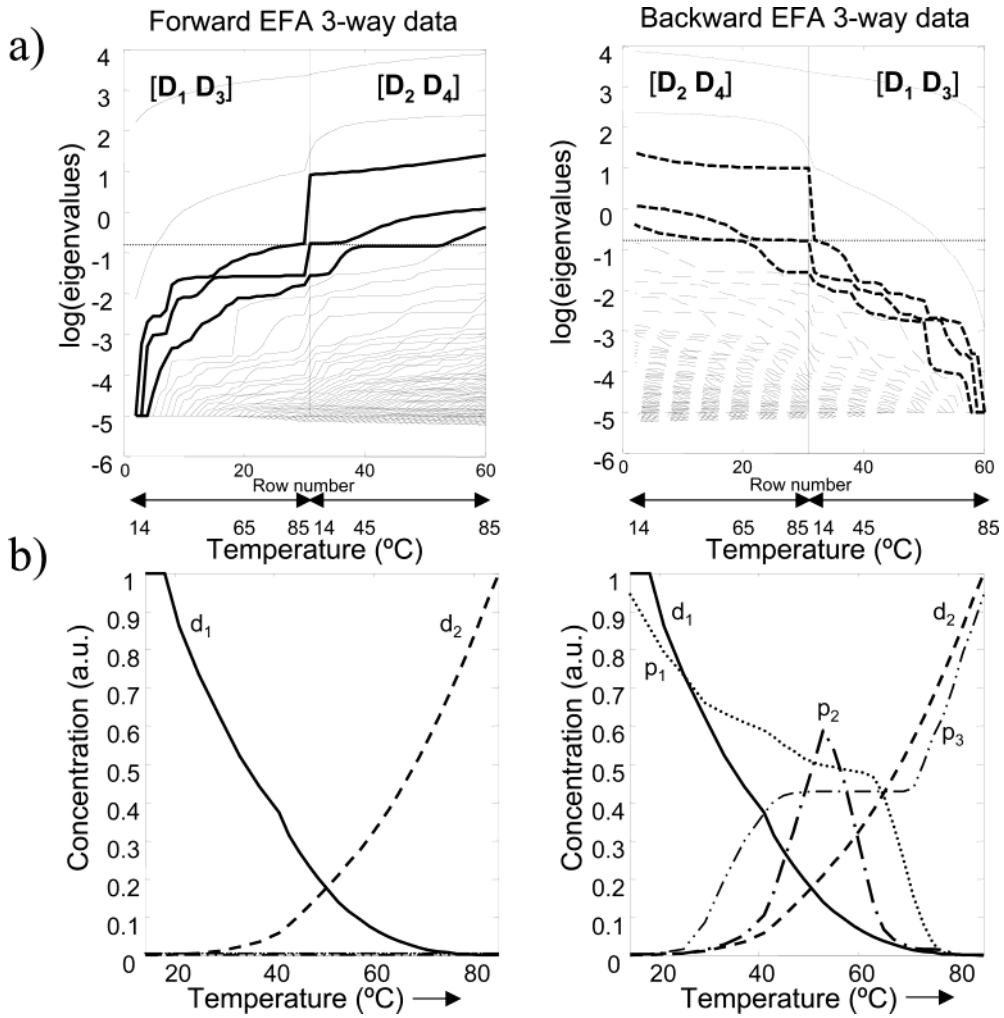


Figure 6. Combined NIR/MIR exploratory analysis. (a) Results from adapted EFA algorithm on the full-rank row- and column-wise augmented [$\mathbf{D_1\ D_3}$;$\mathbf{D_2\ D_4}$] data set. Left plot: forward EFA. Right plot: Backward EFA. The vertical lines separate the information related to the row-wise appended deuterium oxide experiments [$\mathbf{D_1\ D_3}$] and the row-wise appended protein solution experiments [$\mathbf{D_2\ D_4}$]. Unexplained lines and linestyles as in Figure 2. (b) Initial concentration estimates derived from EFA analysis. Left plot: concentration estimates related to [$\mathbf{D_1\ D_3}$]. Right plot: concentration estimates related to [$\mathbf{D_2\ D_4}$]. d_1 and d_2, as in Fifure 2, p_1, p_2, and p_3 are the three protein species.

Figure 6 for EFA analysis and derived initial estimates). Constraints are applied as in previous sections, and one should note the special importance of using local rank constraints setting the absence of the different protein conformations in certain temperature ranges to achieve the successful resolution of such a complex system.

Resolved concentration and spectra profiles are shown in Figure 7a and b, respectively. NIR and MIR provide a common description of the temperature-dependence of D_2O, which proves to be consistent in both pure deuterium and protein experiments. However, the combined analysis of NIR and MIR measurements needs three protein conformations (p_1, p_2, and p_3) to describe the temperature-dependence of β-lactoglobulin structure in D_2O. The concentration profiles and spectra obtained with the combined analysis of NIR and MIR measurements are consistent with the results obtained with the analysis of NIR and MIR measurements, separately, and the pure concentration profiles obtained by unconstrained least-squares regression using the original augmented data set [$\mathbf{D_1\ D_3}$;$\mathbf{D_2\ D_4}$] and the row-wise augmented matrix of pure NIR and MIR spectra fully agree with the resolved concentration profiles obtained by MCR−ALS. These two results, obtained in an independent manner, help to confirm and validate the protein model obtained from MCR−ALS analysis.

Figure 8 shows the complete mechanistic and structural description of the temperature-dependent conformational transitions of β-lactoglobulin through the evolution of their concentration profiles and shapes of their spectra, respectively. The points in each concentration profile have been nonlinearly fitted[49] to clarify the evolution of the protein process (note that no chemical model has been used to perform this fit). The crossing temperature between the concentration profiles of p_1 and p_2 is 46 °C, and between those of p_2 and p_3, is 63 °C. According to the proposed mechanism in the Introduction, the transition at 46 °C corresponds to the conformational change between the native conformation (p_1) and the R-type state (p_2), and the transition at 63 °C would take place between the R-type state (p_2) and the molten globule state (p_3). Only the combined analysis of NIR and MIR measurements allowed for the detection and modeling of the three protein conformations involved in the process that would otherwise not be resolvable because of the high similarity in their pure spectral

(49) AISN Software Inc., TableCurve 2D Windows v4.06.

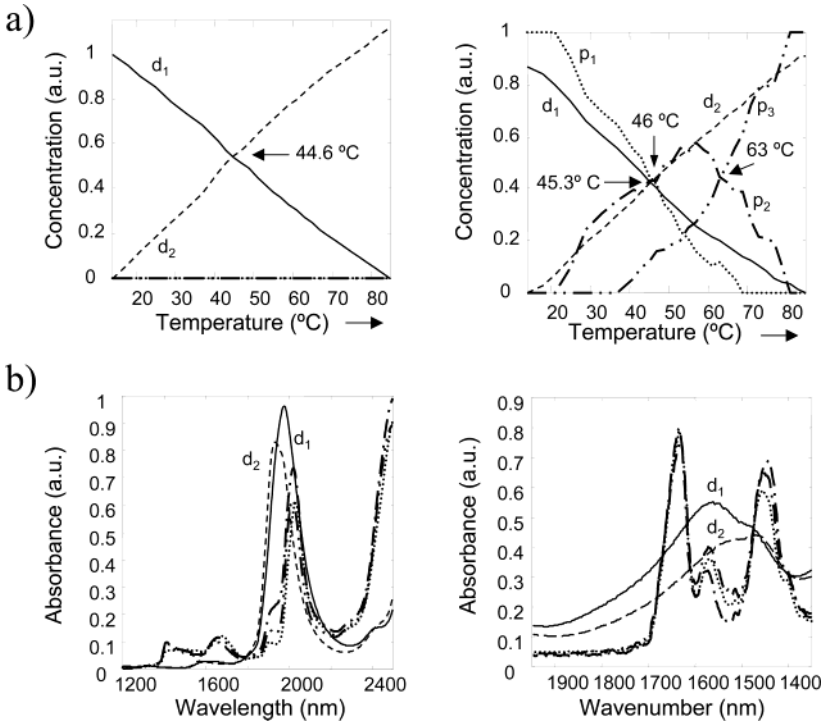


Figure 7. MCR−ALS results from the temperature-dependent monitoring of β-lactoglobulin with NIR and MIR. Thin lines mark deuterium oxide contributions; bold lines mark protein contributions. (a) Resolved concentration profiles. Left plot: related to **D_3**. Right plot: related to **D_4**. d_1, d_2, p_1, p_2, and p_3 as in Figure 6.

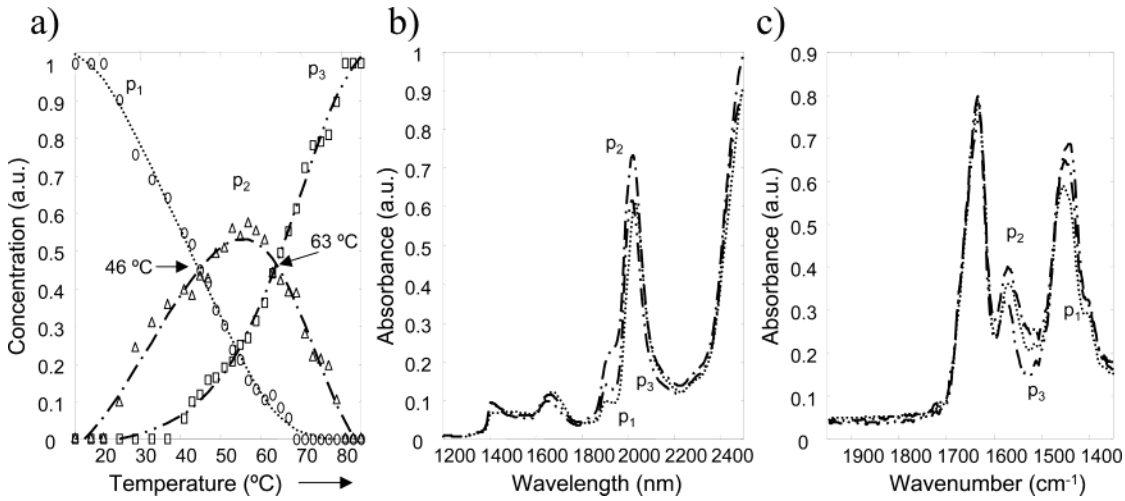


Figure 8. Graphical description of the temperature-dependent conformational transitions of β-lactoglobulin (p_1, native conformation; p_2, R-type state; p_3, molten globule state). (a) Evolution of the concentration profiles. (b) Pure NIR spectra. (c) Pure MIR spectra.

shapes. The intermediate R-type state conformation, which cannot be isolated experimentally, can be fully characterized from a mechanistic and structural point of view.

CONCLUSIONS

The combined MCR−ALS analysis of NIR and MIR measurements has allowed for the complete mechanistic and structural description of the temperature-dependence of β-lactoglobulin through the recovery of the resolved pure concentration profiles and spectra of all the protein conformations involved in the process, including an intermediate. These results support a mechanism proposed in the literature[19,20] that suggests that the thermal denaturation of β-lactoglobulin proceeds by a two-step process evolving from the native conformation to a molten globule state, passing through an R-type intermediate conformation. The application of curve resolution to multispectroscopic data has proven to be an excellent external tool to validate mechanisms of protein processes proposed by using either single process curves or deconvolution-based methods, subject to a major uncertainty in the generation and interpretation of results.

Received for review April 14, 2003. Accepted July 25, 2003.

AC0343883