# Laser-Based Mid-infrared Spectroscopy Enables In-line Detection of Protein Secondary Structure from Preparative Liquid Chromatography

Christopher K. Akhgar[a], Julian Ebner[b], Oliver Spadiut[b],
Andreas Schwaighofer*[a], Bernhard Lendl*[a]

[a]Institute of Chemical Technologies and Analytics, Technische Universität Wien,
Getreidemarkt 9, 1060 Vienna, Austria;
[b]Institute of Chemical, Environmental and Bioscience Engineering, Technische Universität Wien,
Getreidemarkt 9, 1060 Vienna, Austria

## ABSTRACT

External cavity-quantum cascade laser (EC-QCL) based mid-infrared (IR) spectroscopy is an emerging technology for analyzing proteins in aqueous solutions. Higher sensitivity and larger applicable optical path lengths compared to conventional Fourier-transform IR (FTIR) spectroscopy open a wide range of possible applications, including near real-time protein monitoring from complex downstream operations.

In this work, an EC-QCL based mid-IR spectrometer was coupled to a preparative liquid chromatography (LC) system. The large optical path length (25 µm) and the broad tuning range of the laser (1350-1750 cm$^{-1}$) allowed robust spectra acquisition in the most important wavenumber range for protein secondary structure determination. A model system based on size exclusion chromatography (SEC) and three different proteins was employed to demonstrate the advantages of LC-QCL-IR coupling. The recorded spectra showed distinct amide I and II bands across the chromatographic run. Mid-IR spectra, extracted from the three chromatographic peak maxima showed features typical for the secondary structures of the exhibited proteins with high comparability to off-line reference spectra. Band positions and maxima of mid-IR absorbances were compared to a conventional UV detector, revealing excellent agreement of peak shapes and maxima. This work demonstrates that laser-based mid-IR spectroscopy offers the significant advantage of providing almost real-time information about protein secondary structure, which typically has to be obtained by laborious and time-consuming offline analysis. Consequently, coupling of LC and laser-based mid-IR spectroscopy holds high potential for replacing conventional off-line methods for monitoring proteins in complex biotechnological processes.

**Keywords:** Mid-infrared spectroscopy, quantum cascade laser, liquid chromatography, proteins, secondary structure

# 1. INTRODUCTION

Liquid chromatography (LC) continues to be an essential unit operation for purification of pharmaceutical proteins. Typically, the corresponding protein concentrations are monitored by single-wavelength UV/Vis spectroscopy, offering excellent sensitivity, high robustness and a broad linear range. One of the major drawbacks of these detectors is, however, that discrimination and quantitation of different co-eluting proteins is not possible. Mid-infrared (IR) spectroscopy is a powerful method for the analysis of proteins in a label-free and non-destructive manner. The amide I (1700-1600 cm$^{-1}$) and amide II (1600-1500 cm$^{-1}$) bands are the most important mid-IR regions for protein analysis. These bands allow protein quantification and provide distinct information about their secondary structure [1]. A challenge in mid-IR spectroscopy of proteins is the overlap of the amide I band and the significantly more pronounced HOH-bending band of water (centered approx. 1640 cm$^{-1}$). Conventional Fourier-transform IR (FTIR) spectrometers are equipped with low power thermal light sources. Here, the optical path-length in transmission configuration is restricted to <10 µm for protein analysis in order to avoid total absorption of IR radiation through water [2]. Such short path-lengths are not suitable for LC-IR flow-through measurements, as they lead to low robustness and limited sensitivity. As an alternative, attenuated total reflection (ATR)-FTIR spectroscopy was applied for in-line monitoring of LC effluents [3, 4]. This configuration offers higher robustness but does not overcome the limitations regarding sensitivity. Quantum cascade lasers (QCLs) are novel mid-IR light sources with optical powers higher by a factor of $10^4$ and more compared to thermal light sources [5]. Combined with an external cavity (EC), they can be tuned across several hundred wavenumbers, offering high potential for the analysis of protein amide I and amide II bands [6]. It has been shown that these higher spectral power densities allow approximately 4-5 times larger optical path-lengths for protein analysis compared to conventional FTIR instruments [7]. In this framework, academic setups were developed that allow rugged measurements and achieve sensitivities that clearly outperform high-end FTIR spectroscopy [8, 9]. These setups were successfully combined with chemometric methods to analyse individual proteins in complex matrices [10-12]. Recently, a commercially available EC-QCL based mid-IR spectrometer, the ChemDetect Analyzer (Daylight, Solutions) was introduced [13]. This device offers portability and robust spectra acquisition between 1350 and 1770 cm$^{-1}$ with an optical path-length of 25 µm. In this context, LC-QCL-IR coupling represents a novel opportunity for in-line monitoring of proteins with high potential for complementing conventional off-line methods [14]. In this work, the ChemDetect Analyzer was coupled to a lab-scale preparative-LC system to perform in-line monitoring of the protein secondary structure. A case study based on size exclusion chromatography (SEC) employing a mixture of three proteins differing in their secondary structure was performed. The shapes of amide I and amide II bands were monitored across the chromatographic run and compared to reference off-line spectra. Finally, the mid-IR signal was benchmarked against a conventional UV detector and reversed phase-high-performance LC (RP-HPLC) off-line measurements. Here, the position and shape of the SEC peaks was compared and the advantages of the ChemDetect Analyzer regarding sensitivity to secondary structure were investigated.

# 2. EXPERIMENTAL

A schematic of the LC-IR flow path is depicted in Figure 1. Chromatographic separation was performed using an ÄKTA pure system (Cytiva Life Sciences, MA, USA) equipped with a HiLoad 16/60 Superdex 200 pg column, a U9-M UV monitor and a F9-C fraction collector. Ovalbumin ($\geq$90%), $\alpha$-chymotrypsinogen A from bovine pancreas and myoglobin from equine skeletal muscle ($\geq$95%) were purchased from Sigma Aldrich (Steinheim, Germany) and dissolved in 50 mM phosphate elution buffer (pH 7.4) to obtain a mixture with concentrations of 10 mg/mL for each of the three proteins. 0.5 mL of this mixture was injected into the chromatographic system using a 0.5 mL loop, followed by isocratic separation with a constant flowrate of 0.25 mg/mL. Mid-IR based in-line monitoring was performed with a ChemDetect Analyzer (Daylight Solutions Inc., San Diego USA). The equipped EC-QCL was operated between 1350 and 1750 cm$^{-1}$ and was thermally stabilized using an external water-cooling unit with an operation temperature of 17 °C. The chromatographic effluent was measured in a flow cell, composed of 2 diamond windows and a 25 µm thick spacer. Spectra were acquired using the ChemDetect Software package. First, a reference background spectrum was obtained by averaging 121 single scans within a measurement time of 60 s. Thereafter, continuous acquisition during the chromatographic run was performed by averaging 20 scans within a measurement time of 10 s. Reference off-line spectra were acquired by averaging 91 scans during a measurement time of 45 s. Data processing and analysis was carried out with tailored in-house developed code in MatLab R2020b (Mathworks, Inc., Natick, MA, 2020). If necessary, absorption bands of water vapor in the atmosphere were subtracted from the analyte absorption spectra, followed by smoothing of the spectra with a 2$^{nd}$ order Savitzky-Golay filter with a window of 15 points, leading to a spectral resolution of 3.6 cm$^{-1}$.

Protein concentration of the collected fractions was quantified off-line using a RP-HPLC method described by Kopp et al. [15]. In short, a BioResolve RP mAb Polyphenyl column (Waters, MA, USA) was run at 70 °C and a constant flow rate of 0.4 mL/min with ultrapure water supplemented with 0.1% trifluoroacetic acid (TFA) as mobile phase A and acetonitrile supplemented with 0.1% TFA as mobile phase B. A gradient from 25% B to 75% B over 10 min was used and total run time for one measurement was 18 minutes. The injected sample volume was 2 µL and UV-VIS absorbance at 214 nm, 280 nm and 404 nm was recorded over the whole run. In order to quantify the used proteins (Ovalbumin, $\alpha$-chymotrypsinogen A, myoglobin), standard calibrations in the range from 0.0625 g/L to 1 g/L were performed for each protein.



Figure 1. Schematic of the LC-QCL-IR flow path.

# 3. RESULTS AND DISCUSSION

## 3.1 Laser-based mid-IR spectroscopy of in-line monitoring of proteins

The ChemDetect Analyzer was used for in-line monitoring of proteins from LC effluents. The obtained IR absorption spectra in the amide I+II region provide information about the protein secondary structure, thereby offering a significant advantage compared to conventional LC detectors. In SEC, the analytes are separated by their size and shape. Thus, large molecules elute first, whereas small molecules elute later due to increased interaction with the stationary phase. In the present case, ovalbumin, α-chymotrypsinogen A and myoglobulin were selected as model proteins due to different secondary structures and molecular weights. Figure 2A depicts the 3D plot (wavenumber-time-absorbance) of the conducted SEC run. The graph shows a stable baseline and distinct amide I and amide II bands of the three proteins. Figure 2B shows the absorption spectra, of the amide I bands, extracted from the chromatographic peak maxima. The first peak with a maximum at approximately 25 min can be related to ovalbumin due its higher molecular weight of 44.5 kDa [16]. Ovalbumin is composed of both, α-helices and β-sheets, resulting in two IR band maxima at 1656 and 1638 cm$^{-1}$ in the amide I region [17, 18]. The molecular mass of α-chymotrypsinogen A (25.6 kDa) [19] and myoglobin (17 kDa) [20] are more similar, thus the two proteins were not fully separated and show co-eluting peaks with maxima at approximately 60 and 70 min. α-Chymotrypsinogen A is mainly composed of β-sheet secondary structure [21]. The corresponding mid-IR spectrum at 60 min shows
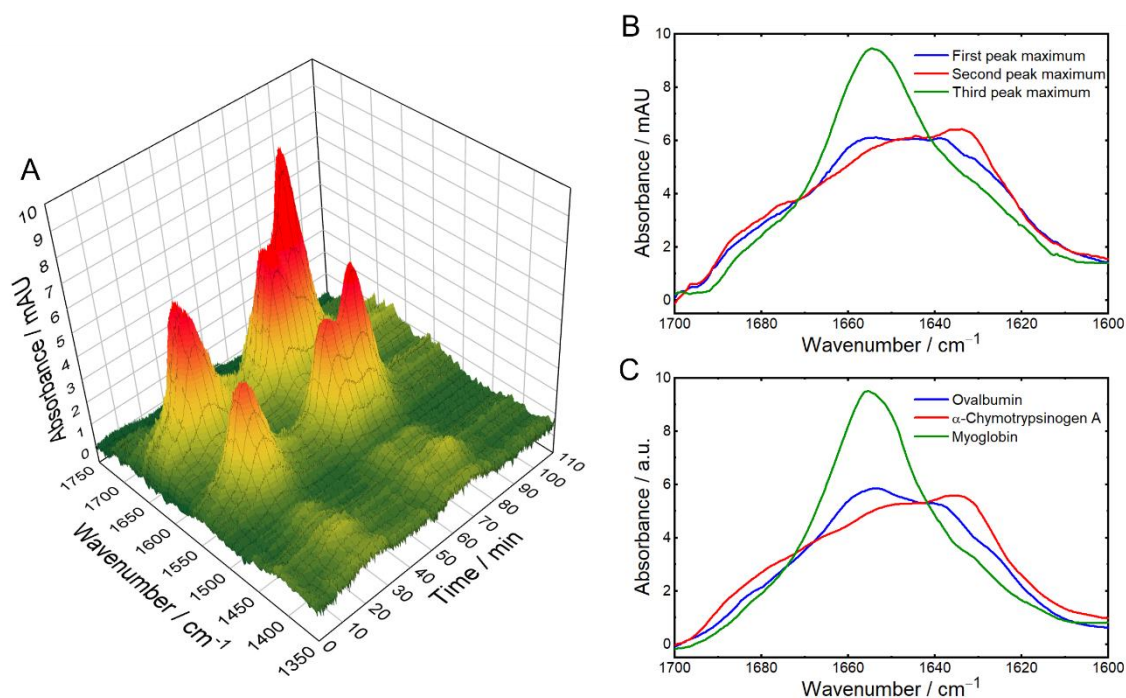


Figure 2. (A) Spectral 3D plot of the chromatographic run, containing ovalbumin, α-chymotrypsinogen A and myoglobulin. (B) IR absorption spectra, extracted from the chromatographic peak maxima (C) Reference off-line spectra of 5 mg/mL protein solutions.

the β-sheet typical amide I band maximum at approximately 1635 cm-1 and a shoulder at 1680 cm-1. Myo mainly contains α-helical secondary structure and shows the characteristic narrow amide I band with a maximum at approximately 1656 cm-1 [22]. Furthermore, the spectra of the peak maxima were compared to reference off-line measurements of pure protein solutions. Figure 2C shows ChemDetect absorption spectra of 5 mg/mL ovalbumin, α-chymotrypsinogen A and myoglobin solutions. Shape and position of the absorption bands show high comparability between in-line and off-line recorded spectra, thus allowing protein identification. The obtained detailed information about protein structure cannot be acquired with conventional LC detectors. Thus, laborious, and time-consuming off-line measurements of the collected fractions are usually performed. Consequently, LC-QCL-IR coupling holds high potential for replacing conventional off-line methods for protein monitoring in complex downstream operations.

### 3.2 Comparison between laser-based mid-IR and UV detector

In order to further demonstrate the advantage of laser-based mid-infrared spectroscopy over conventional LC detectors, the signal of the ChemDetect Analyzer was compared to the response of the UV detector. Figure 3 displays mid-IR absorbances at specific wavenumbers, as well as UV absorbance over the chromatographic run. 1656 cm-1 and 1632 cm-1 were selected in order to present more specific absorbances due to α-helix and β-sheet secondary structure, respectively. The UV signal at 280 nm is the most common wavelength for protein quantification but does not provide information regarding secondary structure. The positions of the chromatographic peaks show excellent agreement between the ChemDetect Analyzer and reference UV detector. The first peak shows comparable absorbance signals for 1656 and 1632 cm-1, because ovalbumin contains both, α-helix and β-sheet elements. The second and third chromatographic peak maxima, from two co-eluting proteins show similar signals at 280 nm.
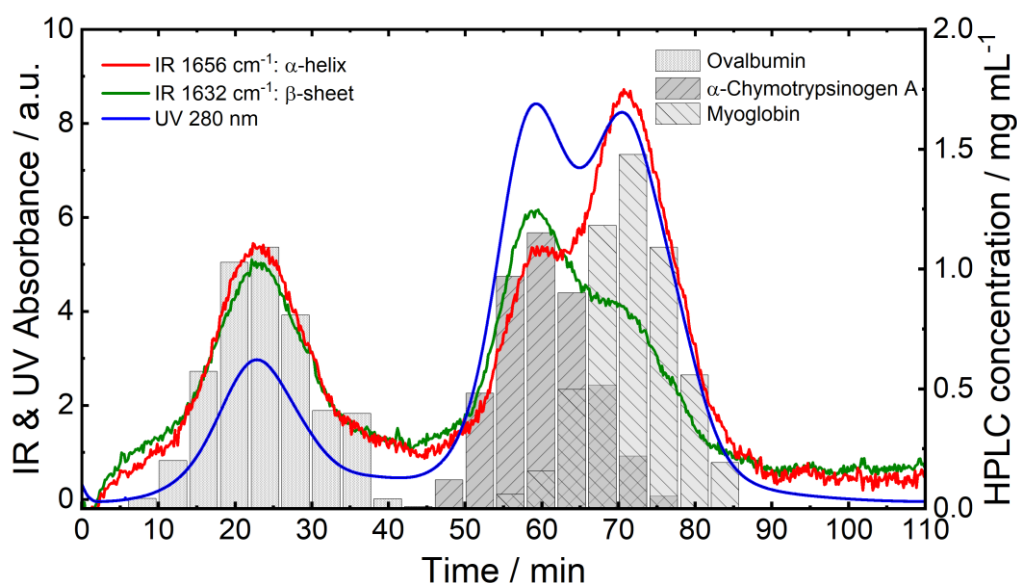


Figure 3. Comparison between signals from mid-IR analyzer at different wavenumbers (green and red lines) and UV detector (blue line) across the chromatographic run and protein reference concentrations, obtained by measuring the collected fractions with reversed-phase HPLC (grey bars).

In contrast, the signals of the ChemDetect Analyzer provide additional information. For the peak maximum at approximately 60 min, the absorbance signal at 1632 cm$^{-1}$ is more pronounced than the 1656 cm$^{-1}$ signal due to the β-sheet secondary structure of α-chymotrypsinogen A. The third maximum shows significantly higher absorbance at 1656 cm$^{-1}$ due to the α-helical structure of myoglobin. Because the UV detector signal at 280 nm does not provide any information about the secondary structure, protein identification must be performed by off-line measurements. The bars in Figure 3 indicate the protein concentrations, determined by RP-HPLC in the collected fractions. These concentrations agree well with the profile of the secondary structure specific information, obtained from ChemDetect Analyzer signal. Consequently, the obtained LC-QCL-IR chromatograms agree well with the conventionally applied detection methods, while offering the additional advantage of providing near real-time information about protein secondary structure.

# 4. CONCLUSION & OUTLOOK

In this work, EC-QCL based mid-IR spectroscopy was successfully coupled to a preparative SEC system for in-line monitoring of proteins. A model system based on three proteins with different secondary structures was applied in order to demonstrate the advantages of the ChemDetect Analyzer over conventional LC detectors. The recorded spectra across protein amide I and amide II bands showed good quality and high comparability to off-line reference IR measurements. Peak positions and maxima showed excellent agreement to a conventional UV detector and HPLC off-line measurements of the collected fractions. Here, laser-based mid-IR spectroscopy offers the significant advantage of providing almost real-time information about protein secondary structure, which enables protein identification that usually has to be obtained by applying laborious and time-consuming offline analysis. In the future, more complex chemometric methods should be applied to allow in-line quantification of proteins from co-eluting chromatographic peaks.

# REFERENCES

[1] A. Barth, "Infrared spectroscopy of proteins," Biochim. Biophys. Acta Bioenerg., 1767(9), 1073-1101 (2007).

[2] H. Fabian, and W. Mäntele, [Infrared Spectroscopy of Proteins] John Wiley & Sons, Hoboken, USA(2006).

[3] S. Großhans, M. Rüdt, A. Sanden et al., "In-line Fourier-Transform Infrared Spectroscopy as a Versatile Process Analytical Technology for Preparative Protein Chromatography," J. Chromatogr. A, 1547, 37-44 (2018).

[4] A. Sanden, S. Suhm, M. Rüdt et al., "Fourier-transform infrared spectroscopy as a process analytical technology for near real time in-line estimation of the degree of PEGylation in chromatography," J. Chromatogr. A, 1608, 460410 (2019).

[5] J. Faist, F. Capasso, D. Sivco et al., "Quantum Cascade Laser," Science, 264(5158), 553-556 (1994).

[6] A. Schwaighofer, and B. Lendl, [Quantum cascade laser-based infrared transmission spectroscopy of proteins in solution] Academic Press, Cambridge, USA, 3 (2020).

[7] M. R. Alcaraz, A. Schwaighofer, C. Kristament et al., "External-Cavity Quantum Cascade Laser Spectroscopy for Mid-IR Transmission Measurements of Proteins in Aqueous Solution," Anal. Chem., 87(13), 6980-6987 (2015).

[8] A. Schwaighofer, M. Montemurro, S. Freitag et al., "Beyond Fourier Transform Infrared Spectroscopy: External Cavity Quantum Cascade Laser-Based Mid-infrared Transmission Spectroscopy of Proteins in the Amide I and Amide II Region," Anal. Chem., 90(11), 7072-7079 (2018).

[9] C. K. Akhgar, G. Ramer, M. Żbik et al., "The Next Generation of IR Spectroscopy: EC-QCL-Based Mid-IR Transmission Spectroscopy of Proteins with Balanced Detection," Anal. Chem., 92(14), 9901-9907 (2020).

[10] J. Kuligowski, A. Schwaighofer, M. R. Alcaraz et al., "External cavity-quantum cascade laser (EC-QCL) spectroscopy for protein analysis in bovine milk," Anal. Chim. Acta, 963, 99-105 (2017).

[11] A. Schwaighofer, J. Kuligowski, G. Quintas et al., "Fast quantification of bovine milk proteins employing external cavity-quantum cascade laser spectroscopy," Food Chem., 252, 22-27 (2018).

[12] M. Montemurro, A. Schwaighofer, A. Schmidt et al., "High-throughput quantitation of bovine milk proteins and discrimination of commercial milk types by external cavity-quantum cascade laser spectroscopy and chemometrics," Analyst, 144(18), 5571-5579 (2019).

[13] A. Schwaighofer, C. K. Akhgar, and B. Lendl, "Broadband laser-based mid-IR spectroscopy for analysis of proteins and monitoring of enzyme activity," Spectrochim. Acta A, 253, 119563 (2021).

[14] C. K. Akhgar, J. Ebner, O. Spadiut et al., "QCL-IR Spectroscopy for In-line Monitoring of Proteins from Preparative Ion-Exchange Chromatography," Anal. Chem., (Submitted).

[15]    J. Kopp, F. B. Zauner, A. Pell *et al.*, "Development of a generic reversed-phase liquid chromatography method for protein quantification using analytical quality-by-design principles," J. Phar. Biomed.*,* 188, 113412 (2020).

[16]    T. Strixner, and U. Kulozik, [7 - Egg proteins] Woodhead Publishing, (2011).

[17]    P. E. Stein, A. G. W. Leslie, J. T. Finch *et al.*, "Crystal structure of ovalbumin as a model for the reactive centre of serpins," Nature*,* 347(6288), 99-102 (1990).

[18]    A. Dong, J. D. Meyer, J. L. Brown *et al.*, "Comparative Fourier Transform Infrared and Circular Dichroism Spectroscopic Analysis of α1-Proteinase Inhibitor and Ovalbumin in Aqueous Solution," Arch. Biochem. Biophys.*,* 383(1), 148-155 (2000).

[19]    P. E. Wilcox, [5 Chymotrypsinogens—chymotrypsins] Academic Press, (1970).

[20]    J. Zaia, R. S. Annan, and K. Biemann, "The correct molecular weight of myoglobin, a common calibrant for mass spectrometry," Rapid Commun. Mass Spectrom.*,* 6(1), 32-6 (1992).

[21]    S. T. Freer, J. Kraut, J. D. Robertus *et al.*, "Chymotrypsinogen: 2,5-Å crystal structure, comparison with α-chymotrypsin, and implications for zymogen activation," Biochemistry*,* 9(9), 1997-2009 (1970).

[22]    J. C. Kendrew, G. Bodo, H. M. Dintzis *et al.*, "A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis," Nature*,* 181(4610), 662-666 (1958).