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

Andreas Schwaighofer a, Mirta R. Alcaraz b,c,d, Laurin Lux a, Bernhard Lendl a,∗

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.
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 [16–18,59].

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 CaF₂ 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 ms. A total of 100 scans were recorded for

![Fig. 1. Experimental setup for pH titration monitored by EC-QCL IR transmission spectroscopy.](image)
by addition of HCl. KCl (2.3. Titration procedure spectra. 150 was performed with 20 mL of medium solution, constituted by water throughout the introduced pH change, a similar titration procedure constant in the range of ±0.1 °C. The temperature was monitored throughout the titration, being 37 × 4451 and 35 × 4451 matrices were obtained for protein and pH value between 6.0 and 12.7. For a complete titration experiment, mathematical constraints, such as non-negativity, unimodality, optimization or through the implementation of chemical or pro under study. However, to achieve chemically meaningful component its general applicability without prior information about the system present system. 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 the implementation of chemical or mathematical constraints, such as non-negativity, unimodality, normalization and closure, among others [47].

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

2.5. Software

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

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

3. Results and discussion

3.1. IR spectra of pH-induced conformational change of β-lactoglobulin

QCL-IR spectra of the pH-induced change of β-LG were recorded between pH 6.0 and pH 12.7. In Fig. 2, the QCL-IR absorption spectra in the amide I and amide II region are shown. The QCL-IR spectra at pH 6.0 shows a band maximum in the amide I region at 1634 cm⁻¹ with a shoulder at 1655 cm⁻¹, and a broad band at approximately 1550 cm⁻¹ with a shoulder at 1520 cm⁻¹ in the amide II region, characteristic for the IR signatures of β-LG featuring a predominantly β-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.

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.

![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).](image-url)
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 known, 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 representational information about the system. For MCR-ALS analysis, unfiltered 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. Fig. 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 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 6.0 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 different optical modes for the available net gain in the laser medium [57]. Furthermore, the numerous optical components of optical setups may introduce periodic spectral features due to fringing effects. For this reason, no denoising procedures were implemented prior to MCR-ALS modelling in order to avoid equivocal results due to artefacts introduced by the denoising step.

After MCR-ALS, a denoising procedure was carefully selected, to decompose the spectroscopic signal from the non-random noise components of the experimental data. Consequently, discrete Meyer wavelet transform (MWT) was applied to denoise the retrieved spectral profiles. Fig. 3B shows a comparison of the raw spectral profiles retrieved by MCR-ALS before and after denoising by MWT and Fig. 3C depicts the residuals obtained from the MWT. Analysis of residuals reveals the presence of low band-width spikes in the spectral profiles that can be attributed to the characteristic bands of water vapor. Since the entire titration procedure requires approximately 5 h, it is experimentally difficult to maintain the same humidity throughout this time period. Furthermore, beside the stochastic noise, interference fringes are present in the entire spectral region and recognizable particularly well in the region between 1580-1615 cm$^{-1}$, that is undisturbed by water vapor gas-phase bands. Interference fringes are sinusoidal pattern on the baseline of the spectrum caused by interference between radiation that has been transmitted directly through an optical element such as windows or sample cells with light that has been reflected internally. Furthermore, the decreasing amplitude of the interference fringes with increasing wavenumbers indicates that the surfaces are not perfectly parallel [58]. Within the experimental setup, this kind of fringes can stem for example from sample cell windows or the Sapphire window employed for attenuating the laser power in the amide II region [19].

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 straightforward evaluation. Consequently, MCR-ALS was employed to unravel the overlapping spectral features in the IR spectra. With this chemometrical 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.
Declaration of competing interest

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.

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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.saa.2019.117636.

References
