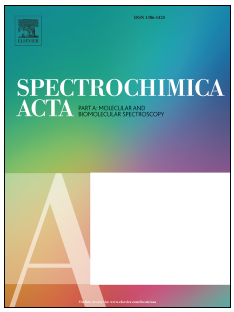
Journal Pre-proof



pH titration of β-lactoglobulin monitored by laser-based Mid-IR transmission spectroscopy coupled to chemometric analysis

Andreas Schwaighofer, Mirta R. Alcaraz, Laurin Lux, Bernhard Lendl

PII: S1386-1425(19)31026-1

DOI: https://doi.org/10.1016/j.saa.2019.117636

Reference: SAA 117636

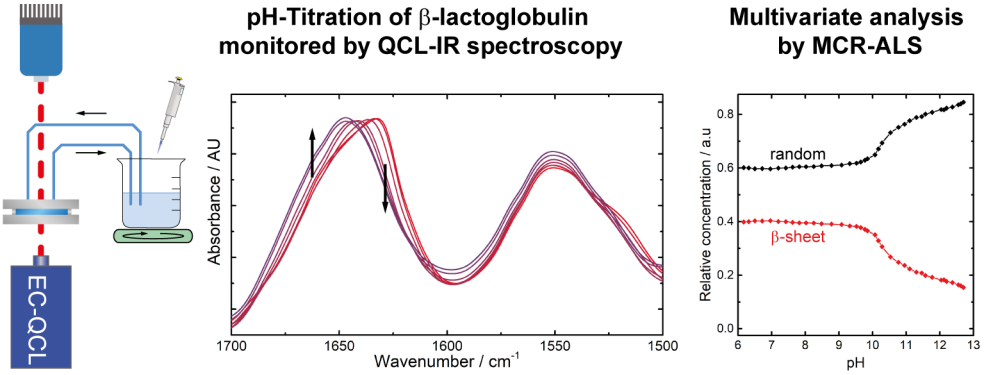
To appear in: *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*

Received Date: 10 July 2019

Revised Date: 27 September 2019

Accepted Date: 7 October 2019

Please cite this article as: A. Schwaighofer, M.R. Alcaraz, L. Lux, B. Lendl, pH titration of β-lactoglobulin monitored by laser-based Mid-IR transmission spectroscopy coupled to chemometric analysis, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* (2019), doi: https://doi.org/10.1016/j.saa.2019.117636.



pH Titration of β-Lactoglobulin Monitored by Laser-based Mid-IR Transmission Spectroscopy coupled to Chemometric Analysis

Andreas Schwaighofer[1], Mirta R Alcaraz[2,3,4], Laurin Lux[1], Bernhard Lendl[1,*]

[1] Institute of Chemical Technologies and Analytics, Vienna University of Technology, Getreidemarkt 9/164-UPA, 1060 Vienna, Austria

[2] Laboratorio de Desarrollo Analítico y Quimiometría (LADAQ), Cátedra de Química Analítica I, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral, Ciudad Universitaria, Santa Fe (S3000ZAA), Argentina

[3] Departamento de Química Inorgánica, Analítica y Química Física, INQUIMAE, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Intendente Güiraldes 2160, Ciudad Universitaria, Pabellón 2, Buenos Aires (C1428EGA), Argentina

[4] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Godoy Cruz 2290 CABA (C1425FQB), Argentina.

*Corresponding author: Email: bernhard.lendl@tuwien.ac.at

Abstract

A novel external cavity-quantum cascade laser (EC-QCL)-based setup for mid-IR transmission spectroscopy in the amide I and amide II region was employed for monitoring pH-induced changes of protein secondary structure. pH titration of β-lactoglobulin revealed unfolding of the native β-sheet secondary structure occurring at basic pH. Chemometric analysis of the dynamic IR spectra was performed by multivariate curve resolution-alternating least squares (MCR-ALS). Using this approach, spectral and abundance distribution profiles of the conformational transition were obtained. A proper post-processing procedure was implemented allowing to extract information about pure protein spectra and spurious signals that may interfere in the interpretation of the system. This work demonstrates the potential and versatility of the EC-QCL-based IR transmission setup for flow-through applications, benefitting from the high available optical path length.

Keywords: mid-infrared, quantum cascade laser, protein structure, pH titration, β-lactoglobulin, chemometrics

1. Introduction

Mid-infrared (mid-IR) spectroscopy is a powerful and well-established analytical technique to study the structure of biological macromolecules, e.g., proteins [1]. Vibrations of the polypeptide repeating units of proteins result in nine characteristic group frequencies in the mid-IR region referred to as amide bands. For investigation of the protein secondary structure, the amide I (1700-1600 cm^{-1} – originating from the C=O stretching and N−H in-phase bending vibration of the amide group) and the amide II bands (1600-1500 cm^{-1} – arising from N−H bending and C−N stretching vibrations) have been recognized to be most characteristic to secondary structure [1,2]. The sensitivity to individual secondary structure elements originates in differing patterns of hydrogen bonding, dipole-dipole interactions and geometric orientations in the α-helices, β-sheets, turns and random coil structures that induce different frequencies of the C=O vibrations [3]. Although the amide I band demonstrates to be most sensitive to conformational changes, it has been shown that additional and more in-depth information about protein secondary structure can be gained by collective analysis of both spectral regions, particularly in combination with chemometric analysis [4-6].

An experimental limitation to investigations of protein secondary structure in aqueous solutions with state-of-the-art Fourier transform infrared (FTIR) spectrometers is constituted by the low feasible path lengths of the transmission cells. This constraint originates from the combination of two factors: first, the HOH-bending band of water near 1645 cm^{-1} with a high molar absorption coefficient, which overlaps with the protein amide I band; and second, the low emission power provided by conventional thermal light sources (globars). As a consequence, path lengths most commonly used for IR transmission measurements of proteins in aqueous solutions are in the range of 8 μm to avoid total IR absorption in the region of the HOH-bending band. This limitation comes along with laborious cell and sample handling as well as the need

for high protein concentration (>10 mg mL^{-1}) [7]. The low feasible path lengths for FTIR transmission measurements of proteins in aqueous solution are a considerable impairment for the robustness of analysis and impede flow-through measurements and high-throughput applications. This prevents semi-automated measurements in process analytical applications, but also basic monitoring of titrations in transmission mode to follow conformation changes of proteins as a function of external perturbation (pH, ligand concentration, denaturant concentration, etc) [8].

With the introduction of quantum cascade lasers (QCL) a significant step forward was made towards resolving these restrictions of low power light sources in mid-IR spectroscopy [9]. They provide polarized and coherent light with spectral power densities several orders of magnitude higher than thermal light [10]. Initially, the advantages of this new light source were predominantly exploited in the gas phase and custom-built setups have gained manifold implementations in process analytical applications as well as in biomedical spectroscopy [11]. In the last decade, a new type of these mid-IR lasers, external cavity-QCLs (EC-QCLs), became commercially available combining high emission powers with spectral tuning ranges of several hundred wavenumbers. The high available emission powers enabled to increase the optical path up to 38 µm for transmission measurements in the protein region [12]. EC-QCL-based IR transmission measurements have been successfully accomplished for the analysis of protein secondary structure [13-15]. Furthermore, the feasibility of protein discrimination and quantitation in commercial bovine milk samples has been demonstrated by QCL-IR spectroscopy and evaluation of the amide I band using partial least squares (PLS) modelling [16-18]. Recently, an EC-QCL-based IR transmission setup was introduced for the analysis of the protein amide I and II regions, which favourably competes with FTIR spectroscopy regarding the signal to noise ratio at equal data acquisition times [19].

β-Lactoglobulin (β-LG) is a predominantly β-sheet protein consisting of 162 amino acids that are folded in three α-helices and nine strands of antiparallel β-sheet, that are wrapped in a way to form an antiparallel β-barrel [20,21]. It is a main constituent of bovine milk and therefore it has been the subject of many studies probing its function and structure in relation to internal and external perturbation factors such as protein concentration [22-24], temperature [25-27], pressure [28], pH [21,29], ionic strength [30], denaturant concentration [31,32], metal ion concentration [33], among others. In this regard, it has been found that β-LG features multiple structural transitions along the entire pH range. The monomeric structure of β-LG prevailing at low pH values starts to dimerize at pH 3. Between pH 4 and pH 5, β-LG undergoes a dimer-to-octamer transition. Throughout these transitions taking place in the acidic pH range, the global secondary structure of the antiparallel β-barrel does not show evident changes [24]. Further increase of the pH value into the alkaline region leads to disruption of the β-sheet secondary structure and formation of disordered secondary structure, as investigated by circular dichroism (CD) spectroscopy [21] and FTIR spectroscopy [34]. Moreover, conformational changes of β-LG induced by pH [24], concentration [35], temperature [25,28,36], pressure [25,28] as well as adsorption [37] were successfully evaluated by FTIR spectroscopy.

Multivariate curve resolution-alternating least square (MCR-ALS) is a widespread iterative soft-modelling technique that allows to discriminate individual contributions of underlying constituents [38,39]. MCR-ALS provides sound and meaningful models with chemically interpretable output in the form of spectral profiles of the compounds and the related abundance distribution profiles along the dynamic process. Due to its flexibility and robustness, it has been successfully used in combination with various analytical techniques, such as chromatography [40], electrophoresis [41], and flow analysis [42], among others. In combination

with spectroscopic techniques, particularly with IR spectroscopy, MCR-ALS has become a highly valuable technique for the study of evolving processes [13,43,44].

In this work, we present a continuous flow-through titration to monitor pH-induced protein unfolding by EC-QCL-based IR spectroscopy. This type of experiment was impeded with conventional FTIR spectroscopy due to experimental problems such as defects in liquid tightness and cell clogging arising from the necessity of the low path length required for IR measurements of proteins in aqueous solution. To this end, β-LG was chosen as a model protein as it depicts a gradual transition from β-sheet to random secondary structure in the alkaline pH range. MCR-ALS analysis was utilized to obtain pure spectral and abundance distribution profiles of the pH-induced transition between native and denaturated secondary structure.

2. Materials and methods

2.1. Reagents and Samples

Sodium hydroxide (NaOH) solution 50 % in water, potassium chloride (KCl) p.a. and β-lactoglobulin (β-LG) from bovine milk (≥85 %) were obtained from Sigma-Aldrich (Steinheim, Germany) and used as purchased. Ultrapure water (18 MΩ) was used for preparation of all solutions and was obtained with a Milli-Q water purification system from Millipore (Bedford, USA).

2.2. EC-QCL Setup

A detailed description of the custom-made EC-QCL setup can be found elsewhere [19]. Briefly, a water-cooled external-cavity quantum cascade laser (Hedgehog, Daylight Solutions Inc., San Diego, USA) was used operating at a repetition rate of 100 kHz and a pulse width of

5 µs. Spectra were recorded in the spectral tuning range of 1730–1470 cm^{-1}, covering the amide I and amide II region of proteins, at a scan speed of 1200 $cm^{-1}s^{-1}$. The mid-IR light was focused on the detector element by a gold plated off-axis parabolic mirror with a focal length of 43 mm. A thermoelectrically-cooled MCT detector operating at -78 °C (PCI-10.6, Vigo Systems S.A., Poland) was used as IR detector, as shown in Fig. 1. All measurements were carried out using a custom-built, temperature-controlled flow cell equipped with two mid-IR transparent CaF_2 windows and 31 µm-thick spacer, at 25 °C. A mesh was employed to attenuate the laser intensity and a wedged sapphire window (2.5 mm thickness) was used to selectively reduce the laser intensity in the amide II region. To reduce the influence of water vapor, the setup was placed in a housing of polyethylene foil and constantly flushed with dry air.

The measured signal was processed by a lock-in amplifier (Stanford Research Systems, CA, USA) and digitized by a NI DAQ 9239 24-bit ADC (National Instruments Corp., Austin, USA). Each single beam spectrum consisting of 6000 data points was recorded during the tuning time for one scan of approx. 250 µs. A total of 100 scans were recorded for background and sample single beam spectra at a total acquisition time of 53 s.

A data processing routine was devised to remove those scans that are shifted more than 0.1 cm^{-1} based on evaluation of the similarity index. Using this approach, approximately 3 % of the recorded scans were sorted out. At last, minor Fourier filtering (cutoff frequency of 150−200) was applied to remove residual noise in the final absorption spectra.

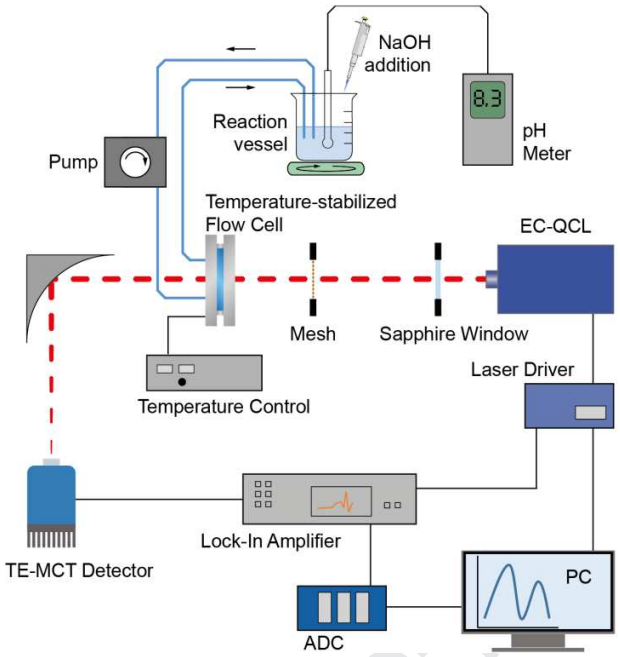


Fig. 1. Experimental setup for pH titration monitored by EC-QCL IR transmission spectroscopy.

2.3. Titration Procedure

400 mg of β-LG were dissolved in 20 mL of water to obtain a protein solution of 20 mg mL^{-1}. Subsequently, the pH value was adjusted to 6.0 by addition of HCl. KCl (final concentration: 100 mmol L^{-1}) was added for stabilization of the ionic strength throughout the titration. pH measurements were carried out using a pH330i (Wissenschaftlich-Technische Werkstätten GmbH, Weilheim, Germany) potentiometer equipped with a Sentix® Mic-D (Wissenschaftlich Technische Werkstätten GmbH, Weilheim, Germany) combined glass electrode and temperature probe. pH titration was performed by adding 10 µL aliquots of 1 M, 2.5 M, 5 M, 10 M and 15.4 M NaOH to 20 mL of the original protein solution to achieve pH increments of approximately 0.1-0.3 between spectra acquisition. In total, 37 IR spectra were obtained covering the pH range from 6.0 to pH 12.7. The solution was continuously pumped through the IR transmission cell with a pump speed of 0.9 mL min^{-1} and stopped while recording

of the IR spectra. The temperature was monitored throughout the titration, being constant in the range of ± 0.1 °C.

For assessing the changes of the HOH-bending band of water throughout the introduced pH change, a similar titration procedure was performed with 20 mL of medium solution, constituted by water containing 100 mmol L^{-1} of KCl. Here, 35 IR spectra were obtained between pH 6.0 and pH 12.9.

For the calculation of all absorption spectra, ultrapure water was taken as a reference.

2.4. MCR-ALS

MCR-ALS is a soft-modelling iterative method that focuses on bilinear decomposition of a data matrix into two submatrices containing chemically meaningful information of contributions of the pure compounds involved in the system [45]. In spectroscopy, the decomposition results provide information about spectral behavior of the individual sample constituents and the related abundance. Here, the MCR-ALS modelling approach was chosen, because it is capable to provide chemically interpretable profiles from a bilinear decomposition of a unique matrix. The majority of bilinear decomposition algorithms are used with quantitative aims and do not offer the possibility to reveal spectral information of the system constituents. Thus, MCR-ALS seems to be the best option to unravel the spectral and chemical behavior of the present system.

One of the most compelling characteristics of MCR-ALS resolution is its general applicability without prior information about the system under study. However, to achieve chemically meaningful component profiles, additional knowledge can be incorporated [46]. This information can be introduced either as initial estimates of the iterative optimization or through

191 the implementation of chemical or mathematical constraints, such as non-negativity, unimodality,
192 normalization and closure, among others [47].

193 In the present work, pH-induced conformational change was monitored in the amide I and
194 amide II spectral region by varying the pH value between 6.0 and 12.7. For a complete titration
195 experiment, 37 × 4451 and 35 × 4451 matrices were obtained for protein and medium solution,
196 respectively, and subjected to MCR-ALS analysis.

197

198 **2.5. Software**

199 Data processing and MCR-ALS analysis were performed in MATLAB R2014
200 (MathWorks, Inc., Natick, MA, 2014). MCR-ALS algorithms were implemented by using MCR-
201 ALS GUI 2.0 graphical interface available at http://www.mcrals.info. Discrete wavelet transform
202 was implemented by using the wavelet package allocated into MATLAB R2014.

203 Daylight Solution driver software (Hedgehog, Daylight Solutions Inc., San Diego, USA)
204 was used for laser control. For data acquisition and temperature control a custom-made LabView-
205 based GUI (National Instruments Corp., Austin, USA) was utilized.

206

207 **3. Results and Discussion**

208 **3.1. IR Spectra of pH-induced Conformational Change of β-Lactoglobulin**

209 QCL-IR spectra of the pH-induced change of β-LG were recorded between pH 6.0 and
210 pH 12.7. In Fig. 2, the QCL-IR absorption spectra in the amide I and amide II region are shown.
211 The QCL-IR spectra at pH 6.0 shows a band maximum in the amide I region at 1634 cm^{-1} with a
212 shoulder at 1655 cm^{-1}, and a broad band at approximately 1550 cm^{-1} with a shoulder at 1520 cm^{-1}
213 in the amide II region, characteristic for the IR signatures of β-LG featuring a predominantly β-
214 sheet secondary structure [5,48]. Upon stepwise increase of the pH value, the maximum of the

amide I band is shifted towards higher wavenumbers and the band shape loses its distinctive form. The IR spectrum at the highest measured pH value shows broad featureless bands with maxima at 1645 cm^{-1} and 1550 cm^{-1} in the amide I and amide II region, respectively, indicating a disordered secondary structure [1,34]. Even though α-helical structure is present in β-LG, no bands could be attributed to this type of secondary structure in the recorded data set, presumably because the induced conformational change is minor compared to the pronounced transition from β-sheet to disordered secondary structure.

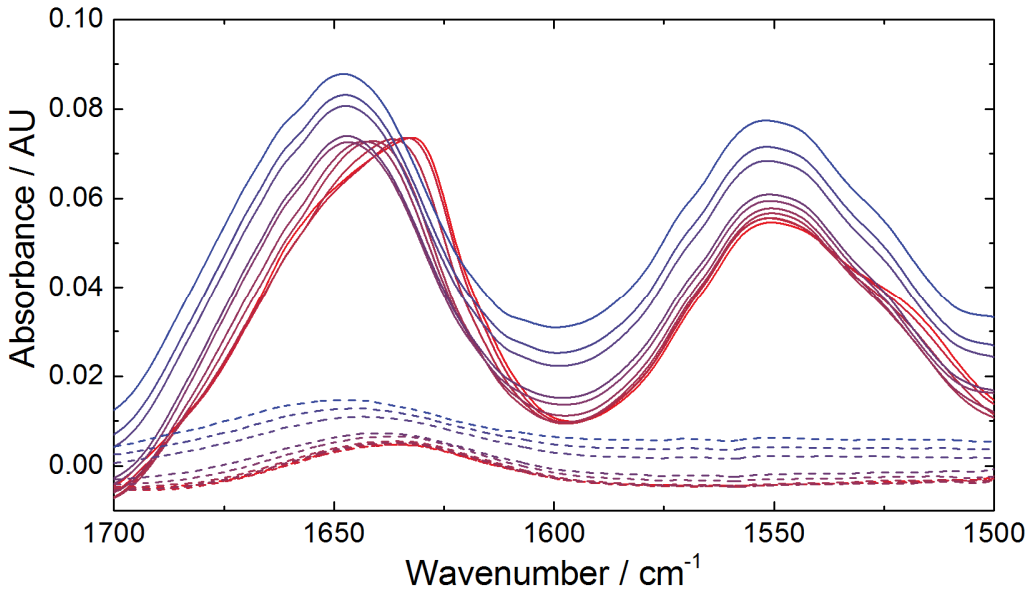


Fig. 2. Fourier-filtered QCL-IR spectra of 20 mg mL^{-1} β-LG recorded as a function of pH between 6.0 (red solid line) and 12.9 (blue solid line) and QCL-IR spectra of water recorded as a function of pH between 6.0 (red dashed line) and 12.7 (blue dashed line).

After reaching the end point of titration at high pH value, HCl was added to test the reversibility of the conformational change however no spectral changes were observed, suggesting the irreversibility of protein unfolding reaction (data not shown) [49].

Fig. 2 also shows the spectral changes of the medium (100 mmol L^{-1} KCl) induced by the variation of the pH value. The spectra reveal an absorption band at 1633 cm^{-1} attributed to KCl-solvated water [50], as well as an increasing drift of the baseline at higher pH values [51]. Consequently, the pH-induced change of the protein secondary structure in the IR spectra is superimposed by spectral changes of the medium solution, which hampers direct analysis of the dynamics of the conformational change. Consequently, MCR-ALS analysis was performed for in-depth analysis of the presented pH-resolved IR spectra and further evaluation of the conformational changes.

3.2. Chemometric Analysis by MCR-ALS

Detailed chemometric analysis of the recorded IR spectra was performed by applying MCR-ALS. This approach has been successfully employed for the analysis of the progression of protein secondary structural change in dynamic spectroscopic data [4,13,52].

For the analysis of complex systems by MCR-ALS, several aspects ought to be considered in order to achieve reliable and meaningful results. One consideration relates to the initialization of the ALS step. Here, the number of components involved in the system and their initial estimates are required. The number of components refers to the spectroscopically active species that explain the system and is usually unknown, therefore, it is estimated by applying singular value decomposition (SVD) or principal component analysis (PCA). Regarding ALS initialization, different first estimates can be used to initiate the modelling. The initial spectral profiles can either reflect the system under study or comprise fully random values. In the first mentioned case, if no spectral information of the pure constituents is available, they can be estimated by means of different methodologies, such as the analysis of the purest variables [53] or the so-called Evolving Factor Analysis (EFA) [54]. Even though it is reported that the nature of the initial estimates does not significantly affect the final result of chemometric modelling

[55], it has been demonstrated that estimates closer to the real information aids achieving the true solution in systems of unknown composition or of extremely complex nature [56]. For evaluation of an evolving system, e.g., protein conformation monitoring in aqueous medium by IR spectroscopy, the estimation of the initial profiles becomes challenging since no information about system composition and information of pure constituents are available in advance.

In pH-titration experiments, a challenge for chemometric modelling of the system is that not only the analyte but also the medium undergoes spectral changes induced by alteration of the pH value, which depend on the surrounding media, among other parameters. This phenomenon, which can be observed in the form of changes in the offset of the baseline or in the band position and shape, arises from the concentration of ions that modify the pH (H^+ and OH^-) as well as the solvation of the ions, KCl in the present case, by the water molecules. This effect can be seen in Fig. 2, where absorption spectra of the medium at different pH are depicted (absorption spectra of the medium were obtained against pure water). In the case of protein titration, the spectral changes of the medium are overshadowed by protein signals, which are more intense than the ones of the medium, precluding the proper extraction of the profiles of all constituents by chemometric decomposition.

In the context of this complex scenario, the following procedure was implemented in the attempt to extract the most reliable, accurate and representative information about the system. For MCR-ALS analysis, un(Fourier-)filtered QCL-IR spectra were employed. First, medium-titration and protein-titration data were individually subjected to MCR-ALS analysis. In both cases, the optimum number of components that explain the system estimated by SVD was evaluated to be three and four, respectively. Abundance distribution profiles obtained from the analysis of the purest variables to each individual data set were utilized as initial estimates of the ALS optimization. For the medium system, non-negativity and spectra normalization constraints were

implemented, while for the protein system, non-negativity, spectra normalization and closure for those components involved in the protein evolution were applied. Once the optimized results were obtained from the separate analyses, the optimized spectral profiles were combined in order to build a unique set of spectral profiles. It should be noted that two of the spectral profiles obtained from the protein titration analysis correspond to the medium. Subsequently, those profiles were replaced with the profiles obtained from the medium-titration analysis. Therefore, five optimized spectral profiles were assembled into a matrix, two for the protein (obtained from protein analysis) and three for the medium (obtained from the medium analysis), and were considered as initial estimates for a subsequent extended MCR-ALS analysis of an augmented data matrix.

The augmented data matrix was built by appending the original medium data matrix to the protein data matrix. In this way, a 72×4451 data matrix was obtained and analyzed by the extended version of MCR-ALS. Here, the previously obtained five optimized spectral profiles were utilized as initial estimates of the ALS step. Non-negativity and spectra normalization were implemented as constraints during optimization. In this manner, it was possible to model the contribution of the medium in presence of protein. It is a pronounced benefit of the employed extended MCR-ALS modelling to analyze the reference and the sample matrix in one single model. This is in particular the case for pH titrations, where pH levels cannot be exactly reproduced for the reference and sample titration. Consequently, absorption spectra at individual pH values cannot be directly calculated without introducing spectral artefacts due to pH differences between the reference and sample spectrum. Figure 3 shows the optimized results achieved by MCR-ALS analysis for the protein evolution through the pH variation.

The MCR-ALS analysis resolved pH-dependent (Fig. 3A) and spectral profiles (Fig. 3B) of the protein secondary structures as well as spectral elements of the medium that change with

increasing pH value. The spectral profile attributed to the β-sheet secondary structure features a band maximum at 1630 cm^{-1} in the amide I region and a prominent shoulder around 1520 cm^{-1} in the amide II region (Fig. 3B). The component attributed to the disordered protein structure shows

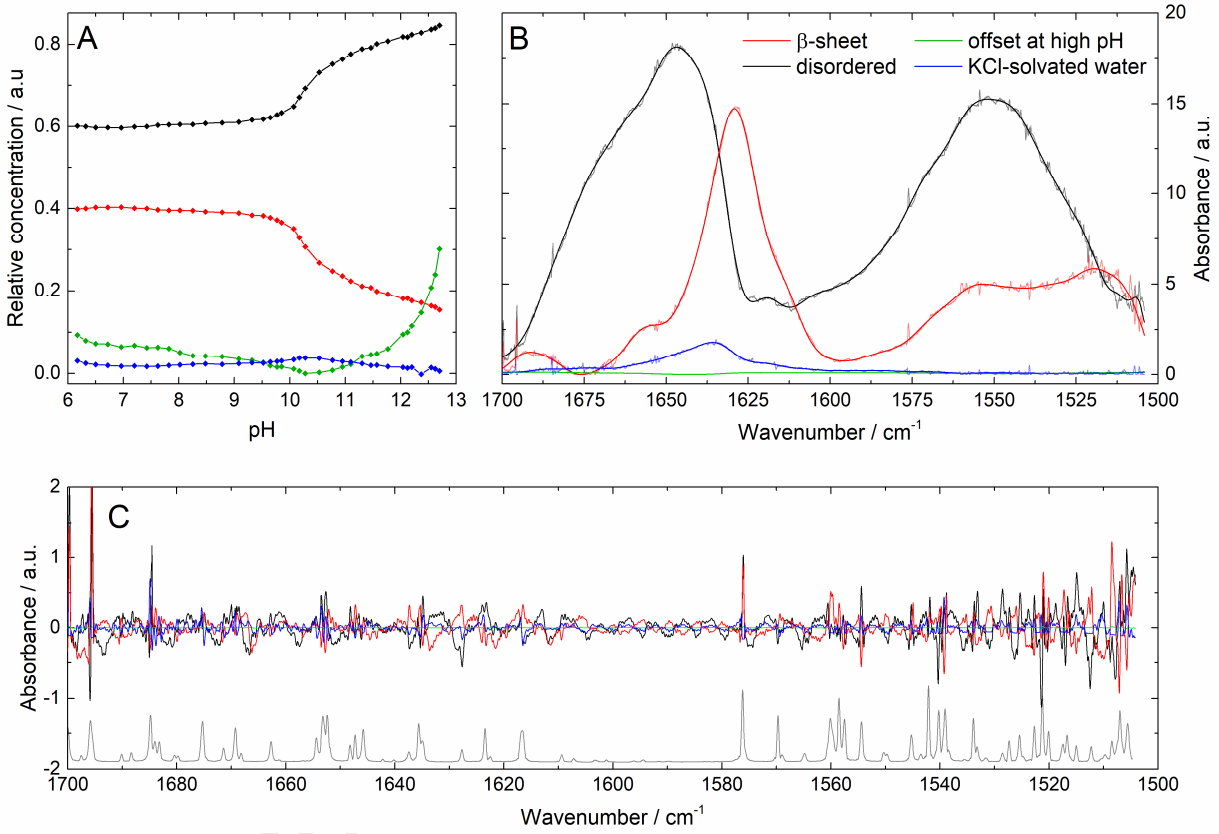


Fig. 3. (A) pH-dependent and (B) spectral profiles retrieved by MCR-ALS for the performed pH-titration of β-lactoglobulin at a concentration of 20 mg mL^{-1} and the pH-titration of the medium. The red and black lines show the β-sheet and disordered secondary structure of the protein, respectively. The green lines indicate the offset of the baseline at elevated pH values and the blue lines exhibit the contribution of KCl-solvated water. Thin lines in (B) are raw spectral profiles received by MCR-ALS, thick lines in (B) were obtained by discrete Meyer wavelet transform

314 (WT). (C) Residuals obtained from the WT for the individual spectral profiles. The grey line
315 shows a (scaled) spectrum of water vapor.

316

band maxima at 1645 cm^{-1} and 1550 cm^{-1} in the amide I and amide II region, respectively. The shapes and maximum positions of the spectral profiles obtained by rigorous application of MCR-ALS analysis fit well to spectral features reported for IR spectra of the respective secondary structure elements. The profiles of the pH-dependent evolution of the two secondary structures (Fig. 3A) reveal that the respective share remains constant between pH 5.0 and 9.5. Upon further increase of the pH value, the fraction of β-sheet secondary structure rapidly decreases while the portion of disordered protein structure increases in the same manner with an inflection point at pH ~10.2. Above pH 11, the rapid conversion of secondary structure is completed, and the change continues at slower and constant rate. This progression of pH induced conversion of β-LG secondary structure agrees well with reported results obtained by CD spectroscopy [21].

The MCR-ALS analysis further revealed pH-dependent (Fig. 3A) and spectral profiles (Fig. 3B) of the medium constituents. One spectral profile with a maximum at 1634 cm^{-1} was attributed to the absorption of KCl-solvated water. This assignment seems reasonable, as the KCl concentration does not change throughout the titration, which is reflected in the constant and flat progression in the pH-dependent profiles (Fig. 3A). Finally, the fourth, rather featureless spectral profile is attributed to the baseline offset in IR spectra that occurs in aqueous solutions at elevated pH values. This is supported by the significant increase of the relative concentration of this spectral profile at pH > 11.5. The fifth retrieved component (not shown) does not demonstrate any particular spectral feature and then could not be attributed to a specific phenomenon; however, it was necessary to include to improve the goodness of fit of the MCR-ALS model.

3.3. Evaluation of spectral profiles

In the previous section, MCR-ALS was employed for chemometric analysis of QCL-IR spectra monitoring the conformational change during pH titration. However, chemometric analysis also allows to retrieve additional in-depth information of the experimentally recorded QCL-IR spectra and reveal details about the experimental setup.

At this point, it should be emphasized that for chemometric models such as MCR-ALS, any characteristic of the analyzed data is of paramount interest and will influence the obtained result. In this regard, the noise plays an important role in the decomposition. In most cases, the noise in the experimental data is introduced by stochastic variations of the system, for example, instrumental and electrical interferences, and consequently presents a random behavior. To facilitate the modelling, random noise can be abolished by implementing denoising procedures, such as Savitzky-Golay smoothing. On the other hand, advanced denoising procedures, as transforms or filtering, may disrupt the random structure of the noise and, then, change the inner structure of the data turning the random into non-random noise. Besides, either smoothing or filtering is not trivial to implement and distortions can be unintentionally introduced to the original signals, which would alter the information comprised in the experimental data. For illustration, Figure S1 shows the results obtained from the MCR-ALS analysis of (Fourier transform) FT-filtered data. It can be observed that only one component was obtained for the medium and the spectral profiles of the protein and the medium are distorted in comparison to the results obtained from the unfiltered data resolution.

Moreover, the experimental data may also contain non-random noise contributions. For example, it is known that spectra obtained by EC-QCL based setups may contain, besides the intrinsic random noise, a fine structure originating from the mode-hops due to competition of

365 different optical modes for the available net gain in the laser medium [57]. Furthermore, the
366 numerous optical components of optical setups may introduce periodic spectral features due to
367 fringing effects. For this reason, no denoising procedures were implemented prior to MCR-ALS
368 modelling in order to avoid equivocal results due to artefacts introduced by the denoising step.

369 After MCR-ALS, a denoising procedure was carefully selected, to decompose the
370 spectroscopic signal from the non-random noise components of the experimental data.
371 Consequently, discrete Meyer wavelet transform (MWT) was applied to denoise the retrieved
372 spectral profiles. Fig. 3B shows a comparison of the raw spectral profiles retrieved by MCR-ALS
373 before and after denoising by MWT and Fig. 3C depicts the residuals obtained from the MWT.
374 Analysis of residuals reveals the presence of low band-width spikes in the spectral profiles that
375 can be attributed to the characteristic bands of water vapor. Since the entire titration procedure
376 requires approximately five hours, it is experimentally difficult to maintain the same humidity
377 throughout this time period. Furthermore, beside the stochastic noise, interference fringes are
378 present in the entire spectral region and recognizable particularly well in the region between
379 1580-1615 cm^{-1}, that is undisturbed by water vapor gas-phase bands. Interference fringes are
380 sinusoidal pattern on the baseline of the spectrum caused by interference between radiation that
381 has been transmitted directly through an optical element such as windows or sample cells with
382 light that has been reflected internally. Furthermore, the decreasing amplitude of the interference
383 fringes with increasing wavenumbers indicates that the surfaces are not perfectly parallel [58].
384 Within the experimental setup, this kind of fringes can stem for example from sample cell
385 windows or the Sapphire window employed for attenuating the laser power in the amide II region
386 [19].

387

388 **4. Conclusion**

EC-QCL based IR transmission spectroscopy was applied for performing a titration to monitor the pH-induced conformational change of β-lactoglobulin between pH 6.0 and 12.7. The experimental feasibility for this measurement was enabled by the large optical transmission path applicable due to the high emission power of quantum cascade lasers. The spectra revealed a change of protein secondary structure in the investigated pH region, but also an increasing offset of the baseline at elevated pH values that impeded straight-forward evaluation. Consequently, MCR-ALS was employed to unravel the overlapping spectral features in the IR spectra. With this chemometric technique, spectral and abundance distribution profiles were obtained by analysis of recorded dynamic IR spectra. Spectral profiles obtained by the MCR-ALS model for the identified secondary structure elements show good comparability with recorded IR spectra regarding band positions. The abundance distribution profiles reveal a transition from β-sheet to disordered secondary structure with a transition point at a pH value of ~10.2, which is in accordance to previously reported investigations of this system by CD spectroscopy. In addition, wavelet transform was successfully implemented to the MCR-ALS spectral profiles allowing to extract significant information about the spurious signals that are intrinsically present in QCL-IR spectra.

In conclusion, the present study demonstrates that laser-based IR transmission spectroscopy is an excellent tool for performing flow-through measurements of proteins in aqueous solution. Furthermore, it was demonstrated that MCR-ALS in combination with IR spectroscopy is a powerful technique for monitoring and interpreting protein folding.

Acknowledgements

This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 780240. M.R.A. acknowledges CONICET for her postdoc fellowship.

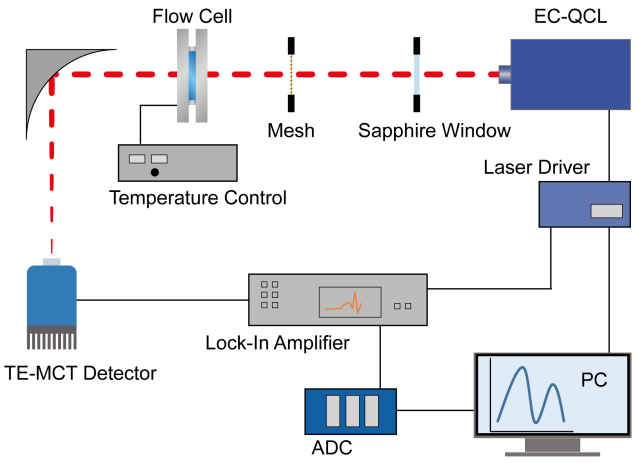
References

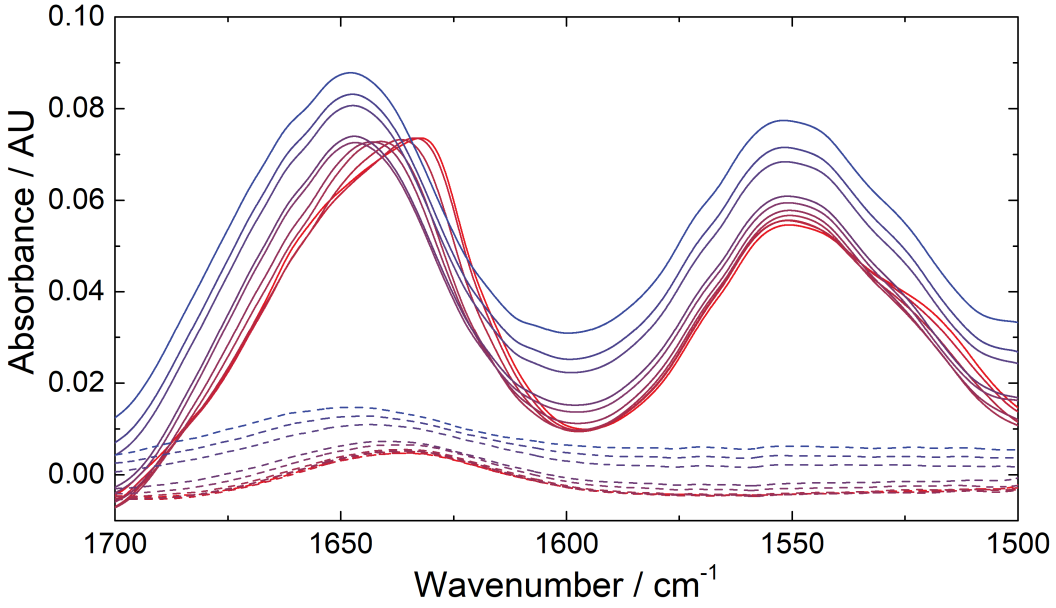
[1] A. Barth, Infrared spectroscopy of proteins, Biochim. Biophys. Acta, Bioenerg. 1767 (2007) 1073-1101.
[2] M.C. Manning, Use of infrared spectroscopy to monitor protein structure and stability, Expert Rev. Proteomics 2 (2005) 731-743.
[3] S. Bal Ram, Basic Aspects of the Technique and Applications of Infrared Spectroscopy of Peptides and Proteins, in: Infrared Analysis of Peptides and Proteins, American Chemical Society, 1999, pp. 2-37.
[4] M.R. Alcaráz, A. Schwaighofer, H. Goicoechea, B. Lendl, Application of MCR-ALS to reveal intermediate conformations in the thermally induced α-β transition of poly-l-lysine monitored by FT-IR spectroscopy, Spectrochim. Acta, Part A 185 (2017) 304-309.
[5] F. Dousseau, M. Pezolet, Determination of the secondary structure content of proteins in aqueous solutions from their amide I and amide II infrared bands. Comparison between classical and partial least-squares methods, Biochemistry 29 (1990) 8771-8779.
[6] S. Navea, R. Tauler, E. Goormaghtigh, A. de Juan, Chemometric tools for classification and elucidation of protein secondary structure from infrared and circular dichroism spectroscopic measurements, Proteins: Struct., Funct., Genet. 63 (2006) 527-541.
[7] H. Fabian, W. Mäntele, Infrared Spectroscopy of Proteins, in: Handbook of Vibrational Spectroscopy, John Wiley & Sons, Ltd, Hoboken, NJ, USA, 2006.
[8] C. Zscherp, A. Barth, Reaction-Induced Infrared Difference Spectroscopy for the Study of Protein Reaction Mechanisms, Biochemistry 40 (2001) 1875-1883.
[9] J. Faist, F. Capasso, D.L. Sivco, C. Sirtori, A.L. Hutchinson, A.Y. Cho, Quantum Cascade Laser, Science 264 (1994) 553-556.
[10] M.J. Weida, B. Yee, Quantum cascade laser-based replacement for FTIR microscopy, Proc. Soc. Photo-Opt. Instrum. Eng. 7902 (2011) 79021C.
[11] A. Schwaighofer, M. Brandstetter, B. Lendl, Quantum cascade lasers (QCLs) in biomedical spectroscopy, Chem. Soc. Rev. 46 (2017) 5903-5924.
[12] M.R. Alcaráz, A. Schwaighofer, C. Kristament, G. Ramer, M. Brandstetter, H. Goicoechea, B. Lendl, External cavity-quantum cascade laser spectroscopy for mid-IR transmission measurements of proteins in aqueous solution, Anal. Chem. 87 (2015) 6980-6987.
[13] M.R. Alcaráz, A. Schwaighofer, H. Goicoechea, B. Lendl, EC-QCL mid-IR transmission spectroscopy for monitoring dynamic changes of protein secondary structure in aqueous solution on the example of beta-aggregation in alcohol-denaturated alpha-chymotrypsin, Anal. Bioanal. Chem. 408 (2016) 3933-3941.
[14] A. Schwaighofer, M.R. Alcaráz, C. Araman, H. Goicoechea, B. Lendl, External cavity-quantum cascade laser infrared spectroscopy for secondary structure analysis of proteins at low concentrations, Sci. Rep. 6 (2016) 33556.

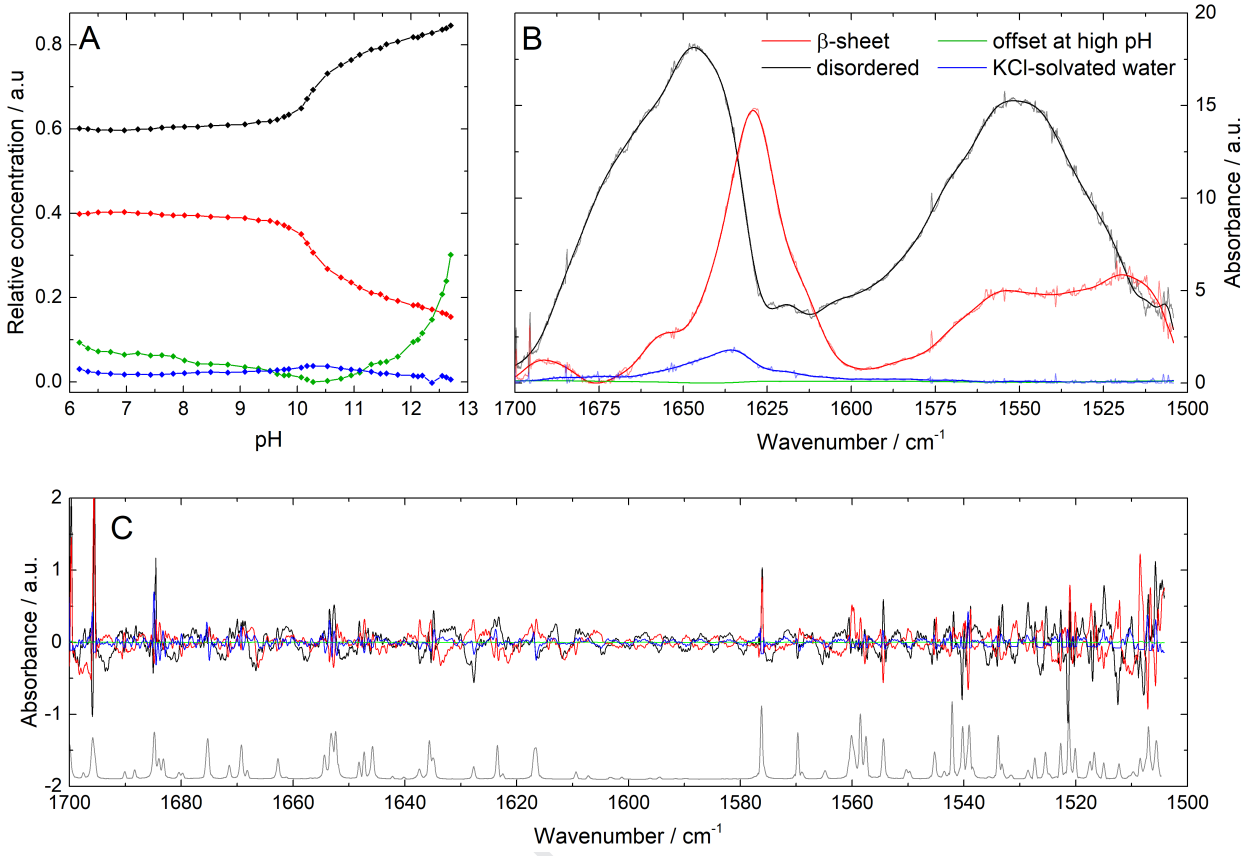
[15] D.J. Wurm, J. Quehenberger, J. Mildner, B. Eggenreich, C. Slouka, A. Schwaighofer, K. Wieland, B. Lendl, V. Rajamanickam, C. Herwig, O. Spadiut, Teaching an old pET new tricks: tuning of inclusion body formation and properties by a mixed feed system in E. coli, Appl. Microbiol. Biotechnol. 102 (2018) 667-676.
[16] J. Kuligowski, A. Schwaighofer, M.R. Alcaráz, G. Quintás, H. Mayer, M. Vento, B. Lendl, External cavity-quantum cascade laser (EC-QCL) spectroscopy for protein analysis in bovine milk, Anal. Chim. Acta 963 (2017) 99-105.
[17] A. Schwaighofer, J. Kuligowski, G. Quintás, H.K. Mayer, B. Lendl, Fast quantification of bovine milk proteins employing external cavity-quantum cascade laser spectroscopy, Food Chem. 252 (2018) 22-27.
[18] A. Schwaighofer, M.R. Alcaráz, J. Kuligowski, B. Lendl, Recent advancements of EC-QCL based mid-IR transmission spectroscopy of proteins and application to analysis of bovine milk, Biomed. Spectrosc. Imaging 7 (2018) 35-45.
[19] A. Schwaighofer, M. Montemurro, S. Freitag, C. Kristament, M.J. Culzoni, B. Lendl, Beyond FT-IR spectroscopy: EC-QCL based mid-IR transmission spectroscopy of proteins in the amide I and amide II region, Anal. Chem. 90 (2018) 7072-7079.
[20] A. Dong, J. Matsuura, S.D. Allison, E. Chrisman, M.C. Manning, J.F. Carpenter, Infrared and circular dichroism spectroscopic characterization of structural differences between beta-lactoglobulin A and B, Biochemistry 35 (1996) 1450-1457.
[21] N. Taulier, T.V. Chalikian, Characterization of pH-induced transitions of beta-lactoglobulin: Ultrasonic, densimetric, and spectroscopic studies, J. Mol. Biol. 314 (2001) 873-889.
[22] X.L. Qi, S. Brownlow, C. Holt, P. Sellers, Thermal-Denaturation of Beta-Lactoglobulin - Effect of Protein-Concentration at Ph-6.75 and Ph-8.05, Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol. 1248 (1995) 43-49.
[23] B. Czarnik-Matusewicz, K. Murayama, Y.Q. Wu, Y. Ozaki, Two-dimensional attenuated total reflection/infrared correlation spectroscopy of adsorption-induced and concentration-dependent spectral variations of beta-lactoglobulin in aqueous solutions, J. Phys. Chem. B 104 (2000) 7803-7811.
[24] J.C. Ioannou, A.M. Donald, R.H. Tromp, Characterising the secondary structure changes occurring in high density systems of BLG dissolved in aqueous pH 3 buffer, Food Hydrocolloids 46 (2015) 216-225.
[25] S. Ngarize, H. Herman, A. Adams, N. Howell, Comparison of changes in the secondary structure of unheated, heated, and high-pressure-treated ss-lactoglobulin and ovalbumin proteins using Fourier transform Raman spectroscopy and self-deconvolution, J. Agr. Food Chem. 52 (2004) 6470-6477.
[26] C. Le Bon, T. Nicolai, D. Durand, Growth and structure of aggregates of heat-denatured beta-Lactoglobulin, Int. J. Food Sci. Technol. 34 (1999) 451-465.
[27] P. Havea, H. Singh, L.K. Creamer, Characterization of heat-induced aggregates of beta-lactoglobulin, alpha-lactalbumin and bovine serum albumin in a whey protein concentrate environment, J. Dairy Res. 68 (2001) 483-497.
[28] G. Panick, R. Malessa, R. Winter, Differences between the Pressure- and Temperature-Induced Denaturation and Aggregation of β-Lactoglobulin A, B, and AB Monitored by FT-IR Spectroscopy and Small-Angle X-ray Scattering, Biochemistry 38 (1999) 6512-6519.
[29] S. Uhrínová, M.H. Smith, G.B. Jameson, D. Uhrín, L. Sawyer, P.N. Barlow, Structural Changes Accompanying pH-Induced Dissociation of the β-Lactoglobulin Dimer†,‡, Biochemistry 39 (2000) 3565-3574.

[30] Y. Yan, D. Seeman, B. Zheng, E. Kizilay, Y. Xu, P.L. Dubin, pH-Dependent Aggregation and Disaggregation of Native β-Lactoglobulin in Low Salt, Langmuir 29 (2013) 4584-4593.
[31] V.N. Uversky, N.V. Narizhneva, S.O. Kirschstein, S. Winter, G. Lober, Conformational transitions provoked by organic solvents in beta-lactoglobulin: Can a molten globule like intermediate be induced by the decrease in dielectric constant?, Folding Des. 2 (1997) 163-172.
[32] D. Hamada, C.M. Dobson, A kinetic study of β-lactoglobulin amyloid fibril formation promoted by urea, Protein Sci. 11 (2002) 2417-2426.
[33] G. Navarra, M. Leone, V. Militello, Thermal aggregation of beta-lactoglobulin in presence of metal ions, Biophys. Chem. 131 (2007) 52-61.
[34] H.L. Casal, U. Kohler, H.H. Mantsch, Structural and Conformational-Changes of Beta-Lactoglobulin-B - an Infrared Spectroscopic Study of the Effect of Ph and Temperature, Biochim. Biophys. Acta 957 (1988) 11-20.
[35] S. Joris, I. John, T. Hans, D. Athene, S. Paul van der, Mass-action driven conformational switching of proteins: investigation of beta-lactoglobulin dimerisation by infrared spectroscopy, J. Phys. D: Appl. Phys. 48 (2015) 384001.
[36] T. Lefevre, M. Subirade, Structural and interaction properties of beta-Lactoglobulin as studied by FTIR spectroscopy, Int. J. Food Sci. Technol. 34 (1999) 419-428.
[37] Y. Fang, D.G. Dalgleish, Conformation of beta-lactoglobulin studied by FTIR: Effect of pH, temperature, and adsorption to the oil-water interface, J. Colloid Interf. Sci. 196 (1997) 292-298.
[38] C. Ruckebusch, L. Blanchet, Multivariate curve resolution: A review of advanced and tailored applications and challenges, Anal. Chim. Acta 765 (2013) 28-36.
[39] R. Tauler, A. de Juan, Multivariate Curve Resolution for Quantitative Analysis, in: Arsenio Muñoz de la Peña, Héctor C. Goicoechea, G.M. Escandar, C.O. Alejandro (Eds.) Data Handling in Science and Technology, Elsevier, 2015, pp. 247-292.
[40] M.R. Alcaráz, G.G. Siano, M.J. Culzoni, A. Munoz de la Pena, H.C. Goicoechea, Modeling four and three-way fast high-performance liquid chromatography with fluorescence detection data for quiatitation of fluoroquinolones in water samples, Anal. Chim. Acta 809 (2014) 37-46.
[41] M.R. Alcaráz, L. Vera-Candioti, M.J. Culzoni, H.C. Goicoechea, Ultrafast quantitation of six quinolones in water samples by second-order capillary electrophoresis data modeling with multivariate curve resolution-alternating least squares, Anal. Bioanal. Chem. 406 (2014) 2571-2580.
[42] M.R. Alcaráz, A.V. Schenone, M.J. Culzoni, H.C. Goicoechea, Modeling of second-order spectrophotometric data generated by a pH-gradient flow injection technique for the determination of doxorubicin in human plasma, Microchem. J. 112 (2014) 25-33.
[43] J. Diewok, A. De Juan, R. Tauler, B. Lendl, Quantitation of mixtures of diprotic organic acids by FT-IR flow titrations and multivariate curve resolution, Appl. Spectrosc. 56 (2002) 40-50.
[44] J. Kuligowski, G. Quintas, R. Tauler, B. Lendl, M. de la Guardia, Background Correction and Multivariate Curve Resolution of Online Liquid Chromatography with Infrared Spectrometric Detection, Anal. Chem. 83 (2011) 4855-4862.
[45] S.C. Rutan, A. de Juan, R. Tauler, 2.15 - Introduction to Multivariate Curve Resolution, in: S.D.B.T. Walczak (Ed.) Comprehensive Chemometrics, Elsevier, Oxford, 2009, pp. 249-259.
[46] R. Tauler, M. Maeder, A. de Juan, Multiset Data Analysis: Extended Multivariate Curve Resolution, in: S.D. Brown, R. Tauler, B. Walczak (Eds.) Comprehensive Chemometrics: Chemical and Biochemical Data Analysis, Elsevier, Oxford, 2009, pp. 473-505.

[47] A. De Juan, R. Tauler, Multivariate Curve Resolution (MCR) from 2000: Progress in Concepts and Applications, Crit. Rev. Anal. Chem. 363-4 (2006) 163-176.
[48] H.L. Monaco, G. Zanotti, P. Spadon, M. Bolognesi, L. Sawyer, E.E. Eliopoulos, Crystal-Structure of the Trigonal Form of Bovine Beta-Lactoglobulin and of Its Complex with Retinol at 2.5-a Resolution, J. Mol. Biol. 197 (1987) 695-706.
[49] L. Sawyer, β-Lactoglobulin, in: P.L.H. McSweeney, P.F. Fox (Eds.) Advanced Dairy Chemistry: Volume 1A: Proteins: Basic Aspects, 4th Edition, Springer US, Boston, MA, 2013, pp. 211-259.
[50] J.J. Max, C. Chapados, Interpolation and extrapolation of infrared spectra of binary ionic aqueous solutions, Appl. Spectrosc. 53 (1999) 1601-1609.
[51] R. Vonach, B. Lendl, R. Kellner, Modulation of the pH in the determination of phosphate with flow injection and Fourier transform infrared detection, Analyst 122 (1997) 525-530.
[52] A. Borges, R. Tauler, A.d. Juan, Application of multivariate curve resolution to the temperature-induced unfolding of α-chymotrypsin, Anal. Chim. Acta 544 (2005) 159-166.
[53] W. Windig, J. Guilment, Interactive self-modeling mixture analysis, Anal. Chem. 63 (1991) 1425-1432.
[54] M. Maeder, A. Zilian, Evolving Factor-Analysis, a New Multivariate Technique in Chromatography, Chemometr. Intell. Lab. 3 (1988) 205-213.
[55] A. de Juan, S.C. Rutan, R. Tauler, 2.19 - Two-Way Data Analysis: Multivariate Curve Resolution – Iterative Resolution Methods, in: S.D.B.T. Walczak (Ed.) Comprehensive Chemometrics, Elsevier, Oxford, 2009, pp. 325-344.
[56] M.R. Alcaraz, A. Aguirre, H.C. Goicoechea, M.J. Culzoni, S.E. Collins, Resolution of intermediate surface species by combining modulated infrared spectroscopy and chemometrics, Anal. Chim. Acta. 1049 (2019) 38-46.
[57] G. Wysocki, R.F. Curl, F.K. Tittel, R. Maulini, J.M. Bulliard, J. Faist, Widely tunable mode-hop free external cavity quantum cascade laser for high resolution spectroscopic applications, Appl. Phys. B: Lasers Opt. 81 (2005) 769-777.
[58] P.R. Griffiths, J.A. de Haseth, Fourier Transform Infrared Spectrometry, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2006.






A
Relative concentration / a.u
pH
B
β-sheet
disordered
offset at high pH
KCl-solvated water
Absorbance / a.u.
Wavenumber / cm-1
C
Absorbance / a.u.
Wavenumber / cm-1

Highlights:

- Mid-infrared (IR) External Cavity-Quantum Cascade Laser (EC-QCL) spectroscopy
- pH titration of β-lactoglobulin
- Measurement and analysis of the protein amide I and amide II region
- Evaluation by multivariate curve resolution-alternating least square (MCR-ALS)
- Unfolding of native β-sheet secondary structure at basic pH

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.