

Accepted Manuscript

External Cavity-Quantum Cascade Laser (EC-QCL) Spectroscopy for Protein Analysis in Bovine Milk

Julia Kuligowski, Andreas Schwaighofer, Mirta Raquel Alcaráz, Guillermo Quintás, Helmut Mayer, Máximo Vento, Bernhard Lendl



PII: S0003-2670(17)30218-0

DOI: [10.1016/j.aca.2017.02.003](https://doi.org/10.1016/j.aca.2017.02.003)

Reference: ACA 235063

To appear in: *Analytica Chimica Acta*

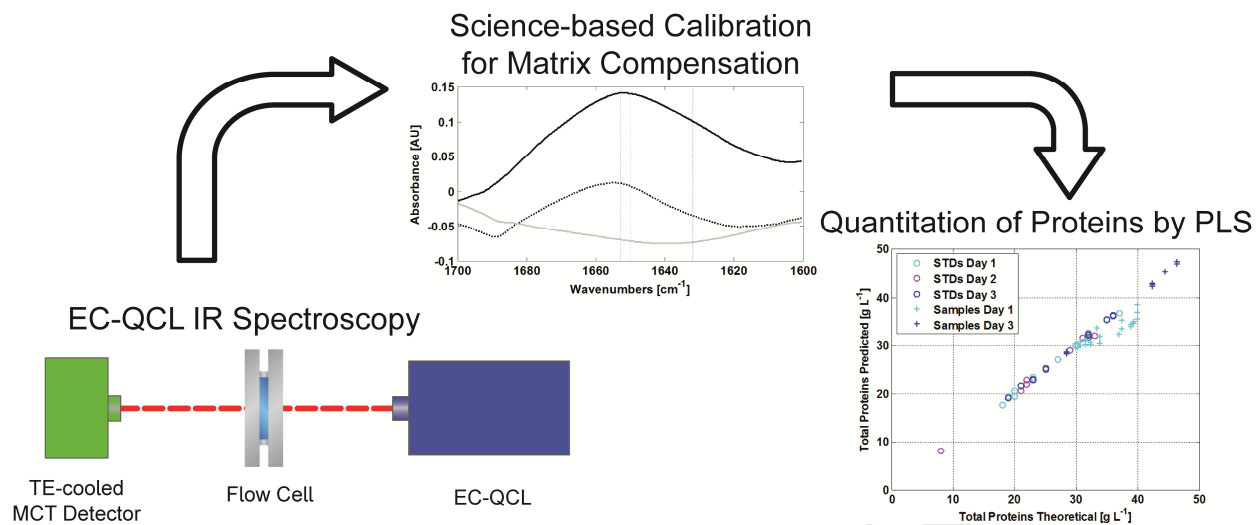
Received Date: 28 September 2016

Revised Date: 23 January 2017

Accepted Date: 2 February 2017

Please cite this article as: 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, *Analytica Chimica Acta* (2017), doi: 10.1016/j.aca.2017.02.003.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



External Cavity-Quantum Cascade Laser (EC-QCL) Spectroscopy for Protein Analysis in Bovine Milk

Julia Kuligowski^{1,2,δ}, Andreas Schwaighofer^{2,δ}, Mirta Raquel Alcaráz³, Guillermo Quintás^{4,5}, Helmut Mayer⁶, Máximo Vento^{1,7}, Bernhard Lendl^{2*}

¹Neonatal Research Unit, Health Research Institute La Fe, Avenida Fernando Abril Martorell 106, 46026 Valencia, Spain

²Institute of Chemical Technologies and Analytics, Vienna University of Technology, Getreidemarkt 9/164-UPA, 1060 Vienna, Austria

³Laboratorio de Desarrollo Analítico y Quimiometría, FBCB, Universidad Nacional del Litoral-CONICET, Ciudad Universitaria, 3000 Santa Fe, Argentina

⁴Human & Environmental Health & Safety (HEHS), Leitai Technological Center, Avenida Fernando Abril Martorell 106, 46026 Valencia, Spain

⁵Unidad Analítica, Health Research Institute La Fe, Avenida Fernando Abril Martorell 106, 46026 Valencia, Spain

⁶Food Chemistry Laboratory, BOKU – University of Natural Resources and Life Sciences, Muthgasse 11, 1190 Vienna, Austria

⁷Division of Neonatology, University & Polytechnic Hospital La Fe, Avenida Fernando Abril Martorell 106, 46026 Valencia, Spain

* bernhard.lendl@tuwien.ac.at

^δ both authors contributed equally

Abstract

The analytical determination of bovine milk proteins is important in food and non-food industrial applications and yet, rather labour-intensive wet-chemical, low-throughput methods have been employed since decades. This work proposes the use of external cavity-quantum cascade laser (EC-QCL) spectroscopy for the simultaneous quantification of the most abundant bovine milk proteins and the total protein content based on the chemical information contained in mid-infrared (IR) spectral features of the amide I band. Mid-IR spectra of protein standard mixtures were used for building partial least squares (PLS) regression models. Protein concentrations in commercial bovine milk samples were calculated after chemometric compensation of the matrix contribution employing science-based calibration (SBC) without sample pre-processing. The use of EC-QCL spectroscopy together with advanced multivariate data analysis allowed the determination of casein, α -lactalbumin, β -lactoglobulin and total protein content within several minutes.

Keywords: *External Cavity-Quantum Cascade Laser (EC-QCL) spectroscopy; mid-infrared (IR); protein analysis; milk analysis; Science-Based Calibration (SBC)*

ACCEPTED MANUSCRIPT

1. Introduction

Milk is considered one of the most nutritionally complete foods available with a recorded production volume of 700×10^6 tonnes per year, about 84% of which is bovine [1]. Due to its nutritional value and its unique functional characteristics, milk is an important raw material in food as well as non-food industry. It is mainly composed of water, lipids, lactose and proteins, and their relative shares widely vary among species, e.g., the percentage range of proteins goes from 10 to 200 g L⁻¹, with a total protein content of ~ 32 g L⁻¹ for bovine milk [2]. Casein (Cas) proteins represent approximately 80 % of the total protein content in bovine milk and the remaining 20 % belong to the whey proteins [3,4]. The caseins fraction, with a protein content of ~ 25 g L⁻¹, includes four gene products (α_{s1} -Cas, α_{s2} -Cas, β -Cas and κ -Cas) and has many applications in the food and non-food industry [5]. It is the central constituent in cheese making and it is used as additive to non-dairy based foods due to its favourable physico-chemical characteristics. Moreover, it is used in non-food applications as coating, adhesive and surfactant [5,6]. The two most prominent and abundant proteins that represent the whey protein fraction are β -lactoglobulin (β -LG) and α -lactalbumin (α -LA) with protein concentrations of ~ 2.7 and ~ 1.2 g L⁻¹, respectively [5]. Nowadays, whey proteins are an important by-product of cheese and casein manufacturers [7]. Its desirable nutritional and functional characteristics have made whey proteins ingredients of choice in the formulation of various foods and beverages. Furthermore, β -LG has been found to have antimicrobial, anticarcinogenic, antiviral and hypocholesterolemic effects, while α -LA from bovine milk seems to be ideally suited as an ingredient for infant nutrition due to its high degree of amino acid homology to human α -LA [8,9]. Whey proteins are also used in the non-food industry including cosmetics as well as for biodegradable packaging and coatings [5,6].

With the growing technological importance of milk proteins, beyond the dairy industry, analytical techniques for the qualitative and quantitative assessment of milk components have gained increasing interest through the years. Due to strict quality regulations, the food

industry is interested in sensitive and accurate characterisation methods, and in addition, inexpensive and rapid on-line analytical techniques are desirable. For analysis of the total protein content in milk, the historic Kjeldahl method is still the internationally recognized reference method for the determination of organic nitrogen in food and beverages, in spite of being a rather labour-intensive wet-chemical method with low throughput [10]. For quantification of individual milk proteins, several methods based on chromatographic [11-15] and electrophoretic [11,16] approaches have been developed, where still time-consuming and wet-chemical sample preparation needs to be performed. Application of mid-infrared (IR) spectroscopy has been adopted for the rapid and non-destructive analysis of the quality and composition of dairy products due to its high-throughput capacity, low cost and simplicity [17-22]. Moreover, a technique based on IR spectroscopy has been accepted by the Association of Official Analytical Chemists (AOAC International) for the analysis of liquid milk [23,24]. Until now, to the best of our knowledge, no mid-IR method for individual quantification of milk proteins has been reported.

Mid-IR spectroscopy detects the fundamental vibrations of covalent bonds in molecules and provides qualitative and quantitative information of the constituents that absorb mid-IR radiation. For protein analysis using mid-IR spectroscopy, the most prominent absorption features are the amide I ($1600\text{-}1700\text{ cm}^{-1}$) and amide II bands ($1500\text{-}1600\text{ cm}^{-1}$), being the amide I band particularly sensitive to secondary protein structure [25]. For quantitative analysis of total protein content of aqueous solutions, most commonly the amide II band is employed at path length of $\sim 50\text{ }\mu\text{m}$ for transmission measurements, usually in combination with multivariate calibration techniques [18,21,23,26]. Discrimination among/between proteins based on secondary structure is preferentially performed by evaluation of the amide I band. However, the development of FTIR-based applications in the $1600\text{-}1700\text{ cm}^{-1}$ spectral region is troublesome, due to the low emission powers of thermal light sources in conventional FTIR spectrometers and the strong absorbance of H_2O centred at 1645 cm^{-1} . Consequently applicable path lengths for transmission measurements in this region are restricted to $<10\text{ }\mu\text{m}$, which compromises sensitivity levels necessary for the study of

biologically relevant concentrations and considerably impair the ruggedness of the application. Particularly for basic research, the use of D₂O is employed replacing H₂O as a solvent, which dramatically restricts the number of applications [25]. Alternatively, the use of an external cavity-quantum cascade laser (EC-QCL) light source with significantly higher emission intensities enables to increase the optical path length to 38 μm for mid-IR measurements in aqueous solutions covering the amide I spectral region. Most recently, EC-QCL-based IR transmission measurements have been successfully accomplished for the analysis of protein secondary structure [27-29].

The aim of the present study is to prove the capability of the EC-QCL setup to simultaneously quantify Cas, α-LA and β-LG contents as well as the total protein content in protein standard mixtures by evaluation of the amide I band using partial least squares (PLS) modelling. Furthermore, the feasibility of protein determinations in commercial bovine milk samples after a chemometric compensation of the matrix signal is demonstrated without the need of sample pre-processing.

2. Material & Methods

2.1 Standards and reagents

Ethanol (96 % v/v, EtOH), sodium hydroxide solution 50 % in water (NaOH), hydrochloric acid 37 % (HCl) ACS reagent, α-lactose monohydrate (≥ 99 %, α-Lac), lyophilized powder of α-LA (≥ 85 %), β-LG (≥ 90 %) and Cas sodium salt, all from bovine milk, were obtained from Sigma-Aldrich (Steinheim, Germany) and used as purchased. Sodium phosphate monobasic dihydrate p.a. (NaH₂PO₄·2H₂O) was obtained from Fluka (Buchs, Switzerland). Ultrapure water was generated by a Milli-Q water purification system from Merck Millipore (Darmstadt, Germany). Pasteurized and homogenized commercial extended shelf life bovine milk was obtained from a local grocery store.

2.2 Preparation of solutions and samples

Individual stock solutions of Cas, α -LA and β -LG at concentrations of 40, 30 and 30 g L⁻¹, respectively, were prepared on each measurement day (i.e. days 1, 2 and 3) by direct weighing and dissolving of the corresponding amount of lyophilized protein in 5 mL of 45 mmol L⁻¹ phosphate buffer at pH 6.6. 10 standard solutions containing binary (Cas and β -LG) or tertiary protein mixtures (Cas, α -LA and β -LG) were prepared by dilution of the corresponding stock solutions in 45 mmol L⁻¹ phosphate buffer at pH 6.6, reaching concentrations ranging between 1-35 g L⁻¹ for Cas and 1-15 g L⁻¹ for α -LA and β -LG. For repeatability assessment, bovine milk samples and fortified bovine milk samples were prepared and measured on two different days. Fortified milk samples were prepared by weighing of the corresponding amount of lyophilized proteins and dissolution in bovine milk covering spiked concentrations of Cas between 1 and 20 g L⁻¹ and α -LA and β -LG between 3 and 8 g L⁻¹. α -Lac solutions were prepared along with the fortified milk samples on measurement days 1 and 3 at a concentration of 48 g L⁻¹ by weighing and dissolution of the appropriate amount of the monohydrate in 45 mmol L⁻¹ phosphate buffer at pH 6.6.

2.3 EC-QCL setup

The custom-made EC-QCL setup has been described in detail earlier [27]. Briefly, the setup is equipped with an EC-QCL (spectral tuning range = 1729.30-1565.06 cm⁻¹; Daylight Solutions Inc., San Diego, USA), a temperature-controlled 38 μ m path length flow cell and a thermoelectrically cooled (-60 °C) HgCdTe (mercury cadmium telluride) detector (MCT-7-TE3; D* = 4 \times 10⁹ cm Hz^{0.5} W⁻¹ at 9.2 μ m; Infrared Associates Inc., USA). The laser was thermoelectrically cooled (head temperature = 18 °C) and operated in pulsed mode at a repetition rate of 100 kHz and a pulse width of 500 ns. A gold plated off-axis parabolic mirror (focal length: 43 mm) was used to focus the mid-IR light on the detector. The measured signal was processed by a two-channel boxcar integrator and digitized by a NI DAQ 9239 24-bit analogue-to-digital converter (National Instruments Corp., Austin, USA) at a sampling rate of 16 kHz. To reduce the influence of water vapour, the setup was placed in a housing of polyethylene foil and constantly flushed with dry air. The entire setup was

controlled by a LabView-based graphical user interface (Lab View 11.0, National Instruments Corp., Austin, USA) with server–client program structure [30].

Each single beam spectrum consisting of 24 000 data points was recorded during the tuning time of 1.5 s. A total of 100 scans (total measurement time: ~500 s) were recorded for background and sample single beam spectra. For IR measurements of milk, deionised water was used as a reference, whereas 45 mmol L⁻¹ phosphate buffer was used as reference for protein standard solutions. The flow cell was cleaned with 2 mL of EtOH, diluted NaOH and HCl and distilled water by the end of each experiment. All solutions were manually injected into the flow cell using a 1 mL disposable syringe.

2.4 Data processing

Data processing was performed in MATLAB 2013b (Mathworks Inc., Natick, USA) using built-in and in-house written routines. PLS Toolbox 8.0 (Eigenvector Research Inc., Wenatchee, USA) was used for data analysis.

2.4.1 EC-QCL data pre-processing

The performed pre-processing routine for EC-QCL raw data has been described in detail elsewhere [27]. The principal item of this procedure is the correlation optimized warping (COW) algorithm that is employed to correct the spectral mismatch of successive scans, caused by irregular shifts in the fine structure of the EC-QCL light source due to mechanical imperfections and triggering issues. COW is a widely used method employed to correct spectral (nuclear magnetic resonance, IR, or Raman) or chromatographic shifts [31-33]. Here, this strategy was applied for the alignment of consecutive scans of each measurement before averaging and to align the background with the sample single beam spectrum before obtaining the corresponding absorbance spectrum. It has been demonstrated that this alignment procedure does neither distort the amplitude nor the wavenumber position of the final absorbance spectrum [34]. Using the aforementioned data processing procedure, the noise level has been significantly reduced for absorbance spectra of proteins [27]. After applying the pre-processing routine, each spectrum, limited to the amide I region between 1600 and 1700 cm⁻¹, comprises a total of 3688 data points.

2.4.2 Background correction of commercial bovine milk spectra employing science-based calibration (SBC)

The SBC method is a multivariate technique that combines the prediction accuracy of the inverse approach with the ease of interpretability of classical models by estimating the spectral signal in a physical way and the spectral noise in a statistical way. SBC is a powerful tool for quantitative analysis with strongly overlapping signals [35-42]. Essentially, this method can be briefly explained as follows: each measured spectrum, \mathbf{x}^T , is described as:

$$\mathbf{x}^T = y \cdot \mathbf{g}^T + \mathbf{x}_n^T \quad (1)$$

where y and \mathbf{g}^T are the abundance and the response spectrum of the matrix signal, respectively. The superscript T denotes a transpose. The noise vector, \mathbf{x}_n^T , comprises the instrumental noise and signals of other interfering compounds. For a matrix \mathbf{X} ($M \times K$), containing M spectra registered at K variables (*i.e.* wavenumbers), the signal can be described by the mean spectrum $\bar{y} \cdot \mathbf{g}^T$ and the root-mean-square expressed as $\sigma_y \cdot \mathbf{g}^T$, being σ_y the standard deviation of y . Likewise, the spectral noise is described by a mean value $\bar{\mathbf{x}}_n^T$ and a covariance matrix Σ ($K \times K$). At constant concentration, differences between spectra in the matrix \mathbf{X} are given as result of the noise. Under these conditions, after mean-centring, the matrix \mathbf{X} only contains information about the spectral noise. Accordingly, the covariance matrix is calculated as:

$$\Sigma \cong \frac{\tilde{\mathbf{X}}^T \tilde{\mathbf{X}}}{M-1} \quad (2)$$

where the “~” tilde indicates that the matrix is mean-centred.

After obtaining the spectral signal of the matrix and calculating the spectral noise, it is possible to determine the optimum regression vector, \mathbf{b}_{opt} , as follows, on the assumption that σ_y^2 tends to ∞ :

$$\mathbf{b}_{\text{opt}} = \frac{\Sigma^+ \mathbf{g}}{\mathbf{g}^T \Sigma^+ \mathbf{g}} \quad (3)$$

wherein the superscript $+$ denotes the pseudo-inverse. For quantitative analysis, the vector \mathbf{b}_{opt} is used to predict the abundance y_{pred} of the matrix in new samples based on the spectral information in $\mathbf{x}_{\text{pred}}^{\text{T}}$.

$$y_{\text{pred}} = \bar{y} + (\mathbf{x}_{\text{pred}} - \bar{\mathbf{x}})^{\text{T}} \cdot \mathbf{b}_{\text{opt}} \quad (4)$$

being \bar{y} the mean value of y and $\bar{\mathbf{x}}$ the mean spectrum of the noise spectra in \mathbf{X} ($M \times K$).

Here, prior to the quantification of Cas, α -LA and β -LG in whole milk, an initial subtraction of the spectral contribution of the milk matrix from the overall IR signal was performed. Accordingly, SBC was employed to achieve a reliable estimation of matrix proportion in milk sample spectra reflected in variations in the absorbance intensity of the H_2O band at 1645 cm^{-1} due to the displacement of water caused by milk matrix compounds relative to the H_2O background spectra. In this work, an α -Lac spectrum was used as response spectrum of the matrix signal and a set of 10 synthetic tertiary protein standard mixtures with a constant concentration of the matrix signal of zero was used as noise matrix for the calculation of the spectral noise.

3. Results & Discussion

3.1 Protein spectra

Figure 1 shows EC-QCL-IR spectra of individual standard solutions of Cas, α -LA and β -LG at concentration levels ranging between 2.5 and 20 g L^{-1} , where differences in spectral profiles of the studied compounds can clearly be observed. It is worth noting that the position and the shape of the amide I band change in dependence of the secondary protein structure. The IR spectrum of Cas contains a broad band with an absorbance maximum at 1650 cm^{-1} that has been attributed to an irregular and extended α -helical secondary structure [43-45]. α -LA is predominantly composed by α -helical structures and shows a band maximum at 1653 cm^{-1} [46,47], whereas β -LG has maximum absorbance at 1632 cm^{-1} with a shoulder located at approximately 1660 cm^{-1} [48]. Figure 1 clearly illustrates that in spite of the spectral differences observed, there is a strong overlap between protein spectra in the

IR region. The large degree of spectral overlap evidences the need for a multivariate approach for protein quantification in a multicomponent system.

3.2 Bovine milk spectra

Figure 2 shows an EC-QCL-IR absorbance spectrum of commercial bovine whole milk using ultrapure water as background. Milk samples were directly introduced into the flow cell without any sample treatment. In IR spectra obtained from milk, distorted spectral features were observed in the amide I region compared to protein standard solutions (see Figure 1). For instance, absorbance band maximum positions in aqueous standard solutions do not match the observed values in blank and spiked whole milk samples.

The chemical complexity of biological fluids, such as milk, is reflected in the IR spectra by the presence of multiple absorbing components, including sugar, proteins and fats, among others. Moreover, the content of these components modifies the relative water content of the samples as compared to the background spectrum, which has a strong effect on the IR spectra in the amide I region due to the aforementioned overlap with the HOH-bending band of water [49]. Therefore, in order to quantify protein contents in milk samples, an initial subtraction of the matrix background contribution was necessary. Here, background correction was a particular challenge because of the limited spectral region accessible with the EC-QCL setup. On the other hand, bovine milk is a complex natural product that cannot be truly simulated in the laboratory by mixing its individual components. Hence, in this work the milk matrix was simplified by an aqueous α -Lac solution that represented the changes in IR signal shapes, in the 1600 to 1700 cm^{-1} region, caused by varying H_2O contents of milk samples (see Figure 2).

In the present work, SBC was employed to estimate the matrix contribution to the IR spectrum in the amide I region in order to correct the spectra prior to quantitative protein analysis. For applying this procedure, two inputs were required: (i) the noise matrix, built from the measurement of a set of standard protein mixtures, containing the spectral information about instrumental noise as well as on components with overlapping spectral features, in this case Cas, α -LA and β -LG and (ii) the α -Lac spectrum, which was used as

matrix signal to model the phenomenon observed in the IR spectra of milk, due to the variations in the matrix composition. Figure 2 depicts the α -Lac spectrum in the selected region, showing a negative absorbance signal with a minimum located at 1640 cm^{-1} when phosphate buffer is used as background, due to the fact that α -Lac partially replaces H_2O . The EC-QCL-IR spectrum of the commercial bovine whole milk sample after correction of the matrix contribution by SBC is also depicted in Figure 2. Using the spectra of standard solutions depicted in Figure 1 as noise matrix, the contribution of the milk matrix could be considerably reduced. After spectral correction, the absorbance band maximum value is located at 1652 cm^{-1} matching with that observed in aqueous standard solutions of proteins (see Figure 1). Besides, the negative baseline offset at 1700 cm^{-1} could be greatly reduced from -0.047 to -0.014 AU in comparison to the offset of $0.0014\pm 0.0015\text{ AU}$ found in spectra of standard mixtures shown in Figure 1. Furthermore, the correlation coefficient calculated between protein standard solutions from Figure 1 and milk spectra in Figure 2 was quantitatively improved from 85 % to 99 % after background correction. In summary, the presented data corroborate the need for correcting the matrix contribution to the overall IR spectra from milk samples for enabling accurate protein determinations. SBC proved to be a powerful and straightforward tool to achieve a high degree of background compensation even when the accessible spectral range is limited.

3.3 Quantitative protein analysis employing PLS regression

As can be seen in Figure 1, there is a high degree of overlap in the amide I region in mid-IR spectra of Cas, α -LA and β -LG. Hence, direct quantification of each protein contribution by measuring band height or area is not feasible. To tackle protein quantification in a multicomponent system, multivariate PLS regression models were optimized for each analyte, as well as for the total protein content defined as the sum of Cas, α -LA and β -LG, which are the most abundant proteins found in bovine milk.

Three calibration data sets from protein standard mixtures were acquired on different experimental days and used for model optimization and cross-validation. The PLS calibration parameters and internal figures of merit are summarized in Table 1. For Cas, β -

LG and total protein models, mean centring was selected as spectral pre-processing. For the determination of α -LA, second derivative spectra in combination with mean centring yielded optimum results. The selection of this additional pre-processing step is justified by the high overlapping between Cas and α -LA bands. For all PLS regression models, between 2 and 6 latent variables (LVs) were employed. Differences in the number of LVs for the same protein on different measurement days are attributed to the different concentration ranges employed in each calibration experiment. Coefficients of determination (R^2) of calibration and leave-one-out cross-validation were in all cases >0.95 and >0.8 , respectively, proving the goodness of fit for the calibration data set. Root-mean-square error of calibration (RMSEC) and cross-validation (RMSECV) ranged between 0.03 and 0.7 g L^{-1} and 0.2 and 1.4 g L^{-1} , respectively, while the CV bias as a measure of accuracy ranged between -0.3 and 0.2 g L^{-1} .

Furthermore, intra- and inter-day recovery values for protein standard solutions at three concentration levels (low, medium and high) were calculated yielding highly accurate and precise results with recoveries ranging between 90 and 105 % and relative standard deviations (RSD) below 16 % for the lowest concentration level. The obtained accuracy and precision levels stayed within the boundaries established in the guidelines of bioanalytical method validation by the US Food and Drug Administration (FDA) [50]. Nominal against predicted protein concentrations obtained from PLS regression models on three different days are represented in Figure 3, illustrating the high correlation between both parameters.

In addition to protein standard mixtures, spectra from blank and fortified bovine milk samples were acquired employing the EC-QCL setup on measurement days 1 and 3. After SBC background correction, concentrations of Cas, α -LA, β -LG and total proteins in milk samples as well as recoveries in spiked milk samples were determined by using the PLS model developed from standard spectra acquired on the same measurement day. Table 3 shows protein concentration values found in commercial bovine whole milk samples and the recovery values obtained for fortified bovine whole milk samples. Nominal against predicted protein concentrations of blank and fortified bovine milk samples are depicted in

Figure 3. It is worthwhile mentioning that although the obtained accuracy and precision levels stayed within the US FDA acceptance criteria for bioanalytical method validation [50], prediction errors for α -LA were higher than for the other studied proteins and the total protein content. This could be related to the low relative concentration of α -LA in the samples in combination with a severe spectral overlapping with other sample constituents.

4. Conclusions

The presented results demonstrate for the first time the capability of an EC-QCL setup for the quantitative determination of different proteins in aqueous solutions by using mid-IR spectra in the amide I region. In combination with SBC-based background correction, Cas, α -LA, β -LG and total proteins contained in commercial bovine whole milk samples were successfully determined by using a PLS model with a confined number of standard solutions employed for model calibration. Furthermore, determinations were carried out within a considerable short time without requiring any sample processing. In light of the results obtained, this high-throughput approach might be of potential interest as a screening tool in the food industry. However, follow-up studies aiming at a thorough validation, which include a comparison with a well-characterized reference procedure, should be performed in order to critically assess the suitability of the method for the determination of proteins in bovine whole milk.

Acknowledgements

Financial support was provided by the Austrian research funding association (FFG) under the scope of the COMET programme within the research project “Industrial Methods for Process Analytical Chemistry - From Measurement Technologies to Information Systems (imPACts)” (contract #843546). M.R.A. gratefully acknowledges the financial support provided by CONICET.

References

- [1] J.A. O'Mahony, P.F. Fox, Milk: An Overview, in: H. Singh, M. Boland, A. Thompson (Eds.) Milk Proteins: From Expression to Food, 2nd Edition 2014, pp. 19-73.
- [2] G.F. Greppi, P. Roncada, R. Fortin, Protein Components of Goat's Milk, in: A. Cannas, G. Pulina (Eds.) Dairy Goats Feeding and Nutrition, CABI, Cambridge, MA, 2008, pp. 71-94.
- [3] A. Haug, A.T. Høstmark, O.M. Harstad, Bovine milk in human nutrition – a review, *Lipids Health Dis.*, 6 (2007) 1-16.
- [4] P.F. Fox, Chapter 1 - Milk: an overview, in: A.T.B. Singh (Ed.) Milk Proteins, Academic Press, San Diego, 2008, pp. 1-54.
- [5] J.L. Audic, B. Chauffer, G. Daufin, Non-food applications of milk components and dairy co-products: A review, *Lait*, 83 (2003) 417-438.
- [6] A. Carr, M. Golding, Functional Milk Proteins Production and Utilization: Casein-Based Ingredients, in: H.P.L. McSweeney, A.J. O'Mahony (Eds.) Advanced Dairy Chemistry: Volume 1B: Proteins: Applied Aspects, Springer New York, New York, NY, 2016, pp. 35-66.
- [7] G.W. Smithers, Whey and whey proteins—From 'gutter-to-gold', *Int. Dairy J.*, 18 (2008) 695-704.
- [8] D.E.W. Chatterton, G. Smithers, P. Roupas, A. Brodkorb, Bioactivity of beta-lactoglobulin and alpha-lactalbumin - Technological implications for processing, *Int. Dairy J.*, 16 (2006) 1229-1240.
- [9] S.M. Kamau, S.C. Cheison, W. Chen, X.M. Liu, R.R. Lu, Alpha-Lactalbumin: Its Production Technologies and Bioactive Peptides, *Compr. Rev. Food Sci. Food Saf.*, 9 (2010) 197-212.
- [10] B. Ribadeau-Dumas, R. Grappin, Milk protein analysis, *Lait*, 69 (1989) 357-416.
- [11] E.D. Strange, E.L. Malin, D.L. Van Hekken, J.J. Basch, Chromatographic and electrophoretic methods used for analysis of milk proteins, *J. Chromatogr. A*, 624 (1992) 81-102.

- [12] G. Bordin, F. Cordeiro Raposo, B. de la Calle, A.R. Rodriguez, Identification and quantification of major bovine milk proteins by liquid chromatography, *J. Chromatogr. A*, 928 (2001) 63-76.
- [13] G. Bobe, D.C. Beitz, A.E. Freeman, G.L. Lindberg, Separation and Quantification of Bovine Milk Proteins by Reversed-Phase High-Performance Liquid Chromatography, *J. Agric. Food Chem.*, 46 (1998) 458-463.
- [14] J. Leonil, V. Gagnaire, D. Molle, S. Pezennec, S. Bouhallab, Application of chromatography and mass spectrometry to the characterization of food proteins and derived peptides, *J. Chromatogr. A*, 881 (2000) 1-21.
- [15] H.K. Mayer, B. Raba, J. Meier, A. Schmid, RP-HPLC analysis of furosine and acid-soluble beta-lactoglobulin to assess the heat load of extended shelf life milk samples in Austria, *Dairy Sci. Technol.*, 90 (2010) 413-428.
- [16] F.-T.A. Chen, A. Tusak, Characterization of food proteins by capillary electrophoresis, *J. Chromatogr. A*, 685 (1994) 331-337.
- [17] M. Carbonaro, A. Nucara, Secondary structure of food proteins by Fourier transform spectroscopy in the mid-infrared region, *Amino Acids*, 38 (2010) 679-690.
- [18] B. Aernouts, E. Polshin, W. Saeys, J. Lammertyn, Mid-infrared spectrometry of milk for dairy metabolomics: A comparison of two sampling techniques and effect of homogenization, *Anal. Chim. Acta.*, 705 (2011) 88-97.
- [19] N.A. Kocaoglu-Vurma, A. Eliardi, M.A. Drake, L.E. Rodriguez-Saona, W.J. Harper, Rapid Profiling of Swiss Cheese by Attenuated Total Reflectance (ATR) Infrared Spectroscopy and Descriptive Sensory Analysis, *J. Food Sci.*, 74 (2009) S232-S239.
- [20] D. Wu, Y. He, S. Feng, D.-W. Sun, Study on infrared spectroscopy technique for fast measurement of protein content in milk powder based on LS-SVM, *J. Food Eng.*, 84 (2008) 124-131.

- [21] Y. Etzion, R. Linker, U. Cogan, I. Shmulevich, Determination of protein concentration in raw milk by mid-infrared Fourier transform infrared/attenuated total reflectance spectroscopy, *J. Dairy Sci.*, 87 (2004) 2779-2788.
- [22] S. Mazurek, R. Szostak, T. Czaja, A. Zachwieja, Analysis of milk by FT-Raman spectroscopy, *Talanta*, 138 (2015) 285-289.
- [23] J.C. Moore, J.W. DeVries, M. Lipp, J.C. Griffiths, D.R. Abernethy, Total Protein Methods and Their Potential Utility to Reduce the Risk of Food Protein Adulteration, *Compr. Rev. Food Sci. Food Saf.*, 9 (2010) 330-357.
- [24] A. International, Official methods of analysis, 17th ed., AOAC International, Gaithersburg, Md, 2005.
- [25] A. Barth, Infrared spectroscopy of proteins, *Biochim. Biophys. Acta, Bioenerg.*, 1767 (2007) 1073-1101.
- [26] H.J. Luinge, E. Hop, E.T.G. Lutz, J.A. Vanhemert, E.A.M. Dejong, Determination of the Fat, Protein and Lactose Content of Milk Using Fourier-Transform Infrared Spectrometry, *Anal. Chim. Acta.*, 284 (1993) 419-433.
- [27] 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.
- [28] M.R. Alcaraz, 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.
- [29] A. Schwaighofer, M.R. Alcaraz, 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.

- [30] C. Wagner, A. Genner, G. Ramer, B. Lendl, Advanced Total Lab Automation System (ATLAS), Modeling, Programming and Simulations Using LabVIEW2011.
- [31] T.G. Bloemberg, J. Gerretzen, A. Lunshof, R. Wehrens, L.M.C. Buydens, Warping methods for spectroscopic and chromatographic signal alignment: A tutorial, *Anal. Chim. Acta.*, 781 (2013) 14-32.
- [32] L. Zheng, D.G. Watson, B.F. Johnston, R.L. Clark, R. Edrada-Ebel, W. Elseheri, A chemometric study of chromatograms of tea extracts by correlation optimization warping in conjunction with PCA, support vector machines and random forest data modeling, *Anal. Chim. Acta.*, 642 (2009) 257-265.
- [33] H. Parastar, M. Jalali-Heravi, R. Tauler, Comprehensive two-dimensional gas chromatography (GC \times GC) retention time shift correction and modeling using bilinear peak alignment, correlation optimized shifting and multivariate curve resolution, *Chemom. Intell. Lab. Syst.*, 117 (2012) 80-91.
- [34] R.H. Jellema, Variable Shift and Alignment, in: B. Walczak, S.D. Brown, R. Tauler (Eds.) *Comprehensive Chemometrics*, Elsevier, Oxford, 2009, pp. 85-108.
- [35] R. Marbach, On Wiener filtering and the physics behind statistical modeling, *J. Biomed. Opt.*, 7 (2002) 130-147.
- [36] R. Marbach, A new method for multivariate calibration, *J. Near Infrared Spectrosc.*, 13 (2005) 241-254.
- [37] R. Marbach, *Methods to significantly reduce the calibration cost of multichannel measurements instruments*, 2003.
- [38] R. Marbach, Multivariate calibration, selectivity, and the SBC method., *Chem. Ing. Tech.*, 82 (2010) 453-466.
- [39] J.D.J. van den Berg, N.D. Vermist, L. Carlyle, M. Holcapek, J.J. Boon, Effects of traditional processing methods of linseed oil on the composition of its triacylglycerols, *J. Sep. Sci.*, 27 (2004) 181-199.

- [40] J. Kuligowski, G. Quintas, S. Garrigues, M. de la Guardia, Chemometric extraction of analyte-specific chromatograms in on-line gradient LC-infrared spectrometry, *J. Sep. Sci.*, 32 (2009) 4089-4095.
- [41] J. Kuligowski, G. Quintas, S. Garrigues, M. de la Guardia, Monitoring of Polymerized Triglycerides in Deep-Frying Oil by On-Line GPC-FTIR Spectrometry Using the Science Based Calibration Multivariate Approach, *Chromatographia*, 71 (2010) 201-209.
- [42] M. Cascant, J. Kuligowski, S. Garrigues, M. de la Guardia, Determination of sugars in depilatory formulations: A green analytical method employing infrared detection and partial least squares regression, *Talanta*, 85 (2011) 1721-1729.
- [43] E.L. Malin, M.H. Alaimo, E.M. Brown, J.M. Aramini, M.W. Germann, H.M. Farrell, P.L.H. McSweeney, P.F. Fox, Solution Structures of Casein Peptides: NMR, FTIR, CD, and Molecular Modeling Studies of α 1-Casein, 1–23, *J. Protein Chem.*, 20 (2001) 391-404.
- [44] D.M. Curley, T.F. Kumosinski, J.J. Unruh, H.M. Farrell, Changes in the secondary structure of bovine casein by Fourier transform infrared spectroscopy: Effects of calcium and temperature, *J. Dairy Sci.*, 81 (1998) 3154-3162.
- [45] H.M. Farrell, E.D. Wickham, J.J. Unruh, P.X. Qi, P.D. Hoagland, Secondary structural studies of bovine caseins: temperature dependence of beta-casein structure as analyzed by circular dichroism and FTIR spectroscopy and correlation with micellization, *Food Hydrocolloids*, 15 (2001) 341-354.
- [46] S.J. Prestrelski, D.M. Byler, M.P. Thompson, Effect of Metal-Ion Binding on the Secondary Structure of Bovine Alpha-Lactalbumin as Examined by Infrared-Spectroscopy, *Biochemistry*, 30 (1991) 8797-8804.
- [47] E.A. Permyakov, L.J. Berliner, α -Lactalbumin: structure and function, *FEBS Lett.*, 473 (2000) 269-274.

[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-Å Resolution, *J. Mol. Biol.*, 197 (1987) 695-706.

[49] Y. Jung, J. Hwang, Near-Infrared Studies of Glucose and Sucrose in Aqueous Solutions: Water Displacement Effect and Red Shift in Water Absorption from Water-Solute Interaction, *Appl. Spectrosc.*, 67 (2013) 171-180.

[50] Food and Drug Administration (FDA) (2001) Guidance for industry: bioanalytical method validation.

ACCEPTED MANUSCRIPT

Legends of Figures

Figure 1. Mid-IR absorbance spectra of Cas, α -LA, β -LG standards covering concentrations between 2.5 and 20 g L⁻¹ and standard mixtures at different proportions ranging between 3 and 30 g L⁻¹, 3 and 10 g L⁻¹ and 3 and 15 g L⁻¹ for Cas, α -LA and β -LG, respectively, using a phosphate buffer spectrum as background. Black lines: major Cas content; dark grey line: major α -LA content; light grey lines: major β -LG content; dashed lines located at 1653 cm⁻¹ (magenta), 1650 cm⁻¹ (cyan) and 1632 cm⁻¹ (blue) reflect the positions of the absorption band maxima in mid-IR spectra of α -LA, Cas and β -LG, respectively.

Figure 2. Mid-IR absorbance spectra of whole milk before (black dotted line) and after (black solid line) SBC based background correction. The spectrum of α -Lac (grey solid line) was employed for compensating the matrix effect caused by varying H₂O contents in bovine milk. Dashed lines located at 1653 cm⁻¹ (magenta), 1650 cm⁻¹ (cyan) and 1632 cm⁻¹ (blue) reflect positions of the absorption band maxima in mid-IR spectra of α -LA, Cas and β -LG, respectively.

Figure 3. Nominal vs. predicted concentrations of Cas, α -LA, β -LG and total proteins in standards (STDs) and milk samples on different measurement days (Day 1, 2 and 3) by employing the described PLS models.

Table 1. PLS calibration parameters and internal figures of merit.

Parameter	Day 1				Day 2			Day 3			
	Cas (1-35 g L ⁻¹)	α -LA (1-10 g L ⁻¹)	β -LG (1-10 g L ⁻¹)	Total (18-37 g L ⁻¹)	Cas (3-30 g L ⁻¹)	β -LG (3-15 g L ⁻¹)	Total (8-35 g L ⁻¹)	Cas (0-30 g L ⁻¹)	α -LA (0-10 g L ⁻¹)	β -LG (0-15 g L ⁻¹)	Total (19-36 g L ⁻¹)
Preprocessing	MC	2 nd Der. + MC	MC	MC	MC	MC	MC	MC	2 nd Der. + MC	MC	MC
LVs	4	2	6	4	2	2	4	4	5	6	4
R² cal	0.998	0.95	0.9999	0.996	0.997	0.99	0.996	0.997	0.998	0.9997	0.998
R² CV	0.99	0.8	0.994	0.97	0.993	0.96	0.98	0.991	0.99	0.998	0.997
RMSEC [g L⁻¹]	0.5	0.7	0.03	0.4	0.5	0.5	0.5	0.6	0.4	0.09	0.2
RMSECV [g L⁻¹]	1.3	1.3	0.2	1.3	0.8	0.8	1.2	1	1.4	0.2	0.3
CV bias [g L⁻¹]	-0.2	0.0005	-0.000014	-0.3	0.003	0.07	0.2	-0.1	0.06	0.05	-0.002

Note: MC = mean centering; 2nd Der. = second derivative calculated using the Savitzky-Golay filter (order: 3, window: 15 points); LV = latent variable; R² = coefficient of determination; cal = calibration; CV = cross validation; RMSEC = root mean square error of calibration given in g L⁻¹

¹; RMSE = $\sqrt{\frac{1}{I} \sum_1^I (c_{\text{nom}} - c_{\text{pred}})^2}$; RMSECV = root mean square error of cross validation.

Table 2. Recoveries values obtained for protein standard mixtures.

Protein	Intra-Day Recovery [%] \pm sd (protein concentration g L ⁻¹)			Inter-Day Recovery [%] \pm sd (protein concentration g L ⁻¹)		
	low	medium	high	low	medium	high
Cas	92 \pm 11 (3)	100 \pm 3 (15)	101 \pm 2 (30)	93 \pm 16 (3)	99 \pm 2 (15)	99 \pm 3 (30)
α-LA	98 \pm 5 (3)	99 \pm 9 (5)	99 \pm 2 (10)	90 \pm 12 (3)	105 \pm 9 (5)	93 \pm 9 (10)
β-LG	100 \pm 2 (3)	100.0 \pm 1.3 (7)	100.1 \pm 0.4 (15)	96 \pm 6 (3)	102 \pm 4 (7)	100.1 \pm 0.4 (15)
Total	99.02 \pm 0.15 (23)	100.25 \pm 0.10 (32)	100.0 \pm 0.2 (36)	100.0 \pm 1.2 (23)	100.7 \pm 0.5 (32)	100.3 \pm 0.4 (36)

Table 3. Protein concentrations found in blank bovine milk samples and recovery values obtained from fortified bovine milk samples.

Protein	Concentration [g L⁻¹] ± sd	Intra-Day Recovery [%] ± sd	Inter-Day Recovery [%] ± sd
Cas	27 ± 3	109.1 ± 1.2	102 ± 10
α-LA	2.0 ± 0.5	88 ± 6	84 ± 6
β-LG	2.0 ± 0.3	93 ± 8	100 ± 10
Total	29.1 ± 1.1	101.3 ± 0.8	97 ± 6

