



TECHNISCHE
UNIVERSITÄT
WIEN

Dissertation

Novel Mid-Infrared based Methods for Milk Analysis and Bioprocess Monitoring

A thesis submitted for the degree of
Doctor of Technical Sciences (Dr. techn.)

at

Technische Universität Wien
Faculty of Technical Chemistry
Institute of Chemical Technologies and Analytics

under supervision of

Univ. Prof. Dr. Bernhard Lendl
Dr. Andreas Schwaighofer

defended by

DI Christopher Karim Akhgar
Mat.Nr. 01140217

Vienna, September 2022

Abstract

Mid-infrared (IR) spectroscopy provides chemically meaningful information by detecting absorption of IR radiation as result of excitation of fundamental molecular vibrations. Compared to many other techniques applied within the biochemical field, it offers advantages such as short measurement times, low costs, minimum sample preparation and non-destructive measurement. In this work, especially the advantages of quantum cascade lasers (QCLs), representing novel high-power mid-IR light sources, were used to develop new mid-IR based methods for milk analysis and bioprocess monitoring.

First, a mid-IR transmission setup for protein analysis in aqueous solutions was developed that achieved a limit of detection (LOD) as low as 0.0025 mg/mL, thus clearly outperforming previous academic setups and high-end Fourier-transform IR (FTIR) spectrometers. This improvement was accomplished by combining a latest-generation external cavity (EC)-QCL with a balanced detection scheme that reduced the noise level by a factor of up to 20 in balanced mode compared to single-channel measurements. The large optical path length of 26 μm allowed for robust spectra acquisition of the amide I + amide II band when analyzing aqueous protein samples. Here, successful identification of protein secondary structures was possible at concentrations of 0.1 mg/mL using measurement times of 45 s. This opens a wide range of possible future applications including determination of different proteins in human milk, which cannot be readily met by conventional FTIR spectroscopy.

Furthermore, mid-IR based fatty acid analysis in milk was demonstrated using a novel approach based on solvent-free lipid separation and attenuated total reflection (ATR)-FTIR spectroscopy. The recorded spectra of the lipid fraction showed distinctly better resolved fatty acid related features compared to transmission spectra of whole milk samples. Partial least squares (PLS) based multivariate calibration equations were calculated by relating ATR-FTIR spectra to gas chromatography-mass spectrometry (GC-MS) reference concentrations, showing excellent results for the most important sum parameters as well as for certain individual fatty acids in both, human milk and bovine milk. Moreover, the developed approach inherently eliminated covariation structures between individual fatty acid concentrations and total fat content, which is a common problem in FTIR based determination of the milk fatty acid profile. All calibration equations were successfully validated by relating important spectral wavenumber regions to spectral features from milk fatty acids. Consequently, the developed method bears great potential for future high-throughput and accurate fatty acid determination in bovine and human milk.

For bioprocess monitoring, a commercially available EC-QCL based mid-IR spectrometer was used that allows robust and sensitive measurements of protein spectra beyond the amide I and II region. This spectrometer was thoroughly benchmarked against a high-end FTIR spectrometer, showing a noticeably smaller footprint, approximately three times larger applicable optical path lengths and similar sensitivity without the need of liquid nitrogen cooling. Subsequently, this spectrometer was coupled to a preparative liquid chromatography (LC) system to enable inline monitoring of the protein secondary structure. Several runs based on ion exchange chromatography (IEX) were monitored. Here a major challenge was caused by IR absorbance bands due to the applied NaCl gradient that overlap with noticeably smaller protein bands. For this purpose, a novel background correction approach was presented, allowing for flexible compensation of different linear and step gradients, revealing high-quality protein spectra. Finally, QCL-IR spectroscopy was employed for monitoring of size exclusion chromatography (SEC) runs with overlapping protein peaks. Here, chemometric evaluation revealed spectra of the pure proteins as well as their concentration

profiles across the run. These results were compared to offline reference spectra and high-performance LC (HPLC) measurements of the collected fractions, showing excellent agreement in both cases. Consequently, EC-QCL based mid-IR spectroscopy represents a highly flexible process analytical technology (PAT) tool for providing inline information about proteins from bioprocesses that was previously only accessible by time- and labor-intensive offline methods.

Keywords: Infrared spectroscopy • Quantum cascade laser • Proteins • Fatty acids • Milk • Liquid chromatography

Kurzfassung

Die Spektroskopie im mittleren Infrarot (MIR) liefert chemische Informationen durch Absorption von Infrarotstrahlung, die von fundamentalen Molekülschwingungen herführt. Verglichen mit vielen anderen Techniken, die in der Biochemie Einsatz finden, bietet die MIR Spektroskopie Vorteile wie kurze Messzeiten, niedrige Kosten, minimale Probenvorbereitung und zerstörungsfreien Einsatz. In dieser Arbeit wurden speziell die Vorteile von Quantenkaskadenlasern (QCLs), welche MIR Lichtquellen mit hoher Leistung sind, eingesetzt um neue MIR Methoden für die Analyse von Milch und die Überwachung von Bioprozessen zu entwickeln.

Zuerst wurde ein MIR Transmissionssetup für die Analyse von Proteinen in wässriger Lösung entwickelt, welches eine Nachweisgrenze (LOD) von 0.0025 mg/mL erreichte und damit zuvor entwickelte Forschungsaufbauten sowie high-end Fourier-Transform-Infrarot (FTIR) Spektrometer klar übertrifft. Diese Verbesserung wurde durch die Kombination eines externen Resonator (EC)-QCL der neuesten Generation und einem „balanced detection“ Schema, welches das Rauschen im „balanced mode“ verglichen mit Einkanal Messungen bis zu einem Faktor von 20 reduzierte. Die große optische Schichtdicke von 26 μm ermöglichte robuste Aufnahme von Spektren im Bereich der Amid I + Amid II Banden von Proteinen. Hier konnten Protein Sekundärstrukturen erfolgreich bei Konzentrationen von 0.1 mg/mL identifiziert werden bei Messzeiten von 45 s. Dies öffnet eine breite Auswahl an möglichen zukünftigen Anwendungen, inklusive Bestimmung von einzelnen Proteinen in Muttermilch, was derzeit mit konventioneller FTIR Spektroskopie nicht möglich ist.

Außerdem wurde eine MIR basierte Methode zur Analyse von Fettsäuren in Milch entwickelt, die lösungsmittelfreie Lipidabtrennung und abgeschwächte Totalreflexion (ATR)-FTIR Spektroskopie kombiniert. Die aufgenommenen Spektren zeigten deutlich besser aufgelöste spektrale Merkmale die auf Fettsäuren zurückzuführen sind als Transmissionsmessungen von Vollmilch. Multivariate Kalibrationen basierend auf Regressionen der partiellen kleinsten Quadrate (PLS) wurden berechnet indem ATR-FTIR Spektren mit Referenzkonzentrationen aus Gaschromatographie-Massenspektrometrie (GC-MS) Messungen in Bezug gebracht wurden. Diese zeigten sehr gute Ergebnisse für die wichtigsten Summenparameter und bestimmte individuelle Fettsäuren in Kuh- und Muttermilch. Außerdem verhindert der entwickelte Ansatz mögliche Kovarianzen zwischen individuellen Fettsäurekonzentrationen und dem Gesamtfettgehalt inhärent, was sonst ein häufiges Problem in FTIR basierter Bestimmung des Milchfettsäureprofils ist. Alle Kalibrationen wurde erfolgreich validiert indem wichtige Wellenzahlbereiche zu spektralen Merkmalen von Fettsäuren in Bezug gebracht wurden. Folglich hat die entwickelte Methode großes Potential für zukünftige akkurate Fettsäurebestimmung von Kuh- und Muttermilch mit hohem Durchsatz.

Für die Überwachung von Bioprozessen wurde ein kommerziell erhältliches EC-QCL basiertes MIR Spektrometer eingesetzt, welches robuste und empfindliche Messungen von Proteinspektren über Protein Amid I und II Region hinaus ermöglicht. Dieses Gerät wurde zuerst gründlich mit hochwertigen FTIR Spektrometer verglichen. Hier wurde gezeigt, dass die deutlich kompaktere Bauweise, die circa dreimal höher anwendbare Schichtdicke und eine ähnliche Empfindlichkeit ohne Kühlung des Detektors mittels Flüssigstickstoff erreicht wird. Danach wurde das Spektrometer mit einem präparativen Flüssigchromatographie (LC) System gekoppelt um inline Messungen der Protein Sekundärstruktur zu ermöglichen. Mehrere chromatographische Läufe basierend auf Ionenauschromatographie (IEX) wurden durchgeführt. Hier wurden Natriumchlorid Gradienten eingesetzt die zu Absorptionsbanden führten, welche mit deutlich kleineren Proteinbanden überlappten. Für diesen Zweck wurde eine neue Hintergrundkorrekturmethode entwickelt, welche flexible Kompensation von unterschiedlichen linearen und Stufengradienten ermöglichte und

dadurch qualitativ hochwertige Proteinspektren offenlegte. Schließlich wurde EC-QCL Spektrometer auch eingesetzt um Trennungen mittels Größenausschluss-Chromatographie (SEC) mit überlappenden Proteinpeaks zu überwachen. Hier wurden durch chemometrische Auswertung Spektren von den reinen Proteinen sowie deren Konzentration über den chromatographischen Lauf erhalten. Diese Ergebnisse wurden mit offline Referenzspektren und Hochleistungsflüssigkeitschromatographie (HPLC) Messungen der gesammelten Fraktionen verglichen, was in beiden Fällen sehr gute Übereinstimmung zeigte. Folglich ist EC-QCL basierte MIR eine hochflexible „Process Analytical Technology“ (PAT) Methode mit der inline Information über Proteine in Bioprocessen erhalten werden kann, welche in der Vergangenheit nur mit zeit- und kostenaufwendigen offline Methoden zugänglich war.

Schlagworte: Infrarotspektroskopie • Quantenkaskadenlaser • Proteine • Fettsäuren • Milch • Flüssigchromatographie

Acknowledgements

First of all, I would like to thank my supervisor Bernhard Lendl for giving me the opportunity to perform my PhD thesis within his group, even though my knowledge about infrared spectroscopy was very limited before the start of this work. I highly appreciate your scientific guidance, the opportunity to attend several international conferences and the chance to work in this well-funded environment that you have created.

I would like to thank all former and current member of the research group for the pleasant working atmosphere as well as numerous activities outside of university (in no particular order):

Alicja, Andreas G., Anton, Bettina, Catarina, Daniel, Davide, Dominik, Elisabeth, Felix, Giovanna, Giulia, Harald, Jakob, Jesus, Karin, Markus, Nikolaus, Paul, Raul, Savda, Shilpa, Stefan, Stephan, Ufuk and Yide. Special thanks goes to Andreas Schwaighofer for introducing me to the world of mid-IR spectroscopy and for always providing scientific as well as non-scientific input whenever I needed help. I clearly learned a lot from your writing and layout skills that have significantly improved my manuscripts, posters and presentations. I would also like to thank Georg, particularly for helping me during the first year of my PhD with electronics, programming and optics as well as for your constant IT support. Thank you Pily, for always providing help regarding administrative issues. Thank you, Marlene, for performing the practical part of you bachelor thesis with me.

I would like to thank Erwin Rosenberg and Vanessa Nürnberger for sharing their knowledge about GC-MS, as well as for performing measurements of fatty acids in milk. Furthermore, I would like to thank Oliver Spadiut and his research group, especially Julian Ebner and Julian Kopp for operating the Äkta system and performing HPLC measurements, as well as for the interesting scientific discussions on protein purification.

I also thank the project partners from *NUTRISHIELD*. Unfortunately, most of our meetings could not take place in person due to the ongoing pandemic. Special thanks goes to Julia Kuligowski, Victoria Ramos-Garcia, Isabel Ten-Doménech and Alba Morena-Gimenez for supplying us with samples and sharing your expertise on the composition and analysis of human milk.

Thank you to my friends outside of university for the good times, for distracting me from my PhD project and for your understanding when I did not have time for you due to a busy schedule at work.

I sincerely thank my girlfriend Milli for always motivating and supporting me during all ups and downs.

Finally, I would like to thank my family, especially my parents for their constant support throughout my education. Without them, none of my studies would have been possible.

Content

<i>Abstract</i>	<i>iii</i>
<i>Kurzfassung</i>	<i>v</i>
<i>Acknowledgements</i>	<i>vii</i>
<i>Content</i>	<i>viii</i>
<i>Abbreviations</i>	<i>x</i>
<i>Symbols</i>	<i>xii</i>
Chapter 1 Introduction	1
1.1 <i>Motivation and scope of the work</i>	1
1.2 <i>Fundamentals of mid-infrared spectroscopy</i>	4
1.2.1 <i>Molecular vibrations</i>	4
1.2.2 <i>Fourier-transform infrared spectroscopy</i>	5
1.2.3 <i>Transmission and attenuated total reflection spectroscopy</i>	7
Chapter 2 Laser-based IR spectroscopy of proteins	11
2.1 <i>Mid-IR transmission spectroscopy of proteins</i>	11
2.2 <i>Quantum cascade lasers</i>	14
2.2.1 <i>QCL design</i>	14
2.2.2 <i>Advantages and disadvantages of QCLs for transmission measurements</i>	15
2.3 <i>State-of-the art in EC-QCL based transmission spectroscopy of proteins</i>	17
2.3.1 <i>First-generation setup (absorbance spectroscopy)</i>	17
2.3.2 <i>Second-generation setup (absorbance spectroscopy)</i>	18
2.3.3 <i>Alternative setups</i>	19
Chapter 3 Composition and analysis of milk	21
3.1 <i>Milk composition</i>	21
3.1.1 <i>Human milk composition and differences to bovine milk</i>	21
3.1.2 <i>Protein composition</i>	23
3.1.3 <i>Fat composition</i>	24
3.2 <i>Established methods for analyzing milk</i>	27
3.2.1 <i>Conventional analytical methods</i>	27
3.2.2 <i>Mid-Infrared based methods</i>	29
3.3 <i>Novel methods developed within this work</i>	34
3.3.1 <i>Development of a balanced EC-QCL-based IR transmission setup for protein analysis</i>	34
3.3.2 <i>Fatty acid analysis</i>	35
Chapter 4 Bioprocess monitoring	39
4.1 <i>Preparative liquid chromatography</i>	39
4.1.1 <i>Instrumentation</i>	39
4.1.2 <i>Separation principles</i>	40
4.2 <i>Conventional Methods for Liquid Chromatography Monitoring</i>	42
4.2.1 <i>Detectors</i>	42
4.2.2 <i>Offline methods</i>	43
4.3 <i>Hyphenation of LC with IR spectroscopy</i>	44
4.3.1 <i>State-of-the art</i>	44
4.3.2 <i>Novel methods developed within this work</i>	45
Chapter 5 Green analytical chemistry	49
5.1 <i>Background</i>	49

5.1.1	The twelve principles of green analytical chemistry (SIGNIFICANCE)	50
5.2	<i>Green metric tools</i>	52
5.2.1	The Analytical GREENness (AGREE) approach	52
5.2.2	Alternative metrics to assess greenness.....	54
5.3	<i>Evaluation of the developed IR based methods according to AGREE metric</i>	55
Chapter 6	Results and introduction to the scientific publications	57
	<i>Publication I</i>	58
	<i>Publication II</i>	58
	<i>Publication III</i>	59
	<i>Publication IV</i>	59
	<i>Publication V</i>	60
	<i>Publication VI</i>	60
	<i>Publication VII</i>	61
	<i>Publication VIII</i>	61
Chapter 7	Conclusion and outlook	63
Bibliography	65
Appendix	75
A.	<i>Scientific publications</i>	75
	Publication I	75
	Publication II	83
	Publication III	98
	Publication IV.....	106
	Publication V.....	124
	Publication VI.....	130
	Publication VII.....	139
	Publication VIII.....	148
B.	<i>Curriculum Vitae</i>	158

Abbreviations

acousto-optic modulator	AOM
alternating least squares	ALS
aluminium antimonide	AlSb
aluminium indium arsenide	AlInAs
ammonia	NH ₃
ammonium hydroxide	NH ₄ OH
ammonium sulfate	(NH ₄) ₂ SO ₄
attenuated total reflection	ATR
barium fluoride	BaF ₂
bicinchoninic acid	BCA
boric acid	H ₃ BO ₃
boron trifluoride	BF ₃
bovine serum albumin	BSA
calcium chloride	CaCl ₂
calcium fluoride	CaF ₂
capillary electrophoresis	CE
copper (II) sulfate	CuSO ₄
correlation optimized warping	COW
cross-validation	CV
cross-validation coefficient of determination	R ² _{CV}
deuterated L-alanine doped triglycine sulfate	DLaTGS
distributed feedback	DFB
enzyme-linked immunosorbent assay	ELISA
external cavity	EC
fatty acid methyl ester	FAME
flame ionization detector	FID
Food and Drug Administration	FDA
Fourier-transform infrared	FTIR
gas chromatography	GC
green analytical chemistry	GAC
high-performance liquid chromatography	HPLC
high-temperature-short time	HTST
hydrochloric acid	HCl
hydrophobic interaction chromatography	HIC
indium arsenide	InAs
indium gallium arsenide	InGaAs
indium phosphide	InP
infrared	IR
ion-exchange chromatography	IEX
isoelectric point	pI
limit of detection	LOD
limit of quantification	LOQ
liquid chromatography	LC
long-chain fatty acids	LCFA
M-(trifluoromethyl)phenyl trimethylammoniumhydroxide	TMTFTH
mass spectrometry	MS

medium-chain fatty acids	MCFA
mercury cadmium telluride	MCT
microfluidic modulation spectroscopy	MMS
milk fat globule	MFG
milk fat globule membrane	MFGM
monounsaturated fatty acids	MONO
multivariate curve resolution	MCR
NO-Bis(trimethylsilyl)trifluoroacetamide	BSTFA
normal phase chromatography	NPC
partial least squares	PLS
polyethylene glycol	PEG
polyethylene terephthalate	PET
polytetrafluorethylene	PTFE
polyunsaturated fatty acids	PUFA
potassium hydroxide	KOH
potassium sulfate	K ₂ SO ₄
prediction coefficient of determination	R ² _p
process analytical technology	PAT
quality by design	QbD
quantum cascade laser	QCL
radioimmunoassay	RIA
reference spectra matrix	RSM
reversed phase chromatography	RPC
root-mean-square error of prediction	RMSEP
root-mean-square error of cross-validation	RMSECV
sample matrix	SM
saturated fatty acids	SAT
self-modelling mixture analysis	SMMA
short-chain fatty acids	SCFA
simple to use interactive self-modelling mixture analysis	SIMPLISMA
size exclusion chromatography	SEC
sodium chloride	NaCl
sodium dodecyl sulfate-polyacrylamide gel electrophoresis	SDS-PAGE
sodium hydroxide	NaOH
sulfuric acid	H ₂ SO ₄
Technische Universität Wien	TUW
trifluoroacetic acid	TFA
trimethylsulfoniumhydroxide	TMSH
ultrahigh temperature	UHT
ultraviolet/visible	UV/Vis
unsaturated fatty acids	UNSAT
vibrational circular dichroism	VCD
World Health Organization	WHO
zinc selenide	ZnSe
α-lactalbumin	α-LA
β-lactoglobulin	β-LG

Symbols

absorbance	$A(\tilde{\nu})$
allowed energy states	V_{iv}
angle of incidence	θ_1
angle of reflection	θ_2
anharmonic constant	x_i
area	A
concentration	c
conductivity	κ
critical angle of incidence	θ_c
effective thickness	d_e
electric dipole moment	μ
frequency	ν_i
intensity in absence of analyte	$I_0(\tilde{\nu})$
interferogram	$B(\delta)$
length	l
light intensity of FTIR	I_{FTIR}
light intensity of QCL	I_{QCL}
limit of detection	LOD
linear absorption coefficient	$a(\tilde{\nu})$
mass	m
measured intensity	$I(\tilde{\nu})$
molar conductivity	Λ_m
molar decadic absorption coefficient	$\varepsilon(\tilde{\nu})$
normal coordinates	Q_i
number of reflections	N
optical path length	d
path length difference	δ
penetration depth	d_p
photon energy	E
Planck constant	h
refractive index	n
resistance	R
speed of light in matter	c_m
speed of light in vacuum	c_0
spring constant	k
thickness	t
transmittance	$T(\tilde{\nu})$
vacuum wavelength	λ_0
vacuum wavenumber	$\tilde{\nu}_0$
vibrational quantum number	ν_i
wavelength	λ
wavenumber	$\tilde{\nu}$

Chapter 1 Introduction

1.1 Motivation and scope of the work

Mid-infrared (IR) spectroscopy is an analytical technique that provides chemically meaningful information by detecting absorption of IR radiation induced by fundamental molecular vibrations. Compared to many other techniques, applied within the biochemical field, it offers advantages such as short measurement times, low-cost, minimum sample preparation and non-destructive operation. Even though mid-IR spectroscopy is nowadays well established for biomedical applications, especially the advances in development of novel light sources such as quantum cascade lasers (QCLs), offer new possibilities for rapid analytical methods that target biologically relevant parameters.¹ This thesis was performed in the research group for Process Analytics at Technische Universität Wien (TUW), headed by Prof. Lendl, where the focus is on developing novel analytical methods in the field of molecular spectroscopy. Here, the aim was to combine previous knowledge, available within the research group, with novel technology in order to develop new mid-IR based methods for milk analysis and bioprocess monitoring.

In the context of milk analysis, the goal was to develop rapid methods for analyzing the protein and fatty acid composition of human milk. This part of the work was performed within the framework of the H2020 project *NUTRISHIELD*, which aims for creating a fact-based personalized nutrition platform for the young. The final project consists of a platform that should include novel analytical tools to analyze a wide range of biomarkers that are related to nutrition and health disorders. Based on the analytical findings, a nutrition assistive mobile app should then provide feedback and steering of people towards better nutrition by taking into account the way individual persons respond to different nutrients and food types. Finally, the scientific validity of the platform should be evaluated in three studies with clinically controlled settings. In order to achieve the desired outcome, the knowledge of 16 partner institutions from 11 different countries with academic and industrial experts in various fields such as laser engineering, device manufacturing, nutritional science, chemistry, computer science and more is combined. The main role of TUW is to develop rapid mid-IR based methods that can subsequently be implemented into portable prototypes by a device manufacturer. Furthermore, TUW is responsible for assessing the validity of the prototypes in the relevant operating environment.

Milk is a complex mixture of water, proteins, fat, lactose and micronutrients, providing all essential nutrients for infants. The protein composition of human milk varies with the stage of lactation and term/preterm delivery, thus additional supplementation is commonly required when donor milk is received.² Conventional methods for human milk analysis are mainly based on chromatographic and/or wet chemical steps, offering reliable results, but coming with severe drawbacks such as application of organic solvents, high costs and significant time consumption. Here, mid-IR spectroscopy should provide a more time and cost-efficient alternative for on-site analysis in the hospital. Within the *NUTRISHIELD* project, laser-based IR spectroscopy was the method of choice for protein analysis. Previously, it was already reported that QCL-based IR spectroscopy can be applied for quantifying the three most abundant proteins in bovine milk.³ As, human milk has significantly lower protein concentrations than bovine milk, more sensitive methods are required than the current state-of-the art. Thus, one of the aims was to develop a novel measurement scheme, based on a latest generation external cavity (EC)-QCL and a balanced detection module, in order to achieve highly sensitive, low noise protein measurements in aqueous solutions. The developed setup should be thoroughly benchmarked against previous setups, as well as high-end Fourier-transform infrared (FTIR) spectroscopy.

Another goal was to develop a method for analyzing the fatty acid profile of human milk, which changes with various factors including mothers diet, stage of lactation, sociodemographic variables and environmental influences.⁴ Publications were available that report fatty acid determination in bovine milk based on FTIR transmission spectroscopy of whole milk, however, these results were either rather poor or likely based on inaccurate chemometric evaluation.⁵ Thus, in this thesis a novel approach should be developed that avoids previously occurring problems. First, method development should be performed using bovine milk due to easier availability, followed by application to human milk in a proof-of-concept study. Finally, a full validation of the method for human milk should be performed by using a larger set of samples that can be divided into training and validation set.

The second part of the thesis deals with mid-IR based methods for monitoring proteins from preparative liquid chromatography (LC). Conventional LC detectors offer excellent sensitivity and a broad linear range, but do not provide information about protein structure or purity. Consequently, labor- and time-intensive offline methods have to be employed in order to analyze the collected fractions. In this thesis, mid-IR spectroscopy should be applied for inline monitoring of LC in order to obtain near real-time information about protein structure, thus type of protein. For this purpose, the ChemDetect Analyzer from Daylight Solutions Inc. (USA), representing a novel, portable, commercially available QCL-based device was applied. Being a novel and newly introduced instrument, the first goal was to benchmark it against FTIR spectroscopy and to have a closer look at its advantages in terms of broader spectral range compared to most other EC-QCL based instrumentation. The next step was to hyphenate the ChemDetect Analyzer to a preparative LC instrument in a proof-of-concept study based on size exclusion chromatography (SEC) and basic data evaluation. Second, a more comprehensive study should be performed by monitoring multiple runs based on ion-exchange chromatography (IEX). Here, a significant challenge is caused by the high absorbance signal due to the applied salt gradient that overlaps with protein signals. In this framework, one of the key targets was to implement a novel chemometric approach, capable of compensating for this matrix effect caused by changes in the hydrogen bonding of aqueous solutions due to the presence of NaCl. Finally, the QCL-IR instrument should be applied for monitoring more complex SEC runs that include overlapping chromatographic peaks. To validate

the methods based on QCL-IR inline detection, all results should be benchmarked against high-performance liquid chromatography (HPLC) offline reference measurements by comparing qualitative as well and quantitative information about proteins.

This work has been written as a cumulative PhD thesis, where the first five introductory chapters provide the most important background information, followed by a brief introduction to the results, already summarized in the scientific publications which resulted from this thesis.⁶⁻¹³ In Chapter 1, the most relevant fundamentals of IR spectroscopy are explained with a focus on the theory of molecular vibrations, FTIR and the two probing modes that were applied within this work: transmission and ATR. As laser-based-IR spectroscopy represented the key technology for protein analysis, an introduction on this topic is provided in Chapter 2. Chapter 3 gives an overview on the composition and analysis of milk, including conventional methods, implemented mid-IR based methods and novel methods developed within this work. In Chapter 4, preparative LC is introduced, followed by a discussion of LC-IR hyphenation and the novel methods applied within this work. An overarching characteristic of the analytical techniques and methods based on mid-IR spectroscopy developed in this thesis is their “greenness” compared to other established techniques and methods, both for protein analysis as well as for lipid analysis. Therefore, an additional chapter has been added which details on how the “greenness” of a method can be determined (Chapter 5). In Chapter 6, the results are presented in form of scientific publications. Results obtained are put into context and discussed and an outlook is given in Chapter 7.

1.2 Fundamentals of mid-infrared spectroscopy

Mid-IR spectroscopy provides qualitative and quantitative information about analytes in a label-free and non-destructive manner by probing fundamental molecular vibrations. This introductory chapter summarizes the most relevant facts about mid-IR spectroscopy, based on Ref¹⁴ and ¹⁵ that provide more in-depth information. In section 1.2.1 the basic principles of molecular vibrations are explained. Section 1.2.2 summarizes FTIR spectroscopy and section 1.2.3 discusses spectra acquisition in transmission as well as using the ATR technique.

1.2.1 Molecular vibrations

IR spectroscopy is based on measuring transitions between vibrational energy states of molecules. Molecular vibrations can occur in different ways such as stretching, in-plane bending and out-of-plane bending. The number of vibrational modes depends on the molecular structure and the number of atoms in the molecule, N . Molecules have $3N$ degrees of freedom, of which three can be assigned to translational motion in mutually perpendicular directions and three to rotations. The remaining $3N - 6$ degrees of freedom represent the maximum number of vibrational modes in non-linear molecules. For linear molecules, the maximum number of vibrational modes equals to $3N - 5$, as rotation about the molecular axis cannot be observed. In a first approximation, each vibrational mode can be described as a harmonic oscillator where atoms vibrate at a certain frequency, ν_i for each mode i . The allowed energy states V_{iv} can be calculated by equation 1.1:

$$V_{iv} = h\nu_i \left(v_i + \frac{1}{2} \right) \quad (1.1)$$

where h is the Planck constant and ν_i the vibrational quantum number ($v_i = 0, 1, 2, \dots$). The harmonic oscillator is a good approximation for lower vibrational quantum numbers, however, when higher states are reached, thus larger displacement of the involved atoms occurs, more accurate description of the energy levels is represented by the anharmonic *Morse* potential:

$$V_{iv} = h\nu_i \left(v_i + \frac{1}{2} \right) + h\nu_i x_i \left(v_i + \frac{1}{2} \right)^2 \quad (1.2)$$

Here, equation 1.1 is extended by an anharmonic term that includes the anharmonicity constant x_i , which typically has values between -0.001 and -0.02. In contrast to the harmonic model where changes of v_i by more than ± 1 are restricted, the anharmonic model allows for $|\Delta v_i| > 1$, thus includes combination of transitions (e.g., $\Delta v_i = 1, \Delta v_j = 1$) and overtone bands ($\Delta v_j = 2, 3, \dots$).

In IR spectroscopy, vibrational transitions are induced by absorption of electromagnetic radiation. One of the prerequisites for this technique is that the photon energy E of the radiation has to match the energy of the transition. The photon energy can be calculated as follows:

$$E = h\nu_i = h \frac{c_0}{\lambda_0} = hc_0 \tilde{\nu}_0 \quad (1.3)$$

where, c_0 is the speed of light in vacuum, λ_0 denotes the vacuum wavelength and $\tilde{\nu}_0$ indicates the vacuum wavenumber. In mid-IR spectroscopy, the wavenumber $\tilde{\nu}$ with the unit cm^{-1} is commonly used as a measure for the energy instead of the frequency or wavelength. The relationship between wavelength λ , frequency, and wavenumber is given in equation 1.4:

$$\lambda = \frac{1}{\tilde{\nu}} = \lambda_0 \cdot n; \quad n = \frac{c_0}{c_m} \quad (1.4)$$

where n is the refractive index and c_m the speed of light in matter. Depending on the energy of a photon, different transitions are excited in the molecule. Based on the located wavenumbers of the excited vibrational modes, the IR range is typically subdivided as follows:

The mid-IR region is located between $4000 - 400 \text{ cm}^{-1}$ ($2.5 - 25 \text{ }\mu\text{m}$) which corresponds to the energy of fundamental transitions from the ground state $v_i = 0$ to the first excited state $v_i = 1$, where different stretching and bending modes of intramolecular transitions can be observed. In the near-IR region ($13300 - 4000 \text{ cm}^{-1}$, $0.75 - 2 \text{ }\mu\text{m}$), overtone and combination bands of these vibrations can be found which are typically less intense than fundamental vibrations. In the far-IR region ($400 - 10 \text{ cm}^{-1}$, $25 \text{ }\mu\text{m} - 1000 \text{ }\mu\text{m}$), mainly skeleton, rotational and lattice vibrations are induced.

Another prerequisite for IR absorption is that electromagnetic radiation leads to a change in the dipole moment μ of the molecule, thus not all theoretical possible transitions are always allowed. At least one component of the electric dipole moment must change, meaning that

$$\frac{\partial \mu}{\partial Q_i} \neq 0 \quad (1.5)$$

where Q_i denotes normal coordinates. For molecules with certain symmetry elements, some vibrational modes can degenerate whereas others are completely forbidden. Even though the large majority of molecules can be probed with IR spectroscopy, certain compounds such as diatomic, symmetric molecules (*e.g.*, H_2 and N_2) are not IR active. IR active compounds show absorption, specific to functional groups such as $-\text{CH}_2$, $-\text{OH}$ and $\text{C}=\text{O}$ in the mid-IR region. These localized vibrational modes arise from large displacement of specific atoms, whereas the remaining molecule is hardly involved. The position of such vibrations can be approximated with a ball spring model:

$$\nu_i = \frac{1}{2\pi} \sqrt{k \left(\frac{1}{m_1} + \frac{1}{m_2} \right)} \quad (1.6)$$

where k is the spring constant, representing the strength of the chemical bond, and m_1 and m_2 define the masses of the involved atoms. Consequently, the positions of stretching vibrations of molecules are not only defined by the involved atoms but also the type of bond. The wavenumber regions in which such functional group specific vibrations are observed is often referred to as *characteristic group frequencies*. In contrast, other modes involving movement of the whole molecule are typically found below 1500 cm^{-1} . This spectral region can be used to distinguish molecules with similar functional groups and structures and is often referred to as the *fingerprint region*. For IR measurements of small molecules in gas phase, narrow structures can be observed due to the large number of rotational vibrational transitions that occur. For liquid phase samples rotational transitions are not observed and molecular vibrations are additionally influenced by the chemical environment, *e.g.*, interactions with the solvent or other dissolved molecules. Thus, liquid phase measurements show significantly less resolved structures and are typically recorded with a spectral resolution of $2\text{-}8 \text{ cm}^{-1}$.

Besides the detailed qualitative information that is provided by mid-IR spectroscopy, quantitative evaluation is possible as well (see section 1.2.3).

1.2.2 Fourier-transform infrared spectroscopy

In general, mid-IR spectrometers consist of a light source, interferometer or dispersive element, sample compartment and detector. Early mid-IR instrumentation that emerged in the 1940s were dispersive instruments, where the IR beam is first focused on the sample, followed by dispersion

by a prism or grating and stepwise spectra acquisition. A major drawback of this technique is the slow spectra acquisition and related stability issues, particularly for biological samples that have to be exposed to the IR radiation during the long measurement times.¹⁶

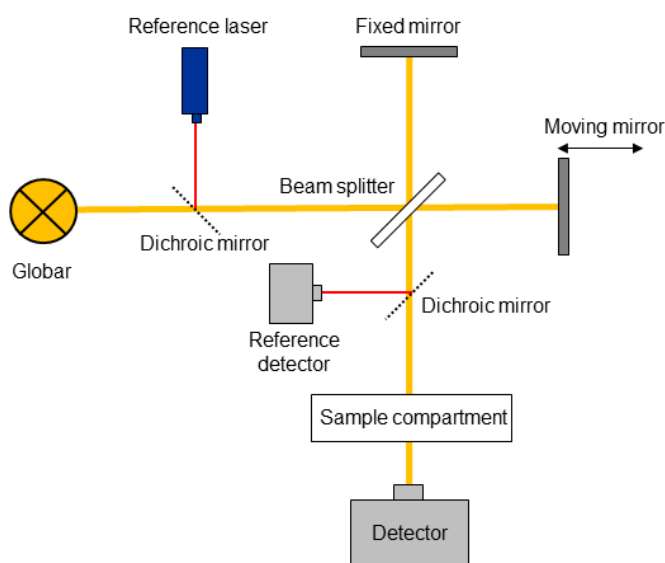


Figure 1: Schematic of a Fourier-transform infrared spectrometer.

Development of the fast Fourier-transform algorithm in the 1960s¹⁷ opened the path for interferometer based spectrometers. Figure 1 shows a schematic of a FTIR spectrometer, which are nowadays the most established instruments. This type of instrumentation is typically equipped with a broadband thermal light source, also referred to as globar, made out of silicon carbide that is heated to approximately 1200 K. The emitted black-body radiation, covering the whole mid-IR region, is guided through a Michelson interferometer,¹⁸ which in its original implementation consists of two mutually perpendicular plane mirrors, with one on a fixed position while the other moves either with a constant velocity (continuous-scan interferometer) or is held at equally spaced points for fixed periods of time (step-scan interferometer). The IR beam is first divided into two equal parts by a beam splitter, before the reflected part is guided to the fixed mirror while the other part transmits to the moving mirror. The retro-reflected beams are then recombined at the beam splitter, leading to either constructive or destructive interferences, depending on the position of the moving mirror. The thereby generated interferogram $B(\delta)$ can be described by equation 1.7, which represents a cosine function of the light intensity $I_0(\tilde{\nu})$, the path length difference between the two mirrors δ and the wavenumber of the incoming radiation:¹⁹

$$B(\delta) = \int_{-\infty}^{\infty} I_0(\tilde{\nu}) \cos(2\pi\tilde{\nu}\delta) d\tilde{\nu} \quad (1.7)$$

The recombined beam is then directed through a sample compartment where some of its energy is absorbed, if IR active analytes are present. The most established probing techniques for liquid phase analysis will be discussed in section 1.2.3. Finally, the beam reaches the detector which is typically either a pyrodetector or photodetector. The recorded detector signal is mathematically transformed into the desired intensity spectrum, also called single beam spectrum by applying an inversed Fourier-transformation:

$$I_0(\tilde{\nu}) = \int_{-\infty}^{+\infty} B(\delta) \cos(2\pi\tilde{\nu}\delta) d\delta \quad (1.8)$$

FTIR instruments offer three major advantages over dispersive instruments. First, the equipped reference laser, emitting at a constant wavelength, and reference detector allow for precise control of the established optical path difference. In this way, high wavenumber precision is achieved, which is referred to as precision or *Connes* advantage. Second, in FTIR spectroscopy all wavenumbers are measured at once, leading to tremendously reduced acquisition times (multiplex or *Fellgett* advantage). Finally, FTIR instruments do not require a slit to achieve resolution, leading to higher optical throughputs (throughput or *Jacquinot* advantage).

1.2.3 Transmission and attenuated total reflection spectroscopy

Many different measurement principles for mid-IR spectroscopy are available, of which transmission and ATR configurations are the most prominent ones for liquid phase samples. As these techniques were applied within this work, a short overview of these is provided in this chapter.

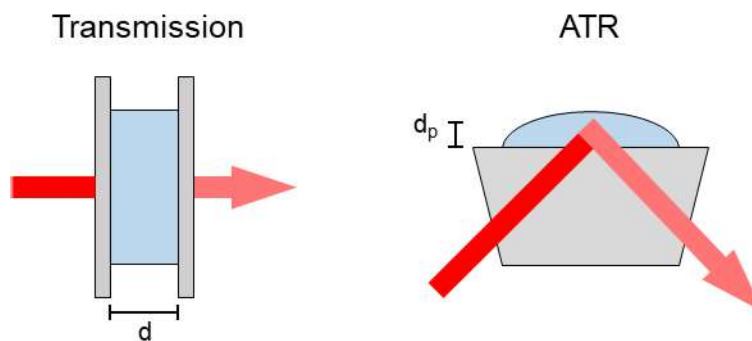


Figure 2: Measurement principles of transmission and attenuated total reflection (ATR) probing modalities.

Figure 2 depicts the principles of transmission and ATR probing modalities. In transmission configuration, the most traditional technique of recording mid-IR spectra, the sample is placed in the IR beam where part of the radiation is absorbed and the remaining light transmits to the detector. The transmittance for each wavenumber $T(\tilde{\nu})$ can be calculated as the ratio between intensity measured in presence $I(\tilde{\nu})$ and in absence of the analyte $I_0(\tilde{\nu})$:

$$T(\tilde{\nu}) = \frac{I(\tilde{\nu})}{I_0(\tilde{\nu})} = e^{-a(\tilde{\nu})d} \quad (1.9)$$

where $a(\tilde{\nu})$ is the linear absorption coefficient and d the length of the optical interaction path between light and sample. The absorbance $A(\tilde{\nu})$ is defined as negative decadic algorithm of the transmittance. This fundamental equation is commonly known as *Beer-Lambert* law:^{20,21}

$$A(\tilde{\nu}) = -\log\left(\frac{I(\tilde{\nu})}{I_0(\tilde{\nu})}\right) = \varepsilon(\tilde{\nu})cd \quad (1.10)$$

where $\varepsilon(\tilde{\nu})$ is the molar decadic absorption coefficient with the unit L/(mol cm) and c the concentration in mol/L. For liquid phase analytes, the standard procedure is to first record a single

beam spectrum of the pure solvent, followed by acquisition of the sample spectrum and subsequent calculation of the absorbance. *Beer-Lambert* law indicates that the absorbance is linearly proportional to the concentration, meaning that it provides quantitative information about an analyte. In general, higher analyte absorbance signals are favorable, as they lead to increased signal-to-noise ratios. As $\varepsilon(\tilde{\nu})$ is a fixed value, the signal can be elevated by increasing the optical path length at a given analyte concentration. In FTIR spectroscopy, the achievable light intensity per wavenumber is limited due to the rather small power output of globars. Typical values measured in the sample compartment of Bruker FTIR spectrometers are in the order of 1-2 tens of $\mu\text{W}/\text{cm}^{-1}$ only. In contrast, QCLs offer significantly higher optical power densities of approximately 4-5 orders of magnitude more, which is favorable for transmission experiments. A more detailed description of QCLs and their advantages is provided in section 2.2. Larger applicable path lengths are desirable as they additionally increase the robustness of flow-through measurements, due to reduced back-pressure when filling the cell, besides their advantages in terms of larger absorbance signals. A significant challenge in FTIR spectroscopy of liquids is caused by mid-IR absorption of solvents themselves, which can easily absorb the low power radiation of globars in certain wavenumber regions completely. As an example, for measurements of aqueous solutions in the wavenumber range of the HOH bending band, the optical path length is restricted to $<10 \mu\text{m}$ for FTIR instruments in order to avoid total light absorption. This issue is of specific relevance for protein analysis and will be further discussed in Chapter 2.

For transmission measurements of liquids, the sample is typically injected into a flow cell, consisting of sample inlet and outlet, a mid-IR transparent window pair and a spacer. Various window materials with different optical, mechanical and thermal properties are available, of which some of the most commonly applied ones are calcium fluoride (CaF_2), barium fluoride (BaF_2) and zinc selenide (ZnSe). The selection of window material depends on the spectral range of interest, chemical properties of the sample and analytical problem to be solved. For measurements of aqueous solutions, CaF_2 is most often used due to beneficial properties such as high ruggedness, low refractive index and mid-IR transparency down to approximately 1000 cm^{-1} .²² The spacer, typically made out of lead, polytetrafluorethylene (PTFE) or polyethylene terephthalate (PET), separates the two windows and defines the optical path length.

As an alternative, ATR²³ is often employed for probing liquid phase analytes. This robust technique is based on total reflection of the IR beam at the interface between an optically denser medium (ATR element) and the optically rarer medium (sample). Here, the electromagnetic wave does not propagate into the sample, however, an evanescent field is formed that interacts with the sample placed on the ATR element.

The red, dotted line in Figure 3A depicts the behavior of an IR beam that travels through the optically denser medium with a refractive index n_1 and an angle of incidence θ_1 and the transmitted beam that travels through the optically rarer medium with a refractive index n_2 and an angle of reflection θ_2 . These variables are mathematically related by Snell's law:

$$n_1 \sin(\theta_1) = n_2 \sin(\theta_2) \quad (1.11)$$

There are however, certain angles of incidence at which no transmitted light can be observed. The critical angle of incidence θ_c can be calculated from equation 1.11 by setting θ_2 to 90° (see Figure 3A, continuous red line):

$$\theta_c = \arcsin\left(\frac{n_2}{n_1}\right) \quad (1.12)$$

For all $\theta_1 > \theta_c$ no light will be transmitted, being referred to as total reflection (see Figure 3b). The thereby formed evanescent field penetrates the medium of lower refractive index. The penetration depth d_p is defined as the depth at which the electrical field falls to the $1/e$ of the electrical field at the interface and can be calculated, depending on the wavelength of the incident light λ , the angle of incidence and the refractive indices of ATR element and sample:²⁴

$$d_p = \frac{\lambda}{2\pi n_1 \sqrt{\sin^2 \theta_1 - \left(\frac{n_2}{n_1}\right)^2}} \quad (1.13)$$

If values for the Bruker Platinum diamond ATR element ($n_1 = 2.42$, $\theta_1 = 45^\circ$) and water ($n_2 = 1.33$) are applied, a d_p of $0.9 \mu\text{m}$ at 1640 cm^{-1} is obtained.

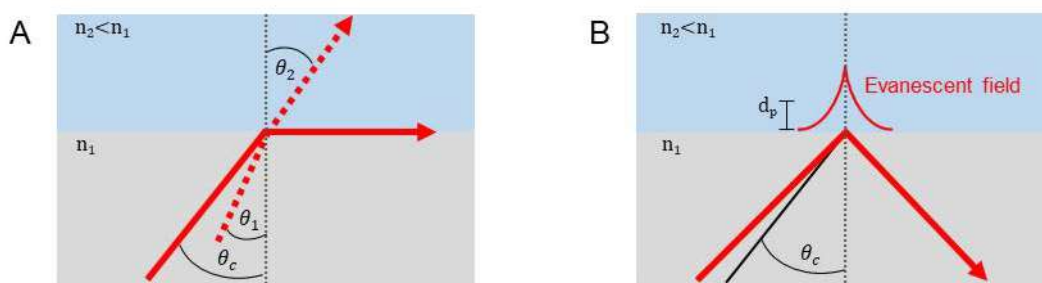


Figure 3: (A) Refraction (dotted red line) and total reflection (solid red line) at the critical angle at the interface between optically denser and rarer medium. (B) Total reflection and generation of an evanescent field.

The degree of interaction between evanescent field and sample can be described by the effective thickness d_e , representing an equivalent to the optical path length d in transmission mode. In brief, d_e can be described as the thickness of a material which gives the same absorbance in transmission configuration at normal incidence as that found in ATR.²⁵ It should be noted that d_e depends on the polarization, whereas d_p is polarization independent. In-depths mathematical descriptions of d_e are beyond the scope of this work and can be found elsewhere.²⁶

Same as in transmission configuration, larger interaction lengths are desirable for many applications in ATR configuration. In the simplest ATR setups, as shown Figure 2, the light is reflected at the interface between optically denser and rarer medium one time. In order to achieve larger interaction lengths, multireflection elements can be used, where the light is reflected several times between the optically denser ATR element and the optically rarer sample medium. Whereas single-bounce elements can for example be realized as prism-shaped ATR elements with triangular cross-sections, ATR elements with trapezoidal cross-sections lead to multireflection behavior. Here, the number of reflections N can be calculated, depending on the thickness t and length l of the crystal:

$$N = \frac{l}{t} \cot(\theta_1) \quad (1.14)$$

In analogy to transmission measurements, the *Beer-Lambert* law can be translated to ATR

measurements for quantitative purposes. Here, the absorbance increases linearly with the number of reflections N , which leads to the following mathematical approximation:²³

$$A(\tilde{\nu}) = \varepsilon(\tilde{\nu})cd_e N \quad (1.15)$$

The selection of the ATR crystal material is of major importance in order to achieve the desired result of an experiment. Typical properties that have to be considered are coverage of the spectral range of interest, the refractive index, chemical stability, inertness with the sample of interest and mechanical properties. According to equation 1.13, the refractive index of the ATR element has to be higher than the one of the sample, whereas higher penetration depths can be achieved at better index match. Some of the most commonly applied materials used as an ATR crystal are diamond, germanium and ZnSe.

Chapter 2 Laser-based IR spectroscopy of proteins

In this thesis mid-IR spectroscopy of proteins is relevant in both, the analysis of milk and in bioprocess monitoring. This chapter summarizes the most important information about QCL based mid-IR spectroscopy of proteins in aqueous solutions and its advantages over FTIR spectroscopy. A more detailed overview on this topic can be found in Ref.²⁷ In section 2.1, the basic principles of mid-IR transmission spectroscopy of proteins are explained. Section 2.2 provides an overview on QCLs, whereas section 2.3 summarizes the state-of-the art of laser-based IR spectroscopy of proteins.

2.1 Mid-IR transmission spectroscopy of proteins

In the 1950s, mid-IR spectroscopy was one of the earliest techniques found to be useful for providing structural information about polypeptides and proteins.²⁸ Nowadays, it represents a powerful technique with minimal sample preparation, short measurement times and relatively low costs for applications ranging from small soluble proteins to large membrane proteins.²⁹

Figure 4 (top) shows the formation of a peptide bond on the example of two alanine amino acids. Here, a covalent bond between carboxyl group and amino group is formed under loss of one water molecule. Proteins are composed of many peptide bonds, leading to the formation of a characteristic amino acid sequence, referred to as primary structure and a peptide backbone which shows typical absorbance bands in the mid-IR region.

Table 1 gives an overview on the nine characteristic amide bands, that are associated with proteins. Of these absorbance bands, the amide I region ($1700 - 1600 \text{ cm}^{-1}$) has been found to be the most sensitive to the secondary structure of the proteins under investigation.³⁰ However, it has been recognized that also the amide II band ($1600 - 1500 \text{ cm}^{-1}$) shows some sensitivity to protein secondary structure.³¹

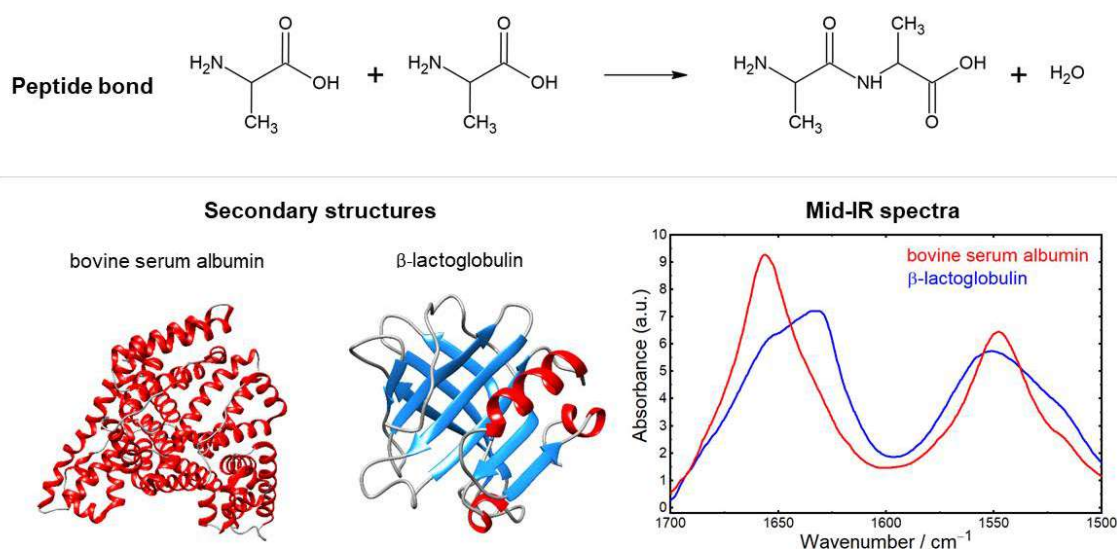


Figure 4: (Top) formation of a peptide bond on the example of two alanine amino acids, (bottom) schematics of the secondary structures of bovine serum albumin and β -lactoglobulin and the corresponding mid-infrared spectra in the protein amide I and amide II region.

In brief, the protein secondary structure can be described as the three-dimensional geometric orientation of local protein segments. Figure 4 (bottom) shows representative 3D models of bovine serum albumin (BSA) and β -lactoglobulin (β -LG) as well as the corresponding IR spectra in the amide I and amide II region. BSA is primarily composed of α -helical secondary structures³² and shows a typical narrow amide I band at 1656 cm^{-1} and a narrow amide II band at approximately 1545 cm^{-1} .^{29,33}

β -LG is mainly composed of β -sheets, showing an amide I band maximum at 1633 cm^{-1} , a shoulder at 1680 cm^{-1} and a broad amide II band at approx. 1550 cm^{-1} .³⁴ A more detailed overview on characteristic amide I band positions for different protein secondary structures is provided in Table 2.

Table 1: Characteristic mid-IR bands of proteins^{16,30}

Designation	Wavenumber / cm^{-1}	Mode of vibration
Amide A	3300	N-H stretching in resonance with amide II overtone
Amide B	3100	
Amide I	1650	C=O stretching, C-N stretching, N-H bending
Amide II	1550	N-H bending, C-N stretching
Amide III	1300	C-N stretching, N-H bending
Amide IV	625	O=C-N bending, mixed with other modes
Amide V	725	N-H out-of-plane bending
Amide VI	600	C-O out-of-plane bending
Amide VII	200	Skeletal torsion

The sensitivity of mid-IR spectroscopy to secondary structural elements can be explained by changes in the interaction of C=O and N-H groups that is induced by different arrangements of hydrogen bonds, dipole-dipole interactions and geometric orientations in α -helices, β -sheets, turns and random coils.³⁵ Quantification of individual secondary structural elements is typically

performed by evaluation of the amide I band by using curve fitting³⁶ or chemometric approaches such as partial least squares (PLS)³⁷ or multivariate curve resolution-alternating least squares (MCR-ALS).³⁸ It has, however, been demonstrated that additional and more comprehensive information can be obtained by combined analysis of the amide I and amide II bands.^{33,39,40}

Table 2: Amide I band position according to protein secondary structure⁴¹

Secondary structure	Average band position / cm^{-1}	Range band position / cm^{-1}
α -helix	1654	1648 – 1657
β -sheet	1633	1623 – 1641
β -sheet	1684	1674 - 1695
turns	1672	1662 – 1686
unordered	1654	1642 – 1660

Most mid-IR transmission measurements of proteins are performed in H_2O , the ubiquitous solvent for biological applications. Here, a pronounced challenge is caused by the strong absorption of the HOH-bending band of water at approximately 1643 cm^{-1} that overlaps with the significantly weaker protein amide I band. For FTIR spectrometers that are equipped with low-power globars, the optical path lengths is restricted to below $10 \mu\text{m}$ in order to avoid total light absorption through water in this spectral region.^{30,35} Such small path lengths lead to multiple experimental constraints. First, assembling of tight transmission cells is a cumbersome task due to high susceptibility of thin spacers to electrostatic charges. Second, according to *Beer-Lambert* law, smaller absorbance signals are obtained at smaller optical path lengths, leading to limited sensitivity. Consequently, for FTIR transmission measurements of aqueous solutions, typically protein concentrations of $>10 \text{ mg/mL}$ have to be used.²⁹ Finally, such small path lengths distinctly impair the robustness of flow-through measurements due to high backpressures, likelihood of cell clogging and vulnerability to introduction of air bubbles during injection of highly viscous samples such as concentrated protein solutions.

As an alternative, D_2O can be used as a solvent for mid-IR based protein measurements, as it has similar physical-chemical properties to H_2O , while offering the advantage that the DOD bending band is located at approx. 1200 cm^{-1} . Here, no overlap with the amide I band occurs, leading to significantly larger applicable path lengths. Besides this experimental advantage, exchange of hydrogen with a deuterium atom in the peptide moieties, leads to shifts of amide I and amide II band positions to lower wavenumbers. However, for most real-life biological applications, investigations of proteins in their native environment is desired, making replacement of H_2O by D_2O inapplicable.^{42,43}

In this thesis, QCL-based IR transmission spectroscopy was the technique of choice for investigating proteins in milk and for bioprocess monitoring. The properties of QCLs are described in the next subchapter, followed by an overview of the current state-of-the art of QCL based transmission spectroscopy of aqueous protein solutions in section 2.3.

2.2 Quantum cascade lasers

The theoretical concept for QCLs was first published in 1971,⁴⁴ followed by experimental verification in 1994.⁴⁵ Due to several clear advantages compared to FTIR spectroscopy, QCL based mid-infrared spectroscopy, nowadays, represents a versatility technique in the bio-medical field.¹ In section 2.2.1, the design of QCLs is briefly explained, including a summary of distributed feedback and external cavity designs. Finally, the advantages of QCLs for mid-IR transmission measurements are explained in section 2.2.2.

2.2.1 QCL design

In general, two conditions have to be met for a laser (light amplification by stimulated emission of radiation) to operate: population inversion of electrons in an active medium and optical feedback by a resonator.⁴⁶ QCLs belong to the group of semiconductor lasers, however, they fundamentally differ from conventional semiconductor lasers such as diode lasers in their operation principle.⁴⁷ For classic semiconductor lasers, light is generated from electron interband transitions, leading to photon energies that correspond to the band gap energy of the utilized gain material. This fundamental limitation restricts their application mainly to the ultraviolet/visible (UV/Vis) and near-IR range.²⁷ In contrast, QCLs, operating in the mid-IR to far-IR region, are not composed of a single bulky semiconductor material but of multiple thin semiconductor layers of different composition, forming a superlattice structure. With an appropriate design of this heterostructure, light emission is achieved through intersubband transitions of electrons. During operation, a single electron can undergo multiple radiative transitions that generate mid-IR photons, whereas for diode lasers each electron results only in a single photon, leading to significantly higher power outputs for QCLs.⁴⁸ As the design of QCLs allows for decoupling of the emitted wavelength from the material band gap, the wavelength is predominantly defined by the thickness of the layers, rather than the material, which leads to possible emission regions over a wide wavenumber range by application of the same material.⁴⁹ In this context, the structure of the gain medium can be designed in a way that emission over several hundred wavenumbers can be achieved. Nowadays, heterostructures based on indium gallium arsenide/alluminium indium arsenide (InGaAs/AlInAs) grown on indium phosphide (InP) substrate are the most advanced technology for mass production requirements,⁵⁰ however, other designs based on indium arsenide/aluminium antimonide (InAs/AlSb) and other have been introduced as well.⁵¹

Besides, the design of the gain material, the laser resonator is of significant importance in order to achieve stimulated emission. Several types of QCLs can be distinguished based on the applied resonator, of which the most established ones, distributed feedback (DFB)-QCLs and EC-QCLs are shown in Figure 5.

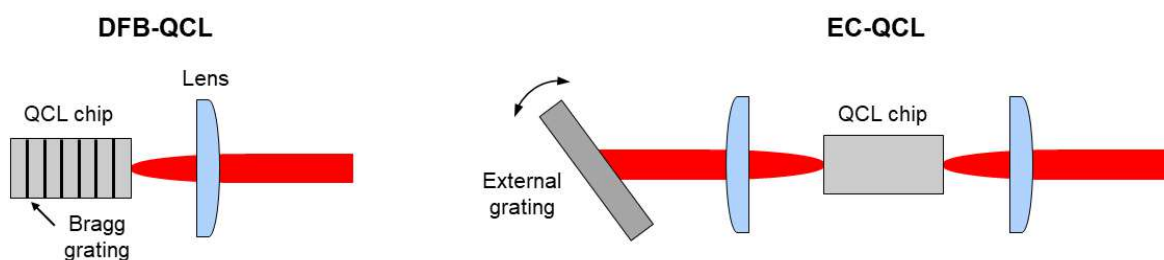


Figure 5: Principle of distributed feedback-QCL (left) and external cavity-QCL (right).

In the DFB configuration, the wavenumber is selected by a Bragg grating that is applied to the gain medium. Here, tunability over a few wavenumbers can be achieved by either internal heating induced by changing the laser current or by an external change of the operation temperature.⁵² Due to this narrow wavelength coverage, DFB-QCLs are mostly used for gas phase applications,⁵³ even though arrays, consisting of multiple lasers can be applied to cover a broad spectral region as required for liquid phase analysis.⁵⁴ In this work, EC-QCLs were used for transmission measurements of aqueous protein solutions, as they allow for tuning over a broad spectral range by using an external diffraction grating. This spectral tuning is achieved by slow motion of the diffraction grating relative to the chip, either continuously (sweep mode) or stepwise by holding the grating for defined periods of time at certain positions (step-and-measure mode). In most applications, sweep mode is preferred, as in this way spectral coverages of several hundred wavenumbers can be achieved with tuning rates of several thousand cm^{-1}/s .

2.2.2 Advantages and disadvantages of QCLs for transmission measurements

Compared to the thermal light sources used in FTIR spectroscopy, QCLs offer increased spectral power densities by a factor of 10^4 and more, which allows to increase the transmission path length. This is particularly beneficial for transmission measurements of aqueous protein solutions, as path lengths of below $10 \mu\text{m}$ typically have to be used in FTIR spectroscopy. The theoretically achievable increase in optical path length for QCL based measurements can be calculated based on *Beer-Lambert* law (equation 1.9).⁵⁵ For this purpose, the thermal light source in FTIR is assumed as the initial intensity ($I_{FTIR} = I_0$), which can be multiplied by the factor k_I to obtain the intensity of the QCL (I_{QCL}). Furthermore, it is assumed that a certain minimum intensity I_D is required that reaches the detector. The maximum achievable path length can subsequently be calculated as follows:

$$d_{max,QCL} = \frac{1}{\epsilon c} \left(\log \frac{I_{FTIR}}{I_D} + \log k_I \right); \quad I_{QCL} = k_I I_{FTIR} \quad (2.1)$$

By assuming a typical power output of $100 \text{ mW}/\text{cm}^{-1}$ for QCLs in pulsed operation and $10 \mu\text{W}/\text{cm}^{-1}$ for the thermal light source of FTIR instruments,⁵⁶ a k_I value of 10^4 is obtained. With these considerations, the maximum achievable path length can be calculated as follows:

$$d_{max,QCL} = \frac{1}{\epsilon c} \left(\log \frac{I_{FTIR}}{I_D} + \log k_I \right) = d_{max,FTIR} + \frac{1}{\epsilon c} \cdot 4; \quad d_{max,FTIR} = \frac{1}{\epsilon c} \cdot \log \frac{I_{FTIR}}{I_D} \quad (2.2)$$

If the values of pure water with a molarity of $55.5 \text{ mol}/\text{L}$ and a molar absorption coefficient of $21.8 \text{ L}/\text{mol cm}$ at 1643 cm^{-1} , representing the maximum of the HOH bending band in the amide I region,⁵⁷ as well as the minimum required intensity for reaching the detector of 850 nW , corresponding to an absorbance of 1.1 , are applied, maximum achievable path lengths of $8.8 \mu\text{m}$ and $41.9 \mu\text{m}$ are obtained for FTIR and QCL based measurements, respectively.²⁷ This larger applicable path length for QCLs leads to significantly increased robustness in sample handling. Moreover, as the absorbance is linearly dependent on the path length, higher signals are obtained as well.

Despite the advantages of QCLs over FTIR for aqueous transmission measurements, some disadvantages have to be considered. For FTIR spectroscopy, the noise level is determined by the detector noise, as the noise introduced by thermal light sources can usually be neglected. In contrast, QCLs contribute to the general noise level, especially when operated in pulsed mode due to pulse-to-pulse intensity fluctuations. In EC-QCL configuration, inconsistent wavelength tuning behavior can contribute to the noise level when multiple single channel spectra are averaged, as it is usually the case. Furthermore, the spectral power density of QCLs is defined by the gain material,

where the highest intensity is reached at a center wavelength, followed by decreasing power towards both sides of the gain curve.⁵⁸ For broadband transmission measurements in highly absorbing matrices such as water, EC-QCLs can easily saturate the detector in one part of the spectrum while dropping below the noise limit at another. Consequently, the challenge of building an EC-QCL based setup with high signal-to-noise ratios is to make the most of the advantages of the light source and at the same time handle its current limitations.

Section 2.3 gives a brief overview about the state-of-the art for QCL-based transmission spectroscopy of proteins, before the start of this thesis. In section 3.3.1, a novel approach developed within this thesis is described that outperforms previous approaches in terms of sensitivity.

2.3 State-of-the art in EC-QCL based transmission spectroscopy of proteins

The methods for protein investigations, applied within this work, are based on lasers from Daylight Solutions and transmission setups that were previously developed within the Lendl group.⁵⁹⁻⁶¹ Consequently, the focus of this chapter is to provide a summary on these setups in section 2.3.1 and 2.3.2, even though alternative measurement schemes will briefly be discussed in section 2.3.3. The methods, employed within this work, with the aim to further improve the current state-of-the-art are described in section 3.3.1 and 4.3.2 in the context of protein analysis of milk samples and in bioprocesses monitoring.

Typical optical setups for laser-based protein analysis consist of EC-QCLs to enable high-power broadband measurements. Commonly, the laser beam has to be adjusted by optical filters in order to adapt the laser emission profile to the linear range of the detector. The beam is then guided through a transmission cell with a large optical path length that is preferably temperature stabilized, followed by a mid-IR detector and a signal amplifier such as a boxcar integrator or a lock-in amplifier.

2.3.1 First-generation setup (absorbance spectroscopy)

The first laser-based setup for broadband protein measurements was published in 2015,⁵⁹ using a first-generation EC-QCL from Daylight Solutions (Über Tuner, purchased 2011) that covered the protein amide I region. A schematic of this setup is shown in Figure 6.

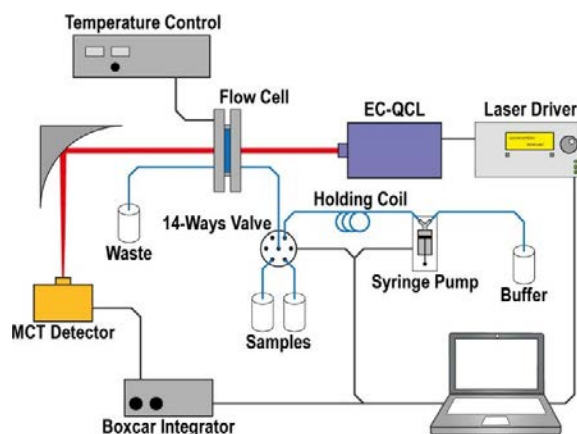


Figure 6: Schematic of the first-generation EC-QCL based setup for measuring proteins in aqueous solutions. Reprinted with permission from Ref⁵⁹. Copyright 2015 American Chemical Society.

Due to the high optical power density of the laser, a transmission path length of 38 μm was achieved. The light intensity was recorded by a thermoelectrically cooled mercury cadmium telluride (MCT) detector and subsequently processed by a boxcar integrator. Single beam spectra were recorded during a tuning time of 1.5 s. A total number of 50 scans was averaged per spectrum in order to reduce the noise level, leading to an acquisition time of ~ 4 min. Inconsistent wavelength tuning behavior of the laser, however, led to spectral mismatch in the fine mode-hop structure of consecutive scans, making a thorough data processing routine necessary. For this purpose, correlation optimized warping (COW), a rubber band like alignment algorithm was used to compensate these flaws, followed by minor Fourier-filtering. Finally, the wavenumber scale of the obtained spectra was determined by calibration against absorbance lines of FTIR water vapor

spectra. Absorbance spectra of proteins with different secondary structures were recorded and compared to FTIR reference spectra of significantly higher concentration levels, showing excellent agreement in band shapes and positions. Calibration curves of different proteins showed linear regression coefficients of $r^2 > 0.99$, indicating excellent applicability of the setup for quantitative purposes. Finally, this setup was successfully applied for investigating dynamic changes of protein secondary structures⁶² and to quantify individual proteins in commercial bovine milk (see section 3.2.2.2).^{63,64} Besides this setup for measuring aqueous solutions, a setup based on the first-generation EC-QCL for protein measurements in D₂O was published, achieving an optical path length of 478 μm .⁶¹ This setup was successfully applied for recording absorbance spectra of model proteins and to monitor dynamic changes in secondary structure by thermal denaturation.

2.3.2 Second-generation setup (absorbance spectroscopy)

During the following years, EC-QCL light sources became commercially available that offer improved characteristics in terms of covered spectral range, tuning rates and motion mechanics of the cavity. In this framework, the second-generation setup was introduced in 2018 (see Figure 7),⁶⁰ using a Hedgehog EC-QCL from Daylight Solutions, tunable between 1470 and 1730 cm^{-1} to cover the protein amide I + II region.

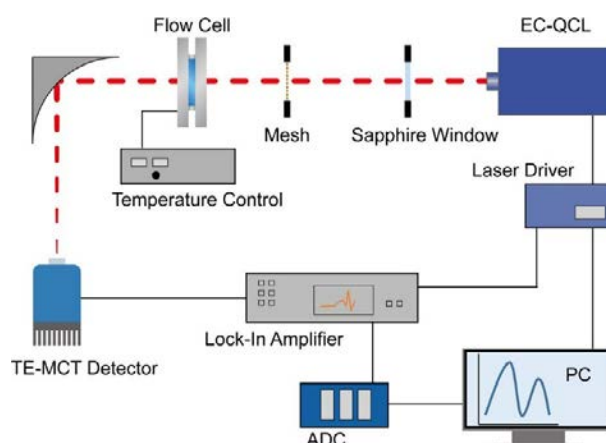


Figure 7: Schematic of the second-generation EC-QCL based setup for measuring proteins in aqueous solutions. Reprinted with permission from Ref⁶⁰. Copyright 2018 American Chemical Society.

Here, a sapphire window had to be applied as an optical filter to selectively attenuate the laser power in the amide II region and to match the high detector signal in this wavenumber range to the smaller signal in the amide I region. As the achievable optical path length always represents a trade-off between emission characteristics of the laser, covered spectral region and dynamic range of the detector, a slightly smaller optical path length of 31 μm , compared to the first-generation setup was employed. The laser was operated with a tuning rate of 1200 cm^{-1}/s , enabling acquisition and averaging of 100 scans within a measurement time of 53 s. The beam was detected by a thermoelectrically cooled MCT detector, followed by signal processing with a lock-in amplifier. Due to significantly improved motion mechanics of the EC, the COW alignment approach was no longer necessary. It was, however, observed that a small number of single scan spectra showed significant difference to the remaining scans. For this purpose, a processing routine based on a similarity index was implemented that removes these outliers prior to scan averaging in order to decrease the overall noise level. This setup was thoroughly benchmarked against FTIR spectroscopy and the first-

generation setup by evaluating relevant performance characteristics. First, mid-IR spectra of model proteins with different secondary structures were recorded across the protein amide I and II region and compared to 20 mg/mL FTIR reference spectra. The obtained spectra showed excellent agreement in terms of band shapes and positions, down to concentration levels of 1 mg/mL. Besides the higher absorbance signals due to the larger applicable optical path lengths for QCL based measurements, also the noise levels have to be considered for a comprehensive evaluation of the method sensitivity. The noise level of transmission measurements can be determined by calculating absorbance spectra of two consecutively recorded background measurements, referred to as 100% lines.⁶⁵ Under ideal conditions, the absorption spectrum would show a flat line at zero absorbance, however, due to the different noise contributions, some fluctuations can be observed. For comparison of QCL-IR spectroscopy to FTIR measurements, 100% lines of water were calculated, followed by calculation of the root-mean-square noise. Subsequently, the limit of detection *LOD* was determined as follows:

$$LOD = \frac{3 \cdot noise}{slope\ of\ calibration\ function} \quad (2.3)$$

An overview of the obtained performance parameters for the first- and second-generation QCL setups, as well as for routine and high-end FTIR spectroscopy are shown in Table 3. The second-generation setup achieved approximately 5 times better LODs within 5 times shorter measurement times, by covering a broader spectral region compared to the first-generation setup. Compared to a routine FTIR spectrometer (Bruker Tensor 37), equipped with a deuterated L-alanine doped triglycine sulfate (DLaTGS) detector, approximately 2 times lower noise levels and 7 times lower LODs were achieved, within comparable acquisition times. Finally, the noise levels of a high-end FTIR spectrometer (Bruker Vertex 80v), equipped with a liquid nitrogen cooled MCT detector could not be met with the laser-based setup. Due to the larger applicable optical path length, however, the LOD was approximately 1.5 times lower for QCL-based measurements. Furthermore, the second generation setup was successfully applied for monitoring secondary structural changes of proteins⁶⁶ and for high-throughput analysis of bovine milk (see section 3.2.2.2).³

Table 3: Performance characteristic of EC-QCL based transmission setups and conventional FTIR spectroscopy.^{27,60}

	Meas. time [s] / scans	RMS-Noise 10 ⁻⁵ [AU]	Path length [μm]	LOD [mg/mL]	Spectral range [cm ⁻¹]	Detector type / temp. [°C] / cooling
QCL setup 1 st gen.	250 / 100	37	38	0.117	1700-1600	MCT / -60 / TE
QCL setup 2 nd gen.	53 / 100	6.2	31	0.025	1700-1500	MCT / -78 / TE
High-end FTIR	53 / 81	2.7	8	0.041	4000-600	MCT / -196 / LN ₂
Routine FTIR	53 / 21	12.1	8	0.182	4000-600	DLaTGS / 25 / T-

2.3.3 Alternative setups

Apart from the described setups developed within the Lendl group, alternative transmission setups and commercial developments were introduced.

In microfluidic modulation spectroscopy (MMS), a rapid exchange of background and sample solution is performed in a 25 μm transmission cell, during operation of a tunable QCL in step-and-measure mode. In this way, good repeatability, low LODs and a broad dynamic range of 0.1 to 200 mg/mL protein solutions were reported.⁶⁷ This technology was also implemented in commercially

available devices, equipped with auto sampler (AQS³ pro and Apollo from RedShift Bio). Furthermore, the ChemDetect Analyzer from Daylight solutions was introduced. This commercially available device is equipped with an EC-QCL, operating in sweep mode, that covers a broad wavenumber range of more than 400 cm⁻¹. In the present thesis, the ChemDetect Analyzer was applied for bioprocess monitoring, thus more detailed information is provided in section 4.3.2. The ChemDetect has been followed by the Culpeo which covers the same spectral range, while additionally being equipped with an integrated module system that enables a straight-forward change between different liquid flow cells to fulfill the requirements for pharmaceutical industry. Most recently, a new approach for compensating the strong absorption of water, allowing EC-QCL based measurements of protein amide I and II band without the use of optical filters, was presented by the group of Y. J. Lee from NIST (Gaithersburg, MA).⁶⁸ In this work, a dual-beam setup, equipped with a spectrum adjuster unit, based on two acousto-optic modulators (AOMs), was introduced. Here, a large optical path length of 38 μm and quantitative protein measurements between 0.1 and 100 mg/mL were achieved. Furthermore, characteristic spectral features of proteins with different secondary structures could be identified down to concentrations of 1 mg/mL.

Besides the clear advantages of QCLs as high-power light sources in transmission configuration for absorbance measurements, their coherent nature enabled developments of non-conventional measurement schemes. In this context, an approach based on Mach-Zehnder interferometry was introduced for liquid phase analysis based on dispersion spectroscopy⁶⁹ and recently translated to protein analysis.⁷⁰

Furthermore, the linearly polarized radiation of QCLs enabled the development of setups for vibrational circular dichroism (VCD) measurements.^{71,72} This technique was successfully applied for measuring pH-induced changes of peptides,⁷³ however, investigations of proteins were not performed yet.

Chapter 3 Composition and analysis of milk

This chapter summarizes the most relevant information on the composition and analysis of milk. In section 3.1, the composition of human milk and its differences to bovine milk are explained, followed by more in-depth discussion of the protein and fat composition of milk. In chapter 3.2, established methods for analyzing milk are discussed with a focus on mid-IR based methods and the reference methods that are of relevance for this work. Chapter 3.3. gives an introduction to the novel mid-IR based methods developed within this work for analyzing proteins and fatty acids in human milk.

3.1 Milk composition

Milk is a complex, dynamic and bioactive fluid with the fundamental role of supplying young mammals with nutrients before they are able to digest solid food.⁷⁴ The synthesis of milk, taking place in mammary glands, defines the characteristic of mammals. Furthermore, milk of other mammals is consumed by humans of all age as a component of varied diet. In this context, milk is considered as one of the most nutritionally complete foods with its main constituents being water, lipids, lactose and proteins. Nowadays, milk is one of the most important agricultural products, with a production volume of 8.5×10^6 tons per annum worldwide, of which approximately 81% is bovine milk, and further growth of expected 1.6% per annum until 2029.⁷⁵

In this thesis, especially the analysis of human milk was of interest, however, due to easier availability of bovine milk, the analytical methods were first developed for this matrix. Thus, the main focus of this section is on providing an overview on the composition of human milk and explaining the main differences to bovine milk. As the methods developed within this work aim to analyze the milk protein and fatty acid composition, these components will be discussed in more detail in section 3.1.2 and 3.1.3, respectively.

3.1.1 Human milk composition and differences to bovine milk

Human milk is a crucial resource for infants as it supplies them with all essential nutritional and immunological requirements, while at the same time assisting healthy development by providing non-nutritive bioactive factors.⁷⁶ According to the World Health Organization (WHO), exclusive breastfeeding is recommended for the first six months of life.⁷⁷

The nutrient composition of human milk derives from three different sources: synthesis in the lactocytes, maternal diet and maternal stores. The milk composition is known to be significantly influenced by prematurity and postnatal age to meet the nutritional demands of newborns at every development stage.⁷⁸

Table 4 shows the macronutrient composition of human milk, depending on the stage of lactation and term/preterm delivery based on a meta study,⁷⁹ compared to the macronutrient composition of bovine milk. The first fluid, produced up to 5 days after birth, referred as colostrum, has a distinct composition, appearance and volume.⁸⁰ Colostrum is rich in immunological components, development factors and epidermal growth factors. Compared to mature milk, with typical macronutrient concentrations ranging from 8 to 12 mg/mL proteins, 32 to 36 mg/mL fat and 67 to 78 mg/mL lactose,⁸¹ the protein concentrations are significantly higher for colostrum, whereas fat and lactose concentrations are slightly lower. Transitional milk, typically occurring from five days to two weeks postpartum, shares some of the characteristics from colostrum. During this period, the protein content decreases, whereas fat and lactose content increase. After these fluctuations in the first two weeks of lactation, the milk composition only slightly changes, before it is considered fully mature between four and six weeks postpartum.⁷⁹

Preterm colostrum consists of higher protein concentrations than term colostrum, however, after this postnatal stage, most of the differences in protein content disappear, resulting in similar protein contents for preterm and term mature milk. In contrast, the lactose composition for preterm colostrum is lower than for term colostrum, before reaching similar concentrations in mature milk.

The main differences of human milk, compared to bovine milk, are the approximately three times lower protein concentrations and slightly lower lactose concentrations. The total fat content for both types of milk is similar.

As the individual protein and fatty acid composition was of major interest for the analytical methods developed within this work, a closer look on their composition is provided in the next two subchapters.

Table 4: Comparison between macronutrient composition of human milk according to postnatal age and bovine milk.^{79,82}

		Concentration in mg/mL			
		Protein	Fat	Lactose	
Human milk	Preterm	Day 1-3	27	22	51
		Week 2	15	35	57
		Mature	10	37	68
	Term	Day 1-3	20	18	56
		Week 2	16	30	62
		Mature	10	34	67
Bovine milk		33	39	46	

Besides the described fluctuations in human milk macronutrients, micronutrient concentrations such as vitamins A, B1, B2, B6, B12, D and iodine can vary depending on maternal diet and body stores. The American Academy of Pediatrics recommended Vitamin K injections during lactation, as this component is always very low in human milk, regardless of the maternal diet.⁸³ Furthermore, vitamin D supplementation during lactation is under investigation, as particularly with limited maternal exposure to sunlight, very low concentrations of that vitamin occur in human milk.⁸⁴

3.1.2 Protein composition

Milk proteins can be separated into two major categories: caseins and whey proteins. Caseins have a traditional importance in cheese making and are classically defined as those proteins, which can be precipitated from bovine milk by acidification to pH 4.6.⁸⁵ In contrast, whey proteins remain in solution under such pH conditions. The ratio between caseins and whey proteins highly varies between species. Bovine milk proteins comprise of approximately 80% caseins, whereas the casein content of human milk proteins varies between approximately 20% and 45% (mature milk), depending on the stage of lactation.⁸⁶ An overview of the most abundant proteins in mature human milk and bovine milk is provided in Table 5.

Table 5: Protein Composition of human milk and bovine milk.⁸⁷

Protein	Concentration mg/mL	
	Human milk	Bovine Milk
α -casein	0	14.6
β -casein	2.2	9.6
κ -casein	0.4	3.6
γ -casein	0	1.6
α -lactalbumin	2.2	1.2
β -lactoglobulin	0	3.0
lactoferrin	1.4	0.3
immunoglobulins	0.8	0.6
lysozyme	0.5	Traces
serum albumin	0.4	0.4
others	0.8	0.6

Caseins can be subdivided into α -casein, β -casein, κ -casein, γ -casein, each of them having their individual physical and functional properties. Whereas the largest part in bovine milk are α -caseins, human milk does not contain these proteins. Due to the differences in its composition, it has been recognized that the human milk casein fraction is more difficult to isolate, requiring acidification to pH 4.3 and addition of calcium chloride (CaCl_2).⁸⁸ In general, caseins are organized in micelles, which are complexes of multiple protein subunits, held together by calcium phosphate. These spherical shaped aggregates with typical diameters of 600 nm for bovine milk and 30-75 nm for human milk,⁸⁹ give milk its characteristic white appearance.⁹⁰ Caseins are considered nutritionally rich by providing the newborns with essential amino acids, calcium and phosphate, while being easily digestible.

The group of whey proteins is very diverse, containing many different proteins with various nutritional and functional properties. α -Lactalbumin (α -LA) appears to be present in milk of all mammals, representing the highest and second highest concentrated whey protein in human milk and bovine milk, respectively. A major difference in the whey protein composition of these two types of milk is the β -LG content, representing the highest concentrated whey protein in bovine milk, whereas it is absent in human milk. In contrast, human whey protein contains a large part of lactoferrin, which is rather low in bovine milk. Furthermore, the whey fraction contains many important enzymes, where some of the most abundant ones are lysozyme and bile-salt stimulated lipase. Whereas caseins are stable to heat treatment, certain whey proteins are not stable under high temperatures. Human milk, stored in milk banks, as well as commercially available bovine milk is typically subjected to different thermal processing steps to ensure appropriate shelf-life. β -LG

starts to denature at approximately 60 °C, whereas α -LA denaturation begins at approximately 75 °C.³ It has been demonstrated, that by determining the concentrations of β -LG and α -LA commercially available bovine milk can be distinguished based on the type of applied thermal heat treatment.⁹¹

3.1.3 Fat composition

Both, human and bovine milk fat is mainly composed of triglycerides (approximately 98%), followed by significantly lower concentrated components such as mono- and diglycerides, cholesterol, phospholipids, glycosphingolipids and free fatty acids.^{92,93} Milk lipids mainly occur as milk fat globules (MFGs), emulsified in water. Figure 8 shows the schematic of a MFG, consisting of a triglyceride rich core that is enveloped by the milk fat globule membrane (MFGM), consisting of two layers. The inner monolayer membrane is mainly composed of phospholipids, whereas the outer membrane comprises a bilayer of amphipathic lipids, primarily phosphatidylcholine, sphingomyelin and cholesterol, cerebrosides, gangliosides, glycosylated proteins, polypeptides and other bioactive components.⁹⁴ In untreated milk, MFGs markedly vary in size with typical diameters ranging from 1 to 10 μm .⁹³ This size is of major importance for milk stability and technological properties, thus commercially available milk is typically homogenized in order to avoid phase separation during storage. Here, the size of MFGs is reduced to more uniform diameters, leading to increased shelf life and a consistent product.

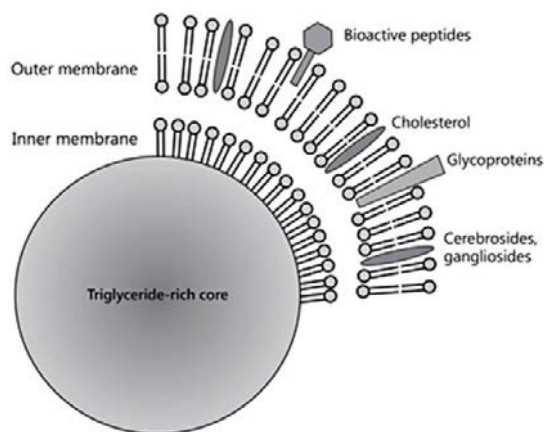


Figure 8: Schematic of a milk fat globule membrane. Adapted with permission from Ref⁹⁴.

Triglycerides are esters, composed of three fatty acids, attached to a glycerol backbone. In contrast, mono- and diglycerides have only one or two fatty acids attached to the glycerol backbone. Fatty acids are the most fluctuating macronutrient of human milk, being influenced by variables such as mothers' diet, genetics, sociodemographic and environmental factors.^{4,95} Based on their chemical structure, they can be grouped, according to their chain lengths or degree of saturation. Figure 9 depicts chemical structures of some of the most abundant fatty acids in milk with different chain lengths and degrees of saturation. Individual fatty acids in milk typically range in length between 4 and 22 carbon atoms, the majority of them having an even number of C atoms. Based on the chain length, fatty acids can be grouped into short-chain fatty acids (SCFA, C4-C10), medium-chain fatty acids (C11-C16) and long-chain fatty acids (C17 and higher). Saturated fatty acids (SAT) have single

bonds between adjacent carbon molecules, whereas unsaturated fatty acids (UNSAT) comprise at least one double bond. Monounsaturated fatty acids (MONO) are referred as those fatty acids with one double bond, whereas polyunsaturated fatty acids (PUFA) have 2 or more double bonds. Depending, on their spatial arrangement, double bonds can be divided into *cis* isomers, where the carbon chain continues on the same side of the bond, and *trans* isomers where the carbon chain continues on the opposite side (see Figure 9).

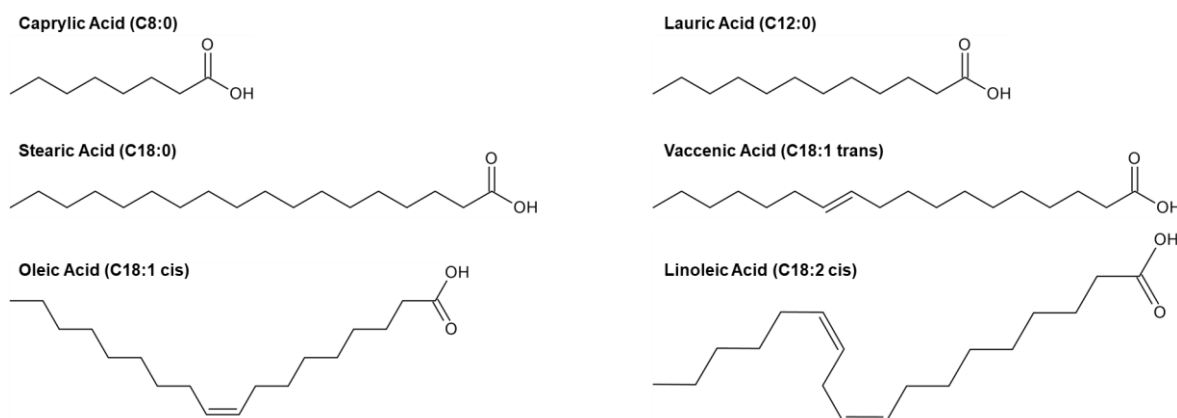


Figure 9: Chemical structures of some fatty acids, occurring in milk.

Milk fat contains hundreds of different fatty acids, whereas only a small number of them are present in concentrations above 1% in both, human and bovine milk fat.^{93,96} Table 6 shows typical fatty acid compositions of human milk and bovine milk.

Table 6: Typical fatty acid profiles of mature human milk from European mothers and bovine milk.⁹⁶⁻⁹⁸

Fatty Acid	Concentration in % of total fat (w/w)	
	Human milk	Bovine Milk
C4:0	<0.2%	4.4
C6:0	<0.2%	2.4
C8:0	0.2	1.4
C10:0	1.6	2.7
C12:0	5.3	3.3
C14:0	5.8	10.9
C16:0	21.4	30.6
C18:0	6.9	12.2
C16:1 <i>cis</i>	2.4	1.0
C18:1 <i>cis</i>	38.6	22.8
C18:2 <i>cis</i>	13.3	1.6
C18:3 <i>cis</i>	0.7	0.7
C18:1 <i>trans</i>	3.1	2.1

With an approximate share of 11%, bovine milk fat contains significant amounts of SCFA, whereas concentrations of these fatty acids in human milk are low. Especially concentrations of butyric acid (C4:0), caproic acid (C6:0) and caprylic acid (C8:0) are very low in human milk. Another major difference between these two types of milk is the number of saturated components. Bovine milk contains approximately 70% SAT, whereas the milk fat of European women typically contains 35-

40% SAT.⁹⁴ For both types of milk, palmitic acid (C16:0) represents the largest fraction of SAT, followed by stearic acid (C18:0), myristic acid (C14:0), lauric acid (C12:0) and capric acid (C10:0). The majority of UNSAT are MONO, with oleic acid (C18:1 *cis*), having the largest share, followed by palmitoleic acid (C16:1 *cis*). PUFA account to approximately 15% of total fat content in human milk, whereas the PUFA concentration in bovine milk is significantly lower with approximately 2%. Here, the most concentrated fatty acid in both types of milk is linoleic acid (C18:2 *cis*) followed by α -linolenic acid (C18:3 *cis*). Finally, *trans* fatty acids, significantly varying with the maternal diet,⁹⁹ represent only a small fraction of the total fat content.

3.2 Established methods for analyzing milk

In order to allow determination of the milk composition, accurate analytical methods are required. This chapter focusses on established methods for protein and fatty acid analysis. In chapter 3.2.1 conventional analytical methods that are based on wet chemical and/or chromatographic steps are discussed. Chapter 3.2.2 deals with more rapid mid-IR based approaches that represent the fundament for the methods developed within this work.

3.2.1 Conventional analytical methods

A large number of different analytical-chemical methods for analyzing diverse parameters in milk, including, micronutrients, hormones, oligosaccharides and various biologically active compounds is available,¹⁰⁰ however, description of these methods is beyond the scope of this thesis. This subchapter focuses on methods for analyzing the protein and fat composition. Some of these methods were used as reference methods, for the novel mid-IR based methods, developed within this work.

3.2.1.1 Analysis of proteins

Colorimetric assays, such as the biuret assay, the Pierce bicinchoninic acid (BCA) assay, the Lowry-Peterson assay and the Bio-Rad Coomassie blue assay are among the most traditional methods for analyzing the total protein content, offering inexpensive analysis based on a spectrophotometer but requiring wet-chemical sample preparation steps. Of these assays, the BCA technique has been found to be best suited for human milk analysis,¹⁰¹ however, it has been pointed out that colorimetric methods may generally overestimate the total protein content in human milk.¹⁰² Nowadays, the Kjeldahl method¹⁰³ represents the gold-standard for determining the total protein content in milk.¹⁰⁴ This method is based on determination of organic nitrogen and subsequent calculation of the protein content. In brief, the sample is digested by a mixture of concentrated sulfuric acid (H_2SO_4) and potassium sulfate (K_2SO_4), where any organic nitrogen, present in the sample, is converted into ammonium sulfate ($(NH_4)_2SO_4$). In order to further enhance this reaction, a catalyst such as copper (II) sulfate ($CuSO_4$) is added. After digestion, the mixture is cooled, followed by addition of sodium hydroxide ($NaOH$) in order to liberate ammonia (NH_3). Finally, the NH_3 steam is distilled into a boric acid (H_3BO_3) solution and titration with hydrochloric acid (HCl) is performed. Some of the major drawbacks of the Kjeldahl method are the required large sample amounts, laborious wet chemical steps, hazardous solvents and low sample-throughput. For this reason, mid-IR based techniques have nowadays become a serious challenger for total protein content determination, offering solvent-free analysis with significantly larger sample throughput (see section 3.2.2).

Classical methods for determining individual proteins in milk are based on immunological techniques, using the ability of antibodies to bind a specific molecule and consequently produce a signal response as a consequence of the occurred binding.¹⁰⁵ Here, one possible technique is radioimmunoassays (RIAs), which are based on radioactively tagged antibodies. RIAs are nowadays rather uncommon, especially in small laboratories, due to risks associated with exposure to radioactive materials.¹⁰⁵ Today, enzyme-linked immunosorbent assays (ELISA), based on enzymatic reactions, are the most often implemented immunoassays. ELISA test kits for detecting a majority of the most abundant milk proteins, including different caseins, α -LA, β -LG and lactoferrin, are commercially available. These kits require wet-chemical sample preparation steps and a spectrophotometer. When analyzing milk with ELISA kits, often very high dilution of the samples is

needed to obtain protein concentrations that are within the calibration range, bearing a major source for errors. Moreover, it has been demonstrated that ELISA-based protein measurements may generally lead to inaccurate results, even if test kits tailored for milk analysis are used.¹⁰⁶ As an alternative, different methods based on HPLC were introduced for analyzing individual proteins in bovine¹⁰⁷⁻¹⁰⁹ and human milk.¹¹⁰ These methods are based on wet chemical sample preparation steps, followed by chromatographic separation with typical run times of 10-30 min. HPLC methods offer analysis of many different proteins within one run by achieving highly accurate results. Finally, capillary electrophoresis (CE), which is also based on wet chemical sample preparation and followed by separation of proteins, was used for analyzing proteins in milk offering similar measurement times as HPLC based methods.¹¹¹

3.2.1.2 Analysis of Fat

Different methods are available for determining the total milk fat content, of which some of the most established ones are the Gerber method, Röse-Gottlieb method and the Babcock test. All of these methods have in common, that they are based on laborious extraction steps, utilizing hazardous organic solvents. For human milk analysis, the Röse-Gottlieb method is most established, where a combination of ammonium hydroxide (NH₄OH), ethanol, ethyl ether and petroleum ether is used to extract the fat fraction, followed by gravimetric determination of the fat content.^{105,112} As a solvent-free alternative, the Creamatocrit technique has been introduced.¹¹³ Here, the sample is centrifugated in a tailored centrifuge, followed by reading of the cream content from a capillary tube, which is linearly related to the fat content in milk. This method offers low-cost and rapid analysis. Similar as for total protein content, the fat content is nowadays often determined by mid-IR spectroscopy, due to a significantly higher sample-throughput than achievable traditional methods (see section 3.2.2).

In this work, especially the fatty acid profile of human milk was of interest. Gas chromatography (GC), with different detection methods such as flame ionization detectors (FID) or mass spectrometry (MS), is the established technique for investigating the fatty acid composition. GC is based on chromatographic separation due to interactions of analytes in a mobile gaseous phase with a solid stationary phase. One prerequisite for this technique is vaporization of the analytes without decomposition, which cannot be achieved for triglycerides. Thus, a derivatization step to fatty acid methyl esters (FAMES) is typically required. Many different derivatization methods, using reagents such as NO-Bis(trimethylsilyl)trifluoroacetamide (BSTFA), boron trifluoride (BF₃) or anhydrous HCl have been proposed, requiring long incubation times and temperatures as high as 95 °C.¹¹⁴⁻¹¹⁶ Some reagents such as HCl, BF₃ and potassium hydroxide (KOH) require additional extraction after the initial derivatization step, whereas M-(trifluoromethyl)phenyl trimethylammoniumhydroxide (TMTFTH) requires time-intensive overnight incubation.^{117,118} In this work, a GC-MS method based on derivatization with 0.2 M trimethylsulfoniumhydroxide (TMSH) in methanol was applied, offering advantages such as minimum sample handling and high-throughput.¹¹⁹ Derivatization of a fatty acid with TMSH is illustrated in Figure 10. Here, methylation occurs through a base-catalyzed transesterification reaction.¹¹⁹ After addition of the reaction agent, this method requires a heating step to approximately 70 °C for 15 min to achieve O-methylation of the carboxyl group of fatty acids.

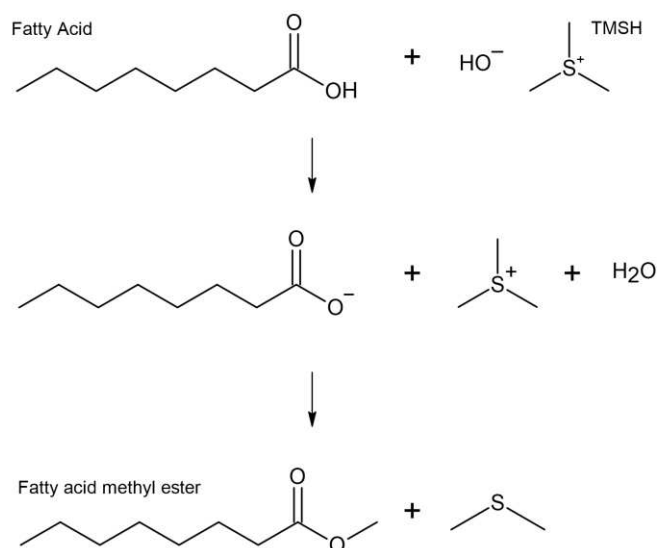


Figure 10: Derivatization of fatty acids to fatty acid methyl esters with trimethyl sulfonium hydroxide (TMSH) based on a base-catalyzed transesterification according to Ref¹¹⁹.

Even though, GC analysis offers high accuracy and maximum sensitivity, some major drawbacks of this technique are the required derivatization step prior to analysis, high cost and significant time consumption with typical run times of 1-2 hours. Thus, analysis is mostly restricted to few samples from larger batches, leading to a high demand for more rapid, low cost, high-throughput methods. Consequently, in this work, novel rapid methods based on mid-IR spectroscopy were developed, that should fulfill the majority of these demands (see section 3.3.2).

3.2.2 Mid-Infrared based methods

Compared to the methods explained in the previous section, mid-IR spectroscopy offers major advantages such as significantly higher sample-throughputs, lack of sample preparation steps that involve organic solvents and simultaneous analysis of multiple milk constituents from different substance classes. For this reason, mid-IR spectroscopy nowadays represents a widely used method for quality analysis of milk.¹²⁰ In section 3.2.2.1, commercially available milk analyzers, their applications and current limitations will be discussed. In section 3.2.2.2, EC-QCL based methods are summarized, that were previously applied for quantifying the most abundant bovine milk proteins.

3.2.2.1 Commercially available milk analyzers

The first mid-IR based analyzer, dedicated to milk analysis (MilcoScan) was introduced in 1975 by Foss Electric AS (Denmark).¹²¹ Since then, rapid advances in commercial instruments occurred, making mid-IR spectroscopy one of the major topics in dairy science.¹²² Nowadays, different analyzers based on FTIR spectroscopy such as the latest-generation MilcoScan and Dairy Spec FT (Bentley Instruments Inc, USA), or cheaper filter-based spectrometers such as the MIRA (Bruker, USA) are available for analyzing bovine milk and many other dairy products. For human milk, the filter-based Human Milk Analyzer (HMA, Miris AB, Sweden) was introduced as a tailored device for measuring the macronutrient composition. Unlike the methods described in chapter 3.2.1, all of these analyzers are standalone units and do not depend on specialized laboratory facilities, making them ideal for on-site analysis by non-experts.

All available mid-IR based milk analyzers are equipped with transmission cells, typically consisting

of CaF_2 windows and optical path lengths of 35-50 μm . As milk has a high viscosity, injection into cells of significantly smaller optical path lengths would tend to formation air bubbles or cell clogging, consequently preventing robust sample handling. Prior to mid-IR transmission measurements, milk samples have to be homogenized to avoid light scattering effects through large MFGs, that could lead to inaccurate results.¹²³ Many commercially available analyzers are equipped with online homogenization units in order to reduce the overall measurement time. In this context, sample throughputs of >1 sample per minute can nowadays be achieved. Figure 11 shows a transmission spectrum of a representative homogenized human milk sample, measured with FTIR spectroscopy and an optical path length of 50 μm . Here, the wavenumber range between 1850 and 2750 cm^{-1} is not shown due to lack of relevant information. The spectral regions between approximately 1600 and 1700 cm^{-1} and >3000 cm^{-1} are not accessible due to total absorption of the applied low-power radiation through water (see chapter Chapter 2). Furthermore, with this approach, the wavenumber region below approximately 1000 cm^{-1} is inaccessible due to the onset of IR absorption by CaF_2 .²² The remaining spectrum shows distinct absorption bands. An overview on the mid-IR absorption bands that typically occur in milk is provided in Table 7.

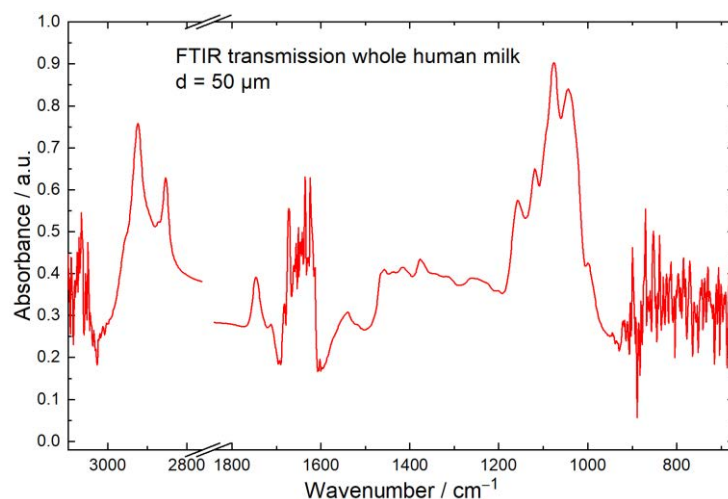


Figure 11: Fourier-transform infrared transmission spectrum of whole human milk, measured with a CaF_2 transmission cell with an optical path length of 50 μm . The wavenumber region between 1850 and 2750 cm^{-1} is not shown due to lack of important information.

For certain parameters such as total protein, total fat and lactose content, specific absorption bands are available, thus all commercial analyzers are capable of targeting those analytes. However, when more in-depth information, such as the concentration of individual proteins or fatty acids should be required, a broader spectral range is needed. Consequently, filter spectrometers are typically restricted to measuring the milk macronutrient composition based on univariate calibrations. When additional parameters are determined by utilizing broader spectral regions with overlapping absorption bands, evaluation is typically performed based on multivariate chemometric methods. The established method for determining individual analytes from multicomponent systems is PLS regressions.¹²² With this method, calibrations can be calculated for analytes of known concentrations (e.g., from reference methods), by considering all expected components of the matrix that have to be present in the measured calibration samples. In order to determine the performance of PLS calibrations, cross-validation (CV) is commonly employed, where the data set is split into different portions of training and testing sets during multiple iterations.¹²⁴ Based on the outcome of CV, the

optimum number of latent variables for building the model can be selected using statistical figures of merit such as the root-mean-square error of cross-validation (RMSECV) and cross-validation coefficient of determination (R^2_{CV}). In order to achieve more thorough validation, the calculated PLS calibrations can be applied to predict the concentrations of an independent external data set (validation set). Here, typically figures of merit such as the prediction coefficient of determination (R^2_P) and root-mean-square error of prediction (RMSEP) are calculated to evaluate model performance.

Table 7: Mid-IR absorbance bands of milk.¹²⁰

Wavenumber / cm^{-1}	Likely Origin	Mode of vibration	Intensity
3008	Fat	<i>cis</i> =CH stretch	w
2959	Fat + minor Protein	CH ₃ asym. stretch	m
2923	Fat + minor Protein	CH ₂ asym. stretch	s
2895	Mixture	CH stretch methine groups	m
2873	Mixture	CH ₃ sym. stretch	m
2852	Fat + minor Protein	CH ₂ sym. stretch	s
1745	Fat + minor Protein	C=O stretch	s
1639-1627	Protein, Interactions with Water + Fat	Amide I	s-w
1570-1516	Protein, Interactions with Fat	Amide II	m-w
1467	Lactose, Fat	OH def/CH ₂ scissoring	m
1440	Protein	CH def.	w
1415	Fat + Lactose	CH rocking	m
1400	Protein	CH def.	w
1379	Fat + Lactose	CCH def.	m
1342	Lactose	CC and CO skeletal	w
1319	Protein	CH ₂ wagging	w
1304	Lactose	OH def.	w
1240	Mixture	Amide III/ OPO asym. stretch/ CO stretch	m
1225	Fat	PO ₂ ⁻ asym. stretch	w
1172	Fat	CO stretch ester	m
1161-1096	Mixture	CO stretch	m
1074-1045	Lactose	CO stretch	s-m
1020	Mixture	CO stretch	m
991	Lactose + Fat	CH oop. stretch	m
966	Fat	trans=CH oop. stretch	w
947	Fat	trans CH oop. stretch, conjugated	w
894	Lactose	CH def	m
786	Lactose	Ring vibration	m
721	Fat	CH ₂ rock	m

Abbreviations: asym.: antisymmetric; sym.: symmetric; def.: deformation; oop.: out-of-plane; w: weak; m: medium; s: strong

Due to inaccessibility of the amide I band, commercially available milk analyzers use the amide II band for protein analysis. Even though major effort has been put into predicting individual protein concentrations in bovine milk, based on FTIR transmission spectra and chemometrics, the reported results were rather poor.¹²⁵ These poor results are likely due to the fact that the amide II band shows only minor sensitivity to protein secondary structure, thus type of protein. Consequently, commercially available IR-based milk analyzers reliably only allow for measuring the total protein content.

For fat analysis, a combination of two mid-IR bands is commonly used, referred to as fat A (C=O

stretch) and fat B (C-H stretch).¹²⁶ Based on these absorbance bands, accurate determination of the total fat content is possible. Certain filter spectrometers such as the HMA, however, only use the C=O stretching band for determination of the total fat content. It has been reported, that the thereby obtained results can significantly differ from reference methods.¹²⁷ For bovine milk, many publications are available that combine commercially available FTIR analyzers, PLS regressions, different methods for data preprocessing and variable selection to determine the fatty acid profile.¹²⁸⁻¹³⁴ Here, good results were obtained when the fatty acid composition was stated as absolute concentration values in whole milk, however, when the results were stated relative to the total fat content, they appear significantly poorer. It has been reported that the determined absolute concentration levels are likely based on covariation structures between individual fatty acids and the total fat content which may change with factors such as breed and feed.⁵ Consequently, mid-IR based fatty acid determination in bovine milk is still a topic of discussion. For human milk, no literature was available that reports mid-IR based methods for measuring the fatty acid composition, before the start of this thesis.

Besides the explained parameters that are of interest in this work, it should be noted that FTIR based milk analyzers are capable of analyzing a large number of additional variables such as total solids, glucose, galactose, density, urea, titratable acidity, free fatty acids, citric acid and others.¹³⁵

3.2.2.2 QCL-IR spectroscopy to detect individual proteins in bovine milk

Recently, the EC-QCL-based academic setups explained in section 2.3 were applied for quantification of individual proteins in bovine milk.¹³⁶ The first method was published in 2017,⁶³ where the first-generation setup (see section 2.3.1) was employed, allowing for measurement times of approximately 10 min per milk sample. Here, the large path length of 38 μm enabled robust spectra acquisition of the amide I band, which is not possible with FTIR based analyzers. Before chemometric quantification, subtraction of contributions from other constituents of the matrix to IR spectra was performed based on multivariate science-based calibrations. Subsequently, PLS models were implemented, allowing for concentration determination of casein, β -LG and total proteins by achieving RMSECV values of approximately 1 mg/mL. In a follow up-study, commercially available bovine milk samples were measured with this method and the obtained protein concentrations compared to HPLC reference values, showing excellent agreement.⁶⁴ Furthermore, it was demonstrated that laser-based IR spectroscopy allows for discrimination of different milk types, according to the amount of experienced heat load, based on the obtained concentration of the temperature sensitive indicator protein β -LG.

Later, the second-generation setup (see chapter 2.3.2), allowing acquisition of spectra across the amide I + amide II region with increased signal-to-noise ratios, was applied for analyzing bovine milk proteins.³ Here, effects of the matrix on the IR spectra was directly considered during calculation of PLS regressions by measuring differently diluted milk samples. Due to increased sensitivity, higher tuning rates of the laser and the larger covered spectral range compared to the first-generation setup, this approach allowed quantification of α -LA, β -LG, casein and total protein content within measurement times of approximately 1 minutes per sample, achieving RMSECVs of approximately 0.25 mg/mL. As illustrated in Figure 12, simultaneous quantification of α -LA and β -LG, both showing sensitivity to high temperatures, allowed for fine discrimination of commercially available bovine milk types.

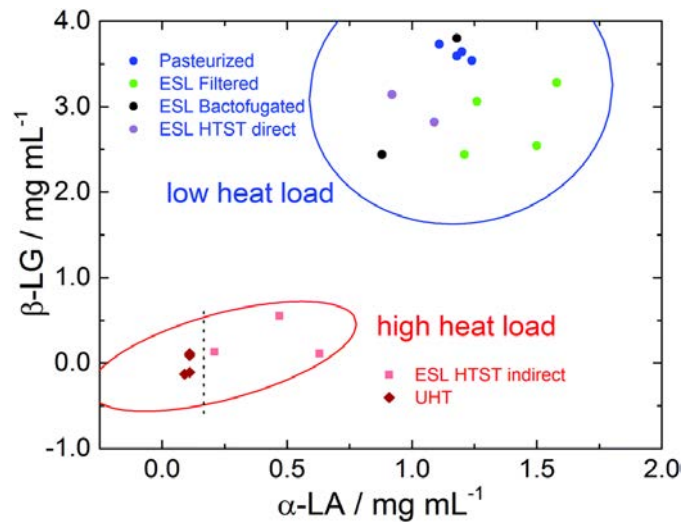


Figure 12: Concentrations of α -lactalbumin (α -LA) plotted vs. concentrations of β -lactoglobulin (β -LG), determined with the second-generation EC-QCL setup. Low concentrations of the two proteins indicate high heat load, whereas higher concentrations indicate low heat load. Reprinted with permission from Ref³.

Here it was demonstrated that milk that experienced high and low heat load can be distinguished. Moreover, for high-temperature-short time (HTST) milk it was further possible to distinguish between applied heating technologies such as direct and indirect heating. Finally, within the high heat load group, it was possible to identify ultrahigh temperature (UHT) milk.

3.3 Novel methods developed within this work

In this work, novel methods for analyzing the protein and fatty acid composition of human milk were developed. For protein analysis, a novel, highly sensitive EC-QCL based setup was implemented (section 3.3.1). For fatty acid analysis, an approach based on solvent-free lipid separation and ATR-FTIR spectroscopy was the method of choice (section 3.3.2).

3.3.1 Development of a balanced EC-QCL-based IR transmission setup for protein analysis

As quantification of individual proteins in milk based on commercially available FTIR instruments showed poor results (see section 3.2.2.1) and laser-based IR spectroscopy allowed determination of individual proteins in bovine milk (see section 3.2.2.2), EC-QCL based IR spectroscopy was the method of choice. Compared to bovine milk, the total protein content in human milk is significantly lower (see section 3.1.1), making analysis of individual protein concentrations more challenging. Another challenge is caused by the secondary structures of the most abundant proteins in human milk, which are more similar than those of bovine milk. Figure 13 shows a comparison of absorbance spectra between the three most abundant proteins in bovine milk and human milk. Casein is mainly composed of unordered secondary structures, β -LG of β -sheets and α -LA of α -helices, consequently, the amide I + II bands of these most abundant proteins in bovine milk show significant differences. In contrast, human milk proteins exhibit more similar protein spectra. Lactoferrin is predominantly composed of α -helices, thus showing similar amide I + II bands as α -LA with only slight differences. Furthermore, unordered proteins such as casein have a similar amide I band maximum to α -helices (see Table 2). Due to these smaller differences in absorbance spectra and the overall lower protein concentrations, higher sensitive methods are required to determine individual protein concentrations in human milk based on IR spectroscopy.

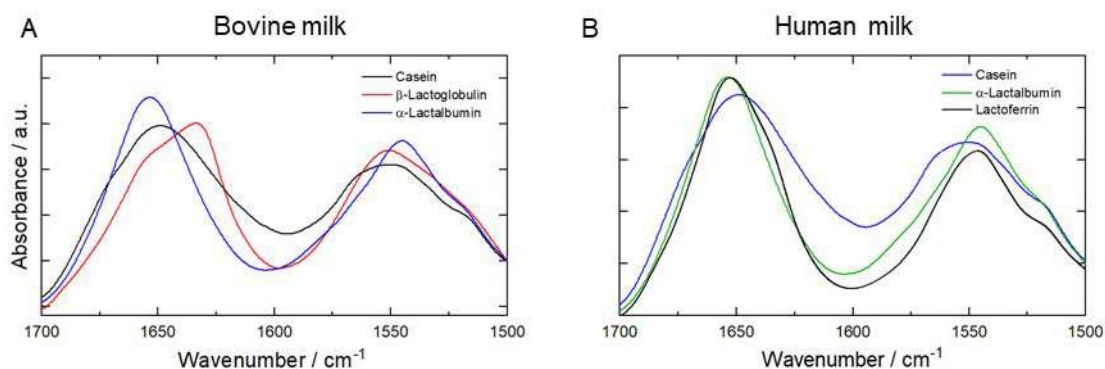


Figure 13: Mid-IR absorbance spectra of protein amide I + II band of the three most abundant proteins in (A) bovine milk and (B) human milk.

These challenges required development of a novel, more sensitive setup for protein transmission measurements in aqueous solutions, compared to the second-generation setup, described in section 2.3.2. For this purpose, the latest generation EC-QCL from Daylight Solutions (Hedgehog) was applied. Compared to the second-generation setup, a novel scheme, based on balanced detection was implemented. Balanced detection was first introduced in 1990,¹³⁷ with the aim of achieving highly sensitive measurements with optical setups that apply noisy lasers. In brief, the laser beam is split into two beams of equal intensities, one light path containing the sample and the

other one a reference solution, while both of them contain the noise produced by the laser. By recording both beams with two matched detectors and subsequent subtraction of the signal, a major part of the noise can be cancelled out. In the context of QCL-based setups, balanced detection was applied for gas-phase measurements, leading to significantly increased sensitivity compared to single channel detectors.^{138,139} Later, balanced detection was used for liquid-phase measurements, in order to reduce the noise level of an EC-QCL based detector for SEC.¹⁴⁰ Most recently, balanced detection was also applied in an ATR setup to achieve background-free low-noise measurements.¹⁴¹

Figure 14 shows a schematic of the setup developed within this work. Here, a thermoelectrically cooled balanced detection module from Vigo (Vigo System S.A., Poland) was employed that is composed of two precisely matched MCT detectors. The laser beam was split into two parts by a 50/50 beam splitter and guided through a two-path transmission cell (one containing the sample or background solution, the other one a blank reference solution), followed by guiding of the two beams onto the MCT detectors. In order to achieve maximum noise suppression, the intensity on the two detection modules should be as similar as possible. For this purpose, a CaF₂ window was used as a neutral-density filter for one of the arms. In contrast to the second-generation setup, a high-end lock-in amplifier from Zurich instruments (Switzerland), allowing for faster data acquisition rates was employed. The setup performance was evaluated by recording protein spectra at different concentration levels and comparing them to FTIR transmission spectra. Moreover, noise levels of balanced channel and signal channel were compared, followed by benchmarking against previous setups and high-end FTIR spectroscopy. The obtained results are presented in Publication I.⁶

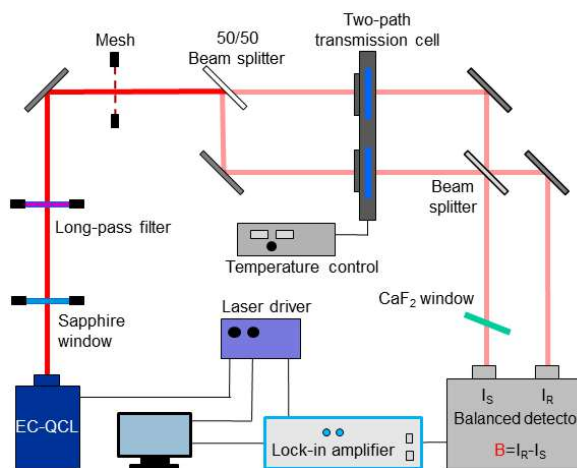


Figure 14: Schematic of the developed balanced detection setup. Adapted from Ref⁶.

3.3.2 Fatty acid analysis

As FTIR based transmission measurements of whole milk showed rather poor results for fatty acid analysis (see section 3.2.2.1), a different approach was applied within this work. It has been recognized, that for whole milk transmission spectra several problems occur that make determination of individual fatty acids a difficult task. First, many lipid related absorbance bands overlap highly with bands from other major milk components such as proteins, carbohydrates and

water (see Table 7). Moreover, CaF_2 , the typical window material used in transmission measurements of milk, has its absorption edge at approximately 1000 cm^{-1} ,²² meaning that fat related IR bands at lower wavenumbers are not accessible using this approach. In this context, it has been shown that an alternative approach, where small amounts of milk samples were transferred into well plates, dried in a desiccator, and subsequently measured in transmission mode, lead to better results.¹⁴² Here, the lipid preconcentration step was expected to contribute to a major part of the gained model improvements. Based on these findings, it was expected that a complete separation of the lipid fraction from the milk sample and subsequent mid-IR measurements of this fraction, would lead to even better results. Thus, the first step towards this novel approach was to find a proper method for lipid separation. Classic solvent extraction methods for milk fat separation are considered as highly reliable for quantitatively removing the lipid fraction and subsequent determination of the total fat content (see section 3.2.1.2). However, these methods require large amounts of potentially hazardous organic solvents and are very time consuming. As an alternative, more rapid methods have been developed that require smaller amounts of organic solvents, allowing separation of the lipid fraction from approximately 20 milk samples within 30 min.¹⁴³

In this work, the lipid separation method of choice was solvent-free and based on two centrifugation steps. This method was first introduced and validated in 2004,¹⁴⁴ followed by modification in 2005.¹⁴⁵ Compared to the traditional methods, this approach does not allow for quantitative separation of the total fat content, however, a pure lipid fraction is obtained that is representative in fatty acid composition for total fat. The centrifugation methods have previously been thoroughly validated by measuring the fatty acid composition of the obtained fat fraction with GC and comparing it to the fatty acid profile of milk fat, obtained by standard, solvent-extraction, showing excellent agreement.^{144,145} As the purpose of the method developed within this work was to obtain the relative fatty acid profile of human milk, quantitative extraction was not required. Here, the modified method from Luna et al.¹⁴⁵ was employed, as it does not require cooling of the centrifuge to $4\text{ }^\circ\text{C}$. In this method, 30 mL milk sample aliquots were transferred into falcon tubes (Figure 15A) and centrifuged at $17,800\times g$ for 30 min at $20\text{ }^\circ\text{C}$ (Figure 15B). The upper layer was then transferred into microtubes (Figure 15C) and centrifuged at $19,300\times g$ for 20 min at the same temperature, resulting into three separate layers: an upper layer consisting of pure lipids, a middle layer of proteins, fat and other water insoluble compounds and a bottom layer of water (Figure 15D).

After centrifugation, the upper pure lipid layer was used for further analysis. One part of the fat was dissolved into dichloromethane and analyzed with a GC-MS reference method, based on derivatization with TMSH (see section 3.2.1.2). For mid-IR measurements, ATR-FTIR was the technique of choice (see section 1.2.3), as one drop of pure milk fat is sufficient to record absorption spectra over the whole mid-IR region. For this purpose, Bruker Tensor 37 FTIR spectrometer, equipped with a MCT detector and Platinum ATR single-bounce element was used. The proof-of-concept of this novel approach was first performed based on 45 bovine milk samples due to easier availability compared to human milk samples. The results of this study are presented in Publication II.⁷ After that, the method was applied for a limited data set of 8 human milk samples, representing the first mid-IR based approach for determination the human milk fatty acid composition. The thereby obtained results are shown in Publication III.⁸ Finally, a full validation for human milk was performed based on 50 samples that were split into training and validation set (see Publication IV).⁹

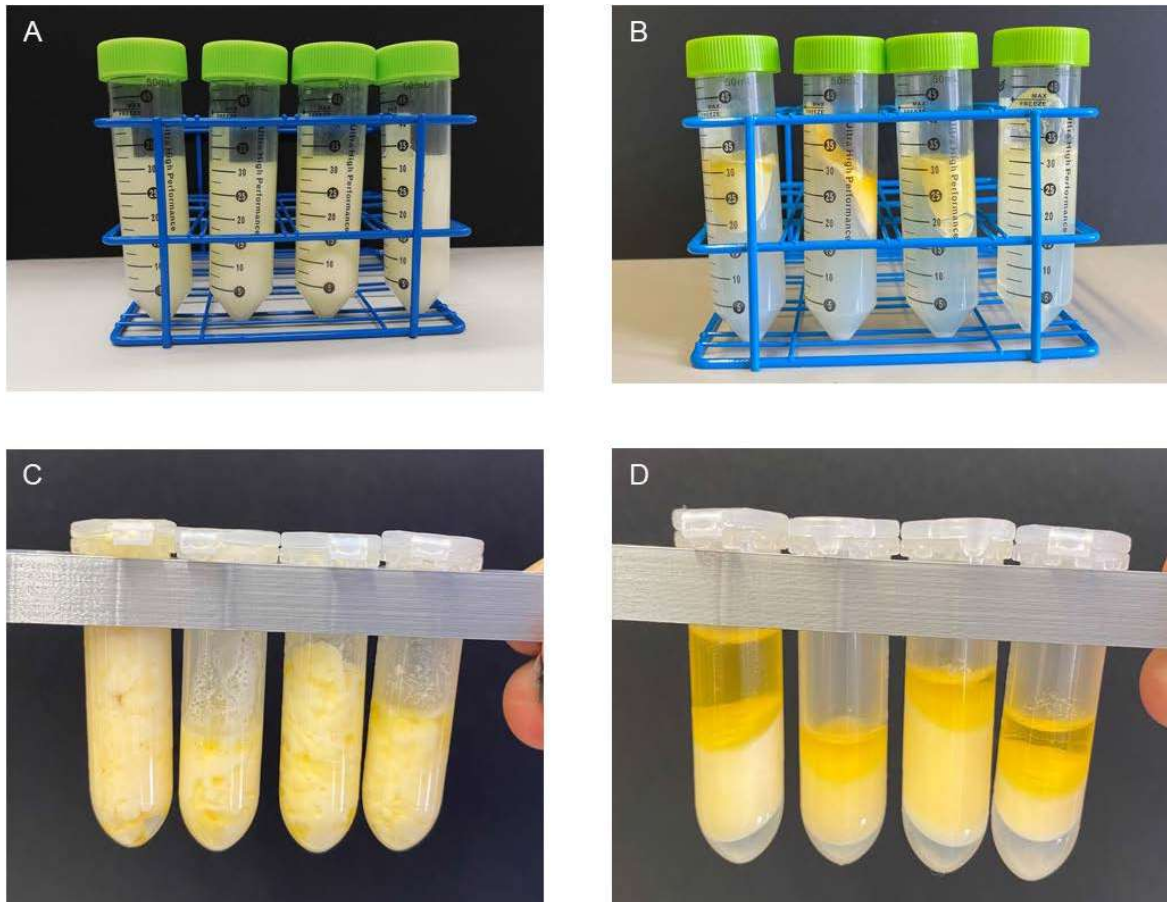


Figure 15: Fat separation from human milk according to the method from Luna et al.¹⁴⁵ (A) Untreated whole human milk, (B) after the first centrifugation step, (C) transferred upper cake before and (D) after centrifugation, showing an upper phase of pure fat, a middle layer of lipids, proteins and other water insoluble components and a lower water phase.

Chapter 4 Bioprocess monitoring

In this work, bioprocess monitoring was performed by applying novel QCL-IR based methods for analyzing proteins from preparative LC. The most relevant background information on this topic is provided in this chapter. Section 4.1 gives a general overview on LC, including instrumentation and separation principles. In section 4.2 conventional analytical methods for monitoring preparative LC, including detectors and offline methods are discussed. Finally, in section 4.3 hyphenation of LC to mid-IR spectroscopy is outlined, including an introduction to the novel analytical methods developed within this work.

4.1 Preparative liquid chromatography

In LC, analytes are separated in a liquid mobile phase by interactions with a solid stationary phase according to different physio-chemical properties. Nowadays, LC represents an essential unit operation in the production of biopharmaceuticals, thus it is part of most protein purification and polishing protocols.¹⁴⁶ In this section the most relevant instrumentation for preparative LC (section 4.1.1) and separation principles (section 4.1.2) are summarized based on Ref¹⁴⁷ that provides more in-depth information.

4.1.1 Instrumentation

All LC based purification systems for both, laboratory and industrial scale bioprocesses comprise several typical components: solvent reservoirs that contain one or multiple different eluents (mobile phase), pump(s) and mixer, a sample introduction device, a separation column (stationary phase), one or multiple detectors, fraction collector and a data acquisition system. A schematic of a typical LC setup is shown in Figure 16.

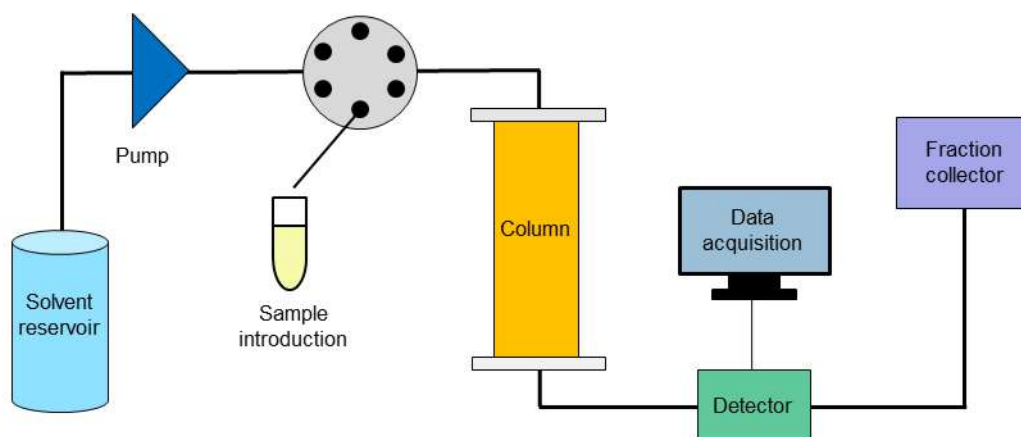


Figure 16: Schematic of a LC setup.

Solvent reservoirs are clean, inert containers, containing the buffers used as mobile phase. The mobile phase is usually pumped through the system by positive displacement pumps in order to maintain accurate flow. For this purpose, different types such as long-stroke piston pumps, reciprocating pumps, peristaltic pumps, diaphragm pumps and others can be employed. The selection of the pump depends on various factors such as the intended application (e.g., laboratory or large scale), desired precision and possible pulsatile flow. Chromatographic separation can either be performed isocratic with a consistent mobile phase across the entire run or by using a gradient, where the eluent composition is changed over time in order to achieve the desired elution of analytes. For generation of gradients, buffer mixers are applied that blend varying proportions of two, or rarely more, solvents. Here, either low-pressure mixtures using a single pump with ratio inlet valves or high-pressure mixtures using independent pumps with adjusted flow rates, are employed. Sample introduction can either be performed through a switching valve with a loop, or in feed operation, using a pump. For laboratory scale instruments, as applied within this work, sample injectors based on loops that can be placed in and out of the eluent flow are most commonly used. Here, the sample loop is filled at atmospheric pressure while the valve is in load position, followed by turning the valve into injection position where the loop is connected to the mobile phase stream, thus the sample is carried to the column.¹⁴⁸ The key part of each LC system is the column, where analytes are retained, thus chromatographic separation occurs. Various column types with different separation principles can be used for protein purification protocols (see section 4.1.2). After chromatographic separation, the LC effluent is collected by a fraction collector, which is in the most rudimentary design comprised of a valve and different vessels. In the simplest case, fractions are collected manually based on individual operator control, however, more sophisticated approaches based on retention times, setting of threshold values for detector signals or mass selective detectors are available.¹⁴⁹ Monitoring of proteins from preparative LC is typically performed by a combination of inline detectors and offline measurements of the collected fractions (see section 4.2).

4.1.2 Separation principles

Numerous different separation principles can be applied for preparative LC, of which some of the most implemented ones are size exclusion chromatography (SEC), ion-exchange chromatography (IEX), normal phase chromatography (NPC), reversed phase chromatography (RPC), hydrophobic interaction chromatographic (HIC) and different types of affinity chromatography. The separation

mechanism is determined by the chemistry of the stationary phase; however, adaption of the mobile phase is additionally required to achieve the desired retention behavior. For proteins, several parameters such as salt concentrations, pH-values and other properties of the mobile phase have to be considered, as they can significantly influence protein stability. In the present work, SEC and IEX were applied as separation principles, followed by inline monitoring of the effluents with EC-QCL-IR spectroscopy.

In SEC, compounds are separated according to their size and shape, which offers several clear advantages compared to other separation principles, including non-denaturing conditions, straight forward operation and isocratic elution.¹⁵⁰ SEC columns are packed with an inert substrate, comprising of pores with different sizes. Based on the size of the analytes, larger molecules can penetrate a smaller number of pores, whereas smaller molecules can penetrate a larger number of pores.¹⁵¹ According to this behavior, larger analytes elute first, whereas smaller analytes are retarded by the column for a longer period of time, thus eluting later. For purification of proteins based on SEC, typically isocratic elution with aqueous buffers at constant flow rates is employed.

In IEX, separation is mainly based on electrostatic interactions between analytes and an oppositely charged functional group on the stationary phase.¹⁵² For proteins, the net surface charge depends on the isoelectric point (pI) and the pH of the mobile phase. At equal pI and pH-values, proteins do not carry a net charge, however, pH-values below the pI will lead to a positive charge, whereas pH-values above the pI lead to negative net charge. Consequently, both, cation- and anion exchange chromatography can be employed by application of adequate eluents. In IEX for protein separation, typically mobile phases based on aqueous buffers are applied. In contrast to SEC, gradient elution is commonly performed, either based on varying salt concentrations or pH-values to achieve chromatographic resolution.¹⁵²

In NPC, the stationary phase is more hydrophilic than the mobile phase, which is typically composed of a mixture of organic solvents. Due to limited solubility of proteins in such solvents, application of NPC in the biotechnological field is restricted to peptides and other small molecules. On the other hand, RPC uses a stationary phase that is more hydrophobic than the mobile phase, which is typically a mixture of water and organic solvents, facilitating sufficient solubility of proteins. Consequently, RPC is widely used in prep LC, as well as in HPLC of proteins (see section 4.2.2). In contrast to SEC and IEX, RPC however often requires more harsh conditions such as high temperatures and low pH values that can denature proteins.¹⁵³ HIC is also based on a hydrophobic stationary phase with smaller densities and hydrophobicity of the ligand compared to RPC. Finally, in affinity chromatography the stationary phase comprises a biologically related agent, capable of specifically retarding the analyte in the mobile phase.¹⁵⁴

4.2 Conventional Methods for Liquid Chromatography Monitoring

4.2.1 Detectors

Many different detectors for monitoring of LC processes are available, of which some of the most implemented ones are ultraviolet/visible (UV/Vis), conductivity, pH, refractive index, fluorescence, light scattering, electrochemical and MS detectors.¹⁴⁷ In this work, UV/Vis and conductivity detectors were applied, consequently the focus of this section is on these two types.

UV/Vis detectors are based on absorption of electromagnetic radiation in the wavelength region between approximately 200 and 800 nm, typically operating in transmission configuration (see section 1.2.3) with common optical path lengths of 1 to 2 mm. In analogy to IR spectroscopy, quantitative evaluation can be performed according to *Beer-Lambert* law. Some of the major advantages of UV/Vis detectors are their broad dynamic range and excellent sensitivity. They, however, typically provide univariate signals based on a selected wavelength, consequently do not give in-depth information about protein structures, as mid-IR spectroscopy does (see section 2.1). The most frequently used wavelength for protein quantification by UV/Vis spectroscopy is 280 nm, which is based on absorption caused by the aromatic ring of the amino acids tryptophan and tyrosine. In this context, proteins of similar molecular weights can have distinctly different absorption coefficients due to varying amounts of these amino acids. The absorbance at 280 nm is additionally influenced by the protein structure. Consequently, factors such as pH-value, temperature and ionic strength, all affecting the structure, have to be considered.^{155,156} Furthermore absorption through the peptide bond (e.g., at 205 nm) can be used for protein quantification.¹⁵⁷ This is particularly useful for proteins that contain small amounts of tryptophan and tyrosine. Proteins with additional chromophores show further absorbance peaks in the UV/Vis region. Typical examples are the iron-containing proteins hemoglobin and myoglobin, which are red in color and have peaks near 600 nm.^{147,158}

Conductivity detectors are commonly used to monitor electrolyte concentrations in the mobile phase, which is particularly useful for IEX runs with gradients of varying salt concentrations. This type of detector is based on measuring the resistance between two electrodes within a flow cell that contains the chromatographic effluent.¹⁵² Here, the conductivity κ can be calculated based on the distance between the electrodes l , the cross-sectional area A and the resistance R of the solution that passes through the cell:¹⁴⁷

$$\kappa = \frac{l}{RA} \quad (4.1)$$

Subsequently, the measured conductivity can be related to the concentration of an electrolyte c according to Kohlrausch's law:

$$\kappa = \Lambda_m c \quad (4.2)$$

where Λ_m is the molar conductivity that varies from salt to salt and depends on factors such as temperature and viscosity of the solution. In the present work, the signal of the conductivity detector was of special interest for developing a method, capable of compensating for undesired mid-IR absorbance bands of the water background resulting from changes in H-bonding in the water due to the applied sodium chloride (NaCl) gradient (see section 4.3.2).

4.2.2 Offline methods

In order to allow for acquisition of more in-depth information such as target protein titer or purity, offline measurements of the collected fractions usually have to be performed. Here, a seemingly endless number of analytical techniques is available, where the selection of a proper method depends on factors such as proteins of interest, required sensitivity, time consumption and degree of automation. Some of the most established methods are based on HPLC, sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), CE, western blotting and biological activity assays such as ELISA,¹² all of them coming with certain advantages and disadvantages. Recently, a generic HPLC method based on RPC was introduced, with the aim of baseline resolving and quantifying a large number of different proteins.¹⁵³ In the present work, this method was used as a reference method for QCL-IR inline measurements, consequently the focus of this section is on this method. In brief, a BioResolve RP mAb Polyphenyl column (Waters, USA) was applied, representing a state-of-the-art stationary phase that provides high separation efficiency, while at the same time minimizing protein on-column adsorption. During the chromatographic runs, the column temperature was kept constant at 70 °C. Water was used as a mobile phase A and acetonitrile as mobile phase B, both of them containing additional 0.1% v/v of trifluoroacetic acid (TFA). A flow rate of 0.4 mL/min was applied during the 18 min run time. Gradient elution was performed, starting with 25% eluent B and a linear increase to 75% within 10 min, followed by holding these conditions for 2 min, before re-equilibrating the column for 6 min at 25% B. UV/Vis detection was applied (see section 4.2.1). This method also offers compatibility to MS detection, in case that more in-depth information is required.¹⁵³ Validation was performed by determination of pivotal parameters such as linearity, limit of quantification (LOQ), repeatability and by comparing the obtained results to SDS-PAGE. Here it was demonstrated, that the HPLC method represents a highly sensitive and accurate tool for analyzing a broad range of proteins with different physio-chemical properties.¹⁵³

4.3 Hyphenation of LC with IR spectroscopy

Conventional analytical methods for monitoring bioprocesses are based on offline analysis of collected samples (see section 4.2.2). During the recent years, quality by design (QbD) principles were established in the pharmaceutical industry with the aim of assuring product quality by understanding and controlling formulation and manufacturing variables.¹⁵⁹ In order to comply with these principles, process analytical technology (PAT) is required that allow inline control and timely adaption of bioprocess parameters. According to the Food and Drug Administration (FDA), PAT is “a system for designing, analyzing and controlling manufacturing through time measurements (i.e., during processing) of critical quality and performance attributes of raw and in-process materials and process, with the goal of ensuring final product quality”.¹⁶⁰ In case of chromatographic protein purification, conventional inline detectors (see section 4.2.1) do not provide the required critical quality and performance attributes, consequently there is a clear demand for novel methods that can close this gap. In this context, mid-IR spectroscopy allows for acquisition of detailed quantitative and qualitative information about proteins (see Chapter 2), thus bearing high potential as a PAT tool for monitoring of LC effluents. In section 4.3.1 the current state-of-the art of LC-IR hyphenation is explained, followed by an introduction to the novel methods developed and applied within this work (section 4.3.2).

4.3.1 State-of-the art

Most studies reporting hyphenation of LC with IR spectroscopy are based on analytical HPLC and FTIR spectroscopy, however, also few studies are available dealing with preparative LC. In general, two types of coupling methodologies can be discriminated: application of solvent-elimination interfaces and flow-cell interfaces.¹⁶¹ Solvent-elimination approaches are based on evaporation of the eluent and deposition of the analyte onto a substrate, followed by IR spectra acquisition. This approach comes with the advantage that spectra are not influenced by absorbance bands of the eluent, which are typically significantly more pronounced than those of analytes. Simultaneous eluent evaporation and analyte deposition, while maintaining the integrity of LC separation is, however, a challenging task due to possible spatial inhomogeneities and the morphology of certain analytes that can change over time.¹⁶¹ Moreover, in preparative purification processes, where the analyte fractions should be collected after separation, this destructive type of sample preparation is inapplicable.¹² In contrast, the approach based on flow-cells offers a more rapid alternative, where spectra acquisition of the chromatographic effluent is continuously performed while maintaining sample integrity. Here, the effluent passes through a flow-cell and IR spectra acquisition is performed in real-time, offering advantages such as instrumental simplicity, low cost and possible use of nonvolatile buffers.¹⁶¹ With this approach, however, strong IR absorption of eluents can require short applicable path lengths that distinctly impair the robustness in flow-through operation, especially when low-power mid-IR light sources are applied (e.g., for FTIR spectrometers). Moreover, pronounced eluent absorbance bands make application of gradient elution a challenging task, requiring complex chemometric methods in order to compensate for varying background spectra.

For protein chromatography, solvent evaporation schemes were developed in the 1990s and successfully applied to determine secondary structures from analytical-scale HPLC effluents.¹⁶²⁻¹⁶⁴ As this approach cannot be applied for preparative LC where intact fractions should be collected, laborious offline methods of the fractions are nowadays still the method of choice in order to allow for protein discrimination. Due to significant challenges for transmission measurements in aqueous

solutions (see Chapter 2), mid-IR spectroscopy of proteins found its way into inline detection of preparative LC processes only very recently.^{165,166} In these works FTIR spectroscopy was combined with multireflection ATR elements to allow for robust spectra acquisition (see section 1.2.3). Here, it was possible to distinguish proteins based on their secondary structure, as well as to monitor other eluting components such as polyethylene glycol (PEG) or Triton X-100. Even though this approach allowed for robust spectra acquisition, the sensitivity is still limited due to small achievable interaction lengths between IR light and sample in the protein amide I region, leading to small absorbance signals. Consequently, this method is only applicable for purification processes that which deal with high protein concentrations.

In this work, the advantages of EC-QCL based mid-IR spectroscopy for transmission measurements of proteins (see section 2.2.2) were utilized in order to achieve, sensitive inline measurements of proteins from preparative LC (see section 4.3.2).

4.3.2 Novel methods developed within this work

In this work, laser-based IR spectroscopy was applied for monitoring of proteins from preparative LC systems with different separation mechanisms. Previously, an EC-QCL based setup was successfully coupled to HPLC in order to allow quantification of different sugars in several beverages by applying a straight-forward separation system based on isocratic elution and an aqueous mobile phase.¹⁶⁷ In the present case, however, LC-QCL-IR hyphenation was significantly more challenging due to pronounced light absorption through the HOH bending band of water that overlaps with the protein amide I band (see Chapter 2) and strong variations in absorption of the eluent when gradient elution was performed.

In order to tackle these challenges, the commercially available ChemDetect Analyzer from Daylight solutions was used, representing an EC-QCL based device that allows for robust spectra acquisition beyond the protein amide I and II region. The equipped laser covers the wavenumber range between 1350 and 1770 cm^{-1} , whereas a diamond transmission cell with an optical path length of 25 μm enables robust measurements in flow-through operation. The number of averaged scans can be manually set by the user in order to fulfill the desired sensitivity. Here, a measurement time of approximately 0.5 s is achieved per scan. In addition, the small footprint and portability of the ChemDetect Analyzer make it an ideal device for hyphenation to other techniques. As no scientific publications on this device were previously available, the first step was to benchmark it against FTIR spectroscopy. Here, band shapes and positions of model proteins with different secondary structures were compared, followed by comparison of the noise levels and LODs. In order to demonstrate the advantages of the larger covered wavenumber range, compared to most other EC-QCL setups, an enzymatic reaction was monitored. Here, the hydrolysis of triacetin to glycerin and acetate, catalyzed by lipase was selected, as distinct absorbances such as the bands at 1733 cm^{-1} and 1374 cm^{-1} , both arising from C=O stretching, are within the tuning range of the laser.^{168,169} The obtained results are presented in Publication V.¹⁰

The next step was to perform LC-QCL-IR hyphenation by using a laboratory scale preparative LC system (Äkta pure system, Cytiva Lifa Sciences, USA), equipped with UV detector, conductivity detector and fraction collector and directly guiding the effluent into the transmission cell of the ChemDetect Analyzer. The flow path of the applied laboratory setup is illustrated in Figure 17.

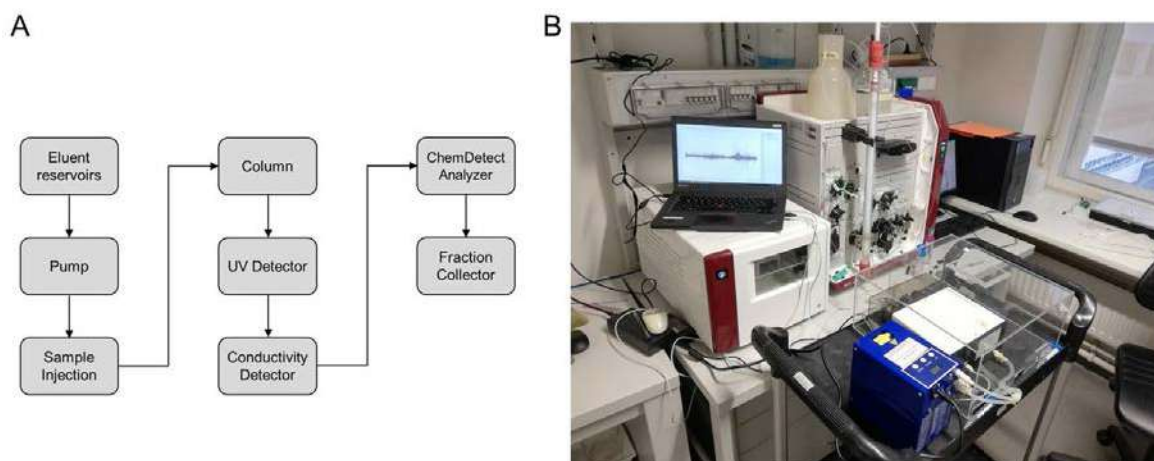


Figure 17: (A) Flow-path and (B) photo of the applied LC-QCL-IR setup. Adapted from Ref¹².

First, the proof-of-concept was shown based on a SEC run, involving three model proteins with different secondary structures under isocratic conditions. Here, basic data evaluation, based on different amide I band shapes was performed to assign the eluting proteins to reference offline spectra. The thereby obtained results are shown in Publication VI.¹¹

Second, more complex runs based on IEX were monitored. Here, a pronounced challenge was caused by the NaCl gradient that had to be applied. This gradient lead to mid-IR absorbance bands that overlap highly with the smaller protein bands, consequently a background compensation approach was necessary. For this purpose, a previously introduced approach based on a reference spectra matrix (RSM)¹⁷⁰ was modified. In the initial version of this method, the spectra acquired during the LC run are viewed as sample matrix (SM), whereas the RSM is additionally recorded during the re-equilibration phase of the chromatographic system or during a blank run. Next, absorbance bands, characteristic for the eluent, are identified that have to be located in different wavenumber regions than analyte bands. With this information, spectra in the SM are corrected with RSM spectra that have the most similar eluent composition. This approach was previously successfully employed for correcting different gradients in a wide range of HPLC-FTIR applications such as analysis of nitrophenols,^{171,172} pesticides^{170,173} and carbohydrates.^{174,175} For EC-QCL measurements of proteins from IEX, however, gradient compensation is more challenging due to the limited available wavenumber region, that does not show absorbance bands caused by the presence of NaCl in the gradient. Consequently, in this thesis a modified approach was developed by additionally incorporating the signal of the conductivity detector. In order to show the high flexibility of this approach, a single blank run was recorded, followed by correcting sample runs of different linear and step gradients.¹² The accomplished results are presented in Publication VII.¹²

Finally, SEC runs were monitored, involving proteins with overlapping chromatographic peaks that cannot be distinguished with conventional UV/Vis detectors. In order to extract information about individual proteins, chemometric methods had to be applied. In this work, combination of self-modeling mixture analysis (SMMA)¹⁷⁶ and MCR¹⁷⁷ was used, offering the advantage that all information can be obtained from the recorded data set without availability of previous knowledge. SMMA is a family of techniques, capable of calculating estimates of pure chemical components and their contributions to a mixture without requiring prior knowledge about the data set. One prerequisite for these techniques is that the spectral variables obey *Beer-Lambert* law in terms of linearly increasing absorbances proportional to the concentration, which is the case for the collected QCL-IR data. In this work, simple to use interactive self-modelling mixture analysis

(SIMPLISMA) was the method of choice where the spectral profiles are calculated based on a purity value.¹⁷⁸ These spectral profiles were subsequently compared to offline reference spectra of pure protein solutions in order to validate the obtained results. Next, MCR, representing a soft modelling technique, capable of solving the bilinear description of the recorded data set, for both spectral and concentration profiles, through composition into two allocated matrices, was employed. Consequently, spectra of pure proteins from overlapping chromatographic peaks, as well as their concentration profiles over the chromatographic run were obtained. The obtained quantitative information was compared to reference concentrations from HPLC offline measurements of the collected fractions (see section 4.2.2). The outcome is presented in Publication VIII.

Chapter 5 Green analytical chemistry

In this work, the developed mid-IR based methods for milk analysis and bioprocess monitoring were compared to established reference methods in terms of accordance with the principles of green analytical chemistry (GAC). In this chapter the topic of GAC is briefly summarized, whereas more detailed information can be found elsewhere.¹⁷⁹⁻¹⁸¹ Section 5.1 provides the most relevant background information, including an description of the twelve principles of GAC. In section 5.2 metric tools for evaluating a method's greenness are discussed with a focus on the approach that was used in this thesis. Finally, the outcome of evaluating the greenness of the methods developed within this thesis is presented in section 5.3.

5.1 Background

In 1995, the first preliminary proposal of GAC was published, entitled as “towards environmentally conscientious analytical chemistry through miniaturization, containment and reagent replacement.”¹⁸² However, this article remained widely unknown among scientists in the analytical community.¹⁸¹ Three years later, the term and concept of green chemistry was introduced, which focused on the reduction of negative impacts from chemical activities on the environment and human health.¹⁸³ In 2000, the term green analytical chemistry emerged,¹⁸⁴ replacing older denominations for the same concept such as environmentally methods, clean analytical chemistry and sustainable analytical chemistry. Since then, GAC has experienced exceptional growth.¹⁸¹ The increasing interest in this field is best illustrated by the number of published articles per year, which increased from 15 to 829 between 2000 and 2018.¹⁸⁰ As main reasons for this extraordinary success, three major categories of motivations have been named: ethical reasons, economic reasons and increased knowledge.¹⁸⁰ Ethical motivations are predominantly related to the increased awareness of the population towards environmental and climate related problems caused through human activities. Furthermore, the safety of operators in analytical laboratories is highly desirable, which is best achieved through replacement of hazardous chemicals with innocuous or less harmful compounds. Economical aspects are related to the fact that GAC methods are generally more cost-efficient than classical methods. Finally, the increase in available knowledge allows for new developments in material science, improved method automatization and novel strategies in waste degradation, which can be used to make analytical procedures greener.

In the beginning of the 21st century, mainly three different ways have been used to decrease the negative environmental impact of analytical methods: (i) reduction of solvent amounts required for sample pre-treatment, (ii) reduction of toxic solvents and reagents during measurements, particularly through automation and miniaturization and (iii) development of direct analytical methods that do not require solvents and reagents.¹⁸⁵ In order to allow translation to everyday work in analytical laboratories, however, more precise guidelines are required. In this context, twelve principles of GAC were proposed that nowadays represent the theoretical foundation for decreasing the environmental impact of analytical methods.¹⁸⁶

5.1.1 The twelve principles of green analytical chemistry (SIGNIFICANCE)

The twelve principles of green chemistry were formulated more than 20 years ago as a guideline on sustainable chemistry.¹⁸³ Even though these principles were mainly designed for the field of synthetic chemistry, some of them such as “avoid chemical derivatives” could be directly translated to analytical chemistry.¹⁸⁰ Other green chemistry principles such as “maximize atom economy”, however, are inadequate for GAC. Consequently, twelve principles tailored for GAC, consisting of several known concepts and some new ideas were proposed in 2013.¹⁸⁶ These principles that nowadays represent the base for making analytical methods more eco-friendly are as follows:^{186,187}

1. Direct analytical techniques should be applied to avoid sample treatment: By avoiding sample treatment and preparation steps, environmental, health and safety issues can be tremendously reduced.
2. Minimal sample size and minimal number of samples are goals: Except for direct analytical methods, which are not always applicable, every analysis begins with sampling. A reduction in the number of samples can be achieved in two major ways: (i) by using non-invasive methods and instruments and (ii) by application of statistics for selecting sampling sites. A reduction of the sample size can be achieved through application of miniaturized analytical methods.
3. In-situ measurements should be performed: The instrument should be placed close to the measurement location to reduce the time between sample collection and analysis, as well as the time between consecutive measurements.
4. Integration of analytical processes and operations saves energy and reduces the use of reagents: The number of unitary steps should be decreased for example by simultaneous performance of multiple steps to save material, energy and time.
5. Automated and miniaturized methods should be selected: Automation results in lower exposure to hazardous solvents and vapors, whereas miniaturized methods require less solvents, reagents, and energy.
6. Derivatization should be avoided: Derivatization is related to additional steps, (hazardous) chemicals and waste generation.
7. Generation of large volume of analytical waste should be avoided and proper management of analytical waste should be provided: Amounts of reagents and single-use devices should be minimized. When possible, reuse, recycling or online decontamination of waste should be applied.
8. Multi-analyte or multi-parameter methods are preferred versus methods using one analyte at a time: In cases where more than one analyte or parameter should be determined, multi-analyte/parameter methods are preferred over time- and resource-

- intensive single analyte/parameter methods.
9. The use of energy should be minimized: Energy consumed by analytical instruments and during sample preparation should be kept as low as possible.
 10. Reagents obtained from renewable source should be preferred: Chemicals derived from renewable resources are highly desirable.
 11. Toxic reagents should be eliminated or replaced: Elimination or replacement of toxic reagents, most of them being organic solvent, by greener alternatives should be applied whenever possible.
 12. The safety of the operator should be increased: The number of threats that are related to the employed chemicals (e.g., toxic, persistent, corrosive) should be reduced. Furthermore, automatization and miniaturization of the methods can reduce risks for operators.

In addition, the SIGNIFICANCE mnemonic was proposed as an alternative expression system for the twelve principles of GAC (see Table 8).¹⁸⁶

Table 8: The twelve principles of green analytical chemistry expressed as SIGNIFICANCE mnemonic.¹⁸⁶

Assigned Letter	Principle
S	Select direct analytical technique
I	Integrate analytical processes and operations
G	Generate as little waste as possible and treat it properly
N	Never waste energy
I	Implement automation and miniaturization of methods
F	Favour reagents obtained from renewable source
I	Increase safety for operator
C	Carry out <i>in-situ</i> measurements
A	Avoid derivatization
N	Note that the sample number and size should be minimal
C	Choose multi-analyte and multi-parameter method
E	Eliminate or replace toxic reagents

Besides the clear advantages of green analytical methods, it should be noted that one drawback is the need for a compromise between GAC principles and performance parameters. In this context, it has been outlined that certain principles such as replacement of toxic reagents or miniaturized sample size and instruments can negatively influence the accuracy, precision, selectivity and sensitivity of analytical methods.¹⁸⁶ However, it is expected that increased knowledge about existing problems and technological development will lead to major improvements in green analytical methods.

5.2 Green metric tools

To evaluate the greenness of analytical methods, metric tools are required. For this purpose, several approaches have been proposed, some of them being tailored for specific techniques (e.g. hazards related to HPLC methods), whereas others are more general and applicable to the majority of analytical procedures. In this work, the Analytical GREENess (AGREE) metric¹⁸⁷ was applied, thus the focus in section 5.2.1 is to provide an overview on this approach, whereas in section 5.2.2 some alternative assessment tools are briefly discussed.

5.2.1 The Analytical GREENess (AGREE) approach

The AGREE metric approach was introduced in 2020,¹⁸⁷ representing a relatively novel GAC assessment tool. Compared to previous approaches that only include a few criteria, the AGREE approach is the first metric that considers all twelve principles of GAC. During development, particular focus has been put onto the following four features: comprehensiveness of input, flexibility of input importance, simplicity of output and clarity of output. This is assured through a straight-forward, open-access software, where the user is guided through twelve tasks, representing the GAC principles. Each user input is then transformed into a 1-0 scale (i.e., best to worst). Furthermore, different weights can be assigned to each criterion based on the importance in the present scenario. Subsequently, the overall score is calculated by considering the input for all twelve principles and the assigned weights. Figure 18 depicts the final output, before and after a generic data input and assignment of weights. In these clock-like graphs, each GAC principle is represented by a segment with a green-yellow-red (i.e., best to worst) color range, whereas the weight of each principle is reflected by the width of the segment. The overall performance is expressed by the color and score in the middle. In this way, one can easily judge the accordance of the analytical method with individual GAC principles, as well as the overall environmental impact. A detailed overview on the assignment procedure for each principle is provided in Table 9.

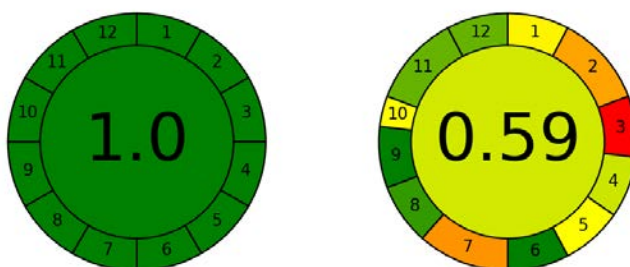


Figure 18: Output of evaluation according to AGREE approach (left) before and (right) after a generic data input and assignment of different weights.

In summary, the AGREE approach represents a straight-forward GAC assessment tool, with some of its main advantages being the coverage of all GAC principles, the user-friendly software and the clarity of the output. However, the lack of considering chemicals, energy and waste involved in the synthesis of the employed reagents and solvents, which can in some cases cause higher environmental impact than their application in the analytical procedure, was named as the major drawback.^{188,189}

Table 9: Calculation procedure for AGREE metric, based on the twelve principles of green analytical chemistry.¹⁸⁷

Principle	Option	Score
1	Remote sensing without sample damage	1.0
	Remote sensing with little physical damage	0.95
	Noninvasive analysis	0.90
	In-field sampling and on-line analysis	0.85
	In-field sampling and direct analysis	0.78
	On-line analysis	0.70
	At-line analysis	0.60
	Off-line analysis	0.48
	External sample treatment and batch analysis (reduced number of steps)	0.30
External sample treatment and batch analysis (large number of steps)	0.0	
2	Amount of sample in g or mL (ln)	$= -0.142 \times \ln + 0.65$
3	In-line positioning of analytical device	1.0
	On-line positioning of analytical device	0.66
	At-line positioning of analytical device	0.33
	Off-line positioning of analytical device	0.0
4	Three or fewer sample preparation steps	1.0
	Four sample preparation steps	0.80
	Five sample preparation steps	0.60
	Six sample preparation steps	0.40
	Seven sample preparation steps	0.20
	Eight or more sample preparation steps	0.0
5	Automatic, miniaturized	1.0
	Semi-automatic, miniaturized	0.75
	Manual, miniaturized	0.50
	Automatic, not miniaturized	0.50
	Semi-automatic, not miniaturized	0.25
	Manual, not miniaturized	0.0
6	No Derivatization applied	1.0
	Derivatization applied (score for particular derivatization agent = DA _i)	$= DA_1 \times DA_2 \times \dots \times DA_n - 0.2$
7	Amount of waste in g or mL (ln)	$= -0.134 \times \ln + 0.6946$
8	Number of analytes per hour (ln)	$= 0.2429 \times \ln - 0.0517$
9	<0.1 kWh per sample	1.0
	0.1-1.5 kWh per sample	0.5
	>1.5 kWh per sample	0.0
10	No reagents are used, or all of them are bio-based	1.0
	Some reagents are bio-based	0.5
	None of the reagents are bio-based	0.0
11	No toxic reagents or solvents are used	1.0
	Amount of toxic reagents or solvents in g or mL (ln)	$= -1.156 \times \ln + 0.5898$
12	Threats that cannot be avoided should be selected from the following list: toxic to aquatic life, bioaccumulative, persistent, highly flammable, highly oxidizable, explosive, corrosive.	
	All threats are avoided	1.0
	One threat present	0.8
	Two threats present	0.6
	Three threats present	0.4
	Four threats present	0.2
	Five or more threats present	0.0

5.2.2 Alternative metrics to assess greenness

A comprehensive overview on existing tools that can be used for assessing the greenness of analytical procedures can be found in a recent review.¹⁸⁸ Many of the available approaches, however, are either not widely used, only applicable for specific purposes or provide only qualitative output. Hence, only few approaches exist that can be considered as serious alternatives to the applied AGREE metric. In this context, particularly the analytical eco-scale should be mentioned, which was proposed in 2012 and since then has been the most widely used GAC assessment tool.¹⁹⁰ This quantitative approach is based on subtraction of penalty points from a total score of 100. In brief, the following three criteria have to be met for an ideal green method: (i) the solvent or reagents do not pose any physical environmental or health hazards, (ii) the energy consumption per sample is below 0.1 kWh and (iii) no waste is generated. One major advantage of the analytical eco-scale is the easy interpretability and comparability of the obtained result in form of a number between 0 and 100. Compared to AGREE, however, a noticeable drawback is the absence of information on the final output regarding steps of the analytical method that caused the penalty points. Furthermore, in this approach the number of hazard pictograms is simply multiplied, whereas the severity is not considered.¹⁸⁸ Later, the analytical eco-scale was extended by a green certificate evaluation, where the assessed method is assigned to a certain class between A (best) to G (worst).¹⁹¹ However, still no information on the nature of the threats is obtained from the final output certificate. In 2018, the green analytical procedure index (GAPI) was proposed as a new tool for determining the greenness of analytical procedures.¹⁹² The output of GAPI is a pictogram, consisting of five pentagrams, each of them referring to one of the following stages of the analytical procedure: sampling, type of method, sample preparation, amounts of solvents/reagents and related hazards, and consumed energy/waste generation. Some of these pentagrams are further subdivided into different segments, representing more detailed aspects of a certain step. The degree of environmental impact is indicated by either green, yellow or red color (i.e., best to worst) of the segments. Compared to earlier approaches, the main advantage of GAPI is the provided output which indicates the steps of the analytical procedure with high environmental impact. In analogy to the AGREE metric, the absence of considering the synthesis of applied chemicals into the assessment procedure was named as the major drawback.¹⁸⁸

5.3 Evaluation of the developed IR based methods according to AGREE metric

In this thesis, the AGREE approach was applied to compare the greenness of the novel mid-IR based methods for milk analysis and bioprocess monitoring with established reference methods. The thereby obtained results are briefly summarized in this section, whereas more detailed information can be found in Publication IV and Publication VIII.

For human milk fatty acid profiling, the developed method based on ATR-FTIR spectroscopy and solvent-free lipid separation was compared to the applied GC-MS reference method and a GC-MS method from literature.¹⁹³ The results of the AGREE evaluation are shown in Figure 19. Based on the total scores of 0.66 for FTIR, 0.33 for the GC-MS reference method, and 0.36 for the GC-MS method from literature, the novel mid-IR based approach shows clear superiority in terms of eco-friendliness.

Furthermore, the presented QCL-IR method for inline monitoring of proteins from SEC was compared to the applied HPLC offline reference method. Figure 20 shows overall scores of 0.84 and 0.43 for QCL-IR spectroscopy and HPLC, respectively, indicating clear advantages of the mid-IR based method regarding its accordance with the GAC principles.

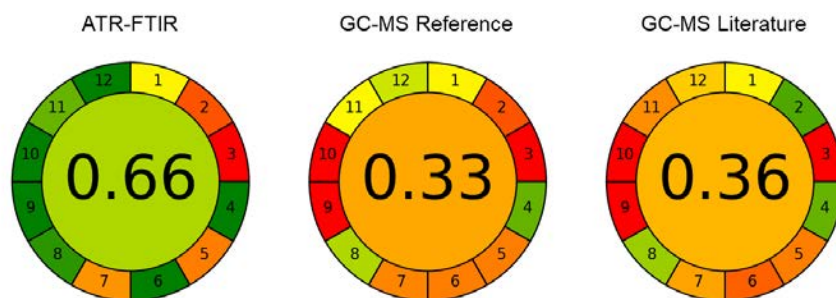


Figure 19: Results for evaluation of methods for human milk fatty acid profiling according to the AGREE approach for (left) the developed ATR-FTIR method, (middle) the applied GC-MS reference method and (right) a GC-MS method from literature.⁹

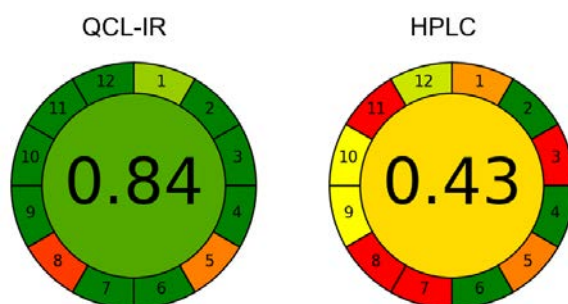


Figure 20: Results for evaluation of methods for monitoring of preparative size exclusion chromatography according to the AGREE approach for (left) inline QCL-IR spectroscopy and (right) offline HPLC analysis.¹³

Chapter 6 Results and introduction to the scientific publications

In this chapter the results obtained within the scientific publications that are part of this cumulative thesis are summarized. Full publications can be found in Appendix A. A list of contributions to the individual publications from low (x), to medium (xx) and high (xxx) is given in Table 10.

Table 10: Contributions to scientific publications from low (x), to medium (xx) and high (xxx).

	Publication							
	I	II	III	IV	V	VI	VII	VIII
Data acquisition	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx
Data analysis	xxx	xxx	xxx	xxx	xx	xxx	xxx	xx
Interpretation	xx	xxx	xxx	xx	x	xxx	xxx	xx
Drafting manuscript	xxx	xxx	xxx	xxx	x	xxx	xxx	xx

Publication I

The Next Generation of IR Spectroscopy: EC-QCL based Mid-IR Transmission Spectroscopy of Proteins with Balanced Detection

Christopher K. Akhgar, Georg Ramer, Mateusz Żbik, Artur Trajnerowicz, Jarosław Pawluczyk, Andreas Schwaighofer and Bernhard Lendl

published in Analytical Chemistry: <https://doi.org/10.1021/acs.analchem.0c01406>

In this publication, an EC-QCL based double-beam mid-IR transmission setup for the analysis of the protein amide I and amide II band in aqueous solutions is presented. By application of a MCT balanced detection module the noise level introduced by the laser light source was minimized. In this way, an LOD as low as 0.0025 mg/mL was achieved. Furthermore, characteristic spectral features of representative proteins with different secondary structures could be measured at concentrations of 0.1 mg mL⁻¹ within 45 s. Compared to high-end FTIR spectroscopy, approximately 8 times better sensitivity was achieved at similar acquisition times. Consequently, the presented setup combines high sensitivity, large optical path lengths as well as short measurement times and thus outperforms previous research type EC-QCL setups as well as commercially available instruments. This opens a wide range of future applications for qualitative and quantitative protein analysis in complex matrices such as human milk, where protein concentrations are below the range accessible by conventional FTIR spectroscopy.

Publication II

Fatty Acid Prediction in Bovine Milk by Attenuated Total Reflection Infrared Spectroscopy after Solvent-Free Lipid Separation

Christopher K. Akhgar, Vanessa Nürnberger, Marlene Nadvornik, Margit Velik, Andreas Schwaighofer, Erwin Rosenberg and Bernhard Lendl

published in Foods: <https://doi.org/10.3390/foods10051054>

In this publication, a novel approach for mid-IR based prediction of bovine milk fatty acid composition is presented, offering various advantages over well-established methods. This approach combines ATR-FTIR spectroscopy with a rapid, solvent-free two-step centrifugation method. ATR-FTIR spectra provided a significantly higher degree of fatty acid-related spectral features compared to conventional transmission spectra of whole milk. PLS regression models compiled of IR spectra with GC-MS reference data showed good prediction accuracies for fatty acid sum parameters as well as certain individual fatty acids. The information content of the calibration equations was validated by identifying the most important spectral features and assigning them to absorbance bands arising from fatty acids in milk. Moreover, the applied approach inherently eliminates covariation structures between individual fatty acids and total fat content, which so far posed a common problem in IR-based milk fat profiling. The demonstrated advantages indicate high potential of the presented method for accurate high-throughput fatty acid profiling in bovine milk.

Publication III

Fatty Acid Determination in Human Milk Using Attenuated Total Reflection Infrared Spectroscopy and Solvent-Free Lipid Separation.

Christopher K. Akhgar, Vanessa Nürnberger, Marlene Nadvornik, Victoria Ramos-Garcia, Isabel Ten-Doménech, Julia Kuligowski, Andreas Schwaighofer, Erwin Rosenberg and Bernhard Lendl

published in Applied Spectroscopy: <https://doi.org/10.1177/00037028211065502>

In this publication, the first approach for mid-IR based determination of the human milk fatty acid composition is presented. Here, the method based on ATR-FTIR spectroscopy and rapid, solvent-free two-step centrifugation was applied. ATR-IR spectra showed a higher degree of fatty acid-related spectral features compared to transmission spectra of whole human milk. PLS models were calculated by relating IR spectra to GC-MS reference concentrations, showing good results for the most important fatty acid sum parameters. Calibration equations were validated by identifying the most important wavenumber regions for the PLS models and assigning them to absorbance bands arising from human milk fatty acids. Based on a set of eight samples, the demonstrated results indicate high potential mid-IR based fatty acid determination in human milk.

Publication IV

Solvent-free Lipid Separation and Attenuated Total Reflectance Infrared Spectroscopy for Fast and Green Fatty Acid Profiling of Human Milk.

Christopher K. Akhgar, Victoria Ramos-Garcia, Vanessa Nürnberger, Isabel Ten-Doménech, Alba Moreno-Giménez, Julia Kuligowski, Erwin Rosenberg, Andreas Schwaighofer and Bernhard Lendl
Submitted to Food Chemistry.

In this work, first approach for mid-IR based simultaneous determination of individual fatty acid concentrations and the most important fatty acid sum parameters in human milk is introduced. The method presented in Publication III was applied to a large data set of 50 samples. PLS models were verified with an independent external validation set, showing excellent results. Based on these results, the effect of different clinical parameters on the human milk fatty acid composition was investigated, showing a change of certain sum parameters over the course of lactation, which agreed well with previous studies based on GC-MS. Evaluation of the method's greenness showed clear advantages in terms of environmental impact compared to established techniques. Consequently, the presented method bears high potential for high-throughput, green fatty acid determination in largescale clinical studies.

Publication V

Broadband laser-based mid-IR spectroscopy for analysis of proteins and monitoring of enzyme activity

Andreas Schwaighofer, [Christopher K. Akhgar](#) and Bernhard Lendl

published in Spectrochimica Acta A: Molecular and Biomolecular Spectroscopy:

<https://doi.org/10.1016/j.saa.2021.119563>

In this publication, the ChemDetect Analyzer, a commercially available EC-QCL-IR spectrometer was benchmarked against FTIR spectroscopy and applied for measuring proteins in water and monitoring of the lipase-catalyzed saponification of triacetin. The large optical path length (25 μm) and the broad tuning range of the laser (1350-1770 cm^{-1}) allowed for robust spectra acquisition beyond the protein amide I + II band. The obtained LOD compared well to research-grade FTIR instrumentation with a liquid nitrogen cooled MCT detector and was approximately 15 times better than for a routine FTIR spectrometer equipped with a pyrodetector. Absorbance spectra of model proteins with different secondary structures compared well in band shapes and positions with reference spectra at concentrations as low as 0.25 mg/mL. The monitored enzymatic reaction demonstrated the advantage of the broad spectral coverage for following complex chemical reactions. The obtained results in combination with the portability and small footprint of the ChemDetect Analyzer open a wide range of possible applications, including inline monitoring of bioprocesses.

Publication VI

Laser-based mid-infrared spectroscopy enables in-line detection of protein secondary structure from preparative liquid chromatography

[Christopher K. Akhgar](#), Julian Ebner, Oliver Spadiut, Andreas Schwaighofer and Bernhard Lendl published in Proceedings of SPIE: <https://doi.org/10.1117/12.2609419>

In this publication, the ChemDetect Analyzer was hyphenated to a LC system to enable inline monitoring of the protein secondary structure. The large optical path length allowed robust spectra acquisition in flow-through operation. A model system based on SEC was applied to demonstrate the advantages of QCL-IR detection. QCL-IR spectra that were recorded across the chromatographic run showed characteristic protein amide I and II bands with excellent agreement to reference offline spectra. Quantitative results compared well with a conventional UV/Vis detector in terms of chromatographic peak maxima and shapes. The reported results demonstrate significant advantages of QCL-IR detectors over conventional detectors with high potential for applying them in more complex studies.

Publication VII

QCL-IR Spectroscopy for In-Line Monitoring of Proteins from Preparative Ion-Exchange Chromatography

Christopher K. Akhgar, Julian Ebner, Oliver Spadiut, Andreas Schwaighofer and Bernhard Lendl published in Analytical Chemistry: <https://doi.org/10.1021/acs.analchem.1c05191>

In this publication, the ChemDetect Analyzer was successfully coupled to a preparative LC system based on IEX. Three different case studies for protein monitoring were employed to demonstrate the high flexibility of LC-QCL-IR hyphenation. A major challenge was caused by overlapping absorbance bands of proteins and the applied NaCl gradient. Here, a novel gradient compensation approach was introduced which can be used for correcting various gradient profiles by requiring only a single blank run. The thereby obtained protein spectra showed excellent agreement to offline reference spectra. Furthermore, concentration profiles obtained from mid-IR measurements agreed well with UV/Vis detection and HPLC offline measurements. Consequently, coupling of LC and EC-QCL based mid-IR spectroscopy holds high potential for complementing laborious and time-consuming offline methods for monitoring proteins in complex biotechnological processes.

Publication VIII

Application of Quantum Cascade Laser-Infrared Spectroscopy and Chemometrics for In-line Discrimination of Co-eluting Proteins from Preparative Size Exclusion Chromatography

Christopher K. Akhgar, Julian Ebner, Mirta Alcaraz, Julian Kopp, Héctor Goicoechea, Oliver Spadiut, Andreas Schwaighofer and Bernhard Lendl published in Analytical Chemistry: <https://doi.org/10.1021/acs.analchem.2c01542>

In this publication, the ChemDetect Analyzer was coupled to a preparative LC system based on SEC to monitor proteins from highly overlapping chromatographic peaks. Two case studies based on three and four different proteins, respectively were performed by using typical conditions employed in biopharmaceutical purification protocols. In order to distinguish co-eluting proteins, chemometric analysis based on SMMA and MCR was performed, providing qualitative as well as quantitative information. The thereby obtained protein spectra showed excellent comparability to offline reference spectra. Concentration profiles across the chromatographic run agreed well with offline HPLC measurements of the collected effluent fractions. The results demonstrate QCL-IR inline detection of proteins from LC for providing information that is typically only accessible by time- and labor-intensive offline methods.

Chapter 7 Conclusion and outlook

In this thesis, several significant contributions could be made to advance mid-IR spectroscopy for milk analysis and bioprocess monitoring. Compared to conventional methods applied within this field, the developed approaches do not require sample preparation steps based on organic solvents or chromatographic separation, thus offering more time-, labor- and cost-efficient analysis. Compared to previous mid-IR based methods, clear improvement was achieved in terms of sensitivity and/or accuracy.

For milk protein analysis, a novel laser-based transmission setup was developed, allowing robust analysis of the amide I and II band in aqueous solutions (see Publication I). The obtained LOD was almost an order of magnitude better than for high-end FTIR spectroscopy as well as for previous EC-QCL based academic setups. This improvement was achieved by combination of a latest-generation EC-QCL and a balanced detection module, comprising of two MCT detectors. In this way, noise levels could be reduced by a factor of up to 20 in balanced mode compared to single channel measurements. This enhanced performance allowed application of EC-QCL-IR spectroscopy for measuring aqueous protein solutions at concentration levels that were previously not accessible with mid-IR spectroscopy. In the future, the developed setup will be applied for measuring human milk samples, followed by chemometric evaluation of individual protein concentrations, as it was previously already demonstrated for bovine milk.³ Furthermore, the setup will be integrated into a portable prototype within the *NUTRISHIELD* project. In this way, rapid onsite analysis in hospitals of donor human milk protein compositions should be enabled which is currently not possible with commercially available IR based milk analyzers.

For milk fatty acid analysis, a novel approach based on ATR-FTIR spectroscopy and solvent-free lipid separation was introduced. Here, it was demonstrated that ATR-FTIR spectra show distinctly better resolved fatty acid related spectral features compared to transmission spectra of whole milk, highlighting the advantage of lipid separation prior to analysis. PLS-based multivariate calibration equations were first calculated for bovine milk by relating IR spectra to GC-MS reference concentrations, showing excellent results for the most important sum parameters and certain individual fatty acids (see Publication II). The same approach was then applied for a limited set of human milk samples (see Publication III), followed by a full validation with a larger set, showing particularly good results for SAT, MONO, PUFA, UNSAT, MCFA, LCFA, C12:0, C16:0, C18:1 *cis* and

C18:2 *cis* (see Publication IV). The information content of the calibration equations was successfully validated by relating important wavenumber regions for PLS calibrations to fatty acid related absorbances based on the SR. Furthermore, covariation structures between individual fatty acid concentrations and total fat content, a common problem in mid-IR based prediction of the milk fatty acid profile, were inherently eliminated.⁵ Subsequently, this method was applied to investigate the effect of different clinical parameters on the human milk fatty acid profile. The presented method has several clear advantages over conventional chromatographic methods as well as commercially available IR based milk analyzers, thus bearing high potential for high-throughput and accurate fatty acid profiling in human milk and bovine milk. In the future, the developed method could be applied in clinical studies for rapid determination of different variables on the human milk fatty acid profile, which was previously only possible by resource-intensive GC-MS methods.

For bioprocess monitoring, the commercially available ChemDetect Analyzer was employed to obtain near-real time information about protein secondary structure, which is typically only accessible by time- and labor-intensive offline methods. First, performance parameters of this EC-QCL based device were compared with FTIR spectroscopy, demonstrating a noticeably smaller footprint, approximately three times larger applicable optical path lengths and similar sensitivity without the need of liquid nitrogen cooling (see Publication V). Subsequently, these advantages were applied for inline monitoring of proteins from a preparative LC system (see Publication VI).¹¹ The thereby recorded spectra showed characteristic protein amide I + II bands and the calculated IR chromatograms agreed well with reference UV/Vis signals. Next, more complex runs based on IEX were monitored, where a major challenge was caused by highly overlapping absorbance bands of proteins and the applied NaCl gradient (see Publication VII). For this purpose, a novel background correction approach was presented, allowing for compensation of different gradient profiles by requiring only a single blank run. In this way, high-quality protein spectra were revealed, showing excellent agreement to reference offline spectra. Finally, SEC runs that showed highly overlapping protein peaks were monitored (see Publication VIII). As conventional LC detectors are not capable of discriminating the individual proteins, this information typically has to be obtained by offline analysis of the collected fractions. In contrast, QCL-IR spectroscopy in combination with chemometrics provided spectral profiles of pure proteins, as well as their concentrations across the chromatographic run. The thereby obtained qualitative and quantitative information agreed well with offline reference methods. Consequently, EC-QCL based mid-IR spectroscopy represents a highly flexible PAT tool for near real-time monitoring of proteins from LC purification processes. When combined with chemometrics, gradient compensation as well as discrimination of co-eluting peaks is possible. In the future, these advantages will be used to monitor further unit operations in downstream processing.

Finally, the methods for mid-IR based human milk fatty acid profiling and inline monitoring of proteins from LC effluents were evaluated in terms of accordance with the twelve principles of GAC (see Publication IV and Publication VIII). For this purpose, the developed approaches were juxtaposed with reference methods based on chromatographic steps by using the AGREE metric. The results for both mid-IR based methods indicated a clear superiority in terms of greenness of the analytical procedures. Particularly the higher sample throughput, lower energy consumption and absence of derivatization agents and other hazardous organic solvents represent major advantages of the methods developed within this work.

Bibliography

- (1) Schwaighofer, A.; Brandstetter, M.; Lendl, B. Quantum cascade lasers (QCLs) in biomedical spectroscopy. *Chem. Soc. Rev.* **2017**, *46*, 5903-5924.
- (2) Valentine, C. J.; Morrow, G.; Fernandez, S.; Gulati, P.; Bartholomew, D.; Long, D.; Welty, S. E.; Morrow, A. L.; Rogers, L. K. Docosahexaenoic Acid and Amino Acid Contents in Pasteurized Donor Milk are Low for Preterm Infants. *J. Pediatr.* **2010**, *157*, 906-910.
- (3) Montemurro, M.; Schwaighofer, A.; Schmidt, A.; Culzoni, M. J.; Mayer, H. K.; Lendl, B. High-throughput quantitation of bovine milk proteins and discrimination of commercial milk types by external cavity-quantum cascade laser spectroscopy and chemometrics. *Analyst* **2019**, *144*, 5571-5579.
- (4) Miliku, K.; Duan, Q. L.; Moraes, T. J.; Becker, A. B.; Mandhane, P. J.; Turvey, S. E.; Lefebvre, D. L.; Sears, M. R.; Subbarao, P.; Field, C. J.; Azad, M. B. Human milk fatty acid composition is associated with dietary, genetic, sociodemographic, and environmental factors in the CHILd Cohort Study. *Am. J. Clin. Nutr.* **2019**, *110*, 1370-1383.
- (5) Eskildsen, C.; Rasmussen, M.; Engelsen, S.; Larsen, L.; Poulsen, N. A.; Skov, T. Quantification of individual fatty acids in bovine milk by infrared spectroscopy and chemometrics: Understanding predictions of highly collinear reference variables. *J. Dairy Sci.* **2014**, *97*, 7940–7951.
- (6) Akhgar, C. K.; Ramer, G.; Žbik, M.; Trajnerowicz, A.; Pawluczyk, J.; Schwaighofer, A.; Lendl, B. The Next Generation of IR Spectroscopy: EC-QCL-Based Mid-IR Transmission Spectroscopy of Proteins with Balanced Detection. *Anal. Chem.* **2020**, *92*, 9901-9907.
- (7) Akhgar, C. K.; Nürnberger, V.; Nadvornik, M.; Velik, M.; Schwaighofer, A.; Rosenberg, E.; Lendl, B. Fatty Acid Prediction in Bovine Milk by Attenuated Total Reflection Infrared Spectroscopy after Solvent-Free Lipid Separation. *Foods* **2021**, *10*, 1054.
- (8) Akhgar, C. K.; Nürnberger, V.; Nadvornik, M.; Ramos-Garcia, V.; Ten-Doménech, I.; Kuligowski, J.; Schwaighofer, A.; Rosenberg, E.; Lendl, B. Fatty Acid Determination in Human Milk Using Attenuated Total Reflection Infrared Spectroscopy and Solvent-Free Lipid Separation. *Appl. Spec.* **2022**, *76*, 730-736.
- (9) Akhgar, C. K.; Ramos-Garcia, V.; Nürnberger, V.; Moreno-Giménez, A.; Kuligowski, J.; Rosenberg, E.; Schwaighofer, A.; Lendl, B. Solvent-free Lipid Separation and Attenuated Total Reflectance Infrared Spectroscopy for Fast and Green Fatty Acid Profiling of Human Milk. *Food Chem.* **Submitted**.
- (10) Schwaighofer, A.; Akhgar, C. K.; Lendl, B. Broadband laser-based mid-IR spectroscopy for analysis of proteins and monitoring of enzyme activity. *Spectrochim. Acta A* **2021**, *253*, 119563.
- (11) Akhgar, C.; Ebner, J.; Spadiut, O.; Schwaighofer, A.; Lendl, B. Laser-based mid-infrared spectroscopy enables in-line detection of protein secondary structure from preparative liquid chromatography. *Proc. Spie.* **2022**, 11957.
- (12) Akhgar, C. K.; Ebner, J.; Spadiut, O.; Schwaighofer, A.; Lendl, B. QCL-IR Spectroscopy for In-

- Line Monitoring of Proteins from Preparative Ion-Exchange Chromatography. *Anal. Chem.* **2022**, *94*, 5583-5590.
- (13) Akhgar, C. K.; Ebner, J.; Alcaraz, M. R.; Kopp, J.; Goicoechea, H.; Spadiut, O.; Schwaighofer, A.; Lendl, B. Application of Quantum Cascade Laser-Infrared Spectroscopy and Chemometrics for In-Line Discrimination of Coeluting Proteins from Preparative Size Exclusion Chromatography. *Anal. Chem.* **2022**, *94*, 11192-11200.
- (14) Chalmers, J. M.; Griffiths, P. R. *Handbook of Vibrational Spectroscopy*; John Wiley & Sons, 2002.
- (15) Griffiths, P. R.; De Haseth, J. A.; Winefordner, J. D. *Fourier Transform Infrared Spectrometry*; John Wiley & Sons, 2007.
- (16) Arrondo, J. L. R.; Muga, A.; Castresana, J.; Goñi, F. M. Quantitative studies of the structure of proteins in solution by fourier-transform infrared spectroscopy. *Prog. Biophys. Mol. Biol.* **1993**, *59*, 23-56.
- (17) Cooley, J. W.; Tukey, J. W. An algorithm for the machine calculation of complex Fourier series. *Math. Comput.* **1965**, *19*, 297-301.
- (18) Michelson, A. A. Visibility of Interference-Fringes in the Focus of a Telescope. *Publ. Astron. Soc. Pac.* **1891**, *3*, 217.
- (19) Perkins, W. D. Fourier transform-infrared spectroscopy: Part I. Instrumentation. *J. Chem. Educ.* **1986**, *63*, A5.
- (20) Lambert, J. H. *Photometria sived de mensura et gradibus luminis, colorum et umbrae*; V.E. Klett, 1760.
- (21) Beer, A. Bestimmung der Absorption des rothen Lichts in farbigen Flüssigkeiten. *Ann. Phys.* **1852**, *162*, 78-88.
- (22) Mayerhöfer, T. G.; Pahlow, S.; Hübner, U.; Popp, J. CaF₂: An Ideal Substrate Material for Infrared Spectroscopy? *Anal. Chem.* **2020**, *92*, 9024-9031.
- (23) Ramer, G.; Lendl, B. Attenuated Total Reflection Fourier Transform Infrared Spectroscopy. In *Encyclopedia of Analytical Chemistry*; John Wiley & Sons, 2013.
- (24) Harrick, N. J.; du Pré, F. K. Effective Thickness of Bulk Materials and of Thin Films for Internal Reflection Spectroscopy. *Appl. Opt.* **1966**, *5*, 1739-1743.
- (25) Chittur, K. K. FTIR/ATR for protein adsorption to biomaterial surfaces. *Biomaterials* **1998**, *19*, 357-369.
- (26) Harrick, N. J. Electric Field Strengths at Totally Reflecting Interfaces. *J. Opt. Soc. Am.* **1965**, *55*, 851-857.
- (27) Schwaighofer, A.; Lendl, B. Quantum Cascade Laser-based IR Transmission Spectroscopy of Proteins in Solution. In *Vibrational Spectroscopy in Protein Research*; Elsevier, 2020.
- (28) Elliott, A.; Ambrose, E. J. Structure of Synthetic Polypeptides. *Nature* **1950**, *165*, 921-922.
- (29) Barth, A. Infrared spectroscopy of proteins. *Biochim. Biophys. Acta Bioenerg.* **2007**, *1767*, 1073-1101.
- (30) Fabian, H.; Mäntele, W. Infrared Spectroscopy of Proteins. In *Handbook of Vibrational Spectroscopy*; John Wiley & Sons, 2006.
- (31) Murphy, B.; D'Antonio, J.; Manning, M.; Al-Azzam, W. Use Of The Amide II Infrared Band Of Proteins For Secondary Structure Determination and Comparability Of Higher Order Structure. *Curr. Pharm. Biotechnol.* **2014**, *15*, 880-889.
- (32) Levitt, M.; Greer, J. Automatic Identification of Secondary Structure in Globular Proteins. *J. Mol. Biol.* **1977**, *114*, 181-239.
- (33) Dousseau, F.; Pezolet, M. 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* **1990**, *29*, 8771-8779.
- (34) Monaco, H. L.; Zanotti, G.; Spadon, P.; Bolognesi, M.; Sawyer, L.; Eliopoulos, E. E. Crystal structure of the trigonal form of bovine beta-lactoglobulin and of its complex with retinol at 2.5 Å resolution. *J. Mol. Biol.* **1987**, *197*, 695-706.
- (35) Singh, B. R. 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.
- (36) Srour, B.; Brüchert, S.; Andrade, S.; Hellwig, P. Secondary Structure Determination by Means of ATR-FTIR Spectroscopy. In *Membrane Protein Structure and Function Characterization: Methods and Protocols*; Springer, 2017.
 - (37) Wilcox, K. E.; Blanch, E. W.; Doig, A. J. Determination of Protein Secondary Structure from Infrared Spectra Using Partial Least-Squares Regression. *Biochemistry* **2016**, *55*, 3794-3802.
 - (38) Shariati-Rad, M.; Hasani, M. Application of multivariate curve resolution-alternating least squares (MCR-ALS) for secondary structure resolving of proteins. *Biochimie* **2009**, *91*, 850-856.
 - (39) Navea, S.; Tauler, R.; Goormaghtigh, E.; De Juan, A. Chemometric Tools for Classification and Elucidation of Protein Secondary Structure from Infrared and Circular Dichroism Spectroscopic Measurements. *Proteins* **2006**, *63*, 527-541.
 - (40) Alcaraz, M. R.; Schwaighofer, A.; Goicoechea, H.; Lendl, B. Application of MCR-ALS to reveal intermediate conformations in the thermally induced alpha-beta transition of poly-L-lysine monitored by FT-IR spectroscopy. *Spectrochim. Acta A* **2017**, *185*, 304-309.
 - (41) Barth, A.; Zscherp, C. What vibrations tell us about proteins. *Q. Rev. Biophys.* **2002**, *35*, 369-430.
 - (42) Yang, H.; Yang, S.; Kong, J.; Dong, A.; Yu, S. Obtaining information about protein secondary structures in aqueous solution using Fourier transform IR spectroscopy. *Nat. Protoc.* **2015**, *10*, 382-396.
 - (43) Kong, J.; Yu, S. Fourier Transform Infrared Spectroscopic Analysis of Protein Secondary Structures. *Acta Biochim. Biophys. Sin.* **2007**, *39*, 549-559.
 - (44) Kazarinov, R. F.; Suris, R. A. Possible amplification of electromagnetic waves in a semiconductor with a superlattice. *Sov. phys. Semiconduct.* **1971**, *5*, 707-709.
 - (45) Faist, J.; Capasso, F.; Sivco, D.; Sirtori, C.; Hutchinson, A.; Cho, A. Quantum Cascade Laser. *Science* **1994**, *264*, 553-556.
 - (46) Siegman, A. E. *Lasers*; University Science Books, 1986.
 - (47) Sorokina, I.; Vodopyanov, K. *Solid-State Mid-Infrared Laser Sources*; Springer, 2003.
 - (48) Faist, J. *Quantum cascade lasers*; Oxford University Press, 2013.
 - (49) Pecharroman-Gallego, R. An Overview on Quantum Cascade Lasers: Origins and Development. In *Quantum Cascade Lasers*; IntechOpen, 2017.
 - (50) Meng, B.; Wang, Q. Broadly tunable single-mode mid-infrared quantum cascade lasers. *J. Opt.* **2015**, *17*, 023001.
 - (51) Kruczek, T.; Fedorova, K.; Sokolovskii, G.; Teissier, R.; Baranov, A.; Rafailov, E. U. InAs/AlSb widely tunable external cavity quantum cascade laser around 3.2 μm . *Appl. Phys. Lett.* **2013**, *102*.
 - (52) Fuchs, E.; Binder, E. M.; Heidler, D.; Krska, R. Structural characterization of metabolites after the microbial degradation of type A trichothecenes by the bacterial strain BBSH 797. *Food Addit. Contam.* **2002**, *19*, 379-386.
 - (53) Duxbury, G.; Langford, N.; McCulloch, M. T.; Wright, S. Quantum cascade semiconductor infrared and far-infrared lasers: from trace gas sensing to non-linear optics. *Chem. Soc. Rev.* **2005**, *34*, 921-934.
 - (54) Lewicki, R.; Witinski, M.; Li, B.; Wysocki, G. Spectroscopic benzene detection using a broadband monolithic DFB-QCL array. *Proc. Spie* **2016**, 9767.
 - (55) Brandstetter, M.; Koch, C.; Genner, A.; Lendl, B. Measures for optimizing pulsed EC-QC laser spectroscopy of liquids and application to multi-analyte blood analysis. *Proc. SPIE* **2014**, 8993, 89931 u.
 - (56) Brandstetter, M.; Genner, A.; Anic, K.; Lendl, B. Tunable external cavity quantum cascade laser for the simultaneous determination of glucose and lactate in aqueous phase. *Analyst* **2010**, *135*, 3260-3265.
 - (57) Venyaminov, S.; Prendergast, F. G. Water (H₂O and D₂O) molar absorptivity in the 1000-4000 cm⁻¹ range and quantitative infrared spectroscopy of aqueous solutions. *Anal. Biochem.* **1997**, *248*, 234-245.

- (58) Dabrowska, A.; David, M.; Freitag, S.; Andrews, A. M.; Strasser, G.; Hinkov, B.; Schwaighofer, A.; Lendl, B. Broadband laser-based mid-infrared spectroscopy employing a quantum cascade detector for milk protein analysis. *Sens. Actuators B Chem.* **2022**, *350*, 130873.
- (59) Alcaraz, M. R.; Schwaighofer, A.; Kristament, C.; Ramer, G.; Brandstetter, M.; Goicoechea, H.; Lendl, B. External-Cavity Quantum Cascade Laser Spectroscopy for Mid-IR Transmission Measurements of Proteins in Aqueous Solution. *Anal. Chem.* **2015**, *87*, 6980-6987.
- (60) Schwaighofer, A.; Montemurro, M.; Freitag, S.; Kristament, C.; Culzoni, M. J.; Lendl, B. 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.* **2018**, *90*, 7072-7079.
- (61) Schwaighofer, A.; Alcaraz, M. R.; Araman, C.; Goicoechea, H.; Lendl, B. External cavity-quantum cascade laser infrared spectroscopy for secondary structure analysis of proteins at low concentrations. *Sci. Rep.* **2016**, *6*.
- (62) Alcaraz, M. R.; Schwaighofer, A.; Goicoechea, H.; Lendl, B. 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-denatured alpha-chymotrypsin. *Anal. Bioanal. Chem.* **2016**, *408*, 3933-3941.
- (63) Kuligowski, J.; Schwaighofer, A.; Alcaraz, M. R.; Quintas, G.; Mayer, H.; Vento, M.; Lendl, B. External cavity-quantum cascade laser (EC-QCL) spectroscopy for protein analysis in bovine milk. *Anal. Chim. Acta* **2017**, *963*, 99-105.
- (64) Schwaighofer, A.; Kuligowski, J.; Quintas, G.; Mayer, H. K.; Lendl, B. Fast quantification of bovine milk proteins employing external cavity-quantum cascade laser spectroscopy. *Food Chem.* **2018**, *252*, 22-27.
- (65) Brandstetter, M.; Lendl, B. Tunable mid-infrared lasers in physical chemosensors towards the detection of physiologically relevant parameters in biofluids. *Sens. Actuators B Chem.* **2012**, *170*, 189-195.
- (66) Schwaighofer, A.; Alcaraz, M. R.; Lux, L.; Lendl, B. pH titration of beta-lactoglobulin monitored by laser-based Mid-IR transmission spectroscopy coupled to chemometric analysis. *Spectrochim. Acta, Part A* **2020**, *226*.
- (67) Ma, E.; Wang, L.; Kendrick, B. Enhanced Protein Structural Characterization Using Microfluidic Modulation Spectroscopy. *Spectroscopy* **2018**, *33*, 46-52.
- (68) Chon, B.; Xu, S.; Lee, Y. J. Compensation of Strong Water Absorption in Infrared Spectroscopy Reveals the Secondary Structure of Proteins in Dilute Solutions. *Anal. Chem.* **2021**, *93*, 2215-2225.
- (69) Lindner, S.; Hayden, J.; Schwaighofer, A.; Wolflehner, T.; Kristament, C.; Gonzales-Cabrera, M.; Zlabinger, S.; Lendl, B. External Cavity Quantum Cascade Laser-Based Mid-Infrared Dispersion Spectroscopy for Qualitative and Quantitative Analysis of Liquid-Phase Samples. *Appl. Spec.* **2019**, 000370281989264.
- (70) Dabrowska, A.; Schwaighofer, A.; Lindner, S.; Lendl, B. Mid-IR refractive index sensor for detecting proteins employing an external cavity quantum cascade laser-based Mach-Zehnder interferometer. *Opt. Express* **2020**, *28*, 36632-36642.
- (71) Lüdeke, S.; Pfeifer, M.; Fischer, P. Quantum-cascade laser-based vibrational circular dichroism. *J. Am. Chem. Soc.* **2011**, *133*, 5704-5707.
- (72) Hermann, D. R.; Ramer, G.; Kitzler-Zeiler, M.; Lendl, B. Quantum Cascade Laser-Based Vibrational Circular Dichroism Augmented by a Balanced Detection Scheme. *Anal. Chem.* **2022**, *94*, 10384-10390.
- (73) Rütter, A.; Pfeifer, M.; Lórenz-Fonfría, V. A.; Lüdeke, S. pH Titration Monitored by Quantum Cascade Laser-Based Vibrational Circular Dichroism. *J. Phys. Chem. B* **2014**, *118*, 3941-3949.
- (74) Van Winkel, M.; Vande Velde, S.; De Bruyne, R.; Van Biervliet, S. Clinical practice. *Eur. J. Pediatr.* **2011**, *170*, 1489-1494.
- (75) OECD/FAO. *OECD-FAO Agricultural Outlook 2020-2029*, 2020.
- (76) Oftedal, O. T. The evolution of milk secretion and its ancient origins. *Animal* **2012**, *6*, 355-368.

-
- (77) WHO; de Onis, M. Breastfeeding in the WHO Multicentre Growth Reference Study. *Acta Paediatr. Suppl.* **2006**, *450*, 16-26.
- (78) Chung, M.-Y. Factors Affecting Human Milk Composition. *Pediatr. Neonatol.* **2014**, *55*, 421-422.
- (79) Gidrewicz, D. A.; Fenton, T. R. A systematic review and meta-analysis of the nutrient content of preterm and term breast milk. *BMC pediatr.* **2014**, *14*, 216.
- (80) Boquien, C.-Y. Human Milk: An Ideal Food for Nutrition of Preterm Newborn. *Front Pediatr.* **2018**, *6*.
- (81) Ballard, O.; Morrow, A. L. Human milk composition: nutrients and bioactive factors. *Pediatr. Clin. N. Am.* **2013**, *60*, 49-74.
- (82) Jenness, R.; Wong, N. P.; Marth, E. H.; Keeney, M. *Fundamentals of Dairy Chemistry*; Springer, 1988.
- (83) Greer, F. R. Do breastfed infants need supplemental vitamins? *Pediatr. Clin. North Am.* **2001**, *48*, 415-423.
- (84) Dawodu, A.; Zalla, L.; Woo, J. G.; Herbers, P. M.; Davidson, B. S.; Heubi, J. E.; Morrow, A. L. Heightened attention to supplementation is needed to improve the vitamin D status of breastfeeding mothers and infants when sunshine exposure is restricted. *Matern. Child Nutr.* **2014**, *10*, 383-397.
- (85) Eigel, W. N.; Butler, J. E.; Ernstrom, C. A.; Farrell, H. M.; Harwalkar, V. R.; Jenness, R.; Whitney, R. M. Nomenclature of Proteins of Cow's Milk: Fifth Revision. *J. Dairy Sci.* **1984**, *67*, 1599-1631.
- (86) Kunz, C.; Lönnerdal, B. Human-milk proteins: analysis of casein and casein subunits by anion-exchange chromatography, gel electrophoresis, and specific staining methods. *Am. J. Clin. Nutr.* **1990**, *51*, 37-46.
- (87) Rangel, A.; Sales, D.; Urbano, S.; Galvao Junior, J. G.; Neto, J.; Macêdo, C. Lactose intolerance and cow's milk protein allergy. *Food Sci. Technol.* **2016**, *36*.
- (88) Kunz, C.; Lönnerdal, B. Human milk proteins: separation of whey proteins and their analysis by polyacrylamide gel electrophoresis, fast protein liquid chromatography (FPLC) gel filtration, and anion-exchange chromatography. *Am. J. Clin. Nutr.* **1989**, *49*, 464-470.
- (89) Calapaj, G. G. An electron microscope study of the ultrastructure of bovine and human casein micelles in fresh and acidified milk. *Int. J. Dairy Sci.* **2009**, *35*, 1-6.
- (90) Lönnerdal, B. O.; Atkinson, S. Nitrogenous Components of Milk: A. Human Milk Proteins. In *Handbook of Milk Composition*; Academic Press, 1995.
- (91) van den Oever, S. P.; Mayer, H. K. Analytical assessment of the intensity of heat treatment of milk and dairy products. *Int. Dairy J.* **2021**, *121*, 105097.
- (92) Jensen, R. G.; Newburg, D. S. Bovine Milk Lipids. In *Handbook of Milk Composition*; Academic Press, 1995.
- (93) Jensen, R. G.; Bitman, J.; Carlson, S. E.; Couch, S. C.; Hamosh, M.; Newburg, D. S. Milk Lipids: A. Human Milk Lipids. In *Handbook of Milk Composition*; Academic Press, 1995.
- (94) Koletzko, B. Human Milk Lipids. *Ann. Nutr. Metab.* **2016**, *69*, 27-40.
- (95) Koletzko, B.; Rodriguez-Palmero, M.; Demmelmair, H.; Fidler, N.; Jensen, R.; Sauerwald, T. Physiological aspects of human milk lipids. *Early Hum. Dev.* **2001**, *65*, 3-18.
- (96) Månsson, H. Fatty acids in bovine milk fat. *Food Nutr. Res.* **2008**, *52*, 1821.
- (97) Morera, S.; Castellote, A. I.; Jauregui, O.; Casals, I.; López-Sabater, M. C. Triacylglycerol markers of mature human milk. *Eur. J. Clin. Nutr.* **2003**, *57*, 1621-1626.
- (98) Koletzko, B.; Mrotzek, M.; Bremer, H. J. Fatty acid composition of mature human milk in Germany. *Am. J. Clin. Nutr.* **1988**, *47*, 954-959.
- (99) Mojska, H.; Socha, P.; Socha, J.; Soplińska, E.; Jaroszewska-Balicka, W.; Szponar, L. Trans fatty acids in human milk in Poland and their association with breastfeeding mothers' diets. *Acta Paediatr.* **2003**, *92*, 1381-1387.
- (100) Toldrá, F.; Nolle, L. M. L. *Handbook of Dairy Foods Analysis*; CRC Press, 2021.
- (101) Keller, R. P.; Neville, M. C. Determination of total protein in human milk: comparison of methods. *Clin. Chem.* **1986**, *32*, 120-123.
-

- (102) Donovan, S. M.; Lönnerdal, B. Development of a human milk protein standard. *Acta. Paediatr. Scand.* **1989**, *78*, 171-179.
- (103) Kjeldahl, J. Neue Methode zur Bestimmung des Stickstoffs in organischen Körpern. *Fres. J. Anal. Chem.* **1883**, *22*, 366-382.
- (104) ISO. Milk and milk products — Determination of nitrogen content — Part 4: Determination of protein and non-protein nitrogen content and true protein content calculation (Reference method). ISO 8968-4:2016, 2016.
- (105) Miller, E. M.; Aiello, M. O.; Fujita, M.; Hinde, K.; Milligan, L.; Quinn, E. A. Field and laboratory methods in human milk research. *Am. J. Hum. Biol.* **2013**, *25*, 1-11.
- (106) Ivens, K. O.; Baumert, J. L.; Taylor, S. L. Commercial Milk Enzyme-Linked Immunosorbent Assay (ELISA) Kit Reactivities to Purified Milk Proteins and Milk-Derived Ingredients. *J. Food Sci.* **2016**, *81*, T1871-1878.
- (107) Bordin, G.; Cordeiro Raposo, F.; de la Calle, B.; Rodriguez, A. R. Identification and quantification of major bovine milk proteins by liquid chromatography. *J. Chromatogr. A* **2001**, *928*, 63-76.
- (108) Strange, E. D.; Malin, E. L.; Van Hekken, D. L.; Basch, J. J. Chromatographic and electrophoretic methods used for analysis of milk proteins. *J. Chromatogr. A* **1992**, *624*, 81-102.
- (109) Mayer, H. K.; Raba, B.; Meier, J.; Schmid, A. RP-HPLC analysis of furosine and acid-soluble β -lactoglobulin to assess the heat load of extended shelf life milk samples in Austria. *Dairy Sci. Technol.* **2010**, *90*, 413-428.
- (110) Ferreira, I. Chromatographic Separation and Quantification of Major Human Milk Proteins. *J. Liq. Chromatogr. Relat.* **2007**, *30*, 499-507.
- (111) de Jong, N.; Visser, S.; Olieman, C. Determination of milk proteins by capillary electrophoresis. **1993**, *652*, 207-213.
- (112) ISO. *Milk and milk products - Extraction methods for lipids and liposoluble compounds*; ISO 1415:2001, 2001.
- (113) Lucas, A.; Gibbs, J. A.; Lyster, R. L.; Baum, J. D. Creamatocrit: simple clinical technique for estimating fat concentration and energy value of human milk. *Br. Med. J.* **1978**, *1*, 1018-1020.
- (114) Kang, J. X.; Wang, J. A simplified method for analysis of polyunsaturated fatty acids. *BMC Biochem.* **2005**, *6*, 5.
- (115) Zhang, S.; Wang, H.; Zhu, M.-J. A sensitive GC/MS detection method for analyzing microbial metabolites short chain fatty acids in fecal and serum samples. *Talanta* **2019**, *196*, 249-254.
- (116) Lepage, G.; Roy, C. C. Direct transesterification of all classes of lipids in a one-step reaction. *J. Lipid Res.* **1986**, *27*, 114-120.
- (117) Firl, N.; Kienberger, H.; Hauser, T.; Rychlik, M. Determination of the fatty acid profile of neutral lipids, free fatty acids and phospholipids in human plasma. *Clin. Chem. Lab. Med.* **2013**, *51*, 799-810.
- (118) Sutherland, K. Derivatization using m-(trifluoromethyl)phenyltrimethylammonium hydroxide of organic materials in artworks for analysis by gas chromatography-mass spectrometry: unusual reaction products with alcohols. *J. Chromatogr. A* **2007**, *1149*, 30-37.
- (119) Gries, P.; Rathore, A. S.; Lu, X.; Chiou, J.; Huynh, Y. B.; Lodi, A.; Tiziani, S. Automated Trimethyl Sulfonium Hydroxide Derivatization Method for High-Throughput Fatty Acid Profiling by Gas Chromatography-Mass Spectrometry. *Molecules* **2021**, *26*.
- (120) Kohler, A.; Afseth, N.; Jørgensen, K.; Randby, Å.; Martens, H. Quality Analysis of Milk by Vibrational Spectroscopy. In *Handbook of Vibrational Spectroscopy*; John Wiley & Sons, 2010.
- (121) El-Salam, M. H. A.; Al-Khamy, A. F.; El-Etriby, H. Evaluation of the Milkoscan 104 A/B for determination of milk fat, protein and lactose in milk of some mammals. *Food. Chem.* **1986**, *19*, 213-224.
- (122) De Marchi, M.; Toffanin, V.; Cassandro, M.; Penasa, M. Invited review: Mid-infrared spectroscopy as phenotyping tool for milk traits. *J. Dairy Sci.* **2014**, *97*, 1171-1186.
- (123) Di Marzo, L.; Cree, P.; Barbano, D. M. Prediction of fat globule particle size in homogenized milk using Fourier transform mid-infrared spectra. *J. Dairy Sci.* **2016**, *99*, 8549-8560.
- (124) Stone, M. Cross-Validatory Choice and Assessment of Statistical Predictions. *J. R. Stat. Soc.*,

- B: Stat. Methodol.* **1974**, *36*, 111-133.
- (125) Rutten, M. J. M.; Bovenhuis, H.; Heck, J. M. L.; Arendonk, J. Predicting bovine milk protein composition based on Fourier transform infrared spectra. *J. Dairy Sci.* **2011**, *94*, 5683-5690.
- (126) Wojciechowski, K. L.; Barbano, D. M. Prediction of fatty acid chain length and unsaturation of milk fat by mid-infrared milk analysis. *J. Dairy Sci.* **2016**, *99*, 8561-8570.
- (127) Giuffrida, F.; Austin, S.; Cuany, D.; Sanchez-Bridge, B.; Longet, K.; Bertschy, E.; Sauser, J.; Thakkar, S. K.; Lee, L. Y.; Affolter, M. Comparison of macronutrient content in human milk measured by mid-infrared human milk analyzer and reference methods. *J. Perinatol.* **2019**, *39*, 497-503.
- (128) Soyeurt, H.; Dardenne, P.; Dehareng, F.; Lognay, G.; Veselko, D.; Marlier, M.; Bertozzi, C.; Mayeres, P.; Gengler, N. Estimating Fatty Acid Content in Cow Milk Using Mid-Infrared Spectrometry. *J. Dairy Sci.* **2006**, *89*, 3690-3695.
- (129) Soyeurt, H.; Dehareng, F.; Mayeres, P.; Bertozzi, C.; Gengler, N. Variation of $\Delta 9$ -Desaturase Activity in Dairy Cattle. *J. Dairy Sci.* **2008**, *91*, 3211-3224.
- (130) Rutten, M. J. M.; Bovenhuis, H.; Hettinga, K.; Valenberg, H. J. F.; Arendonk, J. Predicting bovine milk fat composition using infrared spectroscopy based on milk samples collected in winter and summer. *J. Dairy Sci.* **2009**, *92*, 6202-6209.
- (131) De Marchi, M.; Penasa, M.; Cecchinato, A.; Mele, M.; Secchiari, P.; Bittante, G. Effectiveness of mid-infrared spectroscopy to predict fatty acid composition of Brown Swiss bovine milk. *Animal* **2011**, *5*, 1653-1658.
- (132) Ferrand, M.; Huquet, B.; Barbey, S.; Barillet, F.; Lahalle, F.; Larroque, H.; Leray, O.; Trommenschlager, J. M.; Brochard, M. Determination of fatty acid profile in cow's milk using mid-infrared spectrometry: Interest of applying a variable selection by genetic algorithms before a PLS regression. *Chemom. Intell. Lab. Syst.* **2011**, *106*, 183-189.
- (133) Eijndhoven, M. H. T.; Soyeurt, h.; Dehareng, F.; Calus, M. Validation of fatty acid predictions in milk using mid-infrared spectrometry across cattle breeds. *Animal* **2012**, *7*, 1-7.
- (134) Soyeurt, H.; Dehareng, F.; Gengler, N.; McParland, S.; Wall, E.; Berry, D.; Coffey, M.; Dardenne, P. Mid-infrared prediction of bovine milk fatty acids across multiple breeds, production systems, and countries. *J. Dairy Sci.* **2011**, *94*, 1657-1667.
- (135) Foss. *MilkoScan™ FT3 Brochure*. <https://www.fossanalytics.com/de-de/products/milkoscan-ft3>, accessed August 02 2022.
- (136) Schwaighofer, A.; Alcaráz, M. R.; Kuligowski, J.; Lendl, B. Recent advancements of EC-QCL based mid-IR transmission spectroscopy of proteins and application to analysis of bovine milk. *Biomed. Spectrosc. Imaging* **2018**, *7*, 35-45.
- (137) Hobbs, P. Shot noise limited optical measurements at baseband with noisy lasers. *Proc. SPIE* **1991**, *1376*.
- (138) Reyes-Reyes, A.; Hou, Z.; van Mastrigt, E.; Horsten, R. C.; de Jongste, J. C.; Pijnenburg, M. W.; Urbach, H. P.; Bhattacharya, N. Multicomponent gas analysis using broadband quantum cascade laser spectroscopy. *Opt. Express* **2014**, *22*, 18299-18309.
- (139) Waclawek, J. P.; Kristament, C.; Moser, H.; Lendl, B. Balanced-detection interferometric cavity-assisted photothermal spectroscopy. *Opt. Express* **2019**, *27*, 12183-12195.
- (140) Kübel, J.; Bucka, A.; Höpfner, J.; Zimmermann, H.; Godejohann, M.; Wilhelm, M. A New Quantum Cascade IR-Laser Online Detector: Chemical-Sensitive Size-Exclusion Chromatography Measurement at Unprecedented Low Levels. *Macromol. Rapid Commun.* **2019**, *40*, 1900228.
- (141) Freitag, S.; Baer, M.; Buntzoll, L.; Ramer, G.; Schwaighofer, A.; Schmauss, B.; Lendl, B. Polarimetric Balanced Detection: Background-Free Mid-IR Evanescent Field Laser Spectroscopy for Low-Noise, Long-term Stable Chemical Sensing. *ACS Sens.* **2021**, *6*, 35-42.
- (142) Afseth, N. K.; Martens, H.; Randby, Å.; Gidskehaug, L.; Narum, B.; Jørgensen, K.; Lien, S.; Kohler, A. Predicting the Fatty Acid Composition of Milk: A Comparison of Two Fourier Transform Infrared Sampling Techniques. *Appl. Spec* **2010**, *64*, 700-707.
- (143) Stefanov, I.; Vlaeminck, B.; Fievez, V. A novel procedure for routine milk fat extraction based on dichloromethane. *J. Food Compos. Anal.* **2010**, *23*, 852-855.

- (144) Feng, S.; Lock, A. L.; Garnsworthy, P. C. Technical Note: A Rapid Lipid Separation Method for Determining Fatty Acid Composition of Milk. *J. Dairy Sci.* **2004**, *87*, 3785-3788.
- (145) Luna, P.; Juárez, M.; Fuente, M. A. Validation of a Rapid Milk Fat Separation Method to Determine the Fatty Acid Profile by Gas Chromatography. *J. Dairy Sci.* **2005**, *88*, 3377-3381.
- (146) Deutscher, M. P. Chapter 5 Setting Up a Laboratory. In *Methods in Enzymology*; Academic Press, 2009.
- (147) Carta, G.; Jungbauer, A. *Protein Chromatography. Process Development and Scale-Up*; John Wiley & Sons, 2010.
- (148) LaCourse, W. R. HPLC Instrumentation. In *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering*; Elsevier, 2017.
- (149) Schulenberg-Schell, H.; Tei, A. *Principles in Preparative HPLC - A Primer*; Agilent Technologies, 2015.
- (150) Sun, Y.; Shi, Q. H.; Zhang, L.; Zhao, G. F.; Liu, F. F. 2.47 - Adsorption and Chromatography. In *Comprehensive Biotechnology (Second Edition)*; Academic Press, 2011.
- (151) Striegel, A. M. Chapter 10 - Size-exclusion chromatography. In *Liquid Chromatography (Second Edition)*; Elsevier, 2017.
- (152) Nesterenko, P. N.; Paull, B. Chapter 9 - Ion chromatography. In *Liquid Chromatography (Second Edition)*; Elsevier, 2017.
- (153) Kopp, J.; Zauner, F. B.; Pell, A.; Hausjell, J.; Humer, D.; Ebner, J.; Herwig, C.; Spadiut, O.; Slouka, C.; Pell, R. Development of a generic reversed-phase liquid chromatography method for protein quantification using analytical quality-by-design principles. *J. Pharm. Biomed. Anal.* **2020**, *188*, 113412.
- (154) Hage, D. S.; Anguizola, J. A.; Li, R.; Matsuda, R.; Papastavros, E.; Pfaunmiller, E.; Sobansky, M.; Zheng, X. Chapter 12 - Affinity chromatography. In *Liquid Chromatography (Second Edition)*; Elsevier, 2017.
- (155) Stoscheck, C. M. Quantitation of protein. In *Methods in Enzymology*; Academic Press, 1990.
- (156) Layne, E. C. Spectrophotometric and turbidimetric methods for measuring proteins. **1957**, *3*, 447-454.
- (157) Scopes, R. K. Measurement of protein by spectrophotometry at 205 nm. **1974**, *59*, 277-282.
- (158) Arakaki, L. S. L.; Burns, D. H.; Kushmerick, M. J. Accurate Myoglobin Oxygen Saturation by Optical Spectroscopy Measured in Blood-Perfused Rat Muscle. **2007**, *61*, 978-985.
- (159) Pramod, K.; Tahir, M. A.; Charoo, N. A.; Ansari, S. H.; Ali, J. Pharmaceutical product development: A quality by design approach. *Int. J. Pharm. Investig.* **2016**, *6*, 129-138.
- (160) FDA. *PAT - A Framework for Innovative Pharmaceutical Development, Manufacturing and Quality Assurance*, 2004.
- (161) Kuligowski, J.; Quintas, G.; Guardia, M.; Lendl, B. Liquid Chromatography—Liquid Chromatography—Fourier Transform Infrared. In *Encyclopedia of Analytical Science (Third Edition)*; Elsevier, 2019.
- (162) Turula, V.; Haseeth, J. Evaluation of Particle Beam Fourier Transform Infrared Spectrometry for the Analysis of Globular Proteins: Conformation of β -Lactoglobulin and Lysozyme. *Appl. Spec.* **1994**, *48*, 1255-1264.
- (163) Turula, V. E.; de Haseeth, J. A. Particle Beam LC/FT-IR Spectrometry Studies of Biopolymer Conformations in Reversed-Phase HPLC Separations: Native Globular Proteins. *Anal. Chem.* **1996**, *68*, 629-638.
- (164) Turula, V.; Bishop, R.; Ricker, R.; Haseeth, J. Complete structure elucidation of a globular protein by particle beam liquid chromatography - Fourier transform infrared spectrometry and electrospray liquid chromatography-mass spectrometry. Sequence and conformation of β -lactoglobulin. *J. Chromatogr. A* **1997**, *763*, 91-103.
- (165) Großhans, S.; Rüdts, M.; Sanden, A.; Brestrich, N.; Morgenstern, J.; Heissler, S.; Hubbuch, J. In-line Fourier-Transform Infrared Spectroscopy as a Versatile Process Analytical Technology for Preparative Protein Chromatography. *J. Chromatogr. A* **2018**, *1547*, 37-44.
- (166) Sanden, A.; Suhm, S.; Rüdts, M.; Hubbuch, J. 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* **2019**, *1608*, 460410.
- (167) Beskers, T. F.; Brandstetter, M.; Kuligowski, J.; Quintas, G.; Wilhelm, M.; Lendl, B. High performance liquid chromatography with mid-infrared detection based on a broadly tunable quantum cascade laser. *Analyst* **2014**, *139*, 2057-2064.
- (168) López, N.; Pernas, M. A.; Pastrana, L. M.; Sánchez, A.; Valero, F.; Rúa, M. L. Reactivity of pure *Candida rugosa* lipase isoenzymes (Lip1, Lip2, and Lip3) in aqueous and organic media. influence of the isoenzymatic profile on the lipase performance in organic media. *Biotechnology progress* **2004**, *20*, 65-73.
- (169) Armenta, S.; Tomischko, W.; Lendl, B. A Mid-Infrared Flow-Through Sensor for Label-Free Monitoring of Enzyme Inhibition. *Appl. Spectrosc.* **2008**, *62*, 1322-1325.
- (170) Quintás, G.; Lendl, B.; Garrigues, S.; de la Guardia, M. Univariate method for background correction in liquid chromatography–Fourier transform infrared spectrometry. *J. Chromatogr. A* **2008**, *1190*, 102-109.
- (171) Quintas, G.; Kuligowski, J.; Lendl, B. On-Line Fourier Transform Infrared Spectrometric Detection in Gradient Capillary Liquid Chromatography Using Nanoliter-Flow Cells. *Anal. Chem.* **2009**, *81*, 3746-3753.
- (172) Kuligowski, J.; Quintás, G.; Garrigues, S.; de la Guardia, M. Application of point-to-point matching algorithms for background correction in on-line liquid chromatography–Fourier transform infrared spectrometry (LC–FTIR). *Talanta* **2010**, *80*, 1771-1776.
- (173) Quintas, G.; Kuligowski, J.; Lendl, B. Procedure for automated background correction in flow systems with infrared spectroscopic detection and changing liquid-phase composition. *Appl. Spectrosc.* **2009**, *63*, 1363-1369.
- (174) Kuligowski, J.; Quintás, G.; Garrigues, S.; de la Guardia, M. Determination of critical eluent composition for polyethylenglycols using on-line liquid chromatography–Fourier transform infrared spectrometry. *Anal. Chim. Acta* **2008**, *624*, 278-285.
- (175) Kuligowski, J.; Quintás, G.; Garrigues, S.; de la Guardia, M. New background correction approach based on polynomial regressions for on-line liquid chromatography–Fourier transform infrared spectrometry. *J. Chromatogr. A* **2009**, *1216*, 3122-3130.
- (176) Kucheryavskiy, S.; Windig, W.; Bogomolov, A. Chapter 3 - Spectral Unmixing Using the Concept of Pure Variables. In *Data Handling in Science and Technology*; Elsevier, 2016.
- (177) Tauler, R. Multivariate curve resolution applied to second order data. *Chemom. Intell. Lab. Syst.* **1995**, *30*, 133-146.
- (178) Windig, W.; Guilment, J. Interactive self-modeling mixture analysis. *Anal. Chem.* **1991**, *63*, 1425-1432.
- (179) Armenta, S.; Garrigues, S.; de la Guardia, M.; Esteve-Turrillas, F. A. Green Analytical Chemistry. In *Encyclopedia of Analytical Science (Third Edition)*; Academic Press, 2019.
- (180) de la Guardia, M.; Garrigues, S. Chapter 1 Past, Present and Future of Green Analytical Chemistry. In *Challenges in Green Analytical Chemistry (2)*; The Royal Society of Chemistry, 2020.
- (181) Armenta, S.; Esteve-Turrillas, F. A.; Garrigues, S.; de la Guardia, M. 2.34 - Green Analytical Chemistry. In *Comprehensive Foodomics*; Elsevier, 2021.
- (182) de la Guardia, M.; Ruzicka, J. Towards environmentally conscientious analytical chemistry through miniaturization, containment and reagent replacement. *Analyst* **1995**, *120*, 17N-17N.
- (183) Anastas, P. T.; Warner, J. C. *Green chemistry: Theory and Practice*; Oxford University Press, 1998.
- (184) Namieśnik, J. Trends in Environmental Analytics and Monitoring. *Crit. Rev. Anal. Chem.* **2000**, *30*, 221-269.
- (185) Armenta, S.; Garrigues, S.; de la Guardia, M. Green Analytical Chemistry. *Trends Anal. Chem.* **2008**, *27*, 497-511.
- (186) Gałuszka, A.; Migaszewski, Z.; Namieśnik, J. The 12 principles of green analytical chemistry and the SIGNIFICANCE mnemonic of green analytical practices. *Trends Anal. Chem.* **2013**, *50*, 78-84.

- (187) Pena-Pereira, F.; Wojnowski, W.; Tobiszewski, M. AGREE—Analytical GREENness Metric Approach and Software. *Anal. Chem.* **2020**, *92*, 10076-10082.
- (188) Sajid, M.; Płotka-Wasyłka, J. Green analytical chemistry metrics: A review. *Talanta* **2022**, *238*, 123046.
- (189) Gutiérrez-Serpa, A.; González-Martín, R.; Sajid, M.; Pino, V. Greenness of magnetic nanomaterials in miniaturized extraction techniques: A review. *Talanta* **2021**, *225*, 122053.
- (190) Gałuszka, A.; Migaszewski, Z. M.; Konieczka, P.; Namieśnik, J. Analytical Eco-Scale for assessing the greenness of analytical procedures. *Trends Anal. Chem.* **2012**, *37*, 61-72.
- (191) Gallart-Mateu, D.; Cervera, M. L.; Armenta, S.; de la Guardia, M. The importance of incorporating a waste detoxification step in analytical methodologies. *Anal. Methods* **2015**, *7*, 5702-5706.
- (192) Płotka-Wasyłka, J. A new tool for the evaluation of the analytical procedure: Green Analytical Procedure Index. *Talanta* **2018**, *181*, 204-209.
- (193) Ten-Doménech, I.; Ramos-García, V.; Moreno-Torres, M.; Parra-Llorca, A.; Gormaz, M.; Vento, M.; Kuligowski, J.; Quintás, G. The effect of Holder pasteurization on the lipid and metabolite composition of human milk. *Food Chem.* **2022**, *384*, 132581.

Appendix

A. Scientific publications

Publication I

Akhgar, C. K.; Ramer, G.; Žbik, M.; Trajnerowicz, A.; Pawluczyk, J.; Schwaighofer, A.; Lendl, B. The Next Generation of IR Spectroscopy: EC-QCL-Based Mid-IR Transmission Spectroscopy of Proteins with Balanced Detection. *Anal. Chem.* **2020**, *92*, 9901-9907.

The Next Generation of IR Spectroscopy: EC-QCL-Based Mid-IR Transmission Spectroscopy of Proteins with Balanced Detection

Christopher K. Akhgar, Georg Ramer, Mateusz Żbik, Artur Trajnerowicz, Jarosław Pawluczyk, Andreas Schwaighofer,* and Bernhard Lendl*



Cite This: *Anal. Chem.* 2020, 92, 9901–9907



Read Online

ACCESS |



Metrics & More

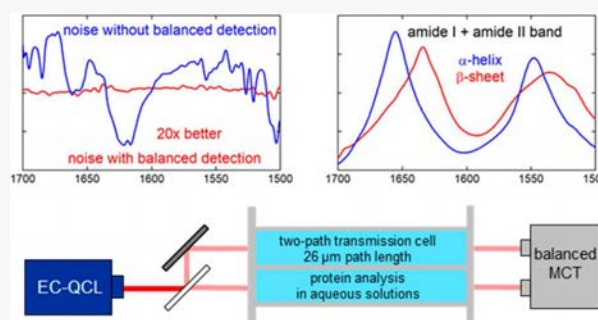


Article Recommendations



Supporting Information

ABSTRACT: We report a mid-IR transmission setup for the analysis of the protein amide I and amide II band in aqueous solutions that achieves a limit of detection as low as 0.0025 mg mL⁻¹ (outperforming our previous results and other state-of-the-art mid-IR-based techniques by almost an order of magnitude). This large improvement is made possible by combining the latest-generation external cavity-quantum cascade laser (EC-QCL) operated at room temperature with an optimized double-beam optical setup that adjusts the path length (26 μm) to ensure robust sample handling. For minimizing the noise introduced by the high-intensity laser light source, a thermoelectrically cooled mercury cadmium telluride balanced detection module was employed. In this way, noise levels better by a factor of up to 20 were achieved compared with single-channel measurements. Characteristic spectral features of proteins with different secondary structures were successfully identified at concentrations as low as 0.1 mg mL⁻¹. Furthermore, a highly linear response was demonstrated for concentrations between 0.05 and 10 mg mL⁻¹. The total acquisition time of the setup can be adapted to fulfill the required sensitivity of the protein measurements and to ensure maximum flexibility for future applications. The presented setup combines high sensitivity, large optical path lengths, and short measurement times and thus outperforms previous research type EC-QCL setups as well as commercially available instruments. This opens a wide range of future applications including protein–ligand interaction studies as well as qualitative and quantitative analyses of proteins in complex matrices such as those found in up- and downstream bioprocess monitoring and similar challenging applications which can not be readily met by conventional FT-IR spectroscopy.



The mid-infrared (mid-IR) region (4000–400 cm⁻¹) provides detailed information about the specific rotational–vibrational transitions of molecules. Polypeptides and proteins show characteristic absorption bands in the IR spectrum that are referred to as amide bands. The most prominent band for secondary structure analysis is the amide I band (1700–1600 cm⁻¹).¹ Additionally, the amide II band (1600–1500 cm⁻¹) has been recognized to be sensitive to secondary structure.² The position and shape of the amide bands show sensitivity to secondary structure due to changes in the interaction of the C=O and N–H groups of the amide bond induced by different arrangements of hydrogen bonding, dipole–dipole interactions, and geometric orientations in the α -helices, β -sheets, turns, and random coil structures.³ For quantification of the individual secondary structure components, curve fitting⁴ or chemometric approaches such as partial least-squares⁵ or multivariate curve resolutions-alternating least-squares (MCR-ALS)⁶ can be applied to the amide I region. However, it has been shown that additional and more comprehensive information can be gained by combined analysis of the amide I and II bands, particularly with chemometric analysis.^{7–9}

Fourier transform infrared (FT-IR) spectroscopy, the most widespread type of mid-IR spectroscopy is a powerful and well-established method for quantitative and qualitative analysis of proteins.^{10,11} FT-IR instruments are commonly equipped with thermal light sources such as Globars that emit a broadband but low-power and constant radiation. For samples with highly absorbing matrix components, the low emission intensity of the light source can lead to limitations of the measurements. A pronounced challenge in mid-IR spectroscopy of proteins is the strong absorption of the HOH-bending band of water near 1643 cm⁻¹ that overlaps with the amide I region of proteins. For conventional methods, the optical path length is restricted to <10 μm, in order to avoid total IR absorption.^{1,12} These low path lengths lead to impaired sensitivity and limited robustness

Received: April 1, 2020

Accepted: June 29, 2020

Published: June 29, 2020



because of the low-intensity IR bands and likely clogging of the cell when measuring a sequence of samples.

More than two decades ago, quantum cascade lasers (QCLs) were introduced as a polarized, coherent and high power light source in the mid-IR region.¹³ They combine stable operation at room temperature and increased spectral power densities (SPDs) by a factor of 10^4 and more compared with thermal light sources.¹⁴ In combination with an external diffraction grating for tuning in a so-called external cavity (EC) design, EC-QCLs offer broadband spectral tuning in excess of several hundred wavenumbers, making them favorable for the analysis of liquid samples. It has been shown that the high available emission power of QCLs enable mid-IR transmission analysis of proteins at an optical path length 4–5 times larger than with conventional FT-IR spectroscopy.^{15–23}

While the high power of EC-QCLs is an advantage over FT-IR spectroscopy, they suffer from noise that FT-IR spectroscopy is immune to, including (1) pulse-to-pulse intensity fluctuations around 2% and (2) wavelength reproducibility issues due to their wavelength tuning mechanism. Furthermore, EC-QCLs have an SPD that is defined by the wavelength-dependent gain of the QC gain material. Hence, EC-QCLs can easily saturate a detector in one part of the spectrum while dropping below the noise limit at another. For FT-IR spectrometers, wavelength stability is given because of Connes' advantage, and the noise level of the thermal light sources can usually be neglected compared with the detector noise. High-intensity laser light sources however contribute to the overall noise level, especially when operated in pulsed mode, because of pulse-to-pulse intensity fluctuations.^{24,25}

The challenge in building a high-sensitivity EC-QCL spectroscopy setup is thus to make the most of the advantages of the source while at the same time removing or handling its current disadvantages.

For example, inconsistent wavelength tuning behavior of the first-generation EC-QCL was overcome by a data processing routine based on correlation optimized warping (COW) and filtering to correct shifts in the fine structure throughout the spectrum.¹⁵ For second generation EC-QCLs, optimized motion mechanics and triggering routines of the external cavity were introduced. Moreover, a higher tuning range and increased scan speed enabled protein measurements in the amide I and amide II region at lower concentrations within shorter measurement times. Significant improvements were made; nevertheless, a mathematical routine was still needed to identify shifted scans which had to be removed prior to signal averaging for achieving best possible performance. Furthermore, the remaining challenges of a highly uneven SPD and large pulse-to-pulse fluctuations were also limiting the potential of EC-QCL transmission-absorption spectroscopy.¹⁹ Here we present an optical design that overcomes some of these challenges to perform EC-QCL protein spectroscopy at unprecedented sensitivity.

The uneven SPD is counteracted by a set of optical attenuators that are matched to the sample matrix absorption and the emission spectrum of the QCL. To suppress laser intensity noise (thermal drifts and $1/f$), balanced detection was proposed by Hobbs.²⁶ Here, the laser beam is split into two beams of equal intensities called the signal and reference beam. Only one of the beams interacts with the sample, but both are detected on matched detectors. The resulting voltages are subtracted from each other to yield the "balanced output". Any change in intensity that afflicts both beams is thus suppressed

in the balanced output. Previously, balanced detection was successfully applied for the reduction of the noise introduced by different types of QCL light sources for gas-phase analysis^{27,28} and to increase the sensitivity of an EC-QCL-based detector for size-exclusion chromatography (SEC).²⁹ In that last implementation, a low dead volume fiber optic-flow cell was used,^{30,31} and balancing was achieved by attenuation of the reference beam by means of a variable aperture to adjust for uneven SPD of the employed EC-QCL.

We demonstrate highly sensitive protein spectroscopy in H_2O using an optimized EC-QCL liquid spectroscopy setup based on improved balanced detection and adjusted SPD. The performance of the system is characterized and benchmarked against previous EC-QCL-based designs showing that it outperforms not only previous EC-QCL setups but also high-end FT-IR and commercial laser-based systems.

■ EXPERIMENTAL SECTION

Reagents and Samples. Hemoglobin from bovine blood, concanavalin A from jack bean, and γ -globulins from bovine blood ($\geq 97\%$) were purchased from Sigma-Aldrich (Steinheim, Germany). Proper amounts of protein powder were dissolved in water. Ultrapure water (resistivity: 18 M Ω cm) was purified with a Milli-Q system from Millipore (Bedford, MA).

Balanced Detection Module. The balanced detection module developed by Vigo (Vigo System S.A., Poland), uses two individual thermoelectrically cooled mid-IR mercury cadmium telluride (MCT) detectors, operated at -73 °C with a 0.5×0.5 mm element size and a detectivity of $D^* = 1.59 \times 10^{10}/1.36 \times 10^{10}$ cm Hz^{0.5} W⁻¹ at $8 \mu\text{m}$ at 200 kHz. These two detectors are precisely matched regarding their transimpedance, voltage responsivity, and detectivity to achieve a high common mode rejection ratio (CMRR). The working principle and electronic circuit are presented in Figure S1 in the Supporting Information. In order to compensate for dark currents (in the order of tens of μA) that are present in MCT structures, manual compensation subcircuits were implemented. The detectors are connected to two identical amplifier channels (signal and reference). Each channel consists of a transimpedance amplifier followed by a fixed gain voltage amplifier. External instrumentation is connected via buffered outputs. To generate "balanced output", signal and reference channels are combined in the subtractor block, where an analogue subtractor generates the difference signal (reference channel–sample channel). To achieve high attenuation of common mode changes in the optical signal, the received power in the signal and reference beam should be made as similar as possible ($I_{\text{signal}}/I_{\text{reference}} \approx 1:1$). Any power difference in the optical signals (differential mode signal) is present at the balanced output, whereas unwanted common mode signals are attenuated.

Experimental Setup for QCL-IR Measurements. Figure 1 shows the schematic of the experimental laser-based setup. The applied EC-QCL (Hedgehog, Daylight Solutions Inc., San Diego U.S.A.) was operated with a laser current of 650 mA, a repetition rate of 1 MHz and a pulse width of 200 ns (duty cycle of 20%). The operating temperature of the water-cooled laser head was set to 19 °C. Spectra were recorded between 1730 and 1470 cm^{-1} , using a scan speed of 3600 $\text{cm}^{-1} \text{s}^{-1}$. The QCL emission spectrum was amended with optical components in order to avoid detector saturation in certain spectral regions (see Adaptation of Laser Emission and Detector Characteristics for details). A wire mesh was used to reduce the

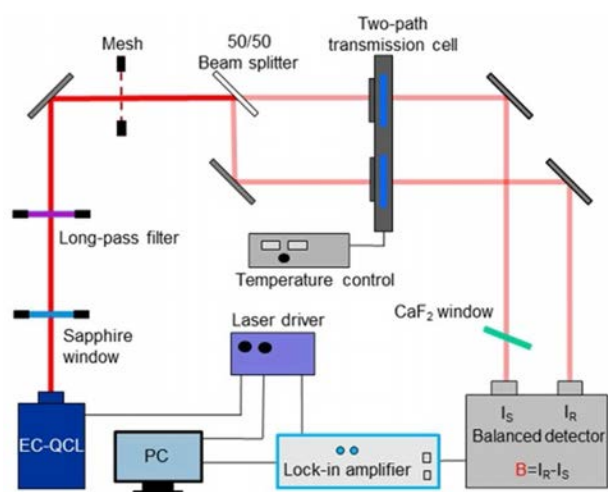


Figure 1. Schematic of the two-path QCL-based mid-IR transmission setup for protein measurements in aqueous solutions.

IR-light intensity equally across the entire scan range. In order to adjust the maximum laser intensity to the dynamic range of the detector,¹⁹ a short-pass filter (wedged sapphire window, 2.5 mm thickness) and a silicon-based long-pass filter were employed to selectively lower the laser intensity in the amide II region and in the high wavenumber range of the amide I band, respectively, leading to a pulsed average power of approximately 5.4 mW at 1650 cm^{-1} . Then, the laser light was divided into two beams by a nominal 50:50 beam splitter (Thorlabs BSW510). The two beams were directed into a custom-built two-path CaF_2 transmission flow cell with optical paths of 26 μm . The temperature of the transmission cell was set to 20 $^\circ\text{C}$ by using a thermoelectric-cooling temperature controller (Meerstetter Engineering GmbH, Rubigen, Switzerland). The thermoelectrically cooled balanced MCT detector module was used to detect the light intensities of the two beams. A CaF_2 window was positioned in front of one of the detecting elements to specifically attenuate the intensity of one laser beam to match the signal intensities on the two detector elements, as outlined below (see **Balanced Detection** for

Reduction of Spectral Noise for details). The optical setup was encased with an acrylic glass housing and continuously flushed with dry air to decrease the influence of water vapor from the atmosphere.

The measured output signals (signal, reference, balanced) were processed and digitized by a lock-in amplifier MFLI (Zurich Instruments, Zurich, Switzerland) with an extended cutoff range of 5 MHz and up to four demodulators (FSM and MD addons). The lock-in demodulator frequency was referenced to the laser repetition rate. As the balanced signal can have positive and negative amplitude, the in-phase (X) signal was recorded. Single-beam spectra were recorded during a tuning time of approximately 72 ms and consisted of 5401 data points. For protein measurements, 300 scans were averaged, leading to a total acquisition time of 45 s. Disposable syringes were used to manually inject approximately 500 μL of solution into the flow cell. Spectral resolutions of 0.5 and 2.6 cm^{-1} were determined for nonfiltered and filtered spectra, by comparing the bandwidth of water vapor spectra to FT-IR spectra acquired at different resolutions. The wavenumber scale of the measured spectra was obtained by calibration against the absorption bands of water vapor. A graphical user interface provided by Daylight Solutions was used to control the laser. Data acquisition was performed using in-house developed Python script (Python 3.7) to access the MFLI API.

Processing of QCL Data. A detailed description of the data preprocessing routine can be found elsewhere.¹⁹ Briefly, the performed preprocessing steps included smoothing, similarity index evaluation, scan averaging, fast Fourier transform (FFT) filtering and calculation of the final absorbance spectra. For the presented setup, scans with a similarity index below 0.94 were removed. This value showed optimal results for balanced detection single channel spectra. An FFT filter with a cutoff frequency of 1500 Hz was applied to the single beam spectra. Before recording a measurement run an initial single channel signal ($I_{\text{ref},0}$) was recorded and stored. This single channel was then added to both background ($I_{\text{bal,bg}}$) and measurement ($I_{\text{bal,smp}}$) balance signal before calculating absorption:

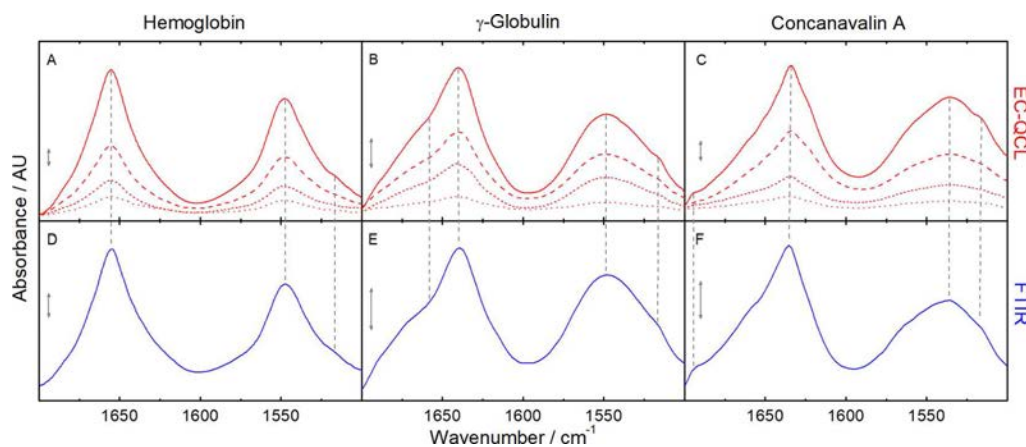


Figure 2. Mid-IR absorbance spectra of 1 (red solid line), 0.5 (red dashed line), 0.25 (red short dashed line), and 0.1 mg mL^{-1} (red dotted line) protein solutions recorded with the EC-QCL setup (A–C) and 20 mg mL^{-1} (blue solid line) protein solutions obtained by high-end FT-IR spectroscopy (D–F). Gray dashed lines highlight the high congruence of the spectral features between the mid-IR spectra acquired by EC-QCL and FT-IR spectroscopy. Gray double-headed arrows indicate the absorbance of 1 mAU for QCL-IR and 10 mAU for FT-IR spectra.

$$A = \log \left(\frac{I_{\text{ref},0} + I_{\text{bal,bg}}}{I_{\text{ref},0} + I_{\text{bal,smp}}} \right) \quad (1)$$

To compare the protein absorbance spectra acquired by FT-IR spectroscopy (s_1) and EC-QCL (s_2) quantitatively, the degree of spectral overlap (s_{12}) was calculated by using the following expression:³²

$$s_{12} = \frac{\|s_1^T s_2\|}{\|s_1\| \|s_2\|} \quad (2)$$

s_{12} can range from 0 (no overlap) to 1 (complete overlap), respectively. Data processing and analysis was carried out with tailored in-house developed code in MatLab R2017a (MathWorks, Inc., Natick, MA 2017).

FT-IR Measurements. A Bruker Vertex 80v FT-IR spectrometer (Ettlingen, Germany) equipped with a Globar (power levels of $<15 \mu\text{W}$ for 1 cm^{-1} spectral range²⁵) and a liquid nitrogen cooled MCT detector ($D^* = 4 \times 10^{10} \text{ cm Hz}^{0.5} \text{ W}^{-1}$ at $9.2 \mu\text{m}$) was used to perform all FT-IR measurements. The samples were manually injected into a flow cell, equipped with two CaF_2 windows and an $8 \mu\text{m}$ -thick spacer. During measurements, the spectrometer was constantly flushed with dry air for at least 10 min prior to data acquisition until water vapor absorption was sufficiently constant. Measurements were performed with a spectral resolution of 2.6 cm^{-1} in double-sided acquisition mode. A Blackman-Harris 3-term apodization function and a zero-filling factor of 2 were used to calculate the final spectra. In total, 266 scans were averaged per spectrum, leading to an acquisition time of 45 s. All measurements were performed at $25 \text{ }^\circ\text{C}$. The software package OPUS 7.2 (Bruker, Ettlingen, Germany) was used for evaluation of spectral data. Water vapor absorption bands were subtracted, if required.

RESULTS AND DISCUSSION

Mid-IR Spectra Recorded with the EC-QCL Setup. IR transmission spectra of proteins with different concentrations in aqueous solutions were recorded by using the developed custom-made EC-QCL setup. The absorbance spectra across the amide I and amide II range were calculated by averaging a total number of 300 scans, enabling qualitative as well as quantitative evaluation. The investigated proteins were selected to exhibit different secondary structures revealing diverse spectral features in the mid-IR spectrum (Figure 2A–C). Hemoglobin (Hemo) is mainly composed of α -helical secondary structure^{33,34} and shows the characteristic band maximum for α -helices at 1656 cm^{-1} in the amide I region as well as a narrow band at approximately 1545 cm^{-1} in the amide II region.^{35,36} γ -Globulin (γ -G1) primarily contains β -sheet structures resulting in an IR band maximum at 1640 cm^{-1} in the amide I region and a broad band with a maximum at approximately 1550 cm^{-1} in the amide II region.³⁷ Concanavalin A (ConA) is predominantly composed of two antiparallel β -sheets³⁸ and displays a distinctive IR spectrum with a band maximum at 1636 cm^{-1} and a sideband at 1694 cm^{-1} in the amide I region as well as a broad band with a maximum at 1536 cm^{-1} in the amide II region.^{7,39,40} To highlight the high-quality spectra obtainable at low concentration, Figure 3 depicts an overlay of the QCL-IR spectra recorded at a protein concentration of 0.1 mg mL^{-1} . The amide I and amide II bands can be clearly distinguished from the noise level. Even though the spectra at such low

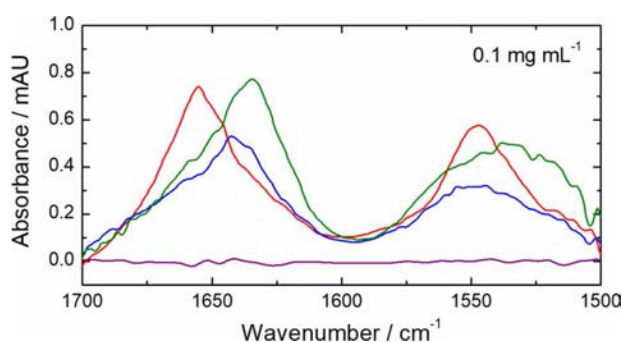


Figure 3. QCL-IR absorbance spectra of hemoglobin (red line), γ -globulin (blue line), and concanavalin A (green line) at concentration levels of 0.1 mg mL^{-1} . For reference, a 100% line of water indicates the noise level (purple line).

concentration levels are not completely smooth, the characteristic band shapes of the individual proteins can be clearly identified.

The spectra obtained in the laser-based setup were compared with FT-IR measurements. Figure 2D–F shows the FT-IR spectra of 20 mg mL^{-1} protein solutions measured at a path length of $8 \mu\text{m}$. Evaluation of the shape and position of the absorbance bands show high comparability between the EC-QCL setup and FT-IR spectroscopy. By using eq 2, the s_{12} values obtained for Hemo, ConA and γ -G1 were 0.99666, 0.99271, and 0.99202, respectively, demonstrating excellent comparability of the spectra obtained with the laser-based IR setup and conventional FT-IR spectroscopy.

For evaluation of the capabilities of the developed setup for quantitation purposes, the height of the band maxima in the amide I region was evaluated for protein solutions with eight different concentrations ranging between 0.05 and 10 mg mL^{-1} . The calibration curves in Figure S2 in the Supporting Information demonstrate high linearity ($r^2 > 0.999$) down to concentrations of 0.05 mg mL^{-1} . Hence, the validity of Lambert–Beer’s law is demonstrated for the method and allows application of the developed setup for quantitative analysis.

Adaptation of Laser Emission and Detector Characteristics. To achieve sensitive measurements in the laser setup, the SPD on the detector needs to be well within the dynamic range of the MCT detector across the full spectral range of the setup and most similar in both arms of the balancing scheme.

Intense absorption of the HOH-bending band of water near 1640 cm^{-1} leads to substantial SPD differences across the spectral range reaching the detector and exceed its dynamic range. Consequently, the path length of the transmission cell and the laser current need to be carefully optimized. Because of the principles of QCL emission, reduction of the emission power by decreasing the laser current was only possible to a certain degree, because the laser current affects both the emission power as well as the spectral emission range. Consequently, additional optical components were introduced to adjust the power level. Figure S3 in the Supporting Information shows the emission power before and after attenuation with the employed optical elements as well as the corresponding transmission after attenuation. First, a wire mesh was used to reduce the laser light intensity equally across the entire spectral emission range. Second, to attenuate the SPD of the EC-QCL in the amide II region and at the high

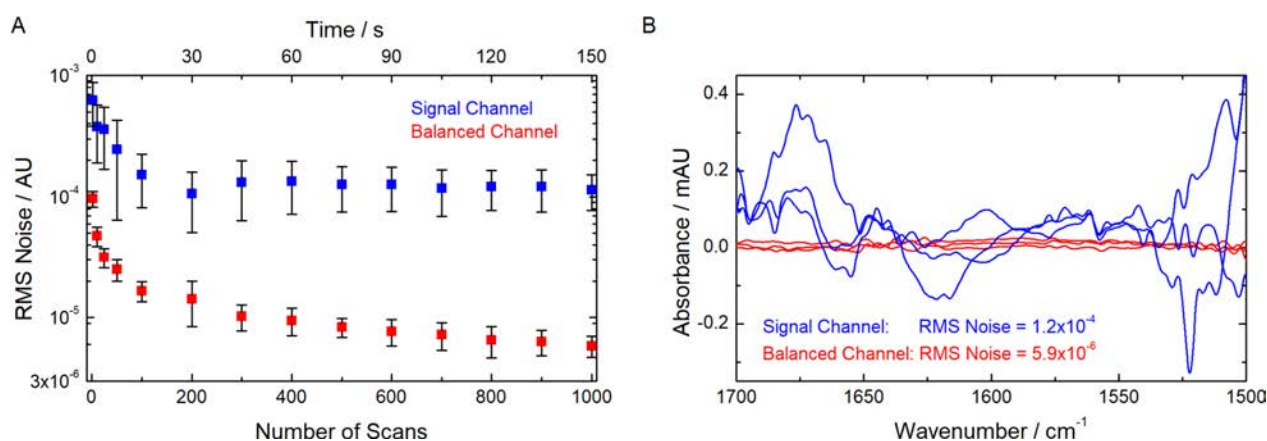


Figure 4. Comparison of the noise level in the signal channel and in the balanced channel of the EC-QCL setup. (A) RMS noise level (average of ten measurements) as a function of time and the total number of scans used for averaging. (B) 100% lines of water obtained by averaging 1000 scan.

Table 1. Performance Parameters for Comparison of EC-QCL-Based IR Transmission Setups and High-End FT-IR Spectroscopy

	meas. time [s]/scans	RMS-noise 10^{-5} [AU]	path length [μm]	LOD [mg/mL]	spectral range [cm^{-1}]	detector type/temp. [$^{\circ}\text{C}$]/cooling
QCL setup 1st gen.	250/100	37	38	0.117	1700–1600	MCT/−60/TE
QCL setup 2nd gen.	53/100	6.2	31	0.025	1700–1500	MCT/−78/TE
QCL balanced setup	45/300	1.0	26	0.0043	1700–1500	MCT/−73/TE
high-end FT-IR	45/266	2.3	8	0.035	4000–600	MCT/−196/LN ₂

wavenumber end of the amide I band, but to preserve high emission powers near the region of high IR absorption of water, wavelength selective filters were employed. A wedged sapphire window was used as a short-wave pass filter to reduce the laser intensity in the amide II region. For reducing the laser intensity at the high wavenumber side of the amide I band, a silicon-based long-wave pass filter was applied. This attenuation of the beam combined with laser current optimization and an optimized path length of 26 μm led to SPD across the entire amide I and amide II region that was well within the dynamic range of the detector.

Balanced Detection for Reduction of Spectral Noise.

In the presented EC-QCL setup, a novel balanced MCT module was employed. This detector provides individual outputs for the intensity levels of the two beam paths as well as the balanced channel, which corresponds to the difference between the two channels. This balanced detection approach enables to reduce the intensity noise inherent in the laser emission spectrum. To achieve maximum noise suppression in the balanced channel, the received power in the two channels should be as equal as possible. As opposed to the work of Kübel et al.,²⁹ who used an aperture to adjust the intensity of the reference beam, we used two almost identical flow cells, filled with sample and solvent, for achieving the most similar intensities across the whole tuning range for both channels. Variations in the splitting ratio of the beam splitter across the tuning range of the laser as well as differences in the optical path length of both transmission cells require additional fine adjustment of one of the two beam paths to fulfill this criterion. For this purpose, a CaF₂ window was used to specifically attenuate the light intensity in one of the two beams. For typical signals (reference, signal, balanced), see Figure S4 in the Supporting Information.

The noise level of the EC-QCL setup was determined using 100% lines from the absorbance spectra of two consecutively recorded measurements with water in both transmission cells. Ideally, the spectrum would reveal a flat line at zero absorbance corresponding to 100% transmittance.⁴¹ The RMS (root-mean-square) noise was calculated between 1700 and 1600 cm^{-1} . Figure 4A shows the average noise level as well as the standard deviation of 10 repeated measurements in the signal channel and in the balanced channel as a function of total number of averaged scans. For the signal channel, the lowest noise level of $\sim 1.2 \times 10^{-4}$ AU was reached by averaging 200 scans. Here, a higher number of scans did not improve the achievable noise level. In contrast, the RMS noise evaluated from balanced channel further decreased with a higher number of scans. Moreover, the variation among repeated measurements is distinctively lower in the balanced channel as indicated by the lower standard deviation (error bars). Finally, by averaging 1000 scans, an average RMS noise of 5.9×10^{-6} AU was reached, which is approximately 20 times lower than the noise level in the signal channel (see Figure 4B). This thorough characterization of the noise level enables high flexibility for future applications. The acquisition time can be modified in accordance to the required sensitivity. While an acquisition time of 150 s can be used, when maximum sensitivity is required, much shorter measurement times are feasible for solutions of high protein content.

Comparison of Setup Performance to FT-IR Spectroscopy and Previous EC-QCL Setups. The noise level of the laser-based balanced detection setup was evaluated and compared with a commercially available high-end FT-IR spectrometer as well as with custom-built IR transmission setups employing first-generation¹⁵ and state-of-the-art EC-QCLs¹⁹ without balanced detection modules. Performance parameters are summarized in Table 1. To evaluate the noise

floor, the RMS noise of 100% lines of water was calculated in the 1700–1600 cm^{-1} region. Mechanical imperfections and triggering issues of the first-generation EC-QCL introduced considerable noise levels in the calculated absorbance spectra. Therefore, COW, a rubberband like alignment method, was used to compensate these flaws, leading to an RMS noise of 26.5×10^{-5} AU at significantly longer measurement times due to limited laser sweep rates. For the second-generation EC-QCL setup, an adapted data processing approach was employed. Here, evaluation by similarity index was used to remove shifted outlier scans among otherwise highly overlapping scans. Compared with the first-generation setup, a SNR improved by a factor of ~ 5 could be achieved within a significantly lower measurement time. For the newly developed balanced detection setup, the measurement time was even further reduced by combining the high sweep rates of the latest generation EC-QCL with a state-of-the-art lock-in amplifier. With this optimized signal-processing scheme, approximately 3.5 times more scans could be recorded during the same acquisition time than in the previously reported setup.¹⁹

For the herein presented setup, an RMS noise of 1.0×10^{-5} AU was achieved at an acquisition time of 45 s using a path length of 26 μm . Compared with the previous EC-QCL setups, the slightly lower applicable path length resulted from the trade-off between the path length and the dynamic range of the detector. While the employed data processing routine was the same as for the second-generation setup, the obtained RMS noise was approximately 6.2 times lower within a slightly shorter measurement time. Furthermore, the performance was compared to an FT-IR spectrometer equipped with a fast interferometer and a liquid nitrogen-cooled MCT detector. Previous QCL setups could not reach the low noise levels of this high-end instrument.

To compare the performance, the same spectral resolution and acquisition time were set for recording 100% lines of water with the laser-based in FT-IR spectrometer. A path length of 8 μm was employed for FT-IR measurements. The final noise level accomplished with the EC-QCL setup was approximately 2.3 times lower than with the high-end FT-IR spectrometer.

Additionally, the limit of detection (LOD) was determined by using

$$\text{LOD} = \frac{3 \times \text{noise}}{\text{slope of the calibration function}} \quad (3)$$

Here, not only the RMS noise is considered but also the linear calibration function calculated from the height of the band maxima of protein solutions with different concentration levels. According to Lambert–Beer's law, a higher signal intensity is reached when larger optical path lengths are applied for transmission measurements. Thus, the short path length of 8 μm that has to be used in FT-IR spectroscopy does not only considerably reduce the robustness of the measurement but also impairs its sensitivity. Hence, the superiority of the laser-based setup is even more pronounced in terms of the LOD than in the noise level. When comparing at the same acquisition time of 45 s, the LOD of the FT-IR measurements was 8 times higher than that of the laser-based balanced detection setup. Finally, the LOD of the laser-based setup can be further decreased to $0.0025 \text{ mg mL}^{-1}$ by increasing the measurement time to 150 s. For the same increase in measurement time, the LOD for FT-IR spectroscopy decreases by approximately the same factor to 0.022 mg mL^{-1} .

CONCLUSIONS AND OUTLOOK

In this work, a new mid-IR transmission setup was introduced for recording protein spectra in the amide I and II region by using an EC-QCL and an MCT balanced detection module. Thorough characterization of the performance revealed noise levels better by a factor of up to 20 in balanced detection mode compared with single-channel measurements. This indicates that the noise introduced by the high-intensity laser light source can be successfully compensated with the balanced detection module. An improved data acquisition scheme allowed recording a high number of scans during short measurement times. The total acquisition time can be adapted in order to fulfill the required sensitivity, allowing high flexibility for IR measurements of proteins. By using an acquisition time of 45 s, characteristic spectral features of proteins with different secondary structures were successfully measured at concentrations as low as 0.1 mg mL^{-1} . Furthermore, quantitative as well as qualitative application of the setup for protein solutions with significantly lower concentration levels is possible when exact secondary structure determination is not required. Compared with high-end FT-IR spectroscopy, the achieved LOD was approximately 8 times lower at similar measurement times. The approach also compares well to other novel, commercially available QCL-based IR spectroscopy systems, such as the AQS3pro (RedShiftBio), which provides a similar performance in terms of LOD and sensitivity to secondary structure at approximately 10-fold measurement time and a reduced spectral coverage (amide I band only).⁴²

The enhanced performance of the presented setup allows application of mid-IR spectroscopy for the analysis of proteins in aqueous solutions at previously prohibitively low concentration levels. Furthermore, the large optical path length of our laser-based approach also allows robust sample handling. Within the covered spectral region, the presented setup outperforms previous EC-QCL-based setups as well as high-end FT-IR spectroscopy and marks the next generation of mid-IR spectroscopy of proteins in aqueous solutions. In the future, this setup will be employed for chemometrics-based protein quantification in complex matrices (e.g., milk) and analysis of bioprocess products, where the protein concentration at certain process steps are below the range accessible by conventional FT-IR spectroscopy.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.analchem.0c01406>.

Working principle of the balanced detection module, calibration curves of proteins, laser emission curve with and without attenuation, single beam spectra of reference, signal and balanced channel (PDF)

AUTHOR INFORMATION

Corresponding Authors

Bernhard Lendl – *Institute of Chemical Technologies and Analytics, Technische Universität Wien, 1060 Vienna, Austria;*
orcid.org/0000-0003-3838-5842; Email: bernhard.lendl@tuwien.ac.at

Andreas Schwaighofer – *Institute of Chemical Technologies and Analytics, Technische Universität Wien, 1060 Vienna,*

Austria; orcid.org/0000-0003-2714-7056;
Email: andreas.schwaighofer@tuwien.ac.at

Authors

Christopher K. Akhgar – Institute of Chemical Technologies and Analytics, Technische Universität Wien, 1060 Vienna, Austria; orcid.org/0000-0001-8266-043X

Georg Ramer – Institute of Chemical Technologies and Analytics, Technische Universität Wien, 1060 Vienna, Austria; orcid.org/0000-0001-8307-5435

Mateusz Żbik – Vigo System S.A., 05-850 Ożarów, Mazowiecki, Poland

Artur Trajnerowicz – Vigo System S.A., 05-850 Ożarów, Mazowiecki, Poland

Jarosław Pawluczyk – Vigo System S.A., 05-850 Ożarów, Mazowiecki, Poland

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.analchem.0c01406>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work has received funding by the European Union's Horizon 2020 research and innovation programme through the NUTRISHIELD project under grant agreement no. 818110. A.S. gratefully acknowledges financial support by the Austrian Science Fund FWF (project no. P32644-N).

REFERENCES



- (1) Fabian, H.; Mäntele, W. Infrared Spectroscopy of Proteins. In *Handbook of Vibrational Spectroscopy*; Griffiths, P., Chalmers, J. M., Eds.; John Wiley & Sons, Ltd.: Hoboken, NJ, 2006.
- (2) Murphy, B.; D'Antonio, J.; Manning, M.; Al-Azzam, W. *Curr. Pharm. Biotechnol.* **2014**, *15*, 880–889.
- (3) Singh, B. R. Basic Aspects of the Technique and Applications of Infrared Spectroscopy of Peptides and Proteins. *Infrared Analysis of Peptides and Proteins*; ACS Symposium Series; American Chemical Society: Washington, DC, 1999; pp 2–37.
- (4) Srour, B.; Brüchert, S.; Andrade, S.; Hellwig, P. Secondary Structure Determination by Means of ATR-FTIR Spectroscopy. In *Membrane Protein Structure and Function Characterization: Methods and Protocols*; Lacapere, J.-J., Ed.; Springer New York: New York, NY, 2017; pp 195–203.
- (5) Wilcox, K. E.; Blanch, E. W.; Doig, A. J. *Biochemistry* **2016**, *55*, 3794–3802.
- (6) Shariati-Rad, M.; Hasani, M. *Biochimie* **2009**, *91*, 850–856.
- (7) Dousseau, F.; Pezolet, M. *Biochemistry* **1990**, *29*, 8771–8779.
- (8) Navea, S.; Tauler, R.; Goormaghtigh, E.; De Juan, A. *Proteins: Struct., Funct., Genet.* **2006**, *63*, 527–541.
- (9) Alcaraz, M. R.; Schwaighofer, A.; Goicoechea, H.; Lendl, B. *Spectrochim. Acta, Part A* **2017**, *185*, 304–309.
- (10) Cooper, E. A.; Knutson, K. Fourier Transform Infrared Spectroscopy Investigations of Protein Structure. In *Physical Methods to Characterize Pharmaceutical Proteins*; Herron, J. N., Jiskoot, W., Crommelin, D. J. A., Eds.; Springer US: Boston, MA, 1995; pp 101–143.
- (11) Kong, J.; Yu, S. *Acta Biochim. Biophys. Sin.* **2007**, *39*, 549–559.
- (12) Yang, H.; Yang, S.; Kong, J.; Dong, A.; Yu, S. *Nat. Protoc.* **2015**, *10*, 382–396.
- (13) Faist, J.; Capasso, F.; Sivco, D.; Sirtori, C.; Hutchinson, A.; Cho, A. *Science* **1994**, *264*, 553–556.
- (14) Schwaighofer, A.; Brandstetter, M.; Lendl, B. *Chem. Soc. Rev.* **2017**, *46*, 5903–5924.
- (15) Alcaraz, M. R.; Schwaighofer, A.; Kristament, C.; Ramer, G.; Brandstetter, M.; Goicoechea, H.; Lendl, B. *Anal. Chem.* **2015**, *87*, 6980–6987.
- (16) Alcaraz, M. R.; Schwaighofer, A.; Goicoechea, H.; Lendl, B. *Anal. Bioanal. Chem.* **2016**, *408*, 3933–3941.
- (17) Schwaighofer, A.; Alcaraz, M. R.; Araman, C.; Goicoechea, H.; Lendl, B. External cavity-quantum cascade laser infrared spectroscopy for secondary structure analysis of proteins at low concentrations. *Sci. Rep.* **2016**, *6*.
- (18) Kuligowski, J.; Schwaighofer, A.; Alcaraz, M. R.; Quintas, G.; Mayer, H.; Vento, M.; Lendl, B. *Anal. Chim. Acta* **2017**, *963*, 99–105.
- (19) Schwaighofer, A.; Montemurro, M.; Freitag, S.; Kristament, C.; Culzoni, M. J.; Lendl, B. *Anal. Chem.* **2018**, *90*, 7072–7079.
- (20) Schwaighofer, A.; Kuligowski, J.; Quintas, G.; Mayer, H. K.; Lendl, B. *Food Chem.* **2018**, *252*, 22–27.
- (21) Montemurro, M.; Schwaighofer, A.; Schmidt, A.; Culzoni, M. J.; Mayer, H. K.; Lendl, B. *Analyst* **2019**, *144*, 5571–5579.
- (22) Schwaighofer, A.; Alcaraz, M. R.; Lux, L.; Lendl, B. pH titration of beta-lactoglobulin monitored by laser-based Mid-IR transmission spectroscopy coupled to chemometric analysis. *Spectrochim. Acta, Part A* **2020**, *226*, 117636.
- (23) Schwaighofer, A.; Lendl, B. Chapter 3 - Quantum cascade laser-based infrared transmission spectroscopy of proteins in solution. In *Vibrational Spectroscopy in Protein Research*; Academic Press, 2020; pp 59–88.
- (24) Hayden, J.; Hugger, S.; Fuchs, F.; Lendl, B. A quantum cascade laser-based Mach-Zehnder interferometer for chemical sensing employing molecular absorption and dispersion. *Appl. Phys. B: Lasers Opt.* **2018**, *124*, 29.
- (25) Brandstetter, M.; Koch, C.; Genner, A.; Lendl, B. *Proc. SPIE* **2014**, *8993*, 89931U.
- (26) Hobbs, P. Shot noise limited optical measurements at baseband with noisy lasers (proceedings Only). *Proc. SPIE* **1990**, *1376*, Laser Noise.216.
- (27) Waclawek, J. P.; Kristament, C.; Moser, H.; Lendl, B. *Opt. Express* **2019**, *27*, 12183–12195.
- (28) Reyes-Reyes, A.; Hou, Z.; van Mastrigt, E.; Horsten, R. C.; de Jongste, J. C.; Pijnenburg, M. W.; Urbach, H. P.; Bhattacharya, N. *Opt. Express* **2014**, *22*, 18299–18309.
- (29) Kübel, J.; Botha, C.; Bucka, A.; Höpfner, J.; Zimmermann, H.; Godejohann, M.; Wilhelm, M. *Macromol. Rapid Commun.* **2019**, *40*, 1900228.
- (30) Lendl, B.; Frank, J.; Schindler, R.; Müller, A.; Beck, M.; Faist, J. *Anal. Chem.* **2000**, *72*, 1645–1648.
- (31) Edelmann, A.; Ruzicka, C.; Frank, J.; Lendl, B.; Schrenk, W.; Gornik, E.; Strasser, G. *J. Chromatogr. A* **2001**, *934*, 123–128.
- (32) Culzoni, M. J.; Goicoechea, H. C.; Ibanez, G. A.; Lozano, V. A.; Marsili, N. R.; Olivieri, A. C.; Pagani, A. P. *Anal. Chim. Acta* **2008**, *614*, 46–57.
- (33) Levitt, M.; Greer, J. *J. Mol. Biol.* **1977**, *114*, 181–239.
- (34) Perutz, M. F.; Rossmann, M. G.; Cullis, A. F.; Muirhead, H.; Will, G.; North, A. C. T. *Nature* **1960**, *185*, 416–422.
- (35) Barth, A. *Biochim. Biophys. Acta, Bioenerg.* **2007**, *1767*, 1073–1101.
- (36) Pan, T.; Peng, A.; Huang, W. *Appl. Mech. Mater.* **2011**, *55–57*, 1168–1171.
- (37) van de Weert, M.; Haris, P. I.; Hennink, W. E.; Crommelin, D. J. A. *Anal. Biochem.* **2001**, *297*, 160–169.
- (38) Becker, J. W.; Reeke, G. N., Jr; Wang, J. L.; Cunningham, B. A.; Edelman, G. M. *J. Biol. Chem.* **1975**, *250*, 1513–1524.
- (39) Alvarez, J.; Haris, P. I.; Lee, D. C.; Chapman, D. *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* **1987**, *916*, 5–12.
- (40) Arrondo, J. L. R.; Young, N. M.; Mantsch, H. H. *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* **1988**, *952*, 261–268.
- (41) Brandstetter, M.; Lendl, B. *Sens. Actuators, B* **2012**, *170*, 189–195.
- (42) Ma, E.; Wang, L.; Kendrick, B. *Spectroscopy* **2018**, *33*, 46–52.

Publication II

Akhgar, C. K.; Nürnberger, V.; Nadvornik, M.; Velik, M.; Schwaighofer, A.; Rosenberg, E.; Lendl, B. Fatty Acid Prediction in Bovine Milk by Attenuated Total Reflection Infrared Spectroscopy after Solvent-Free Lipid Separation. *Foods* **2021**, *10*, 1054.

Article

Fatty Acid Prediction in Bovine Milk by Attenuated Total Reflection Infrared Spectroscopy after Solvent-Free Lipid Separation

Christopher Karim Akhgar¹ , Vanessa Nürnberger², Marlene Nadvornik¹, Margit Velik³, Andreas Schwaighofer¹ , Erwin Rosenberg⁴ and Bernhard Lendl^{1,*}

¹ FG Environmental Analytics, Process Analytics and Sensors, Institute of Chemical Technologies and Analytics, Technische Universität Wien, Getreidemarkt 9, 1060 Vienna, Austria; christopher.akhgar@tuwien.ac.at (C.K.A.); marlene.nadvornik@student.tuwien.ac.at (M.N.); andreas.schwaighofer@tuwien.ac.at (A.S.)

² Competence Center CHASE GmbH, Altenberger Straße 69, 4040 Linz, Austria; vanessa.nuernberger@chasecenter.at

³ Agricultural Research and Education Center Raumberg-Gumpenstein, Altirdning 11, 8952 Irnding-Donnersbachtal, Austria; margit.velik@raumberg-gumpenstein.at

⁴ FG Instrumental and Imaging Analytical Chemistry, Institute of Chemical Technologies and Analytics, Technische Universität Wien, Getreidemarkt 9, 1060 Vienna, Austria; egon.rosenberg@tuwien.ac.at

* Correspondence: bernhard.lendl@tuwien.ac.at; Tel.: +43-1-58801-15140



Citation: Akhgar, C.K.; Nürnberger, V.; Nadvornik, M.; Velik, M.; Schwaighofer, A.; Rosenberg, E.; Lendl, B. Fatty Acid Prediction in Bovine Milk by Attenuated Total Reflection Infrared Spectroscopy after Solvent-Free Lipid Separation. *Foods* **2021**, *10*, 1054. <https://doi.org/10.3390/foods10051054>

Academic Editors: Ben Aernouts, Clement Grelet and Ines Adriaens

Received: 30 March 2021

Accepted: 10 May 2021

Published: 11 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: In the present study, a novel approach for mid-infrared (IR)-based prediction of bovine milk fatty acid composition is introduced. A rapid, solvent-free, two-step centrifugation method was applied in order to obtain representative milk fat fractions. IR spectra of pure milk lipids were recorded with attenuated total reflection Fourier-transform infrared (ATR-FT-IR) spectroscopy. Comparison to the IR transmission spectra of whole milk revealed a higher amount of significant spectral information for fatty acid analysis. Partial least squares (PLS) regression models were calculated to relate the IR spectra to gas chromatography/mass spectrometry (GC/MS) reference values, providing particularly good predictions for fatty acid sum parameters as well as for the following individual fatty acids: C10:0 ($R^2_p = 0.99$), C12:0 ($R^2_p = 0.97$), C14:0 ($R^2_p = 0.88$), C16:0 ($R^2_p = 0.81$), C18:0 ($R^2_p = 0.93$), and C18:1cis ($R^2_p = 0.95$). The IR wavenumber ranges for the individual regression models were optimized and validated by calculation of the PLS selectivity ratio. Based on a set of 45 milk samples, the obtained PLS figures of merit are significantly better than those reported in literature using whole milk transmission spectra and larger datasets. In this context, direct IR measurement of the milk fat fraction inherently eliminates covariation structures between fatty acids and total fat content, which poses a common problem in IR-based milk fat profiling. The combination of solvent-free lipid separation and ATR-FT-IR spectroscopy represents a novel approach for fast fatty acid prediction, with the potential for high-throughput application in routine lab operation.

Keywords: mid-infrared spectroscopy; attenuated total reflection; bovine milk; fatty acids; partial least squares

1. Introduction

Milk is among the fastest growing agricultural commodities, with a worldwide production volume of more than 8.5×10^6 tons per annum, and an expected yearly growth rate of 1.6% until 2029 [1]. Bovine milk accounts for approximately 81% of total milk production, and is considered to be one of the most nutritionally complete foods, with a typical gross composition of 3.9% fat, 3.3% protein, and 4.6% lactose [2]. Milk fat predominantly consists of triglycerides, containing more than 400 different fatty acids, but only 15 of them with relative shares of 1% or higher. The largest fraction are saturated fatty acids (SAT,

approximately 70%), followed by monounsaturated fatty acids (MONO, approximately 25%), and polyunsaturated fatty acids (PUFA, approximately 5%) [3]. Individual fatty acid content in milk is influenced by different factors, such as animal genetics, stage of lactation, and feed intake [4]. Most controversies regarding the health effects of dairy products are associated with lipid composition [5]. SAT especially are often related with harmful effects such as coronary heart disease, while substitution with PUFA might reduce the risk of such disease [6].

Gas chromatography (GC) is the gold standard for milk fatty acid profiling, offering high accuracy combined with maximum sensitivity [7]. Some of the major drawbacks are, however, the essential derivatization step prior to analysis, high costs, and significant time consumption, thus restricting its use for industrial purposes to a few samples from large batches. The demand for rapid, low-cost, high-throughput fatty acid profiling methods is consequently increasing with growing milk production.

Mid-infrared (IR) spectroscopy is a powerful tool for bioanalytical applications [8], which has been demonstrated to present a rapid, label-free alternative to well-established chromatographic methods for the analysis of dairy products [9]. Specific absorption bands, arising from the rotational–vibrational transitions of molecules, allow for compound identification as well as quantification. Important nutritional parameters such as lactose, total fat, and total protein content are routinely detected using the commercially available MilkoScan (Foss, Hillerød, Denmark), a Fourier-transform infrared (FT-IR) spectrometer specifically developed for the analysis of dairy products [10]. Furthermore, novel laser-based mid-IR transmission spectroscopy shows high potential for the quantification of individual proteins in bovine milk [11–14].

Attenuated total reflection (ATR) is a prominent alternative probing technique to transmission mode. Here, the incoming IR light is totally reflected in an optically denser ATR element at the interface with a medium of lower optical density. This leads to an evanescent field that can interact with the sample at typical penetration depths of up to 2 μm per reflection [15]. The sample is placed directly on top of the ATR element, allowing for quick and robust measurements of troublesome liquid matrices, such as oils [16,17].

Substantial effort has been put into investigating the potential of mid-IR transmission spectroscopy for milk fatty acid profiling [18–24]. Here, multivariate chemometric models based on partial least squares (PLS) were established in order to relate mid-IR absorbance spectra acquired from whole milk using the MilkoScan to GC reference data. These studies report good accuracy in predicting the absolute concentrations of certain fatty acids in milk, especially those available in high concentrations, such as C14:0, C16:0, and C18:1. It has been reported that these predictions, however, are most likely based on covariation structures between individual fatty acids and the total fat content, which may change with factors such as breed and feed [25]. When results are stated as relative fatty acid content in milk fat, they appear significantly poorer.

As an alternative approach, dry film FT-IR spectroscopy was introduced [26]. Here, small milk samples were transferred into well plates, dried in a desiccator, and subsequently measured in transmission mode. Multivariate calibrations showed better results than those obtained from direct transmission measurements of whole milk. Here, the lipid preconcentration step was expected to contribute to a major part of the gained prediction improvements. Fine spectral differences associated with fatty acid composition might, however, still be hidden by overlapping absorption bands arising from other major milk components, such as proteins and carbohydrates. Hence, it is of major interest to investigate techniques for lipid separation prior to spectral acquisition, in order to enable improved prediction efficiency and avoid covariation structures with the total fat content.

Classical solvent–solvent extraction methods [27,28] are considered to be reliable for the quantitative separation of lipids from food and animal tissues. Moreover, specific methods for milk fat extraction have been developed, standardized, and are today extensively used in routine lab operation [29,30]. These methods, however, require large amounts of hazardous organic solvents, and are vastly time consuming. A novel, more rapid method

using smaller amounts of organic solvents with shorter exposure times allows for the milk fat separation of approximately 20 samples in 30 min [31]. Alternatively, methods based on two centrifugation steps have been successfully applied to obtain pure milk fat without the use of organic solvents [32,33]. Thorough method validation using GC shows that there is no difference in relative fatty acid composition in the obtained lipid fraction compared to standard solvent–solvent extraction. These methods are consequently ideal for applications that require a representative part of the milk lipids instead of quantitative total fat extraction.

The aim of this study is to show the potential of ATR-FT-IR spectroscopy combined with rapid, solvent-free lipid separation for milk fatty acid profiling. The information content of ATR-IR spectra recorded after lipid separation was compared with the IR transmission spectra of whole milk. By performing multivariate PLS analysis, good prediction accuracy could be obtained for individual fatty acids and relevant sum parameters. PLS calibration equations were optimized based on evaluation of the importance of individual wavenumbers to the multivariate models. Cross-correlations between individual fatty acids and total fat content were inherently avoided by the employed approach. The obtained results indicate several clear advantages over conventional FT-IR transmission spectroscopy of whole milk, revealing high potential for future high-throughput applications.

2. Materials and Methods

2.1. Milk Samples

Forty-five milk samples were collected from the same number of cows in Austria (AREC Raumberg-Gumpenstein, in mid-September 2020), containing two different cattle breeds ($\frac{3}{4}$ Holstein Friesian and $\frac{1}{4}$ Simmental) and three feeding groups. At the time of sample collection, the averages (\pm standard deviation) of the milk yield, days in milk, and lactation number were 20.7 ± 5.73 kg per day, 184 ± 5.73 kg days in milk, and 3.6 ± 2.15 , respectively, ensuring a variety of milk fat composition. The diets of the three feeding groups were based upon ad libitum allowance of a forage mixture, which consisted of 40% grass silage, 30% maize silage, and 30% hay on a dry matter basis. The pelleted concentrate mixture (0%, 20%, and 40% of total feed intake, respectively) consisted of 25% maize, 24% barley, 8% wheat, 8% molasses, 5% wheat bran, 15% soy meal, and 15% rapeseed meal. A pooled sample from morning and evening milk was collected from each cow. Unhomogenized raw milk samples were immediately stored at -80 °C without further conservation until 1 day before fat separation. A homogenized whole milk sample was purchased from an Austrian retailer and used to acquire a mid-IR transmission reference spectrum of whole milk.

2.2. Fat Separation

Milk fat separation was carried out according to the rapid two-step centrifugation method proposed by Feng et al. [32] and modified by Luna et al. [33]. Frozen milk samples were thawed overnight at 4 °C and subsequently tempered at room temperature for at least 20 min. Thirty milliliter aliquots were transferred into falcon tubes and centrifuged at $17,800 \times g$ for 30 min at 20 °C in a Sigma 3–18k centrifuge (Sigma Laborzentrifugen GmbH, Osterode am Harz, Germany). The fat-cake layer was transferred into microtubes and centrifuged at $19,300 \times g$ for 20 min at the same temperature, resulting in three separate layers. The upper lipid layer was removed and used for FT-IR and gas chromatography/mass spectrometry (GC/MS) measurements.

2.3. GC/MS Analysis

Standard solutions, containing 20 mg of milk fat per mL of dichloromethane, were prepared. For the derivatization of fatty acids, an aliquot of 50 μ L of the standard solution was transferred to a pre-cooled 1.5 mL GC vial with a 0.2 mL micro insert. Fifty microliters of internal standard (C17:0) and the same amount of trimethylsulfonium hydroxide (TMSH, 0.25 M in MeOH, Supelco, Vienna, Austria) solution were added, and the vial was capped

immediately. Each vial was vortexed for 5 s, and then heated for 15 min at 70 °C to complete derivatization.

A GC instrument (Shimadzu GC-2010) equipped with a ZB-FAME column (30 m, 0.25 mm I.D., 0.20 µm film thickness; Phenomenex, Aschaffenburg, Germany) coupled with a mass spectrometer (GCMS-QP2010 Plus, Shimadzu, Kyoto, Japan) was used to determine fatty acid content and profile. One microliter samples were injected in split mode (split 100:1) using a Shimadzu AOC-5000 Plus autosampler. The injector temperature was 250 °C. The purge flow was set to 3 mL/min and the column flow to 2.14 mL/min. The oven program was 40 °C initially, held for 3 min, then increased by 10 °C/min to 100 °C, and further increased by 2 °C/min to 200 °C. The transfer line temperature of the mass spectrometer was kept at 200 °C, as was the ion source temperature. After a solvent vent of 2.7 min, the detector voltage was set to 1.05 kV, and the samples were measured in scan mode (35–500 *m/z*).

For quantitative analysis, a method was developed and calibrated using a 37-component FAME mix certified reference material (TraceCERT[®], Supelco, Vienna, Austria). Calibration samples were prepared in the concentration range 20–600 mg/L. Quantitative analysis was based on the evaluation of the quantifier ion peak area for each FAME, provided that the ratio of quantifier and qualifier ions was within acceptable limits. Retention times and qualifier and quantifier ions for each analyte are reported in Table S1 of the Supplementary Materials.

2.4. FT-IR Measurements

ATR-FT-IR measurements were performed using a Bruker Tensor 37 FT-IR spectrometer (Ettlingen, Germany) equipped with a mercury cadmium telluride (MCT) detector ($D^* = 4 \times 10^{10} \text{ cm Hz}^{0.5} \text{ W}^{-1}$ at 9.2 µm). The spectrometer was constantly flushed with dry air in order to reduce the influence of water vapor from the atmosphere. One drop of pure milk fat extract was manually placed onto a Platinum ATR single-bounce element (Bruker, Ettlingen, Germany). Measurements were performed with a spectral resolution of 2 cm^{-1} , between 600 and 4000 cm^{-1} in double-sided, forward-backward acquisition mode. A Blackman-Harris 3-term apodization function and a zero-filling factor of 2 were used to calculate the final spectra. One hundred and twenty-eight scans were averaged per spectrum, leading to an acquisition time of fifty-two seconds. After each spectral acquisition, the ATR surface was cleaned with isopropanol and dichloromethane consecutively until recovery of the baseline signal. Transmission measurements were performed using the same instrument parameters, by injecting homogenized whole milk into a flow cell equipped with two CaF_2 windows and a 37 µm-thick spacer. The software package OPUS 7.2 (Bruker, Ettlingen, Germany) was used for evaluation of the spectral data.

2.5. Data Analysis

Multivariate data analysis was performed in MATLAB R2020a (Mathworks Inc., Natick, MA, USA) using PLS Toolbox 8.9 from Eigenvector Research Inc. (Wenatchee, WA, USA). All ATR-IR absorbance spectra were identically preprocessed by calculation of 2nd derivative spectra, using a Savitzky-Golay filter (window = 15 points) and mean centering. The applied wavenumber range was individually selected for each parameter, based on the selectivity ratio (SR) [34]. Preprocessed FT-IR spectra and GC/MS reference values were used to develop partial least squares (PLS) regression models. Model performance was estimated by applying a contiguous blocks cross-validation with 10 data splits, using the full dataset. Furthermore, external validation was applied by randomly dividing the dataset into a calibration set of 30 samples, and an external validation set, containing 15 samples. Characteristic statistical parameters were calculated to evaluate model performance.

3. Results and Discussion

3.1. Comparison of IR Spectra of Whole Milk and Separated Milk Fat Fraction

Milk fat triglycerides show distinctive mid-IR absorption bands that are influenced by factors such as fatty acid chain length and degree of saturation [35]. Due to this high sensitivity, mid-IR spectroscopy-based prediction of milk fatty acid composition has been reported on multiple accounts and in different implementations. Particularly, the commercially available MilkoScan instrument is widely used for the direct spectral acquisition of whole milk [9]. With this device, FT-IR transmission spectra with a path length of 37 μm are recorded. A limitation of this approach, however, is the limited spectral information in certain wavenumber regions, which can be circumvented by separating the milk fat fraction from the complex milk matrix.

In this work, milk fat was separated according to a rapid two-step centrifugation method [33]. It was shown that the hereby obtained lipid fraction possessed a representative fatty acid composition for the whole milk sample [32,33]. In addition to reduced workload and high throughput, the applied separation method excels in that it completely avoids the use of potentially hazardous or toxic solvents. This characteristic is particularly beneficial for subsequent mid-IR spectroscopy, as small residues of organic solvents can already lead to distinctive absorption bands that hide spectral details of the sample. This is specifically relevant in the present application, because the routinely employed solvents for milk fat extraction often exhibit the same functional groups (e.g., CH_2 , CH_3) and, consequently, IR bands as lipids. Reproducible spectral acquisition after the immediate drying of solvent-based extracts on straight surfaces, such as the ATR crystal, is moreover restricted by the coffee-ring effect [36], which requires complex instrumentation to be avoided [37]. The combination of lipid separation and ATR-FT-IR spectroscopy provides the advantage that very small amounts of sample can be measured in a robust environment by placing them directly onto the active element. In the present study, one drop of the milk fat fraction was sufficient to cover the surface of the ATR crystal and to record representative absorbance spectra. Characteristic mid-IR bands of milk fat are listed in Table 1. Figure 1 displays the typical absorbance spectra of separated milk fat measured in ATR mode (blue) and whole milk recorded in transmission mode with a CaF_2 cell and an optical path length of 37 μm (red). Here, the wavenumber range between 1850 and 2750 cm^{-1} was removed due to lack of information in this region. Visual inspection reveals that the information content of the two IR spectra is significantly different. The whole milk sample also contains vibrational bands from other major components of milk, such as lactose and other carbohydrates (approximately 1000–1480 cm^{-1}) and proteins (amide II band: 1500–1600 cm^{-1}) [38].

Table 1. Characteristic mid-IR absorption bands of milk fat [39].

Wavenumber/ cm^{-1}	Detectable in Whole Milk *	Group	Mode of Vibration	Functional Group
3005	no	C–H	sym. stretch	–C=CH– (cis)
2953	yes	C–H	asym. stretch	– CH_3 (aliphatic)
2922	yes	C–H	asym. stretch	– CH_2 – (aliphatic)
2853	yes	C–H	sym. stretch	– CH_2 – (aliphatic)
1743	yes	C=O	stretch	C=O ester
1655	no	C=C	stretch	C=C (unsaturated)
1462	overlapping	C–H	scissoring	– CH_2 – (aliphatic)
1377	overlapping	C–H	sym. deformation	– CH_3 (aliphatic)
1238	overlapping	C–H	out-of-plane bend	– CH_2 – (aliphatic)
1162	overlapping	C–O	stretch	C–O ester
966	no	C–H	out-of-plane bend	–C=CH– (trans)
722	no	C–H	rocking	– CH_2 – (aliphatic)

* Detection of the absorption band in a whole milk spectrum acquired in transmission mode, using CaF_2 windows and an optical path length of 37 μm . Abbreviations: sym.: symmetric; asym.: asymmetric.

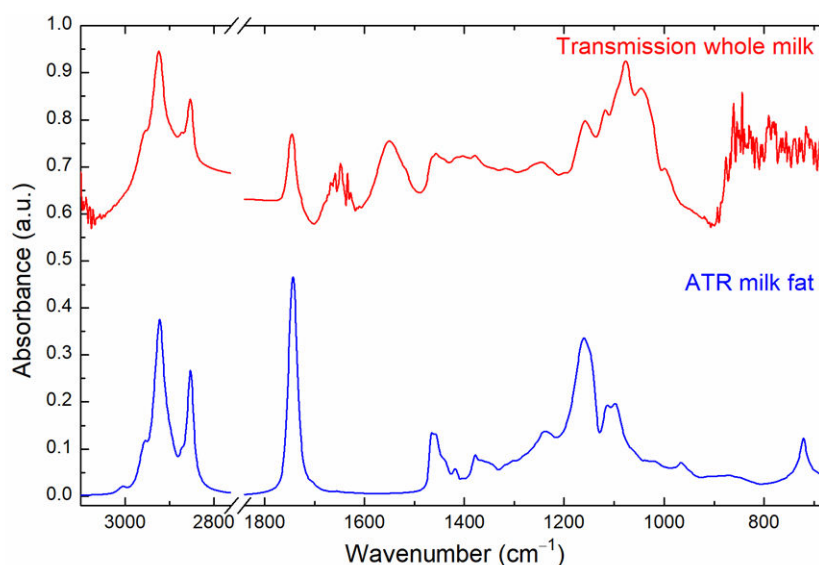


Figure 1. Comparison between the ATR-IR absorbance spectrum of separated milk fat (blue) and whole milk recorded in transmission mode with a CaF_2 cell with an optical path length of $37\ \mu\text{m}$ (red). The spectral range between 1850 and $2750\ \text{cm}^{-1}$ was removed due to lack of relevant information.

The high absorbance of these components adversely affects the evaluation of the significantly lower absorbances originating from fatty acids in these spectral regions. In this context, the low wavenumber region overlapping with carbohydrate absorption bands in particular has proven to be important for the quantitative prediction of individual fatty acids (see Section 3.2.2). Furthermore, for transmission measurements with milk, large transmission paths ($>30\ \mu\text{m}$) are required in order to prevent clogging of the cell because of the high viscosity and complex matrix of milk [40]. However, at these high optical paths, water (HOH bending band: $1643\ \text{cm}^{-1}$) totally absorbs the irradiated IR light, and thus the spectral region between 1600 and $1700\ \text{cm}^{-1}$ is not accessible. Moreover, CaF_2 , the typical window material used for transmission measurements of bovine milk, has its absorption edge at approximately $1000\ \text{cm}^{-1}$ [41], meaning that IR bands at lower wavenumbers are not accessible using this approach. In this inaccessible spectral region for transmission measurements, there are located the C–H out-of-plane band at $966\ \text{cm}^{-1}$ and the C–H rocking band at $722\ \text{cm}^{-1}$, which are well resolved in ATR-IR spectra of lipids. Due to these limitations, for the purpose of the determination of fatty acids, the related spectral features are better resolved in the ATR spectra, thus highlighting the advantage of the lipid separation step.

3.2. Predicting Fatty Acid Content by Mid-IR Spectroscopy

3.2.1. Partial Least Squares Analysis

Individual PLS1 models were calculated to predict the most abundant fatty acids as well as the relevant sum parameters. PLS is a multivariate statistical approach, capable of calculating linear regression models from highly correlated variables, such as those usually found in spectroscopic data [42]. In the present study, the relationship between the recorded ATR absorbance spectra (x-matrix) and the GC/MS reference fatty acid concentration (y-matrix) was calculated. A preprocessing routine, combining second derivative spectra with mean centering, was applied in order to achieve optimal results. Moreover, the included wavenumbers were individually selected for each model (see next chapter). Table 2 provides an overview of the obtained statistical parameters. The root mean square error of calibration (RMSEC) and the calibration coefficient of determination (R^2) were calculated by using the full dataset of available milk samples ($n = 45$) in order to assess the quality of the calibration equations. For the visualization of the calibration equations, Figure 2 shows the relationship between the measured and predicted concentrations on

the examples of unsaturated fatty acids (UNSAT, red) and long-chain fatty acids (LCFA, blue). In the optimal case, all points would lie on the regression line, while those above and below indicate over- and underestimation of ATR-based predictions compared to GC/MS reference values. The small deviation of individual data points from the regression line, as well as the obtained R^2 -values of 0.99, indicate highly linear relationships and very good description of the data by the model. Evaluation of prediction efficiency was performed using contiguous blocks cross-validation with 10 data splits, revealing the root mean square error of cross-validation (RMSECV) and the cross-validation coefficient of determination (R^2_{CV}). Furthermore, external validation was applied by randomly dividing the dataset into a calibration set ($n = 30$) and an external validation set ($n = 15$). Table S2 shows the obtained statistical parameters for the reduced calibration set of 30 samples. The achieved root mean square error of prediction (RMSEP) and prediction coefficient of determination (R^2_P) from the external validation (Table 2) show similar results to the cross-validation, using the whole dataset, indicating high robustness of the calculated prediction models. The optimal number of latent variables (LVs), based on the lowest RMSECV, was between three and eight, which is reasonable for milk fat, which contains a high number of different fatty acids that can cause spectral variability in the system under study. Good prediction efficiencies were obtained for the important health related parameters SAT ($R^2_{CV} = 0.94$, $R^2_P = 0.95$) and UNSAT ($R^2_{CV} = 0.96$, $R^2_P = 0.95$). Further subclassification showed good prediction for MONO ($R^2_{CV} = 0.95$, $R^2_P = 0.94$), while the much lower concentrated PUFA were predicted with moderate accuracy ($R^2_{CV} = 0.61$, $R^2_P = 0.27$). Moreover, sum parameters regarding fatty acid chain length were calculated. Highly concentrated medium-chain fatty acids (MCFA, C12-C16) and LCFA (C17 and higher) were predicted with excellent accuracy ($R^2_{CV} = 0.95/0.98$, $R^2_P = 0.97/0.99$), whereas the lower concentrated group of short-chain fatty acids (SCFA, C4-C10) was predicted with moderate accuracy ($R^2_{CV} = 0.64$, $R^2_P = 0.83$). In the case of individual fatty acid content, excellent predictions ($R^2_{CV} > 0.92$, $R^2_P > 0.93$) were achieved for C10:0, C12:0, C18:0, and C18:1cis, while feasible predictions ($R^2_{CV} > 0.84$, $R^2_P > 0.81$) were obtained for C14:0 and C16:0.

Table 2. Statistical parameters for each individual calibration equation estimating relative individual fatty acid concentration and relevant sum parameters in g/100 g of fat.

Fatty Acid	LVs	Range	g/100 g Fat					
			Full Dataset ($n = 45$)			Split Dataset ($n = 30/15$)		
			RMSEC	RMSECV	R^2	R^2_{CV}	RMSEP	R^2_P
SAT	8	61.6–74.5	0.27	0.66	0.99	0.94	0.8	0.95
MONO	8	19.8–30.3	0.28	0.57	0.99	0.95	0.74	0.94
PUFA	3	2.2–4.2	0.20	0.24	0.73	0.61	0.28	0.27
UNSAT	8	22.1–33.8	0.28	0.58	0.99	0.96	0.74	0.95
SCFA	7	14.2–21.0	0.45	0.78	0.87	0.64	0.67	0.83
MCFA	7	38.1–56.0	0.57	0.96	0.98	0.95	0.85	0.97
LCFA	7	26.4–47.7	0.43	0.76	0.99	0.98	0.65	0.99
C4:0	6	5.4–8.8	0.27	0.42	0.87	0.72	0.49	0.62
C6:0	5	3.1–5.4	0.20	0.31	0.72	0.38	0.24	0.71
C8:0	5	1.5–3.2	0.12	0.16	0.81	0.64	0.11	0.88
C10:0	7	2.1–4.9	0.05	0.11	0.99	0.97	0.10	0.99
C12:0	5	2.0–5.6	0.09	0.16	0.99	0.96	0.19	0.97
C14:0	7	7.4–13.3	0.20	0.49	0.97	0.85	0.48	0.88
C16:0	8	21.1–35.1	0.40	1.05	0.98	0.85	1.4	0.81
C16:1cis	4	1.2–3.9	0.28	0.41	0.73	0.44	0.44	0.39
C18:0	5	5.6–14.6	0.38	0.57	0.97	0.93	0.63	0.93
C18:1cis	8	14.9–27.2	0.22	0.74	0.99	0.92	0.77	0.95

Abbreviations: LVs: latent variables; RMSEC: root mean square error of calibration; RMSECV: root mean square error of cross-validation; RMSEP: root mean square error of prediction; R^2 : calibration coefficient of determination; R^2_{CV} : cross-validation coefficient of determination; R^2_P : prediction coefficient of determination; SAT: saturated fatty acids; MONO: monounsaturated fatty acids; PUFA: polyunsaturated fatty acids; UNSAT: unsaturated fatty acids; SCFA: short-chain fatty acids (C4–C10); MCFA: medium-chain fatty acids (C12–C16); LCFA: long-chain fatty acids (C17 and higher).

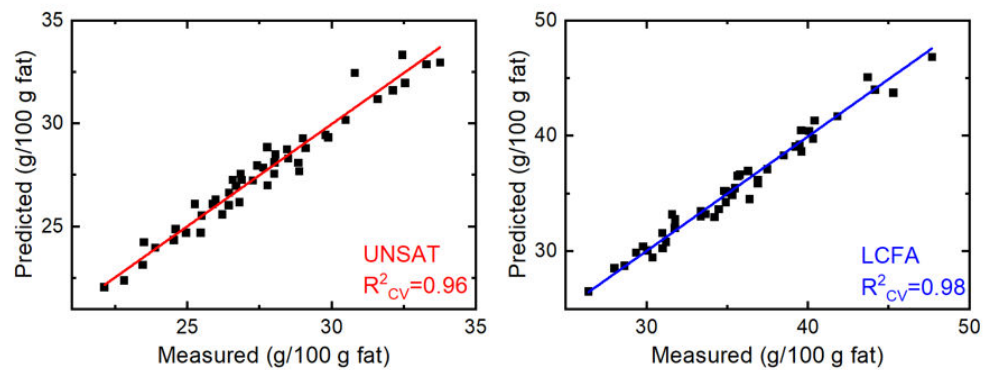


Figure 2. Relationship between measured (GC/MS) and predicted (cross-validation, FT-IR) fatty acid content in g/100 g fat for unsaturated fatty acids (UNSAT, left) and long-chain fatty acids (LCFA, right).

An important parameter that can influence the quality of calibration equations is the number of applied samples. Previously, it has been shown that that use of many different samples can increase the predictability of milk fat composition [20]. For the present study, only a limited set of 45 samples was available. Nevertheless, the achieved results are clearly better than those reported for MilkoScan measurements [18,20] and ATR-FT-IR measurements of whole milk without fat separation [43], when final concentrations are stated in g/100 g fat. Moreover, the presented results are comparable to those obtained from the dry film approach, where a much higher number of samples ($n = 219$) was used [26]. Supposedly, an explanation for the more robust fatty acid prediction enabled by the presented approach is the high accessibility to significant spectral features of the fat fraction compared to (dried) whole milk samples when using ATR-FT-IR spectroscopy, as discussed in the previous section. We expect that even better results can be achieved with the herein presented approach of lipid separation followed by ATR-FT-IR spectroscopy when a larger number of different milk samples are available, indicating high potential for future applications. Finally, it should be noted that additional fatty acids were quantified using the GC/MS reference method, which were present at low concentrations (<2 g/100 g milk fat). However, due to the limited sensitivity of IR spectroscopy and the small sample set, it was not possible to obtain reliable prediction equations for these analytes.

3.2.2. Selection of Wavenumber Range Based on Selectivity Ratio

FT-IR spectra were recorded in the wavenumber range between 600 and 4000 cm^{-1} in order to acquire the maximal amount of information within the mid-IR region. However, for each PLS1 model, the applied spectral region was individually selected, based on the selectivity ratio (SR). The SR is a visualization tool to identify important variables in a multivariate data set for predicting the target variable. A detailed description and mathematical definition can be found elsewhere [34,44]. Briefly, it can be defined as the ratio between explained and unexplained variance for each variable of the model. In the case of mid-IR spectroscopy, the SR is useful for determining specific spectral features with high correlation to the parameter of interest [45].

In this work, the following wavenumber regions without relevant information regarding fatty acid composition were removed for all PLS models: 600–700, 1800–2750, and 3100–4000 cm^{-1} . The remaining wavenumbers were individually selected for each target parameter. Figure 3 displays the included wavenumbers for each calibration model as calculated from the full dataset. Here, brighter regions indicate wavenumbers with low SR, whereas dark areas highlight those with high SR. The figure shows that the significant wavenumbers are highly different between sum parameters that describe fatty acid saturation degree and those that describe chain length. The wavenumber range close to 3005 cm^{-1} has a high SR for SAT, MONO, and UNSAT, while this region is not important for predicting the chain length. This result seems reasonable, because the associated ab-

sorption band arises from the C–H stretching vibration of the cis double bond. Moreover, the spectral region between approximately 2700 and 3000 cm^{-1} , covering several C–H stretching absorption bands, plays an important role in predicting the saturation degree. This relationship was also observed by Christy et al. [46], who predicted the saturation degree in different edible oils using ATR-FT-IR spectroscopy.

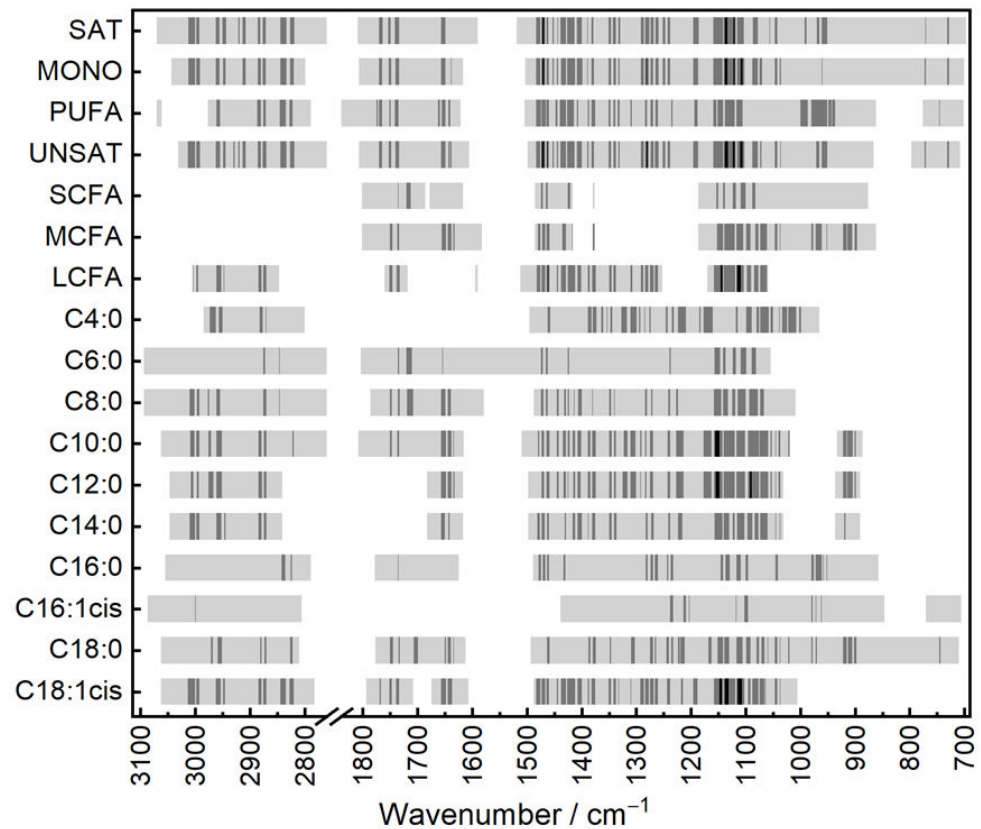


Figure 3. Heatmap, showing spectral regions included for each PLS model in greyscale. Bright grey: selectivity ratio (SR) = 0–0.5; dark grey: SR = 0.5–5; black: SR = 5–15. The spectral range between 1850 and 2750 cm^{-1} was removed due to lack of relevant information. SAT: saturated fatty acids; MONO: monounsaturated fatty acids; PUFA: polyunsaturated fatty acids; UNSAT: unsaturated fatty acids; (C4–C10); MCFA: medium-chain fatty acids (C12–C16); LCFA: long-chain fatty acids (C17 and higher).

The spectral range near 1643 cm^{-1} , not accessible in transmission measurements of whole milk, also contains relevant information regarding the saturation degree. Even though the C=C stretching vibration of unsaturated carbonyl compounds is barely IR-active [47], a weak related absorption band at approximately 1655 cm^{-1} can be observed in the ATR-IR spectra of some lipids. High SR in terms of saturation degree was also obtained in the spectral region between approximately 1050 and 1500 cm^{-1} , which distinctly overlaps with other major components in whole milk transmission spectra. Moreover, the C–H out-of-plane band at 966 cm^{-1} and the C–H rocking band at 722 cm^{-1} , inaccessible in transmission measurements using CaF_2 windows, contain information regarding the saturation degree. PLS models predicting sum parameters concerning chain length showed particularly good results for MCFA and LCFA. Here, the spectral region below 1500 cm^{-1} especially contains several wavenumbers with high SR, indicating important information. The effect of fatty acid chain length on this spectral region was thoroughly investigated by Jones [48], concluding that various small band shifts appear with changing chain length. For predicting individual fatty acids, similar spectral features are important.

Generally, fatty acids with high prediction accuracy, such as C10:0 and C18:1cis, show distinct wavenumber regions with high SR, while fatty acids with weaker prediction, such as C6:0 and C16:1cis, show lower SR.

In conclusion, all spectral regions with medium and high SR can be assigned to absorption bands specific to fatty acids. This verification step is crucial in order to confirm that the information content of calibration equations is based on real absorbance of fatty acids rather than on incidental correlations. While important spectral features are well resolved in the ATR spectra of separated milk fat, a great part of them is hardly accessible or completely inaccessible in whole milk transmission spectra. This evaluation, involving the identification of relevant wavenumbers, thus demonstrates the benefit of using ATR-FT-IR spectroscopy on the milk fat fraction.

3.2.3. Evaluation of Covariation Structures

Due to the performed analysis of the milk fat fraction after separation from the milk matrix, the herein presented approach enables us to state the obtained results in terms of g fatty acid/100 g fat. When comparing the results to other works reporting on mid-IR-based predictions of fatty acids in bovine milk, it should be noted that most authors stated their concentrations in g/100 mL of whole milk. However, it has been demonstrated that this good prediction accuracy is indirect, and primarily based on covariation between individual fatty acids and total fat content, whose dependencies may change with factors such as breed and feed [25]. Covariance is a measure of the degree of association between two random variables [49]. This issue was outlined by showing that PLS models calculated from the milk spectra of a specific cattle breed result in biased predictions when applied to another breed, due to different covariation structures. To further highlight this issue, prediction was performed using a calibration set compiled from skimmed milk spiked with three of the most abundant fatty acids at concentration values avoiding cross-correlations. This approach resulted in significantly poorer models than those obtained for unspiked whole milk. Prediction models, calculated from the same samples, stating the relative fatty acid concentrations in g/100 g fat are, consequently, significantly weaker.

The presented approach, based on ATR-IR measurements, does not have the purpose of providing information regarding total milk fat content, but rather of investigating the relative fatty acid profile. In this way, possible covariation structures to total fat content are inherently eliminated. Consequently, meaningful comparison to published results is only useful for reports where the predicted fatty acid concentrations are also stated relative to milk fat content. Correlations between individual fatty acids are, however, still a great challenge in the spectroscopic prediction of milk fat composition. For this reason, Figure 4 shows a cross-correlation plot of the relative fatty acid content based on the GC results of the applied dataset. For this purpose, Pearson correlation coefficients indicate highly positive correlation (+1), no correlation (0) and highly negative correlation (−1). Here, positively correlated fatty acids are marked in red, whereas those of negatively correlated fatty acids are marked in blue. The plot reveals that most pairwise correlations are small, thus indicating that the predicted concentrations of these fatty acids resulted more from corresponding IR information than from correlations to other fatty acids. Short-chain SAT with similar chain length (i.e., C6:0 and C8:0, C8:0 and C10:0) are, however, highly correlated. These correlations can also be seen in their similar SR profiles in Figure 3. Consequently, grouping similar fatty acids into sum parameters (i.e., SCFA), as was done in this study, is highly beneficial in order to avoid high cross-correlations. Afseth et al. [26] observed comparable correlations between fatty acids in milk with similar chain length, and highlighted that these internal correlations can be used for reliable predictions as long as they are within some degree of certainty valid for future samples.

C4:0	1	0.33	-0.045	-0.48	-0.51	-0.44	-0.35	-0.4	0.48	0.32
C6:0	0.33	1	0.91	0.52	0.49	0.28	-0.12	-0.12	-0.25	-0.45
C8:0	-0.045	0.91	1	0.8	0.78	0.49	-0.056	0.0051	-0.45	-0.58
C10:0	-0.48	0.52	0.8	1	0.98	0.73	0.2	0.069	-0.57	-0.75
C12:0	-0.51	0.49	0.78	0.98	1	0.71	0.2	0.13	-0.62	-0.73
C14:0	-0.44	0.28	0.49	0.73	0.71	1	0.39	-0.046	-0.6	-0.72
C16:0	-0.35	-0.12	-0.056	0.2	0.2	0.39	1	0.39	-0.68	-0.68
C16:1cis	-0.4	-0.12	0.0051	0.069	0.13	-0.046	0.39	1	-0.53	-0.17
C18:0	0.48	-0.25	-0.45	-0.57	-0.62	-0.6	-0.68	-0.53	1	0.68
C18:1cis	0.32	-0.45	-0.58	-0.75	-0.73	-0.72	-0.68	-0.17	0.68	1
	C4:0	C6:0	C8:0	C10:0	C12:0	C14:0	C16:0	C16:1cis	C18:0	C18:1cis

Figure 4. Cross-correlation matrix of pairwise correlations between individual fatty acids from GC/MS reference measurements. Red color indicates positive correlation, whereas blue color indicates negative correlation.

4. Conclusions

In this work, a new mid-IR-based approach for predicting the fatty acid composition of bovine milk was introduced. A rapid, solvent free, two-step centrifugation method was employed in order to obtain representative milk fat fractions. Absorbance spectra of pure lipids were recorded using ATR-FT-IR spectroscopy, and compared to the transmission spectra of whole milk. Fatty-acid-related spectral features were distinctively better resolved in ATR spectra, highlighting the advantage of the preceding lipid separation step. PLS-based multivariate calibration models were calculated in order to relate IR absorbance spectra to relative concentrations of the most abundant fatty acids and sum parameters, obtained via a GC/MS reference method. Prediction efficiency was evaluated by performing cross-validation on the full dataset, as well as by splitting the dataset into a calibration and a validation set. Both methods showed excellent results, indicating high robustness of the models. Particularly high prediction accuracies were obtained for SAT, MONO, UNSAT, MCFA, LCFA, C10:0, C12:0, C14:0, C16:0, C18:0, and C18:1cis. Based on a set of 45 milk samples, the obtained results were clearly better than those reported in literature for whole milk transmission spectra when concentrations were stated in g/100 g fat. The information content of the calibration equations was evaluated by identifying the most important spectral features for predicting individual target variables. Here, relevant wavenumbers were identified based on SR, and successfully assigned to absorbance bands from milk fat. Covariation structures between total fat content and predicted parameters, a common problem in IR-based milk fat profiling, were inherently eliminated with the applied approach. Consequently, the presented method bears several clear advantages over FT-IR transmission spectroscopy of whole milk, revealing its high potential for high-throughput applications. ATR-FT-IR measurements of pure milk fat, including cleaning procedures, can be performed in less than 2 min. When a high number of samples must be analyzed, the two centrifugation steps (30/20 min) can be performed in parallel, and the work flow optimized to provide maximum sample throughput. In the future, the calibration equations might be further improved by using a higher number of different milk samples, whereas additional automatization of the fat separation procedure could facilitate high-throughput operation.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/foods10051054/s1>: Table S1: Retention time, quantifier and qualifier ions for each analyte of the applied GC/MS method; Table S2: Statistical parameters for each individual calibration equation, using a calibration set of 30 samples and a validation set of 15 samples.

Author Contributions: Conceptualization, C.K.A., A.S., E.R. and B.L.; methodology, C.K.A. and V.N.; formal analysis, C.K.A.; investigation, C.K.A., V.N. and M.N.; resources, M.V. and B.L.; data curation, C.K.A. and V.N.; writing—original draft, C.K.A.; writing—review and editing, V.N., M.N., M.V., A.S., E.R. and B.L.; visualization, C.K.A. and A.S.; supervision, A.S., E.R. and B.L.; funding acquisition, B.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work has received funding from the COMET Center CHASE (project No 868615), which is funded within the framework of COMET (Competence Centers for Excellent Technologies) by BMVIT, BMDW, and the Federal Provinces of Upper Austria and Vienna. The COMET program is run by the Austrian Research Promotion Agency (FFG). The authors acknowledge the TU Wien Bibliothek for financial support through its Open Access Funding Program.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The authors thank the following people for their contributions: Thomas Guggenberger and Daniel Eingang (Agricultural Research and Education Center Raumberg-Gumpenstein) for providing bovine milk samples; Julian Ebner and Oliver Spadiut (FG Biochemical Engineering; Institute of Chemical, Environmental and Bioscience Engineering; Technische Universität Wien) for enabling access to the laboratory centrifuge. The authors acknowledge the TU Wien Bibliothek for financial support through its Open Access Funding Program.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. OECD/FAO. *OECD-FAO Agricultural Outlook 2020–2029*; FAO: Paris, France, 2020. [[CrossRef](#)]
2. Jenness, R.; Wong, N.P.; Marth, E.H.; Keeney, M. *Fundamentals of Dairy Chemistry*; Springer: New York City, NY, USA, 1988. [[CrossRef](#)]
3. Månsson, H. Fatty acids in bovine milk fat. *Food Nutr. Res.* **2008**, *52*, 1821. [[CrossRef](#)] [[PubMed](#)]
4. Palmquist, D.L.; Denise Beaulieu, A.; Barbano, D.M. Feed and Animal Factors Influencing Milk Fat Composition. *J. Dairy Sci.* **1993**, *76*, 1753–1771. [[CrossRef](#)]
5. German, J.B.; Gibson, R.A.; Krauss, R.M.; Nestel, P.; Lamarche, B.; van Staveren, W.A.; Steijns, J.M.; de Groot, L.C.; Lock, A.L.; Destailats, F. A reappraisal of the impact of dairy foods and milk fat on cardiovascular disease risk. *Eur. J. Nutr.* **2009**, *48*, 191–203. [[CrossRef](#)] [[PubMed](#)]
6. Micha, R.; Mozaffarian, D. Saturated fat and cardiometabolic risk factors, coronary heart disease, stroke, and diabetes: A fresh look at the evidence. *Lipids* **2010**, *45*, 893–905. [[CrossRef](#)]
7. Christie, W.W. Gas chromatography-mass spectrometry methods for structural analysis of fatty acids. *Lipids* **1998**, *33*, 343–353. [[CrossRef](#)]
8. Schwaighofer, A.; Brandstetter, M.; Lendl, B. Quantum cascade lasers (QCLs) in biomedical spectroscopy. *Chem. Soc. Rev.* **2017**, *46*, 5903–5924. [[CrossRef](#)]
9. De Marchi, M.; Toffanin, V.; Cassandro, M.; Penasa, M. Invited review: Mid-infrared spectroscopy as phenotyping tool for milk traits. *J. Dairy Sci.* **2014**, *97*, 1171–1186. [[CrossRef](#)]
10. Voort, F.R.V.D. Evaluation of Milkoscan 104 Infrared Milk Analyzer. *J. AOAC Int.* **2020**, *63*, 973–980. [[CrossRef](#)]
11. Kuligowski, J.; Schwaighofer, A.; Alcaraz, M.R.; Quintas, G.; Mayer, H.; Vento, M.; Lendl, B. External cavity-quantum cascade laser (EC-QCL) spectroscopy for protein analysis in bovine milk. *Anal. Chim. Acta* **2017**, *963*, 99–105. [[CrossRef](#)]
12. Schwaighofer, A.; Kuligowski, J.; Quintas, G.; Mayer, H.K.; Lendl, B. Fast quantification of bovine milk proteins employing external cavity-quantum cascade laser spectroscopy. *Food Chem.* **2018**, *252*, 22–27. [[CrossRef](#)]
13. Montemurro, M.; Schwaighofer, A.; Schmidt, A.; Culzoni, M.J.; Mayer, H.K.; Lendl, B. High-throughput quantitation of bovine milk proteins and discrimination of commercial milk types by external cavity-quantum cascade laser spectroscopy and chemometrics. *Analyst* **2019**, *144*, 5571–5579. [[CrossRef](#)]
14. Schwaighofer, A.; Alcaráz, M.R.; Kuligowski, J.; Lendl, B. Recent advancements of EC-QCL based mid-IR transmission spectroscopy of proteins and application to analysis of bovine milk. *Biomed. Spectrosc. Imaging* **2018**, *7*, 35–45. [[CrossRef](#)]
15. Ramer, G.; Lendl, B. Attenuated Total Reflection Fourier Transform Infrared Spectroscopy. In *Encyclopedia of Analytical Chemistry*; Meyers, R., RAMTECH Inc., Eds.; John Wiley & Sons: Hoboken, NJ, USA, 2013. [[CrossRef](#)]


16. Sherazi, S.T.H.; Talpur, M.Y.; Mahesar, S.A.; Kandhro, A.A.; Arain, S. Main fatty acid classes in vegetable oils by SB-ATR-Fourier transform infrared (FTIR) spectroscopy. *Talanta* **2009**, *80*, 600–606. [[CrossRef](#)]
17. Karunathilaka, S.R.; Mossoba, M.M.; Chung, J.K.; Haile, E.A.; Srigley, C.T. Rapid Prediction of Fatty Acid Content in Marine Oil Omega-3 Dietary Supplements Using a Portable Fourier Transform Infrared (FTIR) Device and Partial Least-Squares Regression (PLSR) Analysis. *J. Agric. Food Chem.* **2017**, *65*, 224–233. [[CrossRef](#)]
18. Soyeurt, H.; Dardenne, P.; Dehareng, F.; Lognay, G.; Veselko, D.; Marlier, M.; Bertozzi, C.; Mayeres, P.; Gengler, N. Estimating Fatty Acid Content in Cow Milk Using Mid-Infrared Spectrometry. *J. Dairy Sci.* **2006**, *89*, 3690–3695. [[CrossRef](#)]
19. Soyeurt, H.; Dehareng, F.; Mayeres, P.; Bertozzi, C.; Gengler, N. Variation of Δ^9 -Desaturase Activity in Dairy Cattle. *J. Dairy Sci.* **2008**, *91*, 3211–3224. [[CrossRef](#)]
20. Rutten, M.J.M.; Bovenhuis, H.; Hettinga, K.; Valenberg, H.J.F.; Arendonk, J. Predicting bovine milk fat composition using infrared spectroscopy based on milk samples collected in winter and summer. *J. Dairy Sci.* **2009**, *92*, 6202–6209. [[CrossRef](#)]
21. De Marchi, M.; Penasa, M.; Cecchinato, A.; Mele, M.; Secchiari, P.; Bittante, G. Effectiveness of mid-infrared spectroscopy to predict fatty acid composition of Brown Swiss bovine milk. *Animal* **2011**, *5*, 1653–1658. [[CrossRef](#)]
22. Ferrand, M.; Huquet, B.; Barbey, S.; Barillet, F.; Lahalle, F.; Larroque, H.; Leray, O.; Trommenschlager, J.M.; Brochard, M. Determination of fatty acid profile in cow's milk using mid-infrared spectrometry: Interest of applying a variable selection by genetic algorithms before a PLS regression. *Chemom. Intell. Lab. Syst.* **2011**, *106*, 183–189. [[CrossRef](#)]
23. Eijndhoven, M.H.T.; Soyeurt, h.; Dehareng, F.; Calus, M. Validation of fatty acid predictions in milk using mid-infrared spectrometry across cattle breeds. *Animal* **2012**, *7*, 1–7. [[CrossRef](#)]
24. Soyeurt, H.; Dehareng, F.; Gengler, N.; McParland, S.; Wall, E.; Berry, D.; Coffey, M.; Dardenne, P. Mid-infrared prediction of bovine milk fatty acids across multiple breeds, production systems, and countries. *J. Dairy Sci.* **2011**, *94*, 1657–1667. [[CrossRef](#)]
25. Eskildsen, C.; Rasmussen, M.; Engelsen, S.; Larsen, L.; Poulsen, N.A.; Skov, T. Quantification of individual fatty acids in bovine milk by infrared spectroscopy and chemometrics: Understanding predictions of highly collinear reference variables. *J. Dairy Sci.* **2014**, *97*, 7940–7951. [[CrossRef](#)]
26. Afseth, N.K.; Martens, H.; Randby, Å.; Gidskehaug, L.; Narum, B.; Jørgensen, K.; Lien, S.; Kohler, A. Predicting the Fatty Acid Composition of Milk: A Comparison of Two Fourier Transform Infrared Sampling Techniques. *Appl. Spectrosc.* **2010**, *64*, 700–707. [[CrossRef](#)]
27. Folch, J.; Lees, M.; Sloane Stanley, G.H. A simple method for the isolation and purification of total lipides from animal tissues. *J. Biol. Chem.* **1957**, *226*, 497–509. [[CrossRef](#)]
28. Bligh, E.G.; Dyer, W.J. A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* **1959**, *37*, 911–917. [[CrossRef](#)]
29. Lunder, T.L. Simplified Procedure for Determining Fat and Total Solids by Mojonnier Method. *J. Dairy Sci.* **1971**, *54*, 737–739. [[CrossRef](#)]
30. ISO. *Milk and Milk Products—Extraction Methods for Lipids and Liposoluble Compounds*; ISO 1415:2001; ISO Publications: Geneva, Switzerland, 2001.
31. Stefanov, I.; Vlaeminck, B.; Fievez, V. A novel procedure for routine milk fat extraction based on dichloromethane. *J. Food Compos. Anal.* **2010**, *23*, 852–855. [[CrossRef](#)]
32. Feng, S.; Lock, A.L.; Garnsworthy, P.C. Technical Note: A Rapid Lipid Separation Method for Determining Fatty Acid Composition of Milk. *J. Dairy Sci.* **2004**, *87*, 3785–3788. [[CrossRef](#)]
33. Luna, P.; Juárez, M.; Fuente, M.A. Validation of a Rapid Milk Fat Separation Method to Determine the Fatty Acid Profile by Gas Chromatography. *J. Dairy Sci.* **2005**, *88*, 3377–3381. [[CrossRef](#)]
34. Farrés, M.; Platikanov, S.; Tsakovski, S.; Tauler, R. Comparison of the variable importance in projection (VIP) and of the selectivity ratio (SR) methods for variable selection and interpretation. *J. Chemom.* **2015**, *29*, 528–536. [[CrossRef](#)]
35. Kaylegian, K.E.; Lynch, J.M.; Fleming, J.R.; Barbano, D. Influence of fatty acid chain length and unsaturation on mid-infrared milk analysis. *J. Dairy Sci.* **2009**, *92*, 2485–2501. [[CrossRef](#)]
36. Baker, M.J.; Hussain, S.R.; Lovergne, L.; Untereiner, V.; Hughes, C.; Lukaszewski, R.A.; Thiéfin, G.; Sockalingum, G.D. Developing and understanding biofluid vibrational spectroscopy: A critical review. *Chem. Soc. Rev.* **2016**, *45*, 1803–1818. [[CrossRef](#)]
37. Pivonka, D.E.; Kirkland, K.M. Research Strategy for the HPLC/FT-IR Analysis of Drug Metabolites. *Appl. Spectrosc.* **1997**, *51*, 866–873. [[CrossRef](#)]
38. Kohler, A.; Afseth, N.; Jørgensen, K.; Randby, Å.; Martens, H. Quality Analysis of Milk by Vibrational Spectroscopy. In *Handbook of Vibrational Spectroscopy*; Jim, C., Peter, G., Eds.; John Wiley & Sons: Hoboken, NJ, USA, 2010. [[CrossRef](#)]
39. Safar, M.; Bertrand, D.; Robert, P.; Devaux, M.-F.; Genot, C. Characterization of edible oils, butters and margarines by Fourier transform infrared spectroscopy with attenuated total reflectance. *J. Am. Oil Chem. Soc.* **1994**, *71*, 371–377. [[CrossRef](#)]
40. Schwaighofer, A.; Lendl, B. Chapter 3—Quantum cascade laser-based infrared transmission spectroscopy of proteins in solution. In *Vibrational Spectroscopy in Protein Research*; Ozaki, Y., Baranska, M., Lednev, I.K., Wood, B.R., Eds.; Academic Press: Cambridge, MA, USA, 2020; pp. 59–88. [[CrossRef](#)]
41. Mayerhöfer, T.G.; Pahlow, S.; Hübner, U.; Popp, J. CaF₂: An Ideal Substrate Material for Infrared Spectroscopy? *Anal. Chem.* **2020**, *92*, 9024–9031. [[CrossRef](#)]
42. Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130. [[CrossRef](#)]

43. Soulat, J.; Andueza, D.; Graulet, B.; Girard, C.L.; Labonne, C.; Aït-Kaddour, A.; Martin, B.; Ferlay, A. Comparison of the Potential Abilities of Three Spectroscopy Methods: Near-Infrared, Mid-Infrared, and Molecular Fluorescence, to Predict Carotenoid, Vitamin and Fatty Acid Contents in Cow Milk. *Foods* **2020**, *9*, 592. [[CrossRef](#)]
44. Kvalheim, O.M. Variable importance: Comparison of selectivity ratio and significance multivariate correlation for interpretation of latent-variable regression models. *J. Chemom.* **2020**, *34*, e3211. [[CrossRef](#)]
45. V.D. dos Santos, A.C.; Heydenreich, R.; Derntl, C.; Mach-Aigner, A.R.; Mach, R.L.; Ramer, G.; Lendl, B. Nanoscale Infrared Spectroscopy and Chemometrics Enable Detection of Intracellular Protein Distribution. *Anal. Chem.* **2020**, *92*, 15719–15725. [[CrossRef](#)]
46. Christy, A.A.; Egeberg, P.K. Quantitative determination of saturated and unsaturated fatty acids in edible oils by infrared spectroscopy and chemometrics. *Chemom. Intell. Lab. Syst.* **2006**, *82*, 130–136. [[CrossRef](#)]
47. Muik, B.; Lendl, B.; Molina-Diaz, A.; Valcarcel, M.; Ayora-Canada, M.J. Two-dimensional correlation spectroscopy and multivariate curve resolution for the study of lipid oxidation in edible oils monitored by FTIR and FT-Raman spectroscopy. *Anal. Chim. Acta* **2007**, *593*, 54–67. [[CrossRef](#)] [[PubMed](#)]
48. Jones, R.N. The effects of chain length on the infrared spectra of fatty acids and methyl esters. *Can. J. Chem.* **1962**, *40*, 321–333. [[CrossRef](#)]
49. Rice, J.A. *Mathematical Statistics and Data Analysis*, 3rd ed.; Crockett, C., Ed.; Thompson/Brooks/Cole: Belmont, CA, USA, 2007.

Publication III

Akhgar, C. K.; Nürnberger, V.; Nadvornik, M.; Ramos-Garcia, V.; Ten-Doménech, I.; Kuligowski, J.; Schwaighofer, A.; Rosenberg, E.; Lendl, B. Fatty Acid Determination in Human Milk Using Attenuated Total Reflection Infrared Spectroscopy and Solvent-Free Lipid Separation. *Appl. Spec.* **2022**, *76*, 730-736.

Fatty Acid Determination in Human Milk Using Attenuated Total Reflection Infrared Spectroscopy and Solvent-Free Lipid Separation

Applied Spectroscopy
2022, Vol. 76(6) 730–736
© The Author(s) 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/00037028211065502
journals.sagepub.com/home/asp


Christopher K. Akhgar¹ , Vanessa Nürnberger², Marlene Nadvornik¹,
Victoria Ramos-Garcia³, Isabel Ten-Doménech³, Julia Kuligowski³,
Andreas Schwaighofer¹ , Erwin Rosenberg¹  and Bernhard Lendl¹ 

Abstract

This study introduces the first mid-infrared (IR)-based method for determining the fatty acid composition of human milk. A representative milk lipid fraction was obtained by applying a rapid and solvent-free two-step centrifugation method. Attenuated total reflection Fourier transform infrared (ATR FT-IR) spectroscopy was applied to record absorbance spectra of pure milk fat. The obtained spectra were compared to whole human milk transmission spectra, revealing the significantly higher degree of fatty acid-related spectral features in ATR FT-IR spectra. Partial least squares (PLS)-based multivariate regression equations were established by relating ATR FT-IR spectra to fatty acid reference concentrations, obtained with gas chromatography–mass spectrometry (GC-MS). Good predictions were achieved for the most important fatty acid sum parameters: saturated fatty acids (SAT, $R^2_{CV} = 0.94$), monounsaturated fatty acids (MONO, $R^2_{CV} = 0.85$), polyunsaturated fatty acids (PUFA, $R^2_{CV} = 0.87$), unsaturated fatty acids (UNSAT, $R^2_{CV} = 0.91$), short-chain fatty acids (SCFA, $R^2_{CV} = 0.79$), medium-chain fatty acids (MCFA, $R^2_{CV} = 0.97$), and long-chain fatty acids (LCFA, $R^2_{CV} = 0.88$). The PLS selectivity ratio (SR) was calculated in order to optimize and verify each individual calibration model. All mid-IR regions with high SR could be assigned to absorbances from fatty acids, indicating high validity of the obtained models.

Keywords

Mid-infrared spectroscopy, human milk, attenuated total reflection, ATR, partial least squares regression, PLSR, fatty acids

Date received: 28 July 2021; accepted: 18 October 2021

Introduction

Human milk is a complex mixture, containing a multitude of constituents that either can individually or in combination contribute to various health outcomes for the infant.¹ Among the most important factors affecting human milk composition are the stage of lactation and term/preterm delivery.² The typical macronutrient composition of term, mature human milk is 0.9–1.2 g/dL for proteins, 3.2–3.6 g/dL for fat, and 6.7–7.8 g/dL for lactose.³ Fatty acids represent the most variable macronutrients,⁴ known to be additionally influenced by mothers diet, genetics, sociodemographic, and environmental factors.⁵ The milk fat of European women typically contains approx. 35–40% saturated fatty acids (SATs), 45–50% monounsaturated fatty acids (MONO), and 15% polyunsaturated fatty acids (PUFAs).⁶

Quantification of individual fatty acids from human milk is routinely performed by gas chromatography (GC),⁷ offering high accuracy and sensitivity. Some drawbacks are, however,

the associated high costs, significant time consumption, and the essential derivatization step prior to analysis. Due to this work-intensive procedure, there is a need for more rapid, less expensive, and high-throughput methods.

Mid-infrared (IR) spectroscopy provides detailed information about rotational–vibrational transitions of molecules, allowing identification and quantification of a wide range of biochemical components.⁸ Fourier transform IR (FT-IR)

¹Institute of Chemical Technologies and Analytics, Technische Universität Wien, Wien, Austria

²Competence Center CHASE GmbH, Linz, Austria

³Health Research Institute La Fe, Valencia, Spain

Corresponding author:

Bernhard Lendl, Institute of Chemical Technologies and Analytics, Technische Universität Wien, Getreidemarkt 9, Vienna 1060, Austria.
Email: bernhard.lendl@tuwien.ac.at

devices are the most established and widespread instrumentation available operating in this spectral region. For determination of the macronutrient composition of human milk, including total protein, total fat, and lactose content, commercially available mid-IR based devices (e.g., Miris AB human milk analyzer, Sweden) that record absorbance spectra of whole milk in transmission mode are routinely employed at human milk banks. It has been shown that the hereby obtained results are acceptable for most measurement parameters.⁹ Moreover, Raman spectroscopy was applied to analyze the composition of human milk.^{10–12} As an alternative probing mode, attenuated total reflection (ATR) is commonly applied in FT-IR spectroscopy. In this approach, the IR beam passes through an ATR element of high optical density, where it is totally reflected at the interface with an optically less dense sample. Through this reflection, an evanescent field is formed that can penetrate samples at typical depths between 0.5 and 2 μm .¹³ Previously, ATR FT-IR spectroscopy was applied to determine total protein, fat, and carbohydrate content in human milk.¹⁴ The potential of FT-IR spectroscopy for determining the fatty acid composition of bovine milk has been studied extensively.¹⁵ Here, reported methods include probing of whole milk,¹⁶ as well as measuring the extracted lipid fraction after solvent extraction.¹⁷ Method validation showed that solvent-free fat separation combined using ATR FT-IR spectroscopy also allows accurate bovine milk fatty acid profiling.¹⁸ Method validation showed that the thus obtained fraction possesses a representative fatty acid composition for milk from different species.^{19,20}

However, to the best of our knowledge, no studies reporting mid-IR-based determination of human milk fatty acids are available up to now. The aim of the present study is to show the potential of solvent-free lipid separation and ATR FT-IR spectroscopy for measuring the fatty acid composition of human milk. On this account, a rapid two-step centrifugation method was used to receive a pure lipid fraction without using organic solvents. The ATR FT-IR absorbance spectra of the obtained lipid fraction were recorded and compared with spectra of whole human milk, recorded in FT-IR transmission mode. Partial least squares (PLS)-based multivariate calibration models were established in order to relate the ATR FT-IR spectra to fatty acid reference concentrations, measured with gas chromatography–mass spectrometry (GC-MS). The obtained individual PLS models were optimized by evaluating the importance of mid-IR wavenumbers for predicting the target variables. Based on eight human milk samples, the calculated statistical figures of merit clearly indicate high potential toward future high-throughput human milk fatty acid profiling.

Material and Methods

Human Milk Samples

Human milk samples ($N = 8$) were collected following the instructions of the hospital staff. Milk expression was

accomplished with breast milk pumps following the standard operating procedure routinely employed at the hospital and the human milk bank. Prior to extraction, both removable parts of the breast milk pump and collection bottles were sterilized. In addition, mothers must wash their hands with soap and water and nipples with water. Extracted milk was stored at $-20\text{ }^{\circ}\text{C}$ until further processing.

Human Milk Macronutrient Analysis

Human milk macronutrients including fat, carbohydrate, crude and true protein, total solids, and energy were determined using a Miris HMA (Miris AB). Determinations, quality control, and instrument calibration were performed following the standard operation procedure given by the manufacturer.²¹

Lipid Separation

Human milk lipids were separated by employing a previously reported centrifugation method.²⁰ First, 30 mL sample aliquots were unfrozen at $4\text{ }^{\circ}\text{C}$ overnight, followed by storing them at ambient temperature for a minimum of 20 min. Second, they were centrifuged at $20\text{ }^{\circ}\text{C}$ for 30 min at $17\ 800\times g$ in a Sigma 3–18k centrifuge (Sigma Laborzentrifugen GmbH, Germany). The obtained upper phase, containing lipids, proteins, and water was removed and transferred into microtubes, followed by a second centrifugation step at $20\text{ }^{\circ}\text{C}$ for 20 min at $19\ 300\times g$. The thereby obtained pure fat layer of approximately 20–150 mg was removed and analyzed with ATR FT-IR and GC-MS.

Gas Chromatography–Mass Spectrometry Analysis

Sample solutions were prepared from the milk fat, isolated according to the previously described procedure, by dissolving 20 mg per ml dichloromethane. Derivatization of fatty acids was carried out by transferring 50 μL of this solution to a cooled vial (1.5 mL with 0.2 mL low-volume insert). To the sample, 50 μL each of the C17:0 internal standard solutions and the 0.25 M trimethylsulfonium hydroxide (TMSH) in methanol (Supelco, Austria) derivatization reagent were added. Each sample was vortex mixed for 5 s, and the derivatization reaction was brought to completion by heating to $70\text{ }^{\circ}\text{C}$ for 15 min.

Quantitative GC-MS analysis and profiling of fatty acids was performed using a Shimadzu QP2010 Plus gas chromatograph with a single-quadrupole mass spectrometer (Shimadzu, Japan), equipped with a ZB-FAME column (30 m \times 0.25 mm inner diameter \times 0.25 μm film thickness) obtained from Phenomenex (Germany). Split mode injection (split 100:1) of 1 μL sample volumes was performed at $250\text{ }^{\circ}\text{C}$ injector temperature using a Shimadzu AOC-5000 Plus autosampler. The septum purge flow was kept constant at 3 mL/min while the column flow was set at 2.14 mL/min initially under constant linear velocity mode. The oven program was $40\text{ }^{\circ}\text{C}$ (3 min), then increase of $10\text{ }^{\circ}\text{C}/\text{min}$ to $100\text{ }^{\circ}\text{C}$ (no hold), and

further at a rate of 2 °C/min to 200 °C (no hold). Total run time was 59.00 min. Both the transfer line and the ion source temperature were kept at a temperature of 200 °C. Data acquisition was started only after a solvent delay of 2.7 min, when the detector voltage was set to 1.05 kV. Data acquisition was performed in scan mode (m/z 35–350).

Quantitative analysis of FAME was performed by calibrating the system with a TraceCERT Supelco 37 Component FAME mix (Supelco, Austria). For this purpose, calibration levels between 20 and 600 mg/L were prepared. During data evaluation, the peak area of the quantifier ion is evaluated for each target compound, if the quantifier/qualifier ion peak ratio is within the pre-defined acceptance range. Detailed information on the retention times and the quantifier and qualifier ions used for each FAME has been reported previously.¹⁸

FT-IR Measurements

A Bruker Tensor 37 FT-IR spectrometer equipped with a mercury–cadmium–telluride (MCT) detector ($D^* = 4 \times 10^{10} \text{ cm Hz}^{0.5} \text{ W}^{-1}$ at 9.2 μm) was used for all mid-IR measurements. Data acquisition was performed by applying the following settings: spectral resolution of 2 cm^{-1} , double-sided, forward–backward acquisition mode, scan range from 600 to 4000 cm^{-1} , and 52 s measurement time (averaging 128 scans per spectrum). The final spectra were calculated by using a Blackman–Harris three-term apodization function and a zero-filling factor of two. By constantly flushing the spectrometer with dry air, the influence of atmospheric water vapor was reduced. Attenuated total reflection measurements were performed by transferring approximately 5 μL of the extracted lipid fraction onto a Platinum ATR single-bounce element (Bruker). After each measurement, the surface of the ATR element was cleaned with dichloromethane and isopropanol until the initial baseline signal was recovered. Reference measurements were performed by placing homogenized human milk into a CaF_2 transmission cell with an optical path length of 50 μm (as used in the commercially available Miris human milk analyzer). Data was acquired and analyzed with the spectroscopy software Bruker Opus 7.2.

Multivariate Calibration Equations

Multivariate calibration equations were calculated in Matlab R2020a (The Mathworks Inc., USA) using the PLS Toolbox 8.9 from Eigenvector Research Inc. (Wenatchee, USA). First, the recorded spectra of pure milk fat were treated with a pre-processing routine, combining mean centering and calculation of second derivatives by applying a Savitzky–Golay filter with a window of 15 points and polynomial order of two. PLS regression models were created by relating the preprocessed ATR FT-IR spectra to the fatty acid concentrations measured with GC-MS. The included spectral range was individually

selected for each target variable, according to the PLS selectivity ratio (SR).²² The number of latent variables was determined by applying leave-one-out cross-validation (LOOCV).

Results and Discussion

Comparison Between Mid-IR Spectra of Whole Human Milk and Pure Lipid Fraction

In the present study, human milk lipids were separated with a rapid two-step centrifugation method prior to ATR FT-IR analysis.²⁰ Table I gives an overview of typical mid-IR absorbance bands of human milk fat. Figure 1 shows a comparison between a typical absorbance spectrum of pure lipid extract and a transmission spectrum of whole human milk, measured with a CaF_2 cell and a transmission path of 50 μm . The whole milk spectrum includes absorbance bands from other major milk components such as proteins (amide II band: 1500–1600 cm^{-1}) and sugars (approx. 1000–1480 cm^{-1})²³ that overlap with significantly less intense fatty acid bands, thus adversely affecting the evaluation of those. For transmission measurements of whole milk, large optical path lengths (>30 μm) are necessary in order to avoid cell clogging.²⁴ At these large optical path length, however, the spectral region between 1600 and 1700 cm^{-1} (e.g., C=C stretch band at 1655 cm^{-1}) is not accessible, due to total IR absorption through the HOH bending band of water. Furthermore, CaF_2 , the window material that is typically used for transmission measurements of milk, has a strong absorption starting at approximately 1000 cm^{-1} ,²⁵ meaning that absorbance bands at lower wavenumbers (e.g., CH out-of-plane band at 966 cm^{-1} and C–H rocking band at 722 cm^{-1}) are inaccessible with this method. The described limitations of mid-IR transmission spectroscopy of whole milk clearly indicate the advantages of the approach based on ATR FT-IR spectroscopy and lipid separation in order to resolve more spectral details related to fatty acids.

Fatty Acid Quantification

A set of eight human milk samples from lactating mothers with mean (interquartile range, IQR) gestational and postnatal ages of 35 + 4 (7 + 4) weeks + days and 41 (131) days, respectively, was used in this study. Human milk samples were characterized with the Miris HMA providing median (IQR) concentrations of 2.8 (1.1) g/dL for fat, 1.4 (0.7) and 1.1 (0.5) g/dL for crude and true protein, 7.7 (0.6) g/dL for carbohydrate, 12.2 (1.9) g/dL total solid, and 63 (14) kcal/dL for energy.

In order to predict the most abundant fatty acid sum parameters, individual PLS models were calculated by relating the acquired ATR FT-IR spectra to GC-MS reference concentrations. Optimum results were achieved by first treating the spectra of pure milk fat with a pre-processing routine that combines calculation of the second derivative as well as mean centering. For each PLS model, the applied wavenumber range

Table I. Overview of typical mid-IR absorbance bands of human milk fat.²⁹

Wavenumber (cm ⁻¹)	Group	Functional group	Mode of vibration	Detectable in whole milk ^a
3005	C–H	–C=CH– (cis)	Symm. stretch	No
2953	C–H	–CH ₃ (aliphatic)	Asymm. stretch	Yes
2922	C–H	–CH ₂ – (aliphatic)	Asymm. stretch	Yes
2853	C–H	–CH ₂ – (aliphatic)	Symm. stretch	Yes
1743	C=O	C=O ester	Stretch	Yes
1655	C=C	C=C (unsaturated)	Stretch	No
1462	C–H	–CH ₂ – (aliphatic)	Scissoring	Overlapping
1377	C–H	–CH ₃ (aliphatic)	Symm. deformation	Overlapping
1238	C–H	–CH ₂ – (aliphatic)	Out-of-plane bend	Overlapping
1162	C–O	C–O ester	Stretch	Overlapping
966	C–H	–C=CH– (trans)	Out-of-plane bend	No
722	C–H	–CH ₂ – (aliphatic)	Rocking	No

^a Detectability of the absorbance band in a FT-IR spectrum of whole human milk, recorded with a 50 μm CaF₂ flow cell. Symm.: symmetric; asymm.: asymmetric.

was individually selected (see next subchapter). The obtained statistical parameters for each prediction parameter are listed in Table II. The quality of calibration equations can be evaluated by the root mean square error of calibration (RMSEC) and calibration coefficient of determination (R^2). In order to assess the prediction efficiency, LOOCV was applied, providing the root mean square error of cross-validation (RMSECV) and the cross-validation coefficient of determination (R^2_{CV}). The number of latent variables (LVs) was selected by plotting the RMSEC and RMSECV versus the number of applied latent variables (see Figure S1, Supplemental Material). Excellent prediction efficiencies were obtained for SAT ($R^2_{CV} = 0.94$) and unsaturated fatty acids (UNSAT, $R^2_{CV} = 0.91$). Further sub-classification showed good predictions for MONO ($R^2_{CV} = 0.85$) and PUFA ($R^2_{CV} = 0.87$). Prediction results concerning concentrations of fatty acids with different chain length showed excellent results for medium-chain fatty acids (MCFA, C12–C16, $R^2_{CV} = 0.97$) as well as good results for short-chain fatty acids (SCFA, C4–C10, $R^2_{CV} = 0.79$) and long-chain fatty acids (LCFA, C17 and higher, $R^2_{CV} = 0.88$). Figure 2 highlights the quality of the achieved models by displaying the relationship between GC-MS reference concentrations and prediction results from the cross-validation on using SAT and MCFA as examples. The obtained concentrations for the fatty acid sum parameters also agree well with the typical human milk lipid composition of European women.⁴

In the present study, only a limited number of eight human milk samples were available. However, it was shown before that a higher number of samples can clearly improve the prediction efficiencies of fatty acids in bovine milk by IR spectroscopy combined with PLS quantification.¹⁶ Thus, even better results and also predictions for individual fatty acids can be expected when more samples are available. A further

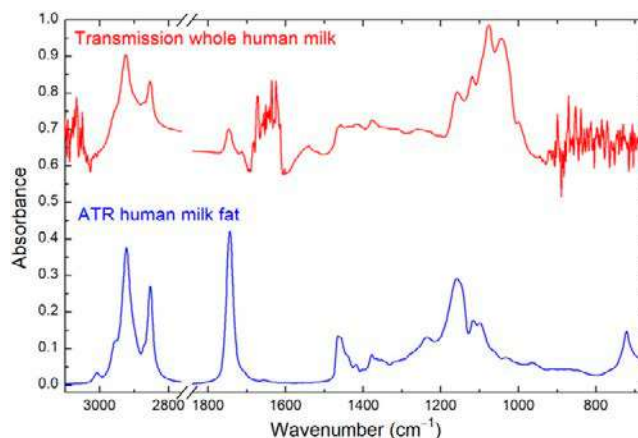


Figure 1. FT-IR absorbance spectrum of whole human milk recorded with a CaF₂ transmission cell (optical path length of 50 μm, red top) and ATR spectrum of the separated lipid fraction (blue below). The wavenumber region between 1850 cm⁻¹ and 2750 cm⁻¹ is not shown due to absence of significant information.

Table II. Statistical figures of merit for each individual PLS model.

Fatty acid	LVs	g/100 g fat				R^2	R^2_{CV}
		Range	RMSEC	RMSECV			
SAT	4	30.0–41.5	0.19	1.1	0.998	0.937	
MONO	3	35.6–54.7	0.82	2.2	0.978	0.846	
PUFA	3	12.7–28.5	0.71	1.7	0.978	0.873	
UNSAT	4	57.0–68.1	0.22	1.3	0.997	0.906	
SCFA	3	2.0–3.7	0.10	0.26	0.968	0.786	
MCFA	4	25.0–34.7	0.17	0.57	0.997	0.969	
LCFA	3	59.2–71.4	0.71	1.5	0.972	0.877	

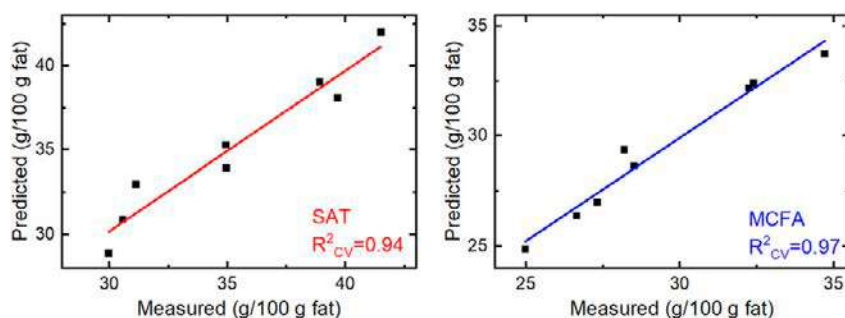


Figure 2. Measured (from GC-MS) versus predicted (from ATR FT-IR spectra and PLS cross-validation) relative fatty acid content on the example of saturated fatty acids (SATs, left) and medium-chain fatty acids (MCFAs, right).

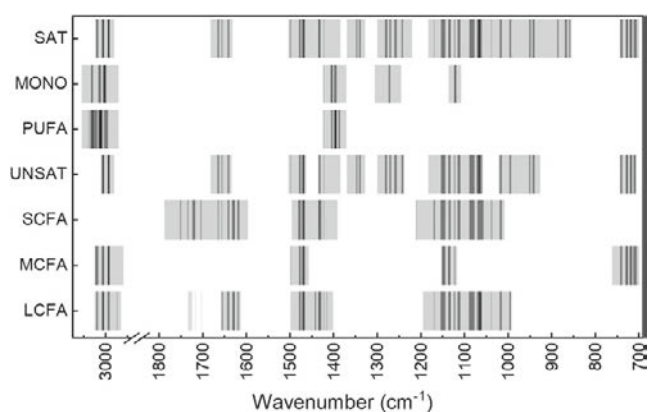


Figure 3. Included wavenumber regions for each individual calibration equation, according to the selectivity ratio (SR). SR of 0–0.5 (bright gray); 0.5–5 (dark gray); and 5–15 (black). The spectral range between 1850 and 2950 cm^{-1} is not shown due to absence of significant information.

important factor that might affect the performance of mid-IR-based determination of the milk fatty acid composition are possible covariation structures between individual fatty acid concentrations and total fat content. It has been reported that the good prediction efficiencies for fatty acids in whole bovine milk are probably based on these covariation structures when absolute fatty acid concentrations are investigated.²⁶ The applied approach inherently eliminates such covariation structures by measuring a representative fraction of pure milk lipids instead of whole milk samples.^{18,26}

Selection of Spectral Regions According to the Selectivity Ratio

In order to optimize each individual PLS model, the selectivity ratio (SR) was calculated. A detailed explanation of the SR can be found in literature.^{22,27} In brief, it can be described as the ratio between explained and unexplained variance for each variable in the dataset, meaning that important variables for predicting target parameters can be identified. Previously, the

SR has been demonstrated to be a strong tool for visualizing mid-IR regions with high correlation to PLS prediction parameters.^{18,28}

In the present study, the included wavenumber range was individually optimized for each PLS model. Figure 3 depicts the included mid-IR regions: wavenumbers with low SR are displayed in bright gray, whereas wavenumbers with high SR are highlighted in black. Moreover, Figures S2–S8, Supplemental Material show loading plots for the individual LVs of each PLS model. The SR and the loadings both indicate that all wavenumbers with high relevance can be related to fatty acid absorption bands listed in Table I. Relating important wavenumbers to specific absorbance bands is the main way to verify that the calculated PLS models are based on absorbance signals from fatty acids instead of coincidental correlations. It should be noted that a large part of these important wavenumbers is either completely inaccessible or highly overlapping with absorbances arising from other human milk constituents in transmission spectra of whole milk. Thus, the advantage of the applied approach based on lipid separation and ATR FT-IR spectroscopy is clearly visible.

Conclusion

In this work, the first approach based on mid-IR spectroscopy for determining the human milk fatty acid composition was presented. A representative lipid fraction was obtained by a solvent-free two-step centrifugation method, which was consequently measured with ATR FT-IR spectroscopy. The advantage of the fat separation step was demonstrated in the higher number of accessible spectral features from fatty acids, compared to transmission spectra of whole milk. Partial least squares models calculated by relating mid-IR spectra to GC-MS reference values showed good prediction accuracies for the important group parameters SAT, MONO, PUFA, SCFA, MCFA, and LCFA. The SR was calculated to assign important wavenumber regions to spectral features arising from fatty acid absorbance. Based on a dataset of eight samples, the achieved statistical figures of merit show the high potential of the applied approach for future high-throughput human milk

fatty acid profiling. We anticipate that even better results can be obtained, when a higher number of samples are available. Then prediction of individual fatty acids can also be expected, as previously shown for bovine milk.¹⁸

Acknowledgments

The authors are grateful to lactating mothers who agreed to participate in this study and to María Gormaz, Amparo Ramón, and Antonia Gálvez from the human milk bank at the University and Polytechnic Hospital La Fe for their support. The authors are also grateful to Prof. Oliver Spadiut and Julian Ebner from the Institute of Chemical, Environmental and Bioscience Engineering (Technische Universität Wien) for providing access to the applied centrifuge.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was funded by the European Union's Horizon 2020 Research and Innovation Programme through the Nutrishield project (<https://nutrishield-project.eu/>) (Grant Agreement No 818110) as well as the COMET Centre CHASE, funded within the COMET – Competence Centers for Excellent Technologies programme by the BMK, the BMDW and the Federal Provinces of Upper Austria and Vienna. The COMET programme is managed by the Austrian Research Promotion Agency (FFG). ITD and JK received funding from the Instituto de Salud Carlos III, Ministry of Economy and Competitiveness, Spain (grant numbers CD19/00176 and CPI6/00034).

Ethical Approval

The study was approved by the Ethics Committee for Biomedical Research of the Health Research Institute La Fe, University and Polytechnic Hospital La Fe (Valencia, Spain) with registry #2019-289-I and all methods were performed in accordance with relevant guidelines and regulations. Written informed consent was obtained from registered, healthy donors admitted after routine screening and interview to the human milk bank at the University and Polytechnic Hospital La Fe.

ORCID iDs

Christopher K. Akhgar  <https://orcid.org/0000-0001-8266-043X>
 Andreas Schwaighofer  <https://orcid.org/0000-0003-2714-7056>
 Erwin Rosenberg  <https://orcid.org/0000-0002-3608-4571>
 Bernhard Lendl  <https://orcid.org/0000-0003-3838-5842>

Supplemental Material

The supplemental material mentioned in the text, consisting of figures, is available in the online version of the journal.

References

1. D. Munblit, D.G. Peroni, A. Boix-Amorós, et al. "Human Milk and Allergic Diseases: An Unsolved Puzzle". *Nutrients*. 2017. 9(8): 894.
2. M.-Y. Chung. "Factors Affecting Human Milk Composition". *Pediatr Neonatol*. 2014. 55(6): 421–422.
3. O. Ballard, A.L. Morrow. "Human Milk Composition: Nutrients and Bioactive Factors". *Pediatr. Clin. N. Am.* 2013. 60(1): 49–74.
4. B. Koletzko, M. Rodríguez-Palmero, H. Demmelmair, et al. "Physiological Aspects of Human Milk Lipids". *Early Hum. Dev.* 2001. 65(Suppl. 2): 3–18.
5. K. Miliku, Q.L. Duan, T.J. Moraes, et al. "Human Milk Fatty Acid Composition Is Associated with Dietary, Genetic, Sociodemographic, and Environmental Factors in the CHILD Cohort Study". *Am. J. Clin. Nutr.* 2019. 110(6): 1370–1383.
6. B. Koletzko. "Human Milk Lipids". *Ann. Nutr. Metab.* 2016. 69(Suppl. 2): 27–40.
7. C. Cruz-Hernandez, S. Goeuriot, F. Giuffrida, et al. "Direct Quantification of Fatty Acids in Human Milk by Gas Chromatography". *J. Chromatogr. A*. 2013. 1284: 174–179.
8. A. Schwaighofer, M. Brandstetter, B. Lendl. "Quantum Cascade Lasers (QCLs) in Biomedical Spectroscopy". *Chem. Soc. Rev.* 2017. 46(19): 5903–5924.
9. G. Fusch, N. Rochow, A. Choi, et al. "Rapid Measurement of Macronutrients in Breast Milk: How Reliable are Infrared Milk Analyzers?" *Clin. Nutr.* 2015. 34(3): 465–476.
10. E.d.C.M. Motta, R.A. Zângaro, L. Silveira Jr. "Quantitative Determination of the Human Breast Milk Macro-Nutrients by Near-Infrared Raman Spectroscopy". *SPIE Photonics West*, 21–26 January 2012, San Francisco, California, USA: Proceeding SPIE, Optical Diagnostics and Sensing XII: Toward Point-of-Care Diagnostics; and Design and Performance Validation of Phantoms Used in Conjunction with Optical Measurement of Tissue IV. 2012. 8229: 82291F.
11. R. Ullah, S. Khan, S. Javaid, et al. "Raman Spectroscopy Combined with a Support Vector Machine for Differentiating Between Feeding Male and Female Infants Mother's Milk". *Biomed. Opt. Express*. 2018. 9(2): 844–851.
12. N. Argov, S. Wachsmann-Hogiu, S.L. Freeman, et al. "Size-Dependent Lipid Content in Human Milk Fat Globules". *J. Agric. Food Chem.* 2008. 56(16): 7446–7450.
13. G. Ramer, B. Lendl. "Attenuated Total Reflection Fourier Transform Infrared Spectroscopy". In: R. Meyers, R. Meyers, editors. *Encyclopedia of Analytical Chemistry*. Hoboken, USA: John Wiley and Sons, 2013.
14. K. Bērziņš, S.D.L. Harrison, C. Leong, et al. "Qualitative and Quantitative Vibrational Spectroscopic Analysis of Macronutrients in Breast Milk". *Spectrochim. Acta, Part A*. 2021. 246: 118982.
15. M. De Marchi, V. Toffanin, M. Cassandro, M. Penasa. "Invited Review: Mid-Infrared Spectroscopy as Phenotyping Tool for Milk Traits". *J. Dairy Sci.* 2014. 97(3): 1171–1186.
16. M.J.M. Rutten, H. Bovenhuis, K. Hettinga, et al. "Predicting Bovine Milk Fat Composition Using Infrared Spectroscopy Based on Milk Samples Collected in Winter and Summer". *J. Dairy Sci.* 2009. 92(12): 6202–6209.

17. I. Stefanov, V. Baeten, O. Abbas, et al. "Evaluation of FT-NIR and ATR FT-IR Spectroscopy Techniques for Determination of Minor Odd- and Branched-Chain Saturated and Trans Unsaturated Milk Fatty Acids". *J. Agric. Food Chem.* 2013. 61(14): 3403–3413.
18. C.K. Akhgar, V. Nürnberger, M. Nadvornik, et al. "Fatty Acid Prediction in Bovine Milk by Attenuated Total Reflection Infrared Spectroscopy After Solvent-Free Lipid Separation". *Foods.* 2021. 10(5): 1054.
19. S. Feng, A.L. Lock, P.C. Garnsworthy. "Technical Note: A Rapid Lipid Separation Method for Determining Fatty Acid Composition of Milk". *J. Dairy Sci.* 2004. 87(11): 3785–3788.
20. P. Luna, M. Juárez, M.A. Fuente. "Validation of a Rapid Milk Fat Separation Method to Determine the Fatty Acid Profile by Gas Chromatography". *J. Dairy Sci.* 2005. 88(10): 3377–3381.
21. Miris. "HMA User Manual". <https://www.mirissolutions.com/media/839b5db2-335e-4170-9284-F2d00a6295d8> [accessed May 26 2021].
22. M. Farrés, S. Platikanov, S. Tsakovski, R. Tauler. "Comparison of the Variable Importance in Projection (VIP) and of the Selectivity Ratio (SR) Methods for Variable Selection and Interpretation". *J. Chemom.* 2015. 29(10): 528–536.
23. A. Kohler, N. Afseth, K. Jørgensen, et al. "Quality Analysis of Milk by Vibrational Spectroscopy". In: C. Jim, G. Peter, editors. *Handbook of Vibrational Spectroscopy*. Hoboken, New Jersey: John Wiley and Sons, 2010.
24. A. Schwaighofer, B. Lendl. "Quantum Cascade Laser-Based Infrared Transmission Spectroscopy of Proteins in Solution". In: Y. Ozaki, M. Baranska, I. Lednev, B. Wood, editors. *Vibrational Spectroscopy in Protein Research*. Cambridge, Massachusetts: Academic Press, 2020. Chap. 3, Pp. 59–88.
25. T.G. Mayerhöfer, S. Pahlow, U. Hübner, J. Popp. "CaF₂: An Ideal Substrate Material for Infrared Spectroscopy?" *Anal. Chem.* 2020. 92(13): 9024–9031.
26. C. Eskildsen, M. Rasmussen, S. Engelsen, et al. "Quantification of Individual Fatty Acids in Bovine Milk by Infrared Spectroscopy and Chemometrics: Understanding Predictions of Highly Collinear Reference Variables". *J. Dairy Sci.* 2014. 97(12): 7940–7951.
27. O.M. Kvalheim. "Variable Importance: Comparison of Selectivity Ratio and Significance Multivariate Correlation for Interpretation of Latent-Variable Regression Models". *J. Chemom.* 2020. 34(4): E3211.
28. A.C.V.D. dos Santos, R. Heydenreich, C. Derntl, et al. "Nanoscale Infrared Spectroscopy and Chemometrics Enable Detection of Intracellular Protein Distribution". *Anal. Chem.* 2020. 92(24): 15719–15725.
29. M. Safar, D. Bertrand, P. Robert, et al. "Characterization of Edible Oils, Butters, and Margarines by Fourier Transform Infrared Spectroscopy with Attenuated Total Reflectance". *J. Am. Oil Chem. Soc.* 1994. 71(4): 371–377.

Publication IV

Akhgar, C. K.; Ramos-Garcia, V.; Nürnberger, V.; Moreno-Giménez, A.; Kuligowski, J.; Rosenberg, E.; Schwaighofer, A.; Lendl, B. *Solvent-free Lipid Separation and Attenuated Total Reflectance Infrared Spectroscopy for Fast and Green Fatty Acid Profiling of Human Milk*. *Food Chem.* Submitted.

Solvent-free Lipid Separation and Attenuated Total Reflectance Infrared Spectroscopy for Fast and Green Fatty Acid Profiling of Human Milk

Christopher Karim AKHGAR^a, Victoria RAMOS-GARCIA^b, Vanessa NÜRNBERGER^{a,c}, Alba MORENO-GIMÉNEZ^b, Julia KULIGOWSKI^b, Erwin ROSENBERG^a, Andreas SCHWAIGHOFER^a and Bernhard LENDL^{a,*}

^aInstitute of Chemical Technologies and Analytics, Technische Universität Wien, Getreidemarkt 9, 1060 Vienna, Austria; christopher.akhgar@tuwien.ac.at (C.K.A); vanessa.nuernberger@tuwien.ac.at (V.N.); egon.rosenberg@tuwien.ac.at (E.R.); andreas.schwaighofer@tuwien.ac.at (A.S.).

^bHealth Research Institute La Fe, Avenida Fernando Abril Martorell 106, 46026 Valencia, Spain. victoria_amos@iislafe.es (V.R.-G.); alba_moreno@iislafe.es (A.M.-G.); julia.kuligowski@uv.es (J.K.).

^cCompetence Center CHASE GmbH, Altenberger Straße 69, 4040 Linz, Austria

*Corresponding author: Bernhard Lendl, Institute of Chemical Technologies and Analytics, Technische Universität Wien, Getreidemarkt 9, 1060 Vienna, Austria. Email: bernhard.lendl@tuwien.ac.at; Tel.: +43 1 58801 15140

Abstract

This study presents the first mid-infrared (IR) based method capable of simultaneously predicting concentrations of individual fatty acids (FAs) and relevant sum parameters in human milk (HM). Representative fat fractions of 50 HM samples were obtained by rapid, two-step centrifugation and subsequently measured with attenuated total reflection IR spectroscopy. Partial least squares models were compiled of the acquired IR spectra with gas chromatography-mass spectrometry (GC-MS) reference data. External validation showed good results particularly for the most important FA sum parameters and the following individual FAs: C12:0 ($R^2_P=0.96$), C16:0 ($R^2_P=0.88$), C18:1cis ($R^2_P=0.92$) and C18:2cis ($R^2_P=0.92$). Based on the obtained results, the effect of different clinical parameters on the HM FA profile was investigated, indicating a change of certain sum parameters over the course of lactation. Finally, assessment of the method's greenness revealed clear superiority compared to GC-MS methods. The reported method thus represents a high-throughput, green alternative to resource-intensive established techniques.

1. Introduction

Human milk (HM) is a complex, dynamic, and bioactive fluid with the fundamental role of supplying infants with all essential nutrients before they are able to digest solid food (Van Winckel, Vande Velde, De Bruyne, & Van Biervliet, 2011). According to the World Health Organization (WHO), exclusive breastfeeding is recommended for the first six months of life, followed by complementary breastfeeding of up to two years or beyond (WHO, 2001). The HM macronutrient composition is known to be significantly influenced by factors such as prematurity or postnatal age in order to fulfil the nutritional demands across all stages of development (Chung, 2014). In this context, the total protein concentration significantly decreases, whereas fat and lactose content increase during the first weeks of lactation, before average concentrations of 1.0 g/dL protein, 3.4 g/dL fat and 6.7 g/dL lactose are reached for mature HM (Gidrewicz & Fenton, 2014).

HM fat occurs as milk fat globules emulsified in water, consisting of a milk fat globule membrane that envelopes a core, rich in triglycerides (TAGs). TAGs, consisting of a glycerol backbone and

three fatty acids (FAs), account for approximately 98% of total HM lipids (Jensen, Bitman, Carlson, Couch, Hamosh, & Newburg, 1995). The FA composition is the most variable macronutrient, known to be significantly influenced by factors such as mothers' genetics and diet, obesity, stage of lactation, socio-demographic variables, and the environment (de la Garza Puentes et al., 2019; Miliku et al., 2019). In this context, it has been demonstrated that the HM FA composition varies significantly between mothers with omnivore, vegan, and vegetarian diets (Sanders & Reddy, 1992). Furthermore, the FA profile of HM from European mothers with western-diet differs in the amount of saturated FAs (SAT), monounsaturated FAs (MONO), and polyunsaturated FAs (PUFA) from lactating mothers with other forms of diet (Koletzko, Thiel, & Abiodun, 1992). With progress of lactation, the relative content of medium-chain FAs (MCFA, C11-C16,) increases, while long-chain FAs (LCFA, C17 and higher) and PUFAs decrease (Boersma, Offringa, Muskiet, Chase, & Simmons, 1991; Floris, Stahl, Abrahamse-Berkeveld, & Teller, 2020).

When not enough own mother's milk is available, pasteurized donor HM (DHM) is preferred over infant formula for feeding of preterm infants, offering immunological advantages and lower rates of necrotising enterocolitis (Poulimeneas et al., 2021). The established method to destroy vegetative bacteria and viruses in DHM, thus ensuring safety for the infants, is "Holder" pasteurization (HoP), where samples are heated to 62.5 °C for 30 min (Colaizy, 2021). The effect of HoP on HM lipids is still a topic of discussion, as available studies report different results, ranging from no influence to a decrease of total fat content and certain FAs (Gao, Miller, Middleton, Huang, McPhee, & Gibson, 2019; Ten-Doménech et al., 2022). Consequently, reliable analytical methods for monitoring the FA profile of DHM are of major interest.

Established methods for determining the FA profile in HM are based on gas chromatography (GC) with different detectors such as flame ionization detectors (FID) or mass spectrometry (MS) (Cruz-Hernandez, Goeuriot, Giuffrida, Thakkar, & Destailats, 2013; Ten-Doménech et al., 2022). These methods offer excellent sensitivity and accuracy, but bear severe drawbacks such as expensive instrumentation, time intensive chromatographic runs of approximately 1-2 hours and a labour-intensive derivatization step prior to analysis, typically involving hazardous chemicals. Particularly, in the context of green analytical chemistry (GAC), there is a growing demand for

more time- and energy-efficient methods that do not involve toxic substances or produce large amounts of waste (Armenta, Garrigues, & de la Guardia, 2008).

Mid-infrared (IR) spectroscopy is capable of providing qualitative and quantitative information about clinically relevant parameters by detecting absorption of IR radiation through fundamental molecular vibrations (Schwaighofer, Brandstetter, & Lendl, 2017). Fourier-transform infrared (FTIR) instruments are nowadays considered as the gold-standard, offering rapid spectra acquisition and high accuracy while covering the whole mid-IR region ($400\text{-}4000\text{ cm}^{-1}$). Compared to GC-MS, this technique offers advantages in terms of reduced measurement times, lower cost, minimum to no sample preparation and non-destructive operation. Commercially available mid-IR analyzers, operating in transmission configuration, tailored for HM analysis (e.g., Miris Ab Human Milk Analyzer (HMA), Sweden) allow for determination of the HM macronutrient composition without the need of specialized staff or expensive laboratory facilities. It has been demonstrated that results for total protein, fat, and lactose content obtained with these analysers are widely acceptable (Fusch et al., 2015). As an alternative to transmission measurements, attenuated total reflection (ATR) is often applied as a more robust probing mode. Here, the IR beam is totally reflected at the interface between an optically denser ATR crystal and the optically rarer sample, leading to the formation of an evanescent wave that can interact with the sample at typical penetration depths of $1\text{-}2\text{ }\mu\text{m}$ (Ramer & Lendl, 2013).

Predicting the FA composition of milk from different mammals by IR spectroscopy is usually performed by relating mid-IR transmission spectra of whole milk to GC reference concentrations by calculation of multivariate partial least squares (PLS) regression models (De Marchi, Toffanin, Cassandro, & Penasa, 2014; Ferrand-Calmels et al., 2014). This approach, however, suffers from the overlap of the target TAG bands with large absorbance bands of water, proteins, and lactose in whole milk transmission spectra. Consequently, the obtained results are dominated by cross-correlations between individual FA concentrations and total fat content (Eskildsen, Rasmussen, Engelsen, Larsen, Poulsen, & Skov, 2014). In this context, it has been demonstrated that fat separation prior to mid-IR spectra acquisition is highly beneficial, as in this way cross-correlations between individual FAs and total fat are inherently eliminated and FA related features are better resolved than for whole milk spectra (Akhgar et al., 2021). Most recently, a solvent-free lipid separation method (Luna, Juárez, & Fuente, 2005) was combined with ATR-FTIR spectroscopy

to show the proof-of-concept for mid-IR based FA profiling in HM based on a limited data set of eight samples (Akhgar et al., 2022). Here, good PLS models were obtained for some important FA sum parameters such as SAT, UNSAT, MCFA, and LCFA.

In the present work, the developed approach was expanded to a larger data set, involving unprocessed DHM samples, as well as samples that underwent HoP. A comprehensive method validation based on splitting of the available samples into a training and a validation set was performed, showing very good prediction results for the most important sum parameters as well as certain individual FAs. Subsequently, the obtained PLS models were applied to investigate the effect of different clinical parameters including the stage of lactation, birth weight, infant's sex, gestational age, maternal age, and HoP on the HM FA profile, which was previously only possible by time- and labour-intensive GC methods. Finally, the greenness of the method was demonstrated based on a comparison to GC-MS reference methods by applying the Analytical GREENness (AGREE) metric approach (Pena-Pereira, Wojnowski, & Tobiszewski, 2020).

2. Material and methods

2.1 Human milk samples

The study was approved by the Ethics Committee for Biomedical Research of the Health Research Institute La Fe, University and Polytechnic Hospital La Fe (Valencia, Spain) with approval number 2019-289-1. All methods were performed in accordance with relevant guidelines and regulations. Written informed consent was obtained from healthy donors admitted after routine screening and interview to the Human Milk Bank at the University and Polytechnic Hospital La Fe.

Human milk samples ($N=50$) with volumes ranging between 50 and 200 mL were collected following the instructions of the hospital staff. Milk expression was accomplished with breast milk pumps following the standard operating procedure routinely employed at the hospital and the Human Milk Bank. Prior to extraction, both, removable parts of the breast milk pump and collection bottles, were sterilized. In addition, mothers must wash their hands with soap and water and skin area that gets in contact with the milk pump with water. HM was extracted at participant's homes and stored at $-20\text{ }^{\circ}\text{C}$ before being transported to the Human Milk Bank. Extracted HM ($N=26$) was stored at $-20\text{ }^{\circ}\text{C}$ until further processing. The remaining HM samples ($N=24$) were pooled HM samples obtained by mixing several aliquots extracted at different days from the same

donor and underwent HoP prior to storage at $-20\text{ }^{\circ}\text{C}$ until further processing. Data including maternal age, infant's postnatal age, infant's sex, infant's birth weight, and gestational age were collected.

2.2 Macronutrient Analysis

Human milk macronutrients (i.e., fat, carbohydrate, crude and true protein, total solids, and energy) were determined using a MIRIS HMA™ (MIRIS AB, Uppsala Sweden). Determinations, quality control, and instrument calibration were performed following the standard operation procedure provided by the manufacturer.

2.3 Solvent-free lipid separation

HM lipid separation was performed according to a previously proposed (Feng, Lock, & Garnsworthy, 2004) and later modified (Luna et al., 2005) solvent-free centrifugation method. First, 30 mL HM sample aliquots were thawed overnight at $4\text{ }^{\circ}\text{C}$, followed by tempering at room temperature for at least 20 min. Subsequently, samples were centrifuged for 30 min at $17800\times g$ and $20\text{ }^{\circ}\text{C}$ in a Sigma 3-18k centrifuge (Sigma Laborzentrifugen GmbH, Germany). The upper cake was transferred into microtubes and centrifuged for 20 min at $19300\times g$ at the same temperature. The obtained upper layer, consisting of pure HM lipids was removed and used for ATR-FTIR and GC-MS measurements.

2.4 ATR-FTIR measurements

FTIR absorption measurements of HM lipids were performed using a Bruker Tensor 37 FTIR spectrometer (Ettlingen, Germany) equipped with a Bruker Optics Platinum ATR module and a liquid nitrogen cooled HgCdTe (mercury cadmium telluride) detector. IR spectra of one drop of pure HM fat were acquired in the range from 600 to 4000 cm^{-1} with a spectral resolution of 2 cm^{-1} in double-sided acquisition mode. A total of 128 scans (52 s acquisition time) were averaged per spectrum, which was calculated using a Blackman-Harris 3-term apodization function and a zero-filling factor of 2. After each HM fat measurement, the ATR crystal was cleaned with Glucopon 600 CS UP solution ($\sim 50\%$ in water, Sigma-Aldrich, Steinheim, Germany) and water to recover the initial baseline signal. In order to reduce the influence of water vapor from the atmosphere on the recorded spectra, the instrument was constantly flushed with dry air. Spectra were acquired and analyzed using the software package OPUS 8.1 (Bruker, Ettlingen, Germany).

2.5 GC-MS reference measurements

For GC-MS reference measurements, solutions of 20 mg/mL pure HM fat in dichloromethane were prepared. 50 μ L of these solutions were transferred into cooled glass vials with micro inserts and mixed with 50 μ L internal standard (C15:0) and 50 μ L of trimethylsulfonium hydroxide (TMSH, 0.25 M in MeOH, Supelco, Vienna) derivatization agent. Subsequently, vials were immediately capped and vortexed for 5 seconds, following by heating to 70 °C for 15 min in order to complete the derivatization reaction.

A Shimadzu GC-2010 Plus Chromatography system coupled to a single-quadrupole mass spectrometer (Shimadzu, Japan), equipped with a ZB-FAME column (30 m, 0.25 mm I.D., 0.20 μ m film thickness; Phenomenex, Germany) was used for all GC-MS measurements. 1 μ L of the derivatized solutions were injected in split mode using a split of 100:1, by a Shimadzu AOC-5000 Plus autosampler. The injector temperature was set to 250 °C. A constant septum flow rate of 3 mL/min was applied with a column flow rate of 2.14 mL/min. The initial oven temperature of 40 °C was held for 3 minutes, then increased by 10 °C per minute to 100 °C, and further increased by 2 °C per minute to 200 °C. Ion source and transferline temperature were both kept constant at 200 °C. After a solvent vent of 2.7 minutes, data acquisition was performed in scan mode (m/z 35-500) with a detector voltage of 1.05 kV.

For quantitative analysis, calibration was performed by preparing appropriate dilutions of a 37 component FAME mix certified reference material (TraceCERT®, Supelco). Quantitative evaluation was performed based on the quantifier ion peak area for each FAME, if the ratio of quantifier and qualifier ions was within acceptable thresholds. Retention times, qualifier and quantifier ions for each analyte are reported elsewhere (Akhgar et al., 2021).

2.6 Data Analysis

PLS models were calculated in Matlab R2020a (Mathworks Inc., Natick, MA, USA), using the PLS Toolbox 8.9 (Eigenvector Research Inc., Wenatchee, WA, USA). All ATR spectra of pure HM lipids were first preprocessed by mean centering and calculation of 2nd derivative spectra, using a 2nd order polynomial Savitzky-Golay filter with a window of 15 points. For each target parameter, the included spectral region was individually selected based on the PLS selectivity ratio (SR) (Farrés, Platikanov, Tsakovski, & Tauler, 2015). First, calibration equations were calculated based on a training set ($N=35$) and evaluated with contiguous blocks cross-validation with 10 data

splits. Second, validation was performed by using these calibrations to predict the concentrations of an external validation set ($N=15$) and calculating characteristic figures of merit.

3. Results and Discussion

3.1 Macronutrient Analysis

A set of 50 HM samples, including 26 HM samples collected prior to HoP and 24 HM samples collected post-pasteurization, was used in this study. The median (interquartile range, IQR) birth weight was 2911 (786) g, whereas maternal, gestational, and postnatal ages were 35 (6) years, 39 + 2 (2 + 2) weeks + days and 158 (213) days, respectively. 25 HM samples were from mothers of female infants. The macronutrient concentration was determined by the mid-IR based Miris HMA™, showing median (IQR) results of 8.0 (0.2) g/dL for carbohydrates, 1.8 (0.8) g/dL for total fat, 0.6 (0.5) g/dL for crude protein, 0.4 (0.4) g/dL for true protein, 10.5 (0.7) g/dL for total solids and 51 (7) kCal/dL for energy. Compared to the macronutrient composition of mature HM from literature, the obtained results for carbohydrates are slightly higher, whereas fat and protein concentrations are lower (Gidrewicz et al., 2014). This study relies on the use of pooled DHM as provided by regular donors enrolled at the hospital HM bank. While this procedure allows the study of samples that are similar to those administered at the hospital, the study protocol did not control for clinical variables possibly affecting the macronutrient composition of HM, such as postnatal age and gestational age (term or preterm delivery), or variables related to sample collection such as expression type, time after last feeding/expression, and time of the day (Ten-Doménech et al., 2022).

3.2 Mid-IR based determination of the fatty acid profile

TAGs show characteristic absorbance bands in the mid-IR region that are influenced by factors such as FA degree of saturation and chain length (De Ruig & Dijkstra, 1975; Kaylegian, Lynch, Fleming, & Barbano, 2009). Milk, however, is a highly complex matrix with various IR absorbance signals originating from other major components such as water, proteins and lactose that overlap with the significantly smaller bands from TAGs and that adversely affect their evaluation (Akhgar et al., 2021; Kohler, Afseth, Jørgensen, Randby, & Martens, 2010). Consequently, in this work, a representative part of the lipid fraction was isolated from HM samples by a solvent-free centrifugation method, before measuring the purified lipid fraction with ATR-FTIR spectroscopy.

Individual PLS calibration models were calculated for each target parameter by relating the recorded ATR-FTIR spectra to GC-MS reference concentrations. The pre-processing step combined mean centering and calculation of second derivatives for all spectra (Akhgar et al., 2021). The recorded ATR-FTIR spectra comprised the entire available mid-IR range (600-4000 cm^{-1}), however, the wavenumber range incorporated in PLS analysis was individually selected for each target parameter based on the PLS selectivity ratio (SR). This is a powerful tool for visualizing features of mid-IR spectra with high correlation to the predicted parameter. In brief, it can be defined as the ratio between explained and unexplained variance for each variable in the data set (Farrés et al., 2015). The selected wavenumber regions for the applied PLS models, as well as their SR are shown in Figure S1.

Table 1 shows the obtained statistical parameters for each PLS model. The presented approach predicts relative FA concentrations in g/100 g fat, which inherently eliminates previous problems regarding covariation structures in mid-IR based FA profiling (Eskildsen et al., 2014). A training set of 35 HM samples was used to calculate calibration models including typical figures of merit such as the root mean square error of calibration (RMSEC) and the calibration coefficient of determination (R^2). Mathematical descriptions of all applied figures of merit are collected in the supplementary information. For internal validation, a contiguous blocks cross-validation with 10 data splits was performed, and the root mean square error of cross-validation (RMSECV) and cross-validation coefficient of determination (R^2_{CV}) were determined. The applied number of latent variables (LVs) was individually selected for each parameter by plotting RMSEC and RMSECV versus the number of LVs (see Figure S2 and Figure S3). Here, the optimum number of LVs was found to be between 2 and 6, which is reasonable for a highly complex mixture such as milk fat, containing a large number of different components that can influence the spectra (Akhgar et al., 2021).

Figure 1 shows the relationship between GC-MS reference concentrations and ATR-FTIR based predictions of PLS cross-validation on the examples of SAT and LCFA. Under ideal conditions, all points would fall on the straight regression line, whereas higher and lower points indicate over- and underestimation of mid-IR measurements compared to the reference values. In the present case, the high R^2_{CV} values of 0.98 and 0.95 as well as the small deviations of data points from the regression lines indicate excellent performance of the models. Relationships for the other PLS

models can be found in Figure S4. For a more comprehensive validation, the calculated PLS models were used to determine the concentrations of an external validation set, comprising 15 HM samples. In order to evaluate the accuracy of the predicted concentrations, the root mean square error of prediction (RMSEP) and the prediction coefficient of determination (R^2_P) were calculated. Here, excellent results were obtained for the sum parameters SAT ($R^2_P=0.97$), MONO ($R^2_P=0.94$), PUFA ($R^2_P=0.93$) and UNSAT ($R^2_P=0.97$) that provide information on the degree of saturation and are of particular interest as their concentration changes depending on the mothers diet (Koletzko et al., 1992). Furthermore, excellent results were obtained for MCFA ($R^2_P=0.94$) and LCFA ($R^2_P=0.94$) that provide information regarding FA chain-length and these sum parameters are of special concern as their concentrations change with the stage of lactation (Boersma et al., 1991; Floris et al., 2020). Results for the group of short-chain fatty acids (SCFA, C10 and lower, $R^2_P=0.25$) were significantly poorer, which is likely based on the fact that the concentration of these FAs is very low in HM. Regarding the prediction of individual FAs, excellent results were obtained for C12:0, C18:1cis and C18:2cis ($R^2_P \geq 0.92$), whereas moderate predictions were obtained for C14:0, C16:0 and C18:0 ($R^2_P \geq 0.71$). It should be noted that particularly good prediction results were obtained for the two most abundant cis UNSAT, however, the results for C16:1cis were significantly poorer ($R^2_P=0.19$), which is likely based on the lower concentration of this FA which dropped below the limit of quantitation of the GC-MS reference method for certain samples.

The number of available samples can significantly influence the quality of PLS models. In the present study, the number of samples was noticeably increased compared to a previous study, where the herein applied approach for HM FA analysis was developed and tested (Akhgar et al., 2022). Compared to these previous results, particularly the PLS models for SAT, MONO, PUFA, UNSAT and LCFA could be clearly improved. Furthermore, it was previously not possible to determine the concentrations of individual FAs. In the present work, some of the most abundant FAs were predicted with high accuracy, consequently representing the first mid-IR based approach capable of determining the most important FA sum parameters and individual FAs in HM simultaneously.

Table 1: Results of partial least squares (PLS) regressions for determining relative fatty acid concentrations and sum parameters in g/100g fat.

Fatty Acid	LVs	Range	g/100 g Fat				Validation set (N=15)	
			Training set (N=35)				RMSEP	R ² _p
			RMSEC	R ²	RMSECV	R ² _{cv}		
SAT	6	21–40	0.42	0.99	0.68	0.98	0.54	0.97
MONO	5	29–51	1.1	0.96	1.4	0.94	1.3	0.94
PUFA	4	16–47	1.3	0.96	1.5	0.95	1.4	0.93
UNSAT	6	60–79	0.42	0.99	0.68	0.98	0.54	0.97
SCFA	4	0.40–1.7	0.13	0.78	0.18	0.62	0.15	0.25
MCFA	4	17–36	0.99	0.96	1.3	0.93	0.65	0.94
LCFA	4	63–82	0.87	0.97	1.1	0.95	0.70	0.94
C8:0	2	0.0–0.15	0.022	0.59	0.026	0.52	0.017	0.29
C10:0	4	0.40–1.5	0.12	0.79	0.16	0.63	0.15	0.20
C12:0	6	1.1–7.2	0.18	0.99	0.30	0.96	0.19	0.96
C14:0	6	1.5–7.9	0.26	0.98	0.44	0.94	0.63	0.71
C16:0	6	12–22	0.78	0.92	1.1	0.84	1.1	0.88
C16:1cis	4	0.0–1.2	0.29	0.32	0.42	0.017	0.31	0.19
C18:0	4	1.8–5.9	0.41	0.77	0.55	0.60	0.47	0.75
C18:1cis	6	28–51	1.1	0.96	1.6	0.92	1.4	0.92
C18:2cis	4	16–47	1.2	0.97	1.5	0.95	1.5	0.92

Abbreviations: LVs: latent variables; RMSEC: root mean square error of calibration; R²: calibration coefficient of determination; RMSECV: root mean square error of cross-validation; R²_{cv}: cross-validation coefficient of determination; RMSEP: root mean square error of prediction; R²_p: prediction coefficient of determination; SAT: saturated fatty acids; MONO: monounsaturated fatty acids; PUFA: polyunsaturated fatty acids; UNSAT: unsaturated fatty acids; SCFA: short-chain fatty acids (C10 and lower); MCFA: medium-chain fatty acids (C11–C16); LCFA: long-chain fatty acids (C17 and higher).

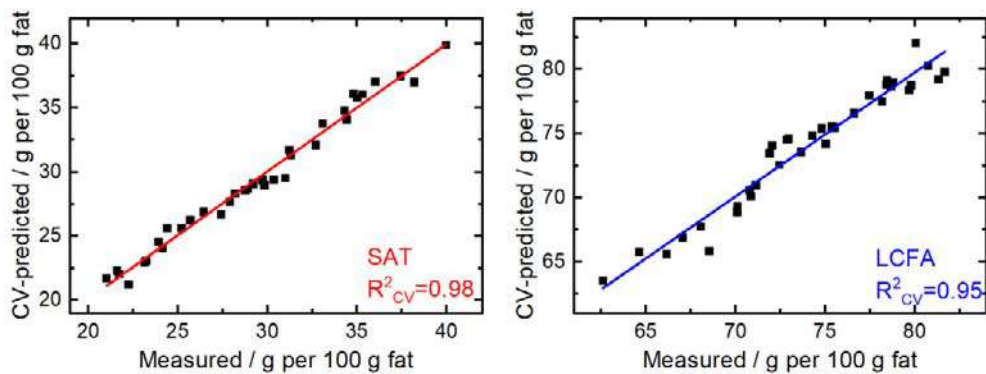


Figure 1: Measured (GC-MS) versus predicted (ATR-FTIR, cross-validation) fatty acid content for the example of saturated fatty acids (SAT, left) and long-chain fatty acids (LCFA, right).

3.3 Investigation of clinically relevant parameters

The presented method based on ATR-FTIR spectroscopy was applied to investigate possible effects of clinical parameters on the HM FA profile. For this purpose, the PLS models from Table 1 were employed to predict the FA profile in all available milk samples. First, changes in the FA profile over the course of lactation were monitored by plotting the obtained concentrations against

the days after delivery. Figure 2 shows the results for MCFA, LCFA and PUFA, as these sum parameters were previously found to significantly change over the course of lactation in a meta study with pooled data analysis (Floris et al., 2020). The results for the remaining sum parameters are shown in Figure S5. The linear fit, as well as the p -value (two-tailed) of 0.02 for MCFA indicate a significant increase of this sum parameter over the course of lactation. These findings agree well with the pooled data analysis (Floris et al., 2020). Furthermore, Figure 2 shows a significant decrease of LCFA with progressing lactation, which also was found in the above mentioned meta-study (Floris et al., 2020). Finally, the herein observed reduction of PUFA was also reported in the meta-study.

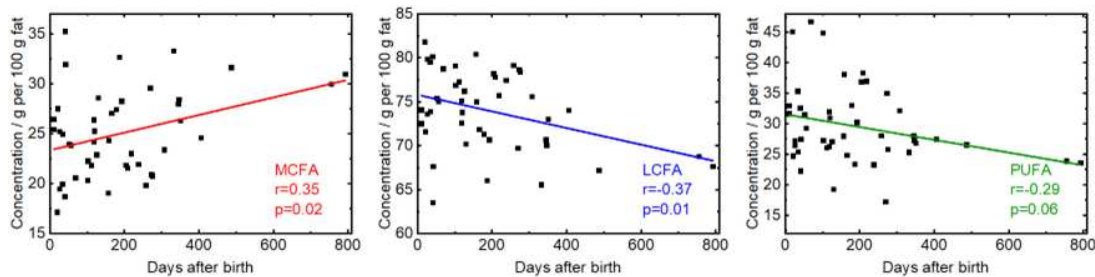


Figure 2: Concentration of medium-chain fatty acids (MCFA), long-chain fatty acids (LCFA) and polyunsaturated fatty acids (PUFA) determined with ATR-FTIR spectroscopy over the course of lactation.

In the next step, the FA profile was compared between HM samples prior to HoP and post-pasteurization. The boxplots in Figure S6 show no significant differences for the most important FA sum parameters in pre- and post-pasteurized DHM. These results agree well with most previous studies that investigated the effect of HoP on the relative HM FA profile reporting no significant difference (Delgado, Cava, Delgado, & Ramírez, 2014; Moltó-Puigmartí, Permanyer, Castellote, & López-Sabater, 2011; Peila et al., 2016; Pitino et al., 2019). In a recent study based on direct derivatization of FAs from HM and subsequent GC-MS analysis, a small but significant decrease of certain absolute FA concentrations (particularly SAT, LCFA, MONO, PUFA) was found following HoP (Ten-Doménech et al., 2022). However, the design of this study involved the analysis of paired pre- and post-pasteurization DHM samples, which eliminates biological background variation that might exist between donors and samples and therefore allows a more sensitive evaluation of changes introduced during HoP.

Finally, evaluation of the parameters gestational age, maternal age, birth weight and infant's sex with the FA profile did not result in any significant correlations or between group differences.

In summary, the presented mid-IR based method bears high potential for studying the effect of environmental parameters or correlations with demographic, anthropometric, or clinical variables on the HM FA profile in clinical studies. Particularly, the short measurement times allow for the analysis of a large number of samples vastly exceeding the throughput of GC methods.

3.4 Greenness evaluation

The presented method based on ATR-FTIR spectroscopy and solvent-free lipid separation was evaluated in terms of accordance with green analytical chemistry (GAC) principles (Armenta et al., 2008). For this purpose, the previously introduced Analytical GREENness (AGREE) metric approach (Pena-Pereira et al., 2020) was applied, representing a straight-forward assessment tool that considers the twelve principles of GAC (SIGNIFICANCE) (Gałuszka, Migaszewski, & Namieśnik, 2013). To compare the environmental impact with conventional methods for HM FA determination, ATR-FTIR spectroscopy was juxtaposed with the applied GC-MS reference method and with a GC-MS method from literature, where the fatty acids are derivatized directly from whole HM (Ten-Doménech et al., 2022). The results are shown in Figure 3 in form of clock-like graphs comprising of twelve segments that represent the SIGNIFICANCE criteria. Here, the performance of each principle is represented by a green-yellow-red (i.e., best to worst) colour range, whereas the overall performance is expressed by the colour and score (0=worst, 1=best) in the middle. The selected options are shown in detail in Table S1. Based on the total scores of 0.66 for FTIR, 0.33 for the GC-MS reference method, and 0.36 for the GC-MS method from literature, the presented mid-IR based method shows clear superiority in terms of environmental impact. Compared to the GC methods, its main advantages are high-throughput, reduced energy consumption, the lack of a derivatization step and the absence of hazardous organic solvents. Furthermore, it bears high potential for additional automatization and reduction of waste by replacing the applied disposables such as centrifugation tubes by reusable material. Consequently, the greenness of the presented method based on ATR-FTIR spectroscopy and solvent free-lipid separation can be further improved in the future.

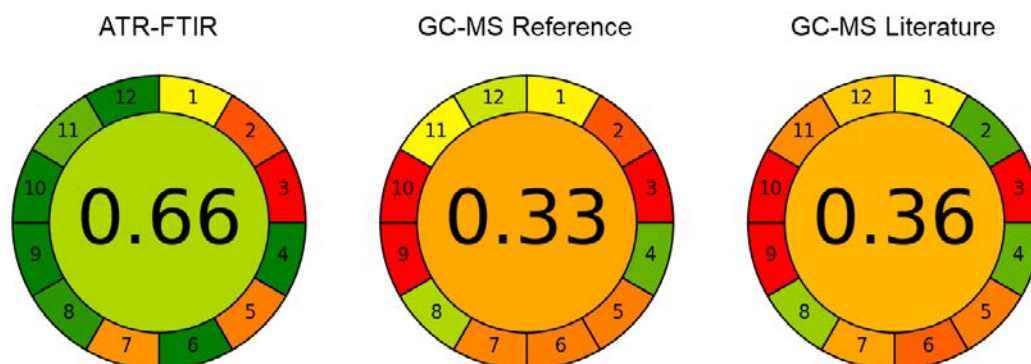


Figure 3: Results of evaluation according to the Analytical GREENness metric approach for ATR-spectroscopy with solvent free lipid separation, the applied GC-MS reference method and a GC-MS method from literature (Ten-Doménech et al., 2022). The a green-yellow-red (i.e., best to worst) colour range represents the performance of each principle.

4. Conclusion

In this study, the first mid-IR based approach for predicting the individual FA profile of HM was presented. The method comprises a solvent-free lipid separation step for obtaining a representative milk fat fraction which is then analyzed with ATR-FTIR spectroscopy. A diverse sample set composed of both, unprocessed and pasteurized DHM samples was employed for method development and validation. PLS models were established by relating the obtained IR spectra with GC-MS reference concentrations. Compared to a previous study with a limited data set, PLS models for the most relevant FA sum parameters were noticeably improved (Akhgar et al., 2022). Furthermore, excellent results were obtained for the individual FAs C12:0, C16:0, C18:0, C18:1cis and C18:2cis. Subsequently, these models were applied to investigate possible correlations between different clinical parameters and the HM FA profile. Here, an increase of MCFA over the course of lactation was observed, whereas LCFA and PUFA decreased, which agreed well with previous studies based on GC-MS data (Floris et al., 2020). Finally, the greenness of the presented method was compared to conventional GC-MS methods. Here, beneficial properties such as low energy consumption, lack of a derivatization step and the solvent-free nature resulted in a clear superiority of the mid-IR based method. Consequently, ATR-FTIR combined with solvent-free lipid separation represents a high-throughput, green analytical method allowing for accurate determination of the HM FA composition, which was previously only possible by time-consuming and resource-intensive GC analysis. Finally, the availability of a rapid analytical method for FA profiling is highly relevant in the clinical field to be able to conduct largescale studies within a

reasonable timeframe and workload, as the effect of diverse factors on the FA composition is still topic of discussion.

Funding resources

This work was funded by the European Union's Horizon 2020 Research and Innovation Programme through the Nutrishield project (<https://nutrishield-project.eu/>) [Grant Agreement No 818110] as well as the COMET Center CHASE (project No 868615), which is funded within the framework of COMET (Competence Centers for Excellent Technologies) by BMVIT, BMDW, and the Federal Provinces of Upper Austria and Vienna. The COMET program is run by the Austrian Research Promotion Agency (FFG). This research was funded by *Instituto de Salud Carlos III* and co-funded by the European Union (grant numbers CP16/00034 and CPII21/00003). The authors acknowledge TU Wien Bibliothek for financial support through its Open Access Funding Programme.

Data availability statement

The acquired ATR-FTIR spectra of 50 HM samples can be found online under: <https://doi.org/10.5281/zenodo.7064893>. Further data are available on request from the corresponding author.

Appendix A. Supplementary data

Supplementary data to this article can be found online.

Conflict of interest statement

The authors declare no competing financial interests.

References

- Akhgar, C. K., Nürnberger, V., Nadvornik, M., Ramos-Garcia, V., Ten-Doménech, I., Kuligowski, J., . . . Lendl, B. (2022). Fatty Acid Determination in Human Milk Using Attenuated Total Reflection Infrared Spectroscopy and Solvent-Free Lipid Separation. *Applied Spectroscopy*, 76(6), 730-736. <https://doi.org/10.1177/00037028211065502>.
- Akhgar, C. K., Nürnberger, V., Nadvornik, M., Velik, M., Schwaighofer, A., Rosenberg, E., & Lendl, B. (2021). Fatty Acid Prediction in Bovine Milk by Attenuated Total Reflection Infrared Spectroscopy after Solvent-Free Lipid Separation. *Foods*, 10(5), 1054. <https://www.mdpi.com/2304-8158/10/5/1054>.
- Armenta, S., Garrigues, S., & de la Guardia, M. (2008). Green Analytical Chemistry. *TrAC Trends in Analytical Chemistry*, 27(6), 497-511. <https://doi.org/https://doi.org/10.1016/j.trac.2008.05.003>.
- Boersma, E. R., Offringa, P. J., Muskiet, F. A., Chase, W. M., & Simmons, I. J. (1991). Vitamin E, lipid fractions, and fatty acid composition of colostrum, transitional milk, and mature milk: an international

- comparative study. *The American Journal of Clinical Nutrition*, 53(5), 1197-1204. <https://doi.org/10.1093/ajcn/53.5.1197>.
- Chung, M.-Y. (2014). Factors Affecting Human Milk Composition. *Pediatrics & Neonatology*, 55(6), 421-422. <https://doi.org/10.1016/j.pedneo.2014.06.003>.
- Colaizy, T. T. (2021). Effects of milk banking procedures on nutritional and bioactive components of donor human milk. *Seminars in Perinatology*, 45(2), 151382. <https://doi.org/https://doi.org/10.1016/j.semperi.2020.151382>.
- Cruz-Hernandez, C., Goeuriot, S., Giuffrida, F., Thakkar, S. K., & Destailhats, F. (2013). Direct quantification of fatty acids in human milk by gas chromatography. *Journal of Chromatography A*, 1284, 174-179. <https://doi.org/https://doi.org/10.1016/j.chroma.2013.01.094>.
- de la Garza Puentes, A., Martí Alemany, A., Chisaguano, A. M., Montes Goyanes, R., Castellote, A. I., Torres-Espínola, F. J., . . . López-Sabater, M. C. (2019). The Effect of Maternal Obesity on Breast Milk Fatty Acids and Its Association with Infant Growth and Cognition-The PREOBE Follow-Up. *Nutrients*, 11(9). <https://doi.org/10.3390/nu11092154>.
- De Marchi, M., Toffanin, V., Cassandro, M., & Penasa, M. (2014). Invited review: Mid-infrared spectroscopy as phenotyping tool for milk traits. *Journal of Dairy Science*, 97(3), 1171-1186. <https://doi.org/10.3168/jds.2013-6799>.
- De Ruig, W. G., & Dijkstra, G. (1975). Characterization and Identification of Triglycerides by Infrared Spectroscopy. *Fette, Seifen, Anstrichmittel*, 77(6), 211-216. <https://doi.org/https://doi.org/10.1002/lipi.19750770603>.
- Delgado, F. J., Cava, R., Delgado, J., & Ramírez, R. (2014). Tocopherols, fatty acids and cytokines content of holder pasteurized and high-pressure processed human milk. *Dairy Science & Technology*, 94(2), 145-156. <https://doi.org/10.1007/s13594-013-0149-y>.
- Eskildsen, C., Rasmussen, M., Engelsen, S., Larsen, L., Poulsen, N. A., & Skov, T. (2014). Quantification of individual fatty acids in bovine milk by infrared spectroscopy and chemometrics: Understanding predictions of highly collinear reference variables. *Journal of Dairy Science*, 97(12), 7940-7951. <https://doi.org/10.3168/jds.2014-8337>.
- Farrés, M., Platikanov, S., Tsakovski, S., & Tauler, R. (2015). Comparison of the variable importance in projection (VIP) and of the selectivity ratio (SR) methods for variable selection and interpretation. *Journal of Chemometrics*, 29(10), 528-536. <https://doi.org/10.1002/cem.2736>.
- Feng, S., Lock, A. L., & Garnsworthy, P. C. (2004). Technical Note: A Rapid Lipid Separation Method for Determining Fatty Acid Composition of Milk. *Journal of Dairy Science*, 87(11), 3785-3788. [https://doi.org/10.3168/jds.S0022-0302\(04\)73517-1](https://doi.org/10.3168/jds.S0022-0302(04)73517-1).
- Ferrand-Calmels, M., Palhière, I., Brochard, M., Leray, O., Astruc, J. M., Aurel, M. R., . . . Larroque, H. (2014). Prediction of fatty acid profiles in cow, ewe, and goat milk by mid-infrared spectrometry. *Journal of Dairy Science*, 97(1), 17-35. <https://doi.org/https://doi.org/10.3168/jds.2013-6648>.
- Floris, L. M., Stahl, B., Abrahamse-Berkeveld, M., & Teller, I. C. (2020). Human milk fatty acid profile across lactational stages after term and preterm delivery: A pooled data analysis. *Prostaglandins, Leukotrienes and Essential Fatty Acids*, 156, 102023. <https://doi.org/https://doi.org/10.1016/j.plefa.2019.102023>.
- Fusch, G., Rochow, N., Choi, A., Fusch, S., Poeschl, S., Ubah, A. O., . . . Fusch, C. (2015). Rapid measurement of macronutrients in breast milk: How reliable are infrared milk analyzers? *Clinical Nutrition*, 34(3), 465-476. <https://doi.org/https://doi.org/10.1016/j.clnu.2014.05.005>.
- Gałaszka, A., Migaszewski, Z., & Namieśnik, J. (2013). The 12 principles of green analytical chemistry and the SIGNIFICANCE mnemonic of green analytical practices. *Trends in Analytical Chemistry*, 50(Complete), 78-84. <https://doi.org/10.1016/j.trac.2013.04.010>.
- Gao, C., Miller, J., Middleton, P. F., Huang, Y.-C., McPhee, A. J., & Gibson, R. A. (2019). Changes to breast milk fatty acid composition during storage, handling and processing: A systematic review. *Prostaglandins, Leukotrienes and Essential Fatty Acids*, 146, 1-10. <https://doi.org/https://doi.org/10.1016/j.plefa.2019.04.008>.
- Gidrewicz, D. A., & Fenton, T. R. (2014). A systematic review and meta-analysis of the nutrient content of preterm and term breast milk. *BMC Pediatr*, 14, 216. <https://doi.org/10.1186/1471-2431-14-216>.

- Jensen, R. G., Bitman, J., Carlson, S. E., Couch, S. C., Hamosh, M., & Newburg, D. S. (1995). Milk Lipids: A Human Milk Lipids. In R. G. Jensen (Ed.), *Handbook of Milk Composition* (pp. 495-542). San Diego: Academic Press.
- Kaylegian, K. E., Lynch, J. M., Fleming, J. R., & Barbano, D. (2009). Influence of fatty acid chain length and unsaturation on mid-infrared milk analysis. *Journal of Dairy Science*, *92*, 2485-2501. <https://doi.org/10.3168/jds.2008-1910>.
- Kohler, A., Afseth, N., Jørgensen, K., Randby, Å., & Martens, H. (2010). Quality Analysis of Milk by Vibrational Spectroscopy. In C. Jim & G. Peter (Eds.), *Handbook of Vibrational Spectroscopy*. Hoboken, USA: John Wiley & Sons.
- Koletzko, B., Thiel, I., & Abiodun, P. O. (1992). The fatty acid composition of human milk in Europe and Africa. *J Pediatr*, *120*(4, Part 2), S62-S70. [https://doi.org/https://doi.org/10.1016/S0022-3476\(05\)81238-7](https://doi.org/https://doi.org/10.1016/S0022-3476(05)81238-7).
- Luna, P., Juárez, M., & Fuente, M. A. (2005). Validation of a Rapid Milk Fat Separation Method to Determine the Fatty Acid Profile by Gas Chromatography. *Journal of Dairy Science*, *88*(10), 3377-3381. [https://doi.org/10.3168/jds.S0022-0302\(05\)73021-6](https://doi.org/10.3168/jds.S0022-0302(05)73021-6).
- Miliku, K., Duan, Q. L., Moraes, T. J., Becker, A. B., Mandhane, P. J., Turvey, S. E., . . . Azad, M. B. (2019). Human milk fatty acid composition is associated with dietary, genetic, sociodemographic, and environmental factors in the CHILd Cohort Study. *The American Journal of Clinical Nutrition*, *110*(6), 1370-1383. <https://doi.org/10.1093/ajcn/nqz229>.
- Moltó-Puigmartí, C., Permanyer, M., Castellote, A. I., & López-Sabater, M. C. (2011). Effects of pasteurisation and high-pressure processing on vitamin C, tocopherols and fatty acids in mature human milk. *Food Chemistry*, *124*(3), 697-702. <https://doi.org/https://doi.org/10.1016/j.foodchem.2010.05.079>.
- Peila, C., Moro, G. E., Bertino, E., Cavallarín, L., Giribaldi, M., Giuliani, F., . . . Coscia, A. (2016). The Effect of Holder Pasteurization on Nutrients and Biologically-Active Components in Donor Human Milk: A Review. *Nutrients*, *8*(8), 477. <https://doi.org/10.3390/nu8080477>.
- Pena-Pereira, F., Wojnowski, W., & Tobiszewski, M. (2020). AGREE—Analytical GREENness Metric Approach and Software. *Analytical Chemistry*, *92*(14), 10076-10082. <https://doi.org/10.1021/acs.analchem.0c01887>.
- Pitino, M. A., Alashmali, S. M., Hopperton, K. E., Unger, S., Pouliot, Y., Doyen, A., . . . Bazinet, R. P. (2019). Oxylipin concentration, but not fatty acid composition, is altered in human donor milk pasteurised using both thermal and non-thermal techniques. *British Journal of Nutrition*, *122*(1), 47-55. <https://doi.org/10.1017/S0007114519000916>.
- Poulimeneas, D., Bathrellou, E., Antonogeorgos, G., Mamalaki, E., Kouvari, M., Kuligowski, J., . . . Yannakoulia, M. (2021). Feeding the preterm infant: an overview of the evidence. *Int J Food Sci Nutr*, *72*(1), 4-13. <https://doi.org/10.1080/09637486.2020.1754352>.
- Ramer, G., & Lendl, B. (2013). Attenuated Total Reflection Fourier Transform Infrared Spectroscopy. In R.A. Meyers, *Encyclopedia of Analytical Chemistry*. Hoboken, USA: John Wiley & Sons.
- Sanders, T. A. B., & Reddy, S. (1992). The influence of a vegetarian diet on the fatty acid composition of human milk and the essential fatty acid status of the infant. *J Pediatr*, *120*(4, Part 2), S71-S77. [https://doi.org/https://doi.org/10.1016/S0022-3476\(05\)81239-9](https://doi.org/https://doi.org/10.1016/S0022-3476(05)81239-9).
- Schwaighofer, A., Brandstetter, M., & Lendl, B. (2017). Quantum cascade lasers (QCLs) in biomedical spectroscopy. *Chemical Society Reviews*, *46*(19), 5903-5924. <https://doi.org/10.1039/c7cs00403f>.
- Ten-Doménech, I., Ramos-García, V., Moreno-Torres, M., Parra-Llorca, A., Gormaz, M., Vento, M., . . . Quintás, G. (2022). The effect of Holder pasteurization on the lipid and metabolite composition of human milk. *Food Chemistry*, *384*, 132581. <https://doi.org/https://doi.org/10.1016/j.foodchem.2022.132581>.
- Van Winckel, M., Vande Velde, S., De Bruyne, R., & Van Biervliet, S. (2011). Clinical practice. *European Journal of Pediatrics*, *170*(12), 1489-1494. <https://doi.org/10.1007/s00431-011-1547-x>.
- WHO. (2001). Resolution WHA54.2, Infant and young child nutrition. *Fifty-fourth World Health Assembly*. Geneva.

Publication V

Schwaighofer, A.; Akhgar, C. K.; Lendl, B. Broadband laser-based mid-IR spectroscopy for analysis of proteins and monitoring of enzyme activity. *Spectrochim. Acta A* **2021**, *253*, 119563.



Contents lists available at ScienceDirect

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

journal homepage: www.elsevier.com/locate/saa

Broadband laser-based mid-IR spectroscopy for analysis of proteins and monitoring of enzyme activity



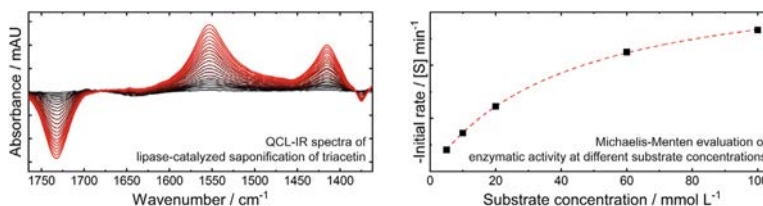
Andreas Schwaighofer*, Christopher K. Akhgar, Bernhard Lendl*

Research Division of Environmental Analytics, Process Analytics and Sensors, Institute of Chemical Technologies and Analytics, Technische Universität Wien, Getreidemarkt 9, 1060 Vienna, Austria

HIGHLIGHTS

- Broadband mid-infrared (IR) external cavity-quantum cascade laser (EC-QCL) spectroscopy.
- Measurement and analysis of the protein amide I and amide II region.
- IR spectra of proteins measured at concentrations as low as 0.25 mg mL⁻¹.
- Monitoring of Lipase-catalyzed saponification of triacetin.
- Laser-based IR spectroscopy poses a fast, sensitive and rugged analysis technique.

GRAPHICAL ABSTRACT



Broadband EC-QCL based mid-IR spectroscopy

- ✓ Spectral coverage: 400 cm⁻¹
- ✓ Protein amide I+II region
- ✓ High SNR
- ✓ Fast, compact, rugged

ARTICLE INFO

Article history:

Received 22 December 2020
 Received in revised form 19 January 2021
 Accepted 27 January 2021
 Available online 7 February 2021

Keywords:

Mid-infrared
 Quantum cascade laser
 Protein secondary structure
 Enzyme activity monitoring

ABSTRACT

Laser-based infrared (IR) spectroscopy is an emerging key technology for the analysis of solutes and for real-time reaction monitoring in liquids. Larger applicable pathlengths compared to the traditional gold standard Fourier transform IR (FTIR) spectroscopy enable robust measurements of analytes in a strongly absorbing matrix such as water. Recent advancements in laser development also provide large accessible spectral coverage thus overcoming an inherent drawback of laser-based IR spectroscopy.

In this work, we benchmark a commercial room temperature operated broadband external cavity-quantum cascade laser (EC-QCL)-IR spectrometer with a spectral coverage of 400 cm⁻¹ against FTIR spectroscopy and showcase its application for measuring the secondary structure of proteins in water, and for monitoring the lipase-catalyzed saponification of triacetin. Regarding the obtained limit of detection (LOD), the laser-based spectrometer compared well to a research-grade FTIR spectrometer employing a liquid nitrogen cooled detector. With respect to a routine FTIR spectrometer equipped with a room temperature operated pyroelectric detector, a 15-fold increase in LOD was obtained in the spectral range of 1600–1700 cm⁻¹. Characteristic spectral features in the amide I and amide II region of three representative proteins with different secondary structures could be measured at concentrations as low as 0.25 mg mL⁻¹. Enzymatic hydrolysis of triacetin by lipase was monitored, demonstrating the advantage of a broad spectral coverage for following complex chemical reactions. The obtained results in combination with the portability and small footprint of the employed spectrometer opens a wide range of future applications in protein analysis and industrial process control, which cannot be readily met by FTIR spectroscopy without recurring to liquid nitrogen cooled detectors.

© 2021 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

* Corresponding authors.

E-mail addresses: andreas.schwaighofer@tuwien.ac.at (A. Schwaighofer), bernhard.lendl@tuwien.ac.at (B. Lendl).

<https://doi.org/10.1016/j.saa.2021.119563>

1386-1425/© 2021 The Author(s). Published by Elsevier B.V.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Mid-infrared (mid-IR) spectroscopy is a powerful and versatile technique for the analysis of structure and dynamics of polypeptides and proteins [1–3]. Vibrations of the polypeptide repeat unit result in nine characteristic group frequencies in the mid-IR region that are referred to as amide bands. Among those, the amide I band (1700–1600 cm^{-1}) and amide II band (1600–1500 cm^{-1}) were shown to be most sensitive to protein structure. 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 and N-H vibrations. The resulting characteristic band shapes and positions can be then correlated with the specific secondary structure folding [4].

Fourier-transform IR (FTIR) spectroscopy is the established and most widespread instrumentation in this spectral region. FTIR spectrometers are commonly equipped with thermal light sources such as Globars that emit broadband and constant but rather low power radiation across the mid-IR range. The low emission intensity can lead to limitations regarding measurements of analytes present in a highly absorbing matrix such as water. For mid-IR spectroscopy of proteins, there is the pronounced challenge of the HOH-bending band of water near 1643 cm^{-1} that overlaps with the protein amide I band. As a consequence, the optical path is restricted to $< 10 \mu\text{m}$ for FTIR transmission measurements in order to avoid total IR absorption [2,3]. These short path lengths lead to impaired sensitivity because of the lower band absorbances and limited robustness due to higher probability of cell clogging.

More than two decades ago, quantum cascade lasers (QCLs) were introduced as a polarized, coherent and high power light source in the mid-IR region [5]. They allow stable operation at room temperature and provide spectral power densities higher than a factor of 10^4 compared with thermal light sources [6]. Since the commercial availability of external cavity-QCLs (EC-QCLs), which offer broadband spectral tuning in the range of several hundred wavenumbers, this type of light source has been increasingly used for studies of liquid samples [7]. It was shown that the high available emission power of QCLs enable mid-IR transmission measurements using an optical path 4–5 times larger than with conventional FTIR spectroscopy [8]. For analysis of proteins, academic setups employing EC-QCLs were reported for investigation of the amide I region [9,10] as well as amide I + II regions [11], finally surpassing the performance of FTIR spectroscopy in terms of limit of detection at similar measurement times [12]. These techniques were employed for analysis of protein structure after chemical [13], thermal [10] and pH-induced [14] denaturation. Recently there was also introduced a commercially available QCL-based IR spectroscopy system, (AQS3pro, RedShiftBio) that covers only the amide I spectral range and provides better performance than FTIR spectroscopy in terms of LOD at approximately 10-fold measurement time [15].

A drawback for laser-based IR setups so far was the limited accessible spectral range. Even though EC-QCLs with spectral coverages of several hundred wavenumbers are achievable [16,17], and even larger tuning ranges of $>1000 \text{cm}^{-1}$ provided by a single device can be obtained by beam combination of multiple EC-QCL modules, uneven spectral tuning densities of QCLs pose a problem for implementation of broadband laser-based IR transmission setups [11,12]. For analysis of proteins, a broader spectral range allows not only to monitor protein structure changes in the amide I + II bands but also to reveal IR signatures of substrates and products of enzymatic reactions. Furthermore, particularly for chemometric analysis additional and more detailed information can be gained by analysis of a larger spectral range [14,18–20].

In this work we employ a commercial broadband laser-based IR spectrometer for protein secondary structure analysis. The performance of the instrument is benchmarked against FTIR spectroscopy. Furthermore, we utilize the large available spectral region (400 cm^{-1}) to follow the enzymatic hydrolysis of triacetin by lipase.

2. Materials and methods

2.1. Reagents and samples

Bovine serum albumin, lysozyme from chicken egg white, and γ -globulins from bovine blood ($\geq 97\%$), lipase from *Candida rugosa* (Type VII, ≥ 700 unit/mg solid) were purchased from Sigma-Aldrich (Steinheim, Germany). Triacetin was purchased from Merck (Darmstadt, Germany). For protein secondary structure measurements, proper amounts of protein powder were dissolved in water. For enzymatic activity measurements, stock solutions of 5 mg mL^{-1} lipase and 250 mM triacetin were prepared. Ultrapure water (resistivity: 18 $\text{M}\Omega \text{cm}$) was purified with a Milli-Q system from Millipore (Bedford, MA).

2.2. Laser-based IR spectroscopy

Laser-based IR spectra were recorded with a ChemDetect Analyzer (Daylight Solutions Inc., San Diego, USA), equipped with a 25 μm diamond transmission flow cell and an EC-QCL providing a spectral coverage between 1350 and 1770 cm^{-1} . External water cooling for the laser head was set to 17 $^\circ\text{C}$. For protein secondary structure measurements, 91 scans were recorded and averaged at a measurement time of 45 s. For enzymatic activity measurements, the enzyme and substrate stock solutions were mixed to obtain the desired concentrations. Subsequently, the resulting solutions were thoroughly vortexed and injected into the sampling cell. At first, a reference spectrum was recorded and afterwards sample spectra were recorded for 20 min. at a time-interval of 20 s (60 scans). The spectrometer was flushed with dry air to decrease the influence of water vapor from the atmosphere. If necessary, absorption bands of water vapor in the atmosphere were subtracted during post-processing. The recorded spectra were treated by Savitzky – Golay smoothing (order: 2, window: 15 points), which resulted in a spectral resolution of 3.6 cm^{-1} , determined by comparison of the band width to FTIR spectra of water vapor. Spectra recording was performed with the ChemDetect software package. Data processing and analysis was conducted with in-house code developed in MatLab R2020b (MathWorks, Inc., Natick, MA, 2014).

2.3. FTIR spectroscopy

FTIR absorption measurements were performed using a Bruker Vertex 80v FTIR spectrometer (Ettlingen, Germany) equipped with a liquid nitrogen cooled HgCdTe (mercury cadmium telluride) detector ($D^* = 4 \times 10^{10} \text{cm Hz}^{0.5} \text{W}^{-1}$ at 9.2 μm) and a Bruker Tensor 37 FTIR spectrometer equipped with a DLATGS (deuterated lanthanum α -alanine doped triglycine sulfate) detector ($D^* = 6 \times 10^8 \text{cm Hz}^{0.5} \text{W}^{-1}$ at 9.2 μm). The samples were manually injected into a flow cell, equipped with two CaF_2 windows and an 8 μm -thick spacer. During measurements, the spectrometer was constantly flushed with dry air for at least 10 min prior to data acquisition until water vapor absorption was sufficiently constant. Spectra were acquired with a spectral resolution of 3.6 cm^{-1} in double-sided acquisition mode. A total of 341 (Vertex operated at 80 kHz scanner velocity) and 48 (Tensor operated at 10 kHz scanner velocity) scans were averaged per spectrum (acquisition time: 45 s), which was calculated using a Blackman-Harris 3-term

apodization function and a zero-filling factor of 2. All spectra were acquired at 25 °C. Spectra were analyzed using the software package OPUS 7.2 (Bruker, Ettlingen, Germany). Water vapor absorption bands were subtracted, if required.

3. Results and discussion

3.1. Mid-IR spectra of proteins recorded with the laser-based IR spectrometer

Laser-based IR spectra of proteins with different secondary structures were recorded at varying concentrations. The broad tuning range of the employed spectrometer enabled recording IR spectra beyond the amide I and amide II regions and allowed to perform qualitative and quantitative evaluation. The examined proteins were chosen to exhibit different secondary structures which show characteristic spectral features in the investigated wavenumber range (Fig. 1A-C). Bovine serum albumin (BSA) mainly contains α -helical secondary structure and shows the corresponding band maximum at 1656 cm^{-1} in the amide I region as well as a narrow band at 1545 cm^{-1} in the amide II region [21,22]. γ -Globulin (γ -Gl) is predominantly composed of β -sheet structures resulting in an amide I band maximum at 1640 cm^{-1} and a broad amide II band with a maximum at 1550 cm^{-1} [23]. Lysozyme (Lys) comprises both α -helices and β -sheets leading to an amide I band maximum at 1656 cm^{-1} with shoulders at \sim 1640 cm^{-1} and \sim 1675 cm^{-1} and a narrow band with a maximum at 1545 cm^{-1} in the amide II region [23,24].

The spectra obtained by the laser-based spectrometer were compared with FTIR measurements (Fig. 1D-F). Evaluation of absorbance band shapes and positions reveal excellent agreement between laser-based and FTIR spectroscopy. It should be highlighted, that even minor spectral features such as the shoulders of the amide II band at approx. 1525 cm^{-1} (BSA, Lys) are well resolved in laser-based IR spectra. Even though not directly correlated to protein secondary structure assignment, the band with a maximum at 1465 cm^{-1} attributed to C-H bending vibrations show good congruence between the two spectroscopy methods [25].

For quantitative evaluation, the height of the band maxima in the amide I region was evaluated for protein solutions with seven different concentrations ranging between 0.1 and 10 mg mL^{-1} . The calibration curves in Fig. S1 show high linearity ($r^2 > 0.999$) down to a concentration of 0.1 mg mL^{-1} .

3.2. Performance comparison of the laser-based IR spectrometer to FTIR spectroscopy

After qualitative comparison of protein absorbance spectra, the overall performance of the laser-based spectrometer was benchmarked against FTIR instruments. For this purpose, typically the limit of detection (LOD)

$$LOD = \frac{3 \times \text{Noise}_{RMS}}{\text{slope of the calibration line}}$$

is employed. This expression not only considers the noise level achieved by an instrument, but also the linear calibration function calculated from the height of the band maxima of protein solutions with different concentrations. According to Lambert – Beer's law, higher absorbance is obtained, when larger optical path lengths are employed for transmission measurements. Consequently, the short path length of 8 μm that needs to be used in FTIR spectroscopy not only considerably reduces the robustness of the measurement but also impairs its sensitivity. The RMS (root mean square)-noise is assessed from a 100% line, which is the absorbance spectrum calculated from two identical single channel spectra [26]. To this end, spectra were recorded with a water-filled transmission cell, followed by evaluation of the spectral region between 1600 and 1700 cm^{-1} . Under ideal conditions, the outcome would be a flat line at 100% transmittance, corresponding to zero absorbance. For the comparison of performance characteristics, a high-end FTIR spectrometer with a fast interferometer and a liquid nitrogen-cooled MCT detector as well as a routine FTIR instrument with a pyroelectric detector were used as a reference. To ensure valid comparison, the same data acquisition time and spectral resolution were set for measuring the 100% lines with the different instruments. For the laser-based IR spectrometer, the spectral resolution was determined by comparison of the band widths of the rotational bands of water vapor to FTIR spectra.

The noise level obtained for high-end FTIR spectroscopy is better by a factor of approx. 2 compared to the laser-based IR spectrometer (see Table 1). However, because of the larger employable path length (factor \sim 3), the LOD of the laser-based IR spectrometer is better by a factor of approx. 1.5. Since the Chem-Detect spectrometer does not require liquid nitrogen cooling, the performance is best compared to a routine FTIR spectrometer. Here, the less sensitive pyroelectric detector and the associated slower scanner velocity result in a lower number of recorded scans

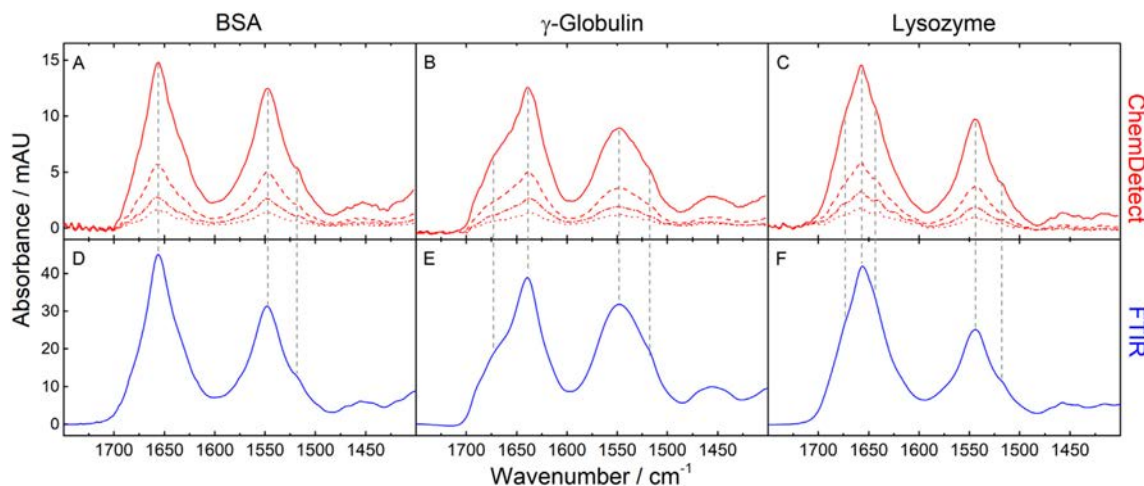


Fig. 1. Mid-IR absorbance spectra of 2.5 (red solid line), 1.0 (red dashed line), 0.5 (red dash-dotted line) and 0.25 mg mL^{-1} (red dotted line) protein solutions acquired with the laser-based spectrometer (A-C) and 20 mg mL^{-1} (blue solid line) protein solution recorded by FTIR spectroscopy (D-F). Gray dashed lines highlight the high overlap of the spectral features between the mid-IR spectra acquired by the laser-based spectrometer and FTIR spectroscopy.

Table 1

Comparison of characteristic parameters and performance between the laser-based IR spectrometer and conventional FTIR spectroscopy.

	Meas. time [sec]/scans	RMS-noise ^a 10 ⁻⁵ [AU]	Path length [μm]	LOD [mg mL^{-1}]	Length \times width \times height/Volume [dm^3]
ChemDetect	45/91	4.89	25	0.026	$2.9 \times 2.3 \times 1.1/7$
High-end FTIR (Vertex 80v)	45/341	2.25	8	0.034	$8.5 \times 7.1 \times 2.8/169$
Routine FTIR (Tensor 37)	45/48	25.36	8	0.383	$6.6 \times 4.5 \times 2.8/83$

^a Determined in the spectral region between 1600 and 1700 cm^{-1} from the 100% line of a water-filled transmission cell. The noise level is stated in absorbance units (AU).

during the same measurement time and consequently in a higher noise level. Hence, the lower achieved noise level (factor ~ 5) of the laser-based IR spectrometer combined with the larger applicable optical path results in a LOD that is better by a factor of ~ 15 than the one of the routine FTIR spectrometer. Finally, it should be noted that FTIR spectroscopy offers the advantage of spectral coverage across the entire mid-IR region.

3.3. Monitoring of an enzymatic reaction

The laser-based IR spectrometer was employed to monitor an enzymatic reaction. For this purpose, the hydrolysis of the triacetate ester triacetin to glycerin and acetate was chosen (Fig. 2A), which is catalyzed by the enzyme lipase [27]. Triacetin was used as substrate because it is sufficiently soluble in water. Fig. 2 shows the dynamic laser-based IR spectra recorded at a substrate concentration of 100 mM. In the accessible spectral region, multiple changes occur that can be attributed to the consumption of substrate and formation of the enzymatic reaction product over time. The negative bands at 1733 cm^{-1} and 1374 cm^{-1} are assigned to the C=O stretching of the ester moiety and symmetric CH_3 deformation vibration, respectively, of the substrate triacetin [28]. The emerging bands at 1553 cm^{-1} and 1415 cm^{-1} are attributed to the asymmetric and symmetric stretching vibrations of the carboxylate group in acetate [29].

For monitoring of the enzymatic activity at different substrate concentrations, the decrease in absorbance at 1733 cm^{-1} was evaluated. The position of this carbonyl stretching band is not prone to pH variations in contrast to the band position of the reaction product [30]. In Fig. 3A, the band heights were plotted versus the reaction time for substrate concentrations between 5 and 100 mM. In the next step, the initial slope of the progression curve was determined by a linear fit between 3 and 5 min of the reaction time.

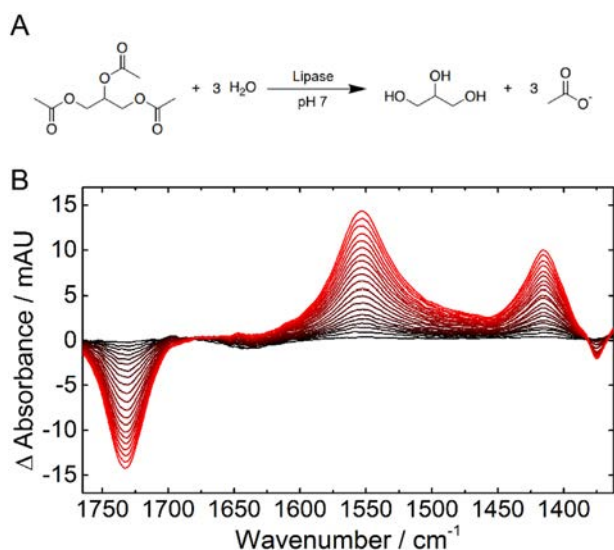


Fig. 2. (A) Scheme of the hydrolysis reaction of triacetin to glycerin and acetate. (B) Laser-based IR spectrum of the enzymatic reaction of lipase with 100 mM triacetin between 0 min (black) and 20 min (red) reaction time.

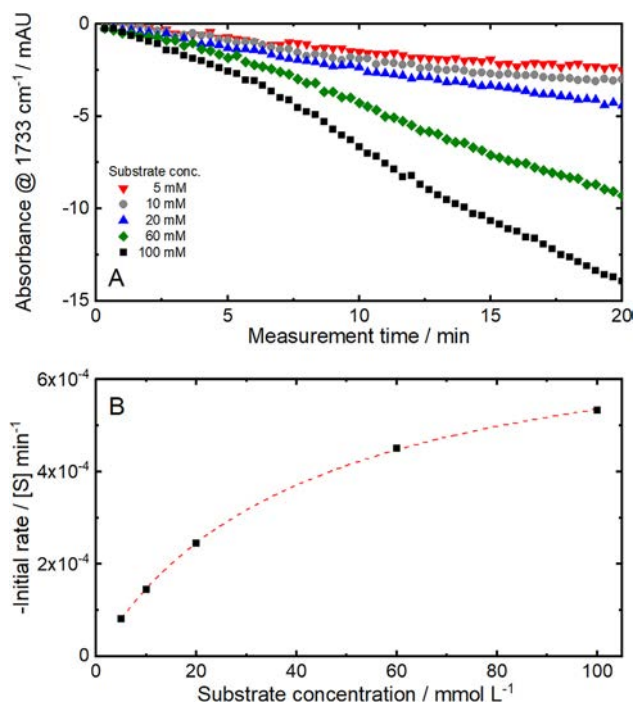


Fig. 3. (A) Progression curve of the enzymatic reaction performed at different substrate concentrations. (B) Initial rate vs substrate concentration follows the Michaelis-Menten relationship (dashed red line).

Subsequently the initial rate was plotted versus the substrate concentration and showed a good fit ($r^2 > 0.9998$) to the Michaelis-Menten equation (Fig. 3B), which is typically used to describe enzyme reactions under conditions of zero-order kinetics at the beginning of the reaction [31]. Based on this evaluation, a Michaelis-Menten constant (K_M) of 41.8 mmol L^{-1} was determined, which agrees with reported values of (K_M) for lipase catalyzed hydrolysis reactions [32].

4. Conclusion

In this work, we employed a commercial broadband laser-based IR spectrometer for analysis of protein secondary structure. Characteristic spectral features of proteins with different secondary structure were successfully measured at concentrations as low as 0.25 mg mL^{-1} . The ChemDetect spectrometer is equipped with a thermoelectrically-cooled detector and requires a small external water-cooling unit to thermally stabilize the laser head during operation. This equipment allows for low maintenance and durable operation under room-temperature environment. Performance comparison with a FTIR instrument under comparable working conditions (i.e. room-temperature operable pyroelectric detector) revealed a LOD better by a factor of ~ 15 for the laser-based IR spectrometer at similar spectra acquisition times. A further important parameter for versatile and application-oriented usage of a spectrometer is its portability in terms of weight and footprint. Also regarding these parameters, the laser-based spectrometer is highly attractive (see Table 1).

The ChemDetect spectrometer was further employed to monitor dynamic spectral changes during the hydrolysis of triacetin by the enzyme lipase. In this regard, the large accessible spectral range proved to be of importance for following the progression of multiple absorbance bands of the substrate as well as the product throughout the enzymatic reaction. Here, the spectral coverage of 400 cm^{-1} exceeds the possibilities previously accomplishable by laser-based IR spectroscopy employing for the analysis of liquids. Furthermore, for monitoring dynamic processes, fast measurement times are an important factor. In this context, the ChemDetect spectrometer enables spectra acquisition rates $< 1\text{ Hz}$ allowing to observe fast reaction progressions that take place within a few minutes. In the present study multiple spectra are co-added in order to improve the noise level by spectra averaging.

The results presented here confirm laser-based IR spectroscopy as a fast, sensitive and rugged analysis technique. In this respect, the employed ChemDetect spectrometer poses a versatile and ready-to-use device that can be readily employed in diverse fields of applications. Envisioned future application are industrial process control and hyphenation with chromatographic separation systems.

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 COMET Center CHASE (project No 868615), which is funded within the framework of COMET (Competence Centers for Excellent Technologies) by BMVIT, BMDW, and the Federal Provinces of Upper Austria and Vienna. The COMET program is run by the Austrian Research Promotion Agency (FFG). The authors acknowledge the TU Wien Bibliothek for financial support through its Open Access Funding Programme.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.saa.2021.119563>.

References

- [1] A. Barth, Infrared spectroscopy of proteins, *Biochim. Biophys. Acta, Bioenerg.* 1767 (2007) 1073–1101.
- [2] H. Fabian, W. Mantele, *Infrared Spectroscopy of Proteins*, Handbook of Vibrational Spectroscopy, John Wiley & Sons Ltd, Hoboken, NJ, USA, 2006.
- [3] H. Yang, S. Yang, J. Kong, A. Dong, S. Yu, Obtaining information about protein secondary structures in aqueous solution using Fourier transform IR spectroscopy, *Nat. Protoc.* 10 (2015) 382–396.
- [4] 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.
- [5] J. Faist, F. Capasso, D.L. Sivco, C. Sirtori, A.L. Hutchinson, A.Y. Cho, Quantum Cascade Laser, *Science* 264 (1994) 553–556.
- [6] A. Schwaighofer, M. Brandstetter, B. Lendl, Quantum cascade lasers (QCLs) in biomedical spectroscopy, *Chem. Soc. Rev.* 46 (2017) 5903–5924.
- [7] A. Schwaighofer, B. Lendl, Quantum cascade laser-based infrared transmission spectroscopy of proteins in solution, in: Y. Ozaki, M. Baranska, I.K. Lednev, B.R. Wood (Eds.), *Vibrational Spectroscopy in Protein Research*, Academic Press, 2020, pp. 59–88.
- [8] M. Brandstetter, L. Volgger, A. Genner, C. Jungbauer, B. Lendl, Direct determination of glucose, lactate and triglycerides in blood serum by a tunable quantum cascade laser-based mid-IR sensor, *Appl. Phys. B: Lasers Opt.* 110 (2013) 233–239.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] C.K. Akhgar, G. Ramer, M. Žbik, A. Trajnerowicz, J. Pawluczyk, A. Schwaighofer, B. Lendl, The Next Generation of IR Spectroscopy: EC-QCL-Based Mid-IR Transmission Spectroscopy of Proteins with Balanced Detection, *Anal. Chem.* 92 (2020) 9901–9907.
- [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-denatured alpha-chymotrypsin, *Anal. Bioanal. Chem.* 408 (2016) 3933–3941.
- [14] 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, *Spectrochim. Acta, Part A* 226 (2020) 117636.
- [15] L.L. Liu, L. Wang, J. Zonderman, J.C. Rouse, H.-Y. Kim, Automated, High-Throughput Infrared Spectroscopy for Secondary Structure Analysis of Protein Biopharmaceuticals, *J. Pharm. Sci.* 109 (2020) 3223–3230.
- [16] S. Riedi, A. Hugl, A. Bismuto, M. Beck, J. Faist, Broadband external cavity tuning in the 3–4 μm window, *Appl. Phys. Lett.* 103 (2013) 031108.
- [17] A. Hugl, R. Maulini, J. Faist, External cavity quantum cascade laser, *Semicond. Sci. Technol.* 25 (2010) 083001.
- [18] M. Montemurro, A. Schwaighofer, A. Schmidt, M.J. Culzoni, H.K. Mayer, B. Lendl, High-throughput quantitation of bovine milk proteins and discrimination of commercial milk types by external cavity-quantum cascade laser spectroscopy and chemometrics, *Analyst* 144 (2019) 5571–5579.
- [19] 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.
- [20] 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.
- [21] J. Grdadolink, Y. Marechal, Bovine serum albumin observed by infrared spectrometry. I. Methodology, structural investigation, and water uptake, *Biopolymers* 62 (2001) 40–53.
- [22] 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.
- [23] M. van de Weert, P.I. Haris, W.E. Hennink, D.J.A. Crommelin, Fourier transform infrared spectrometric analysis of protein conformation: Effect of sampling method and stress factors, *Anal. Biochem.* 297 (2001) 160–169.
- [24] M. Levitt, J. Greer, Automatic Identification of Secondary Structure in Globular Proteins, *J. Mol. Biol.* 114 (1977) 181–239.
- [25] V. Balan, C.T. Mihai, F.D. Cojocar, C.M. Uritu, G. Dodi, D. Botezat, I. Gardikiotis, Vibrational Spectroscopy Fingerprinting in Medicine: from Molecular to Clinical Practice, *Materials* 12 (2019) 2884.
- [26] P.R. Griffiths, J.A. de Haseth, *Fourier Transform Infrared Spectrometry*, John Wiley & Sons Inc, Hoboken, NJ, USA, 2006.
- [27] N. Lopez, M.A. Pernas, L.M. Pastrana, A. Sanchez, F. Valero, M.L. Rua, Reactivity of pure *Candida rugosa* lipase isoenzymes (Lip1, Lip2, and Lip3) in aqueous and organic media. Influence of the isoenzymatic profile on the lipase performance in organic media, *Biotechnol. Progr.* 20 (2004) 65–73.
- [28] S. Armenta, W. Tomischko, B. Lendl, A Mid-Infrared Flow-Through Sensor for Label-Free Monitoring of Enzyme Inhibition, *Appl. Spectrosc.* 62 (2008) 1322–1325.
- [29] M. Kansiz, J.R. Gapes, D. McNaughton, B. Lendl, K.C. Schuster, Mid-infrared spectroscopy coupled to sequential injection analysis for the on-line monitoring of the acetone-butanol fermentation process, *Anal. Chim. Acta* 438 (2001) 175–186.
- [30] P. Domínguez de María, J.M. Sánchez-Montero, J.V. Sinisterra, A.R. Alcántara, Understanding *Candida rugosa* lipases: An overview, *Biotechnol. Adv.* 24 (2006) 180–196.
- [31] H. Bisswanger, Enzyme assays, *Perspect. Sci.* 1 (2014) 41–55.
- [32] Z.D. Knezevic, S.S. Siler-Marinkovic, L.V. Mojovic, Kinetics of lipase-catalyzed hydrolysis of palm oil in lecithin/izoctane reversed micelles, *Appl. Microbiol. Biot.* 49 (1998) 267–271.

Publication VI

Akhgar, C.; Ebner, J.; Spadiut, O.; Schwaighofer, A.; Lendl, B. Laser-based mid-infrared spectroscopy enables in-line detection of protein secondary structure from preparative liquid chromatography. *Proc. Spie.* **2022**, 11957.

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

^aInstitute of Chemical Technologies and Analytics, Technische Universität Wien,
Getreidemarkt 9, 1060 Vienna, Austria;

^bInstitute 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

Copyright 2022 Society of Photo-Optical Instrumentation Engineers (SPIE). One print or electronic copy may be made for personal use only. Systematic reproduction and distribution, duplication of any material in this publication for a fee or for commercial purposes, and modification of the contents of the publication are prohibited.

The published version of this accepted manuscript is available under: C. K. Akhgar, E. Julian, O. Spadiut, A. Schwaighofer and B. Lendl, "Laser-Based Mid-infrared Spectroscopy Enables In-line Detection of Protein Secondary Structure from Preparative Liquid Chromatography," Proc. SPIE 11957, Biomedical Vibrational Spectroscopy 2022: Advances in Research and Industry, 119570G (2 March 2022), DOI: 10.1117/12.2609419.

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 $<10 \mu\text{m}$ 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 ($\geq 90\%$), α -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 mL/min. 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 2nd 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, α -chymotrypsinogen A, myoglobin), standard calibrations in the range from 0.0625 g/L to 1 g/L were performed for each protein.

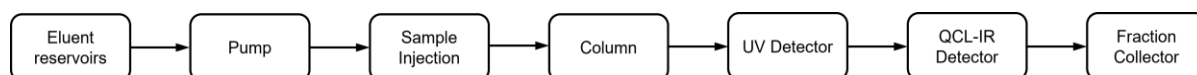


Figure 1. Schematic of the LC-QCL-IR flow path.

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 myoglobin 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

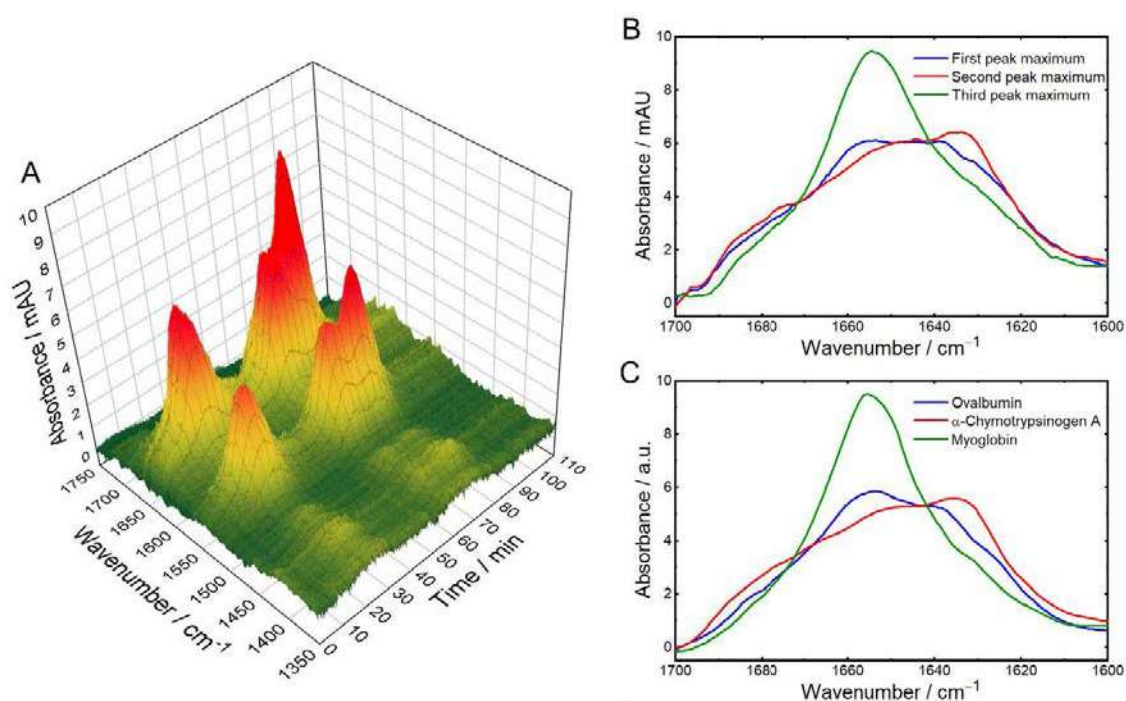


Figure 2. (A) Spectral 3D plot of the chromatographic run, containing ovalbumin, α -chymotrypsinogen A and myoglobin. (B) IR absorption spectra, extracted from the chromatographic peak maxima (C) Reference off-line spectra of 5 mg/mL protein solutions.

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 .

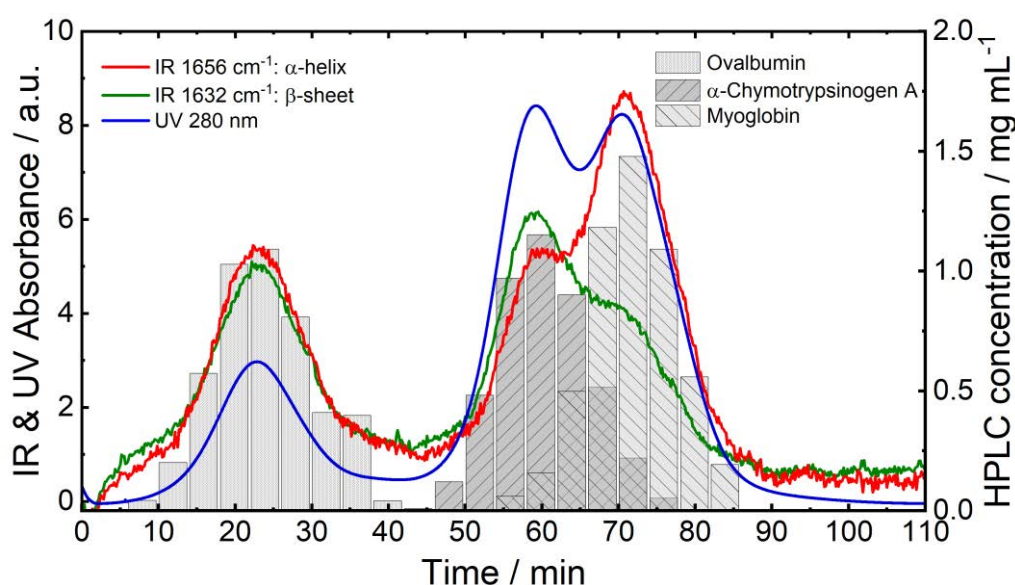


Figure 3. Comparison between signals from mid-IR analyzer at different wavenumbers (green and red lines) and UV detector (blue line) across the chromatographic run and protein reference concentrations, obtained by measuring the collected fractions with reversed-phase HPLC (grey bars).

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.

ACKNOWLEDGEMENTS

This work has received funding from the COMET Center CHASE (project No 868615), which is funded within the framework of COMET (Competence Centers for Excellent Technologies) by BMVIT, BMDW, and the Federal Provinces of Upper Austria and Vienna. The COMET program is run by the Austrian Research Promotion Agency (FFG). Further funding was received by the European Union's Horizon 2020 research and innovation program through NUTRISHIELD project under grant agreement no. 818110 and by the Austrian Research Promotion Agency (FFG), project number 874206. A.S. acknowledges funding by the Austrian Science Fund FWF (project no. P32644-N).

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. Rüdts, 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. Rüdts *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. Žbik *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 α 1-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).

Publication VII

Akhgar, C. K.; Ebner, J.; Spadiut, O.; Schwaighofer, A.; Lendl, B. QCL-IR Spectroscopy for In-Line Monitoring of Proteins from Preparative Ion-Exchange Chromatography. *Anal. Chem.* **2022**, *94*, 5583-5590.

QCL–IR Spectroscopy for In-Line Monitoring of Proteins from Preparative Ion-Exchange Chromatography

Christopher K. Akhgar, Julian Ebner, Oliver Spadiut, Andreas Schwaighofer,* and Bernhard Lendl*

Cite This: *Anal. Chem.* 2022, 94, 5583–5590

Read Online

ACCESS |



Metrics & More



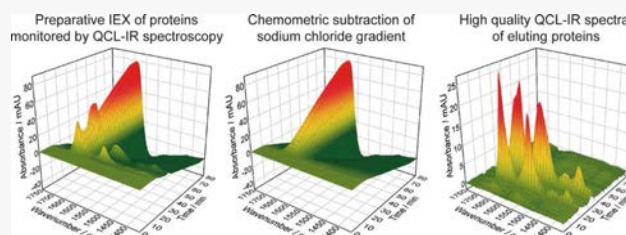
Article Recommendations



Supporting Information

ABSTRACT: In this study, an external cavity-quantum cascade laser-based mid-infrared (IR) spectrometer was applied for in-line monitoring of proteins from preparative ion-exchange chromatography. The large optical path length of 25 μm allowed for robust spectra acquisition in the broad tuning range between 1350 and 1750 cm^{-1} , covering the most important spectral region for protein secondary structure determination. A significant challenge was caused by the overlapping mid-IR bands of proteins and changes in the background absorption of water due to the NaCl gradient.

Implementation of advanced background compensation strategies resulted in high-quality protein spectra in three different model case studies. In Case I, a reference blank run was directly subtracted from a sample run with the same NaCl gradient. Case II and III included sample runs with different gradient profiles than the one from the reference run. Here, a novel compensation approach based on a reference spectra matrix was introduced, where the signal from the conductivity detector was employed for correlating suitable reference spectra for correction of the sample run spectra. With this method, a single blank run was sufficient to correct various gradient profiles. The obtained IR spectra of hemoglobin and β -lactoglobulin were compared to off-line reference measurements, showing excellent agreement for all case studies. Moreover, the concentration values obtained from the mid-IR spectrometer agreed well with conventional UV detectors and high-performance liquid chromatography off-line measurements. LC–QCL–IR coupling thus holds high potential for replacing laborious and time-consuming off-line methods for protein monitoring in complex downstream processes.



Preparative liquid chromatography (prep-LC) remains an essential unit operation in downstream processing of complex biopharmaceuticals.¹ The most widespread detection method for monitoring the corresponding protein concentrations is single-wavelength UV/vis spectroscopy, offering a broad dynamic range and excellent sensitivity. The obtained univariate signal, however, does not give any information about the protein structure and purity during elution.² Thus, the collected fractions have to be additionally analyzed by laborious and time-consuming off-line methods in order to measure critical quality attributes (CQAs). Commonly used off-line methods are high-performance liquid chromatography (HPLC), sodium dodecyl sulphate-polyacrylamide gel electrophoresis, Western blotting and biological activity assays, such as enzyme-linked immunosorbent assays. In recent years, quality by design (QbD) principles were established in the pharmaceutical manufacturing sector.³ In order to comply with these regulatory guidelines, process analytical technology (PAT) tools are required, allowing for in-process control and timely adaption of process parameters. Therefore, analytical methods able to provide information about CQAs in an in-line or on-line measurement setup are required.^{4,5}

Mid-infrared (IR) spectroscopy provides detailed information about rotational–vibrational transitions of proteins. The established technique in this spectral region is Fourier

transform infrared (FTIR) spectroscopy, which is routinely used for quantitative and qualitative analysis of infrared absorption spectra. The most important mid-IR regions for protein quantification and secondary structure determination are the amide I (1700–1600 cm^{-1})^{6,7} and amide II (1600–1500 cm^{-1})⁸ band. Coupling of IR spectroscopy and LC were successfully demonstrated,^{9,10} in most cases using organic solvent gradients. However, flow-through mid-IR transmission measurements of proteins in aqueous matrix have several challenges. One of these constraints arises from the overlap of the HOH-bending band of water near 1643 cm^{-1} with the protein amide I band. A second limitation is the low emission power provided by thermal light sources (Globars) that are used in FTIR spectrometers. Thus, using FTIR spectroscopy, the optical pathlength for aqueous protein solutions is restricted to <10 μm for transmission measurements in order to avoid total IR light absorption of water.^{11,12} These low path-

Received: November 30, 2021

Accepted: March 21, 2022

Published: March 30, 2022



lengths are unsuitable for in-line flow-through measurements due to low robustness and limited sensitivity. Consequently, mid-IR spectroscopy found its way into in-line detection of preparative protein chromatography effluents only recently.^{13,14} In these studies, attenuated total reflection-FTIR spectroscopy was applied, offering robust spectra acquisition, but limited sensitivity. As an alternative approach, complex solvent-removal setups were developed in order to enable protein secondary structure analysis in LC effluents.¹⁵ Here, the eluent is evaporated, while the analyte is (almost) simultaneously deposited onto a suitable substrate prior to FTIR spectra acquisition. Typical challenges for solvent- evaporation interfaces are, however, the morphology of certain analytes that can change over time and possible spatial inhomogeneity.⁹ Moreover, this destructive type of sample preparation prevents fractionation of the effluent after detection.

Quantum cascade lasers (QCLs) are new light sources in the mid-IR spectral region that provide polarized and coherent light with spectral power densities higher by a factor of 10^4 or more compared with thermal light sources.^{16,17} In combination with an external cavity (EC), QCLs enable tuning over several hundred wavenumbers, thus offering high potential for mid-IR transmission spectroscopy of liquids. It was shown that the high available spectral power of EC-QCLs opens a wide range of applications, including robust in-line detection of LC effluents.¹⁸ For protein analysis, the higher spectral power densities enabled to increase the optical path lengths for transmission measurements by a factor of 3–4, thus considerably improving the ruggedness by significantly lowering the backpressure at the cell.¹⁹ In this framework, academic setups were reported that applied EC-QCLs to investigate the amide I^{20,21} and amide I + amide II band,^{22,23} finally achieving a limit of detection almost 10 times lower than high-end FTIR spectroscopy at similar measurement conditions.²⁴ These setups were successfully combined with different chemometric methods in order to quantify individual protein content in complex mixtures^{25–28} and to monitor changes in the protein secondary structure after denaturation.^{21,29,30} Most recently, a commercial EC-QCL-based spectrometer (ChemDetect Analyzer, Daylight Solutions) was introduced, offering robust and sensitive spectra acquisition across the wavenumber region between 1350 and 1770 cm^{-1} .³¹

Another challenge in in-line LC-IR is the compensation of possible eluent absorbance bands, which can be higher by several orders of magnitude compared to actual analyte bands. Even though direct subtraction of a background spectrum is possible under isocratic conditions, background correction during gradient elution is more challenging. For this purpose, a method based on a “reference spectra matrix” (RSM) was introduced.³² In this approach, the spectra acquired in-line during the LC run are viewed as “sample matrix” (SM), whereas the RSM is recorded during a blank run or the re-equilibration phase of the system. Based on this acquired information, spectral regions that show absorbance bands characteristic for eluent composition are identified, which are located in different spectral regions when compared to the analyte spectrum. Consequently, each spectrum in the SM is corrected with the RSM spectrum that has the closest eluent composition. This method was successfully applied for gradient compensation in a wide range of different reversed-phase

HPLC-FTIR applications, including analysis of carbohydrates,^{33,34} nitrophenols,^{35,36} and pesticides.^{32,37}

In the present work, the ChemDetect Analyzer was coupled to a lab-scale prep-LC system. The large optical path length of the transmission cell of 25 μm enabled in-line monitoring of proteins without solvent evaporation steps. This LC–QCL–IR coupling was employed to analyze systems based on ion-exchange chromatography (IEX) and different NaCl gradient profiles. For this application, the laser-based spectrometer offers advantages regarding acquisition of protein spectra in aqueous solution. However, due to the limited spectral region of QCL-IR spectra, gradient compensation is more challenging than with FTIR spectra, which cover the entire mid-IR region. In the studied analytical problems, absorption bands from the salt gradient highly overlap with protein absorptions. Three different case studies were performed, covering various real-life conditions used in chromatographic protein downstream processing. In Case I, a reference blank run was directly subtracted from a sample run with the same linear gradient. For Case II and Case III, two different gradient profiles were employed when compared to the one of the reference blank run. Here, a modified RSM-based approach was devised by incorporating the signal of the conductivity detector. With this approach, each spectrum in the SM was corrected with the RSM spectrum that had the closest conductivity value. Thus, a singular measurement of the RSM blank run was sufficient to correct sample runs with different salt gradient profiles, revealing high quality protein spectra. The obtained results demonstrate the successful coupling of a laser-based IR spectrometer with a LC system and present a novel approach for LC-IR-based gradient correction of highly overlapping eluent and analyte mid-IR spectra, indicating high flexibility for future in-line monitoring of the protein secondary structure in chromatographic effluents.

■ EXPERIMENTAL SECTION

Reagents and Samples. Tris, hydrochloric acid (HCl), and NaCl used for eluents of the prep-LC were purchased from Carl Roth (Karlsruhe, Germany). Hemoglobin (Hemo) from bovine blood and β -lactoglobulin (β -LG) from bovine milk ($\geq 90\%$) were purchased from Sigma-Aldrich (Steinheim, Germany). For LC-IR and protein secondary structure reference measurements, proper amounts of lyophilized protein powder were dissolved in the corresponding buffer. Ultrapure water (MQ) was from a Milli-Q system from Merck Millipore (Darmstadt, Germany). HPLC-Grade acetonitrile (ACN) and trifluoroacetic acid (TFA) were both purchased from AppliChem (Darmstadt, Germany).

LC–QCL–IR Setup. The flow path of LC-IR measurements is illustrated in Figure 1. An ÄKTA pure system (Cytiva Life Sciences, MA, USA) equipped with an U9-M UV monitor, a C9 conductivity monitor, and a F9-C fraction collector was used for the prep-LC runs. All runs were performed with a 1 mL HiTrap Capto Q column (Cytiva Life Sciences, MA, USA). Laser-based mid-IR spectra were acquired with a ChemDetect Analyzer (Daylight Solutions Inc., San Diego, USA). The delay volume between the conductivity detector and ChemDetect Analyzer was determined by injection of 1 M NaCl solution.

Preparative Chromatography Conditions. For all four prep-LC runs, the setup described in Figure 1 was used. The flowrate was kept constant at 75 cm^3/h , with Buffer A (50 mM Tris/HCl, pH 8.5) being used for equilibration and wash and a

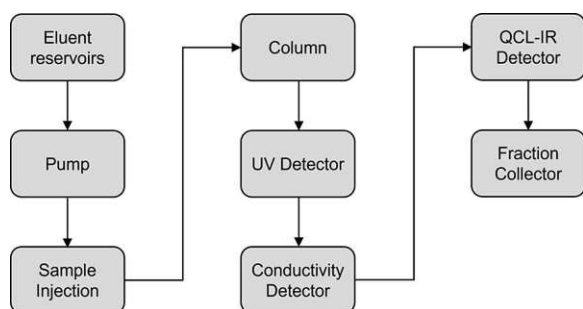


Figure 1. Schematic of the flow path in the LC-QCL-IR setup.

gradient of Buffer A and Buffer B (50 mM Tris/HCl, pH 8.5, 1 M NaCl) being used for elution. Injections were performed using a 1 mL loop, and absorbance at 280 nm as well as conductivity were recorded for all runs. Fractions of 1 mL were collected over the whole run-time and the respective protein concentrations of each fraction measured using the described reversed phase (RP)-HPLC method. The specific elution profiles are shown in Figure S1 and described in detail below.

Blank run: a volume of 1 mL Buffer A was injected. A linear gradient elution was performed over 60 min (30 CVs) from 0% Buffer B to 100% Buffer B.

Case I: as load, 1 mL of 10 mg/mL Hemo and 10 mg/mL β -LG in Buffer A was injected. A linear gradient elution identical to the blank run was performed (0–100% Buffer B in 60 min).

Case II: as load, 1 mL of 5 mg/mL Hemo and 5 mg/mL β -LG in Buffer A was injected. A linear gradient elution was performed over 30 min (15 CVs) from 0% Buffer B to 100% Buffer B.

Case III: as load, 1 mL of 5 mg/mL Hemo and 5 mg/mL β -LG in Buffer A was injected. A step gradient elution was performed with 25% Buffer B 6 min after injection, 50% Buffer B 20.5 min after injection, and 100% Buffer B 30.5 min after injection.

EC-QCL-Based Mid-IR Measurements. The ChemDetect Analyzer, equipped with an EC-QCL (tunable between 1350 and 1750 cm^{-1}) and a diamond transmission flow cell (25 μm optical path length), was used for acquisition of laser-based mid-IR spectra. An external water-cooling unit was operated at 17 $^{\circ}\text{C}$ in order to ensure thermal stabilization of the laser head during operation. Spectra acquisition was performed with the ChemDetect Software package. For flow-through LC-QCL-IR measurements, first, a reference background spectrum was recorded by averaging 121 scans (60 s), followed by continuous acquisition of 20 averaged scans, resulting in a measurement time of 10 s per spectrum. Protein secondary structure off-line reference measurements were acquired by averaging 91 scans during a measurement time of 45 s.

HPLC Reference Measurements. As an off-line analytical method for the collected fractions, a previously published RP-HPLC method was applied.³⁸ In short, a BioResolve RP mAb Polyphenyl column (Waters, MA, USA) was used with MQ as mobile phase A and ACN as mobile phase B, both supplemented with 0.1% v/v TFA. Column temperature was kept constant at 70 $^{\circ}\text{C}$ and an injection volume of 2 μL was used for all samples. Total run time for one injection was 18 min with a flow of 0.4 mL/min and UV/vis absorbance at 214, 280, and 404 nm was recorded for the whole run. A linear gradient from 25% B to 75% B (10 min) was performed, followed by 2 min with 75% B, after which the column was equilibrated with 25% B for 6 min.

Quantification of IR and UV Signals. Protein concentrations (c) across the chromatographic run were calculated from mid-IR and UV signals, according to the Beer–Lambert law

$$c = \frac{A}{\epsilon d} \quad (1)$$

where A denotes the recorded absorbance values, d is the transmission path, and ϵ indicates the molar decadic absorption coefficient.

For laser-based IR spectroscopy, values for A were obtained by integrating the spectra across the amide II region (1500–1600 cm^{-1}). The absorption coefficients (ϵ) for the two proteins were received by integrating the same area region from off-line recorded reference spectra with known protein concentrations.

For quantification of UV/vis spectra, absorption coefficients of the proteins were obtained by performing reference measurements using a Cary 50 Bio UV/vis spectrometer (Agilent Technologies, Santa Clara, California, USA). Spectra were recorded using the Cary WinUV software. Cuvettes with a path length of 10 mm were used to measure 0.5 mg/mL protein solutions. The thereby obtained absorption coefficients at 280 nm agreed well with those calculated via ExPASy ProtParam tool³⁹ and were used to calculate protein concentrations.

Data Processing. Data processing and analysis was conducted with in-house codes developed in Matlab R2020b (MathWorks, Inc., Natick, MA, 2020). During the first preprocessing step, absorbance bands of water vapor from the atmosphere were automatically corrected by subtraction of a water vapour reference spectrum, if necessary. Gradient correction was performed by direct blank run subtraction (Case I) and a modified procedure based on RSM (Case II and Case III).³² Finally, absorption spectra were smoothed with a second order Savitzky–Golay filter (window = 15 points). The spectral resolutions of 1.2 and 3.6 cm^{-1} were determined for smoothed and unsmoothed spectra by comparing the bandwidth of water vapor spectra to FTIR reference spectra.

RESULTS AND DISCUSSION

Relation between Salt Gradient, Conductivity, and Mid-IR Spectra. In-line monitoring of prep-LC effluents was performed with the ChemDetect Analyzer as well as with conventional UV/vis and conductivity detectors (Figure 1). In addition to the routinely recorded signals, laser-based mid-IR spectroscopy offered robust spectra acquisition in the most important wavenumber range for protein secondary structure analysis.

In the present LC–QCL–IR application, IEX was employed as the separation mechanism. Here, the pI of the analyte and the pH of the mobile phase are decisive for the degree of retention. Bound analytes are commonly eluted by utilizing a gradient of increasing salt concentration. Conductivity detection is a widespread method for measuring the salt concentration in IEX.⁴⁰

Figure 2 displays the signals of the mid-IR and conductivity detector on the example of an IEX blank run with a linear NaCl gradient, and thus, the corresponding signal (A) linearly increases with the NaCl concentration.

The blue line in Figure 2C shows the distinctive mid-IR spectrum of NaCl (blue line) with the NaCl-free starting buffer as the reference. Increasing absorbance can be observed

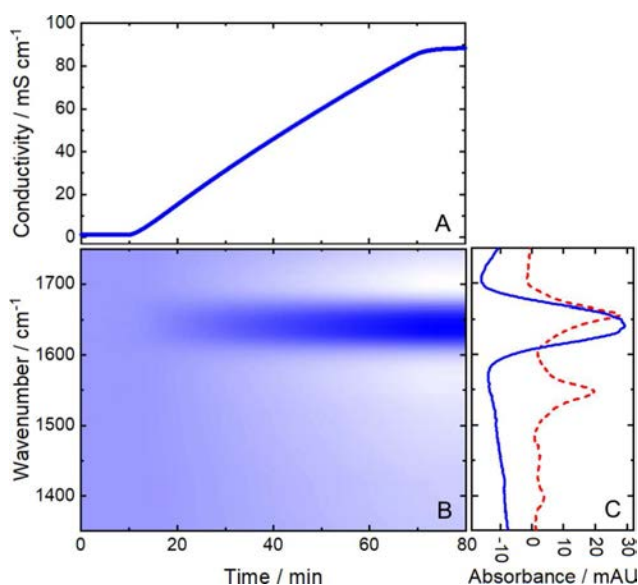


Figure 2. Detector signals from an ion-exchange preparative-liquid chromatography blank run with a NaCl gradient (0–1 M between 10 and 70 min). (A) Conductivity detector response. (B) Heatmap of the corresponding laser-based mid-IR spectra. White areas indicate negative absorbance and dark blue areas positive absorbance. (C) Absorbance spectra of 0.3 M NaCl (retention time: 31.3 min, blue line) and 5 mg/mL hemoglobin (red, dashed line).

between approximately 1610 and 1680 cm^{-1} , while the absorbance in the remaining spectrum decreases with higher salt concentrations. Even though Na^+ and Cl^- are not IR active, reorganization of the water molecules due to the presence of these ions is responsible for a change in the mid-IR spectrum compared to the NaCl-free initial buffer.⁴¹ The maximum of the HOH bending band of water (at approximately 1643 cm^{-1}) is not shifted in position; however, its intensity increases, accompanied by narrowing of the band with increasing NaCl concentrations.⁴² The negative absorbance in the other spectral regions can be related to partial displacement of H_2O molecules with salt.

Figure 2B shows a 2D heat map depicting the response of the NaCl gradient recorded by the ChemDetect analyzer.

White areas indicate negative absorbance, whereas dark blue areas highlight positive absorbance.

The red dashed line in Figure 2C shows the IR spectrum of Hemo, illustrating the spectral overlap of protein bands with the one of NaCl (blue line). Particularly, the amide I band, representing the most important mid-IR region for protein secondary structure determination, is impeded by the modified HOH bending band of water. The described overlap of analyte and changing solvent band indicate significant challenges when employing the ChemDetect Analyzer for in-line monitoring of IEX. In order to obtain feasible mid-IR protein spectra, the influence of the changing salt concentration has to be eliminated. Consequently, feasible gradient compensation strategies are discussed and demonstrated based on model protein systems in the following subchapters.

Case I: Direct Blank Run Subtraction to Enable In-Line Monitoring of Proteins. In this approach, for background compensation, an analyte-free reference blank run was performed with the same gradient as the sample run. Based on the retention time, the reference spectra were subsequently subtracted from the sample spectra.

In the sample run, Hemo and β -LG were included, due to their difference in the pI and secondary structure. Figure 3 shows the spectral 3D plots of the (A) uncorrected sample run and the (B) reference blank run, as well as the (C) corrected sample run after direct background subtraction. The mid-IR spectra of the blank run were already discussed above. Even though the amide I and amide II bands are visible in the uncorrected sample run, the effect of the NaCl gradient on the absorbance spectra is clearly dominating. Thus, no reliable information about the protein secondary structure can be obtained. In contrast, the corrected sample run shows a stable baseline and distinct protein spectra. Figure 4 depicts the absorbance spectra extracted from the peak maxima of the differently corrected sample runs as well as off-line recorded reference IR spectra of the investigated proteins. For Case I (blue lines), the first chromatographic peak with a maximum at approximately 17 min can be related to Hemo due to its higher pI of 7.1.⁴³ Hemo mainly contains α -helical secondary structures and shows the characteristic narrow amide I band with a maximum at approximately 1656 cm^{-1} and a narrow amide II band at 1545 cm^{-1} .^{7,44,45} β -LG is predominantly composed of β -sheet secondary structures⁴⁶ and has a pI of

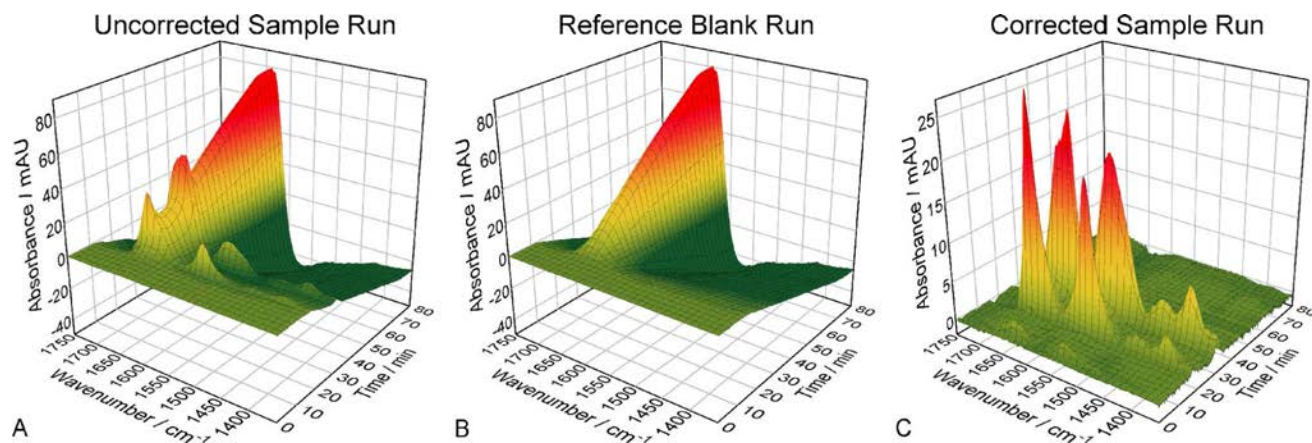


Figure 3. Spectral 3D plots for Case I: (A) uncorrected sample run, (B) reference blank run, and (C) corrected sample run after direct background subtraction.

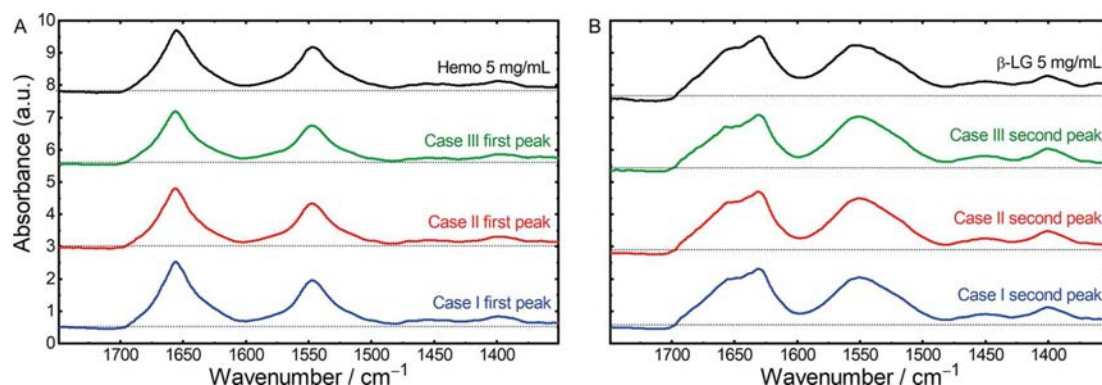


Figure 4. Laser-based mid-IR absorbance spectra, extracted from the (A) first and (B) second peak maxima of the corrected chromatographic runs from Case I (blue lines), Case II (red lines), Case III (green lines), and off-line recorded reference IR spectra (black lines) of 5 mg/mL hemoglobin and β -lactoglobulin. The individual spectra were offset for better visibility. The pointed black lines indicate zero absorbance for every spectrum.

5.1⁴⁷ and thus elutes second under these IEX conditions. The corresponding absorbance spectra show broader amide I and II bands with β -sheet typical maxima at 1632 and 1550 cm^{-1} and a shoulder at 1680 cm^{-1} .^{22,48,49} The obtained high-quality protein spectra thus indicate excellent long-term stability of the ChemDetect Analyzer.

A drawback of this direct gradient compensation approach is the need for a blank run that is performed identical to the sample run. Consequently, any adaptations in the gradient profile require acquisition of an additional blank run and very high reproducibility of the chromatographic system is required in general. The additional time consumption of the blank run and the re-equilibration phase of the column make this approach methodically rigid and impractical to realize for bench-scale as well as industrial applications. Hence, a more flexible strategy is presented in the next subchapter.

Case II and III: Flexible Gradient Compensation with Adapted RSM. Even though direct blank run subtraction revealed excellent protein spectra, more flexible and less laborious gradient compensation strategies are favorable. Gradient compensation based on RSM requires only a single blank run, which can be used for correcting various gradient profiles.³² This approach was successfully applied for HPLC-FTIR, where the entire mid-IR range and eluent specific absorbance bands were available. These specific bands were used as mobile phase identification parameter (IP) in order to select the blank spectrum to be subtracted. This approach, however, is not feasible for monitoring proteins in IEX due to lack of specific absorbance bands and limited sensitivity of FTIR spectroscopy. Due to significantly higher spectral power densities of the EC-QCL, the ChemDetect Analyzer overcomes the typical limitations of FTIR spectroscopy and offers robust and highly sensitive flow-through measurements in the most important wavenumber range for protein analysis.

In the present study, a modified RSM method based on an external IP is proposed and successfully applied. For this purpose, the signal from the conductivity detector was related to the mid-IR spectra. The measured conductivity represents an adequate IP, due to its high dependency on the NaCl concentration, whereas the influence of the protein concentration on the conductivity is negligible (Figure S2). For this purpose, the blank run with the linear NaCl gradient (same as in Case I) was taken as the reference. In order to evaluate the applicability of this approach, two chromatographic sample runs with different gradients compared to the one from the

reference run were conducted: Case II included a linear gradient with a steeper profile, while Case III featured a 3-step gradient. A specific conductivity value was assigned to each mid-IR spectrum in the SM and RSM, based on the retention time. Figure 5 depicts the principle of the applied compensation method on the example of a baseline and protein spectrum in Case III. Based on the comparison of the measured conductivity in the (C) sample run and (A) the

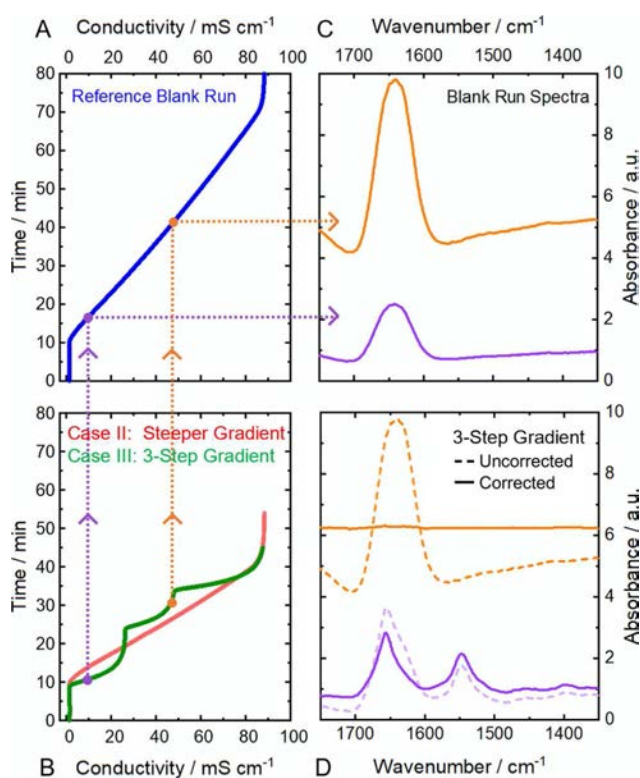


Figure 5. Principle of the applied RSM-based background compensation method. (A) Conductivity detector signals of the reference blank run and (B) sample runs of Case II and Case III. Dotted lines indicate the relation between selected conductivity values and the corresponding mid-IR spectra of background (orange) and protein (purple) of Case III. The absorbance spectra of the (C) blank run were subtracted from (D) the 3-step gradient sample run spectra, based on the most similar conductivity value, revealing distinctive protein spectra (purple) and a flat baseline (orange).

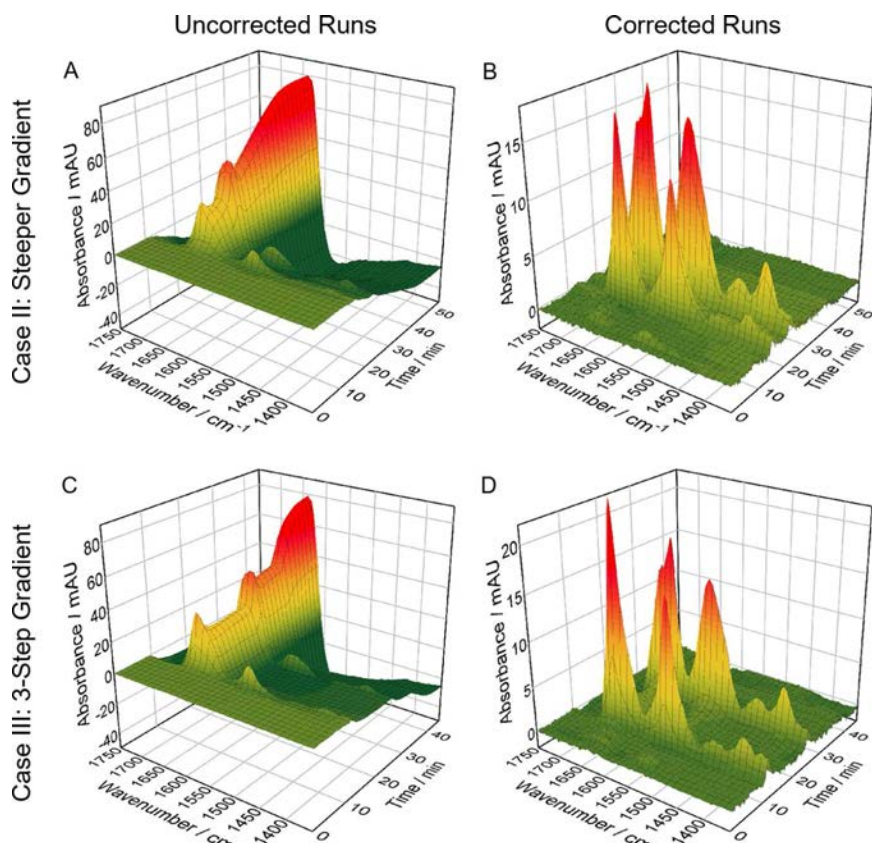


Figure 6. Spectral 3D plots for Case II and Case III: (A,C) uncorrected sample runs and (B,D) corrected sample runs after background correction with adapted RSM.

reference blank run, the corresponding (B) blank run spectra are identified. Finally, the (D) sample spectra are corrected by subtracting the identified blank run spectra with the most similar conductivity value.

Figure 6 displays the spectral 3D plots of the uncorrected (A,C) and corrected (B,D) sample runs. The uncorrected run from Case II constitutes a similar profile as the reference blank run with a shorter measurement time, whereas the run from Case III shows a clearly different profile due to the 3-step gradient. Both corrected runs comprise stable baselines and excellent protein spectra, comparable to those from direct blank run subtraction (Figure 4). Consequently, acquisition of a singular blank run is sufficient for correcting sample runs of widely different gradient profiles. Due to high flexibility and little additional time consumption, the applied gradient compensation approach shows high potential for industrial application.

Comparison of Mid-IR In-Line Measurements to Conventional Off-Line Analysis. In order to verify the signal of the ChemDetect Analyzer, protein concentrations calculated from mid-IR absorbance were compared to those calculated from the UV detector signal. Figure 7 displays the concentrations, determined from the ChemDetect Analyzer (red line) and UV detector (black line) over the chromatographic run from Case III. In mid-IR spectroscopy, the total protein content is best represented by the amide II band because it is less influenced by water absorption than the amide I band. Thus, the wavenumber region between 1500 and 1600 cm^{-1} was integrated to obtain the absorbance values. For comparison, protein concentrations were also calculated from

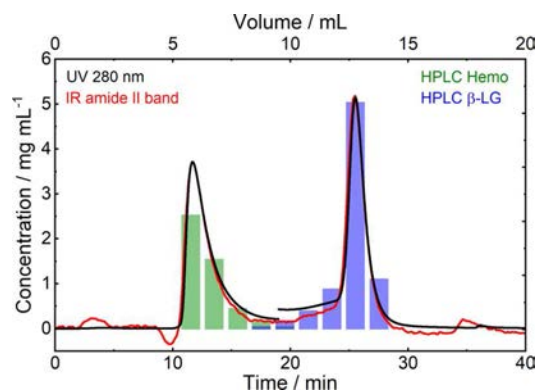


Figure 7. Comparison of protein concentrations obtained from mid-IR amide II band (red line) and UV detector signal at 280 nm (black line) across the chromatographic run from Case III and protein reference concentrations obtained by measuring the collected fractions with reversed-phase HPLC (green and blue bars).

the UV signals at 280 nm recorded by the UV detector of the LC instrument. UV absorption at this wavelength detects aromatic protein residues and disulfide bonds and is most commonly used to monitor proteins in chromatographic applications.¹ Absorption coefficients for Hemo and β -LG for IR and UV spectroscopy were obtained from reference measurements with known protein concentrations. Because these values are different for the two proteins, for conversion of the recorded signals to concentration values, the chromatogram was split into two parts and the absorption coefficients of

Hemo were used from 0 to 19 min, whereas the coefficients of β -LG were applied between 19 and 40 min. Figure 7 depicts the highly overlapping protein concentrations derived from the signals of ChemDetect and UV detector in the time axis as well as for the protein concentration, demonstrating that equivalent quantitative signals can be obtained by the two methods.

For protein identification, IR spectra from the two chromatographic peak maxima were extracted and compared to pure Hemo and β -LG reference spectra, respectively. Figure 4 reveals excellent agreement between in-line and off-line ChemDetect measurements. Absorbance band positions as well as band shapes show excellent comparability between mid-IR spectra of the chromatographic peak maxima and those of the pure reference solutions. In contrast, with conventional LC instrumentation employing an UV detector, real-time information about the protein secondary structure, and consequently type of protein cannot be obtained. In practical applications where the protein identity or purity are usually unknown, highly biased results would be obtained from direct quantitation of the UV signal because absorption coefficients at 280 nm can be different by more than an order of magnitude between proteins.⁵⁰ Then, protein identification must be performed by off-line RP-HPLC measurements of the collected fractions. In this regard, the green bars shown in Figure 7 indicate the concentration of the protein that was identified as Hemo by RP-HPLC, and the blue bars show the concentration of β -LG. Consequently, the obtained LC-QCL-IR chromatograms agree well with the conventionally applied quantification methods, while offering the additional advantage of providing real-time information about the protein secondary structure.

CONCLUSIONS AND OUTLOOK

In this work, a laser-based mid-IR spectrometer was successfully applied for in-line monitoring of proteins from preparative IEX. A major challenge was the highly overlapping absorbance bands of proteins and NaCl gradient that dominated the recorded sample run spectra. Advanced background compensation strategies based on adapted RSM were developed and their implementation in three different case studies resulted in high-quality protein spectra. In Case I, a reference blank run was directly subtracted from a sample run with the same gradient profile. The obtained protein spectra indicated excellent long-term stability of the ChemDetect Analyzer. Case II and III included sample runs with a steeper linear gradient and a 3-step gradient, respectively. Here, a novel gradient compensation approach based on RSM and conductivity IP was introduced. With this method, a single blank run was sufficient for compensating the NaCl gradient in sample runs with distinctively different profiles. The thereby obtained protein spectra showed excellent comparability to off-line reference measurements.

It was shown that the protein concentrations evaluated from signals obtained by the ChemDetect Analyzer are equivalent to UV spectroscopy at 280 nm, which is the standard quantification method for proteins in chromatographic systems. Furthermore, compared to conventional LC detectors, the laser-based mid-IR spectrometer offers the major advantage of providing real-time information about the protein secondary structure, comparable to high-end off-line measurements. The ChemDetect Analyzer thus holds high potential for complementing laborious and time-consuming off-line methods and provides an easily accessible in-line method. In combination with the presented adapted RSM method to

compensate for varying gradient profiles, QCL-IR presents a powerful tool for in-process monitoring and control. Especially, in the light of QbD principles, a near real-time PAT tool able to give information about protein secondary structures and corresponding CQAs presents high potential. In the future, LC-QCL-IR coupling can be employed for chemometrics-based analysis of possible impurities and individual quantification of co-eluting proteins.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.analchem.1c05191>.

Elution profiles of the recorded chromatographic runs, conductivity detector signal of reference blank run and sample run with the same NaCl gradient, and IR spectra and calibration lines for hemoglobin and β -lactoglobulin (PDF)

AUTHOR INFORMATION

Corresponding Authors

Andreas Schwaighofer – Institute of Chemical Technologies and Analytics, Technische Universität Wien, 1060 Vienna, Austria; orcid.org/0000-0003-2714-7056; Email: andreas.schwaighofer@tuwien.ac.at

Bernhard Lendl – Institute of Chemical Technologies and Analytics, Technische Universität Wien, 1060 Vienna, Austria; orcid.org/0000-0003-3838-5842; Email: bernhard.lendl@tuwien.ac.at

Authors

Christopher K. Akhgar – Institute of Chemical Technologies and Analytics, Technische Universität Wien, 1060 Vienna, Austria; orcid.org/0000-0001-8266-043X

Julian Ebner – Institute of Chemical, Environmental and Bioscience Engineering, Technische Universität Wien, 1060 Vienna, Austria

Oliver Spadiut – Institute of Chemical, Environmental and Bioscience Engineering, Technische Universität Wien, 1060 Vienna, Austria

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acs.analchem.1c05191>

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. C.K.A. and J.E. contributed equally.

Funding

Open Access is funded by the Austrian Science Fund (FWF).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work has received funding from the COMET Center CHASE (project no. 868615), funded within the COMET—Competence Centers for Excellent Technologies programme by the BMK, the BMDW, and the Federal Provinces of Upper Austria and Vienna. The COMET programme is managed by the Austrian Research Promotion Agency (FFG). Additional funding was provided by the European Union's Horizon 2020 research and innovation program through NutriShield project under grant agreement no. 818110. This research was further

funded by the Austrian Research Promotion Agency (FFG) (project no. 874206) and by the Austrian Science Fund FWF (project no. P32644-N).

REFERENCES

- (1) Carta, G.; Jungbauer, A. *Protein Chromatography. Process Development and Scale-Up*; John Wiley & Sons: Hoboken, USA, 2010; Vol. 5.
- (2) Gupta, M. *Methods for Affinity-Based Separations of Enzymes and Proteins*; Birkhäuser: Basel, Switzerland, 2002.
- (3) Ali, J.; Pramod, K.; Tahir, M. A.; Charoo, N. A.; Ansari, S. H. *Int. J. Pharm. Invest.* **2016**, *6*, 129–138.
- (4) Esmonde-White, K. A.; Cuellar, M.; Uerpmann, C.; Lenain, B.; Lewis, I. R. *Anal. Bioanal. Chem.* **2017**, *409*, 637–649.
- (5) Ebner, J.; Humer, D.; Klausser, R.; Rubus, V.; Pell, R.; Spadiut, O.; Kopp, J. *Bioengineering* **2021**, *8*, 78.
- (6) Singh, B. R. Basic Aspects of the Technique and Applications of Infrared Spectroscopy of Peptides and Proteins. *Infrared Analysis of Peptides and Proteins*; American Chemical Society: Washington D.C., USA, 1999; pp 2–37.
- (7) Barth, A. *Biochim. Biophys. Acta Bioenerg.* **2007**, *1767*, 1073–1101.
- (8) Murphy, B.; D'Antonio, J.; Manning, M.; Al-Azzam, W. *Curr. Pharm. Biotechnol.* **2014**, *15*, 880–889.
- (9) Kuligowski, J.; Quintás, G.; Guardia, M.; Lendl, B. Liquid Chromatography—Liquid Chromatography—Fourier Transform Infrared. *Encyclopedia of Analytical Science*, 3rd ed.; Elsevier: Amsterdam, Netherlands, 2019; pp 75–85.
- (10) Kuligowski, J.; Quintás, G.; Garrigues, S.; Lendl, B.; de la Guardia, M.; Lendl, B. *Trends Anal. Chem.* **2010**, *29*, 544–552.
- (11) Fabian, H.; Mäntele, W. *Infrared Spectroscopy of Proteins. Handbook of Vibrational Spectroscopy*; John Wiley & Sons: Hoboken, USA, 2006.
- (12) Yang, H.; Yang, S.; Kong, J.; Dong, A.; Yu, S. *Nat. Protoc.* **2015**, *10*, 382–396.
- (13) Großhans, S.; Rüdert, M.; Sanden, A.; Brestrich, N.; Morgenstern, J.; Heissler, S.; Hubbuch, J. *J. Chromatogr. A* **2018**, *1547*, 37–44.
- (14) Sanden, A.; Suhm, S.; Rüdert, M.; Hubbuch, J. *J. Chromatogr. A* **2019**, *1608*, 460410.
- (15) Turula, V. E.; de Haseth, J. A. *Anal. Chem.* **1996**, *68*, 629–638.
- (16) Faist, J.; Capasso, F.; Sivco, D. L.; Sirtori, C.; Hutchinson, A. L.; Cho, A. Y. *Science* **1994**, *264*, 553–556.
- (17) Schwaighofer, A.; Brandstetter, M.; Lendl, B. *Chem. Soc. Rev.* **2017**, *46*, 5903–5924.
- (18) Beskers, T. F.; Brandstetter, M.; Kuligowski, J.; Quintás, G.; Wilhelm, M.; Lendl, B. *Analyst* **2014**, *139*, 2057–2064.
- (19) Schwaighofer, A.; Lendl, B.; Ozaki, Y.; Baranska, M.; Lednev, I. Quantum cascade laser-based infrared transmission spectroscopy of proteins in solution. In *Vibrational Spectroscopy in Protein Research*; Wood, B., Ed.; Academic Press: Cambridge, USA, 2020; pp 59–88.
- (20) Alcaráz, M. R.; Schwaighofer, A.; Kristament, C.; Ramer, G.; Brandstetter, M.; Goicoechea, H.; Lendl, B. *Anal. Chem.* **2015**, *87*, 6980–6987.
- (21) Schwaighofer, A.; Alcaraz, M. R.; Araman, C.; Goicoechea, H.; Lendl, B. *Sci. Rep.* **2016**, *6*, 33556.
- (22) Schwaighofer, A.; Montemurro, M.; Freitag, S.; Kristament, C.; Culzoni, M. J.; Lendl, B. *Anal. Chem.* **2018**, *90*, 7072–7079.
- (23) Chon, B.; Xu, S.; Lee, Y. J. *Anal. Chem.* **2021**, *93*, 2215–2225.
- (24) Akhgar, C. K.; Ramer, G.; Žbik, M.; Trajnerowicz, A.; Pawluczyk, J.; Schwaighofer, A.; Lendl, B. *Anal. Chem.* **2020**, *92*, 9901–9907.
- (25) Kuligowski, J.; Schwaighofer, A.; Alcaráz, M. R.; Quintás, G.; Mayer, H.; Vento, M.; Lendl, B. *Anal. Chim. Acta* **2017**, *963*, 99–105.
- (26) Schwaighofer, A.; Kuligowski, J.; Quintás, G.; Mayer, H. K.; Lendl, B. *Food Chem.* **2018**, *252*, 22–27.
- (27) Schwaighofer, A.; Alcaráz, M. R.; Kuligowski, J.; Lendl, B. *Biomed. Spectrosc. Imag.* **2018**, *7*, 35–45.
- (28) Montemurro, M.; Schwaighofer, A.; Schmidt, A.; Culzoni, M. J.; Mayer, H. K.; Lendl, B. *Analyst* **2019**, *144*, 5571–5579.
- (29) Alcaráz, M. R.; Schwaighofer, A.; Goicoechea, H.; Lendl, B. *Anal. Bioanal. Chem.* **2016**, *408*, 3933–3941.
- (30) Schwaighofer, A.; Alcaraz, M. R.; Lux, L.; Lendl, B. *Spectrochim. Acta, Part A* **2020**, *226*, 117636.
- (31) Schwaighofer, A.; Akhgar, C. K.; Lendl, B. *Spectrochim. Acta, Part A* **2021**, *253*, 119563.
- (32) Quintás, G.; Lendl, B.; Garrigues, S.; de la Guardia, M. *J. Chromatogr. A* **2008**, *1190*, 102–109.
- (33) Kuligowski, J.; Quintás, G.; Garrigues, S.; de la Guardia, M. *Anal. Chim. Acta* **2008**, *624*, 278–285.
- (34) Kuligowski, J.; Quintás, G.; Garrigues, S.; de la Guardia, M. *J. Chromatogr. A* **2009**, *1216*, 3122–3130.
- (35) Quintás, G.; Kuligowski, J.; Lendl, B. *Anal. Chem.* **2009**, *81*, 3746–3753.
- (36) Kuligowski, J.; Quintás, G.; Garrigues, S.; de la Guardia, M. *Talanta* **2010**, *80*, 1771–1776.
- (37) Quintás, G.; Kuligowski, J.; Lendl, B. *Appl. Spectrosc.* **2009**, *63*, 1363–1369.
- (38) Kopp, J.; Zauner, F. B.; Pell, A.; Hausjell, J.; Humer, D.; Ebner, J.; Herwig, C.; Spadiut, O.; Slouka, C.; Pell, R. *J. Pharm. Biomed. Anal.* **2020**, *188*, 113412.
- (39) Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S. e.; Wilkins, M. R.; Appel, R. D.; Bairoch, A. Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*; Walker, J. M., Ed.; Humana Press: Totowa, NJ, 2005; pp 571–607.
- (40) Haddad, P. R. Conductivity Detection. *Journal of Chromatography Library*; Elsevier, 1990, Chapter 9; pp 245–289.
- (41) Max, J.-J.; Chapados, C. *J. Chem. Phys.* **2001**, *115*, 2664–2675.
- (42) Max, J.-J.; Trudel, M.; Chapados, C. *Appl. Spectrosc.* **1998**, *52*, 234–239.
- (43) Zhan, G.; Li, C.; Luo, D. *Bull. Korean Chem. Soc.* **2007**, *28*, 1720–1724.
- (44) Levitt, M.; Greer, J. *J. Mol. Biol.* **1977**, *114*, 181–239.
- (45) Perutz, M. F.; Rossmann, M. G.; Cullis, A. F.; Muirhead, H.; Will, G.; North, A. C. T. *Nature* **1960**, *185*, 416–422.
- (46) Monaco, H. L.; Zanotti, G.; Spadon, P.; Bolognesi, M.; Sawyer, L.; Eliopoulos, E. E. *J. Mol. Biol.* **1987**, *197*, 695–706.
- (47) Engelhardt, K.; Lexis, M.; Gochev, G.; Konnerth, C.; Miller, R.; Willenbacher, N.; Peukert, W.; Braunschweig, B. *Langmuir* **2013**, *29*, 11646–11655.
- (48) van de Weert, M.; Haris, P. I.; Hennink, W. E.; Crommelin, D. J. A. *Anal. Biochem.* **2001**, *297*, 160–169.
- (49) Dousseau, F.; Pezolet, M. *Biochemistry* **1990**, *29*, 8771–8779.
- (50) Pace, C. N.; Vajdos, F.; Fee, L.; Grimsley, G.; Gray, T. *Protein Sci.* **1995**, *4*, 2411–2423.

Publication VIII

Akhgar, C. K.; Ebner, J.; Alcaraz, M.R.; Kopp, J.; Goicoechea, H.; Spadiut, O.; Schwaighofer, A.; Lendl, B. Application of Quantum Cascade Laser-Infrared Spectroscopy and Chemometrics for In-Line Discrimination of Coeluting Proteins from Preparative Size Exclusion Chromatography. *Anal. Chem.* **2022.** 94, 11192-11200.

Application of Quantum Cascade Laser-Infrared Spectroscopy and Chemometrics for In-Line Discrimination of Coeluting Proteins from Preparative Size Exclusion Chromatography

Christopher K. Akhgar, Julian Ebner, Mirta R. Alcaraz, Julian Kopp, Héctor Goicoechea, Oliver Spadiut, Andreas Schwaighofer,* and Bernhard Lendl*



Cite This: *Anal. Chem.* 2022, 94, 11192–11200



Read Online

ACCESS |



Metrics & More

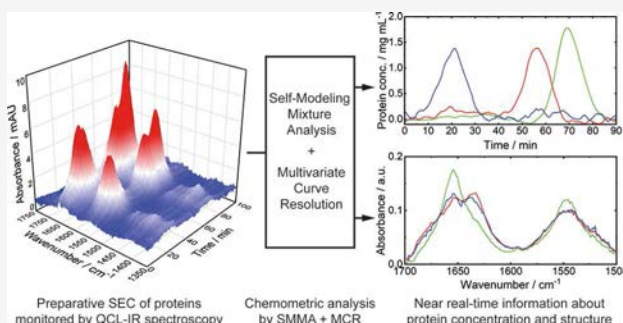


Article Recommendations



Supporting Information

ABSTRACT: An external-cavity quantum cascade laser (EC-QCL)-based flow-through mid-infrared (IR) spectrometer was placed in line with a preparative size exclusion chromatography system to demonstrate real-time analysis of protein elutions with strongly overlapping chromatographic peaks. Two different case studies involving three and four model proteins were performed under typical lab-scale purification conditions. The large optical path length (25 μm), high signal-to-noise ratios, and wide spectral coverage (1350 to 1750 cm^{-1}) of the QCL-IR spectrometer allow for robust spectra acquisition across both the amide I and II bands. Chemometric analysis by self-modeling mixture analysis and multivariate curve resolution enabled accurate quantitation and structural fingerprinting across the protein elution transient. The acquired concentration profiles were found to be in excellent agreement with the off-line high-performance liquid chromatography reference analytics performed on the collected effluent fractions. These results demonstrate that QCL-IR detectors can be used effectively for in-line, real-time analysis of protein elutions, providing critical quality attribute data that are typically only accessible through time-consuming and resource-intensive off-line methods.



Preparative SEC of proteins monitored by QCL-IR spectroscopy

Chemometric analysis by SMMA + MCR

Near real-time information about protein concentration and structure

Protein purification and polishing protocols typically include diverse process unit operations based on liquid chromatography (LC).¹ This technique separates analytes in a liquid mobile phase by interactions with a solid stationary phase according to different physio-chemical properties. In size exclusion chromatography (SEC), compounds are separated by their size and shape which offers several advantages, such as straight-forward operation, nondenaturing conditions, and isocratic elution, compared to other separation principles.² Protein concentrations in chromatographic effluents are routinely monitored in-line by univariate UV/vis or evaporative light scattering detectors, offering excellent sensitivity, high robustness, and a broad linear range. A major drawback of these detectors is, however, that the obtained signals do not provide information that allows for the discrimination or quantitation of different coeluting proteins. Critical quality attributes (CQAs), thus, have to be obtained by analyzing the collected fractions off-line. During process development, this can lead to significant time delays. Moreover, it hinders development based on quality by design (QbD) principles. QbD requires the application of process analytical technology (PAT) tools, facilitating in-process monitoring and in-process control. In-line or on-line measurements providing real-time or

near real-time information on CQAs are required to allow timely adaption of set-points during the purification step.

Mid-infrared (IR) spectroscopy is a well-established technique for nondestructive analysis of diverse compounds, including polypeptides and proteins.³ Conventional Fourier-transform IR (FT-IR) spectrometers are equipped with thermal light sources that emit low-power radiation across the entire mid-IR region (400–4000 cm^{-1}). Even though LC-FT-IR hyphenation was successfully demonstrated for the analysis of numerous analytes including nitrophenols,^{4,5} carbohydrates,^{6,7} and pesticides,^{8,9} mid-IR flow-through measurements of proteins remain challenging. The most important IR bands for protein secondary structure determination and quantitation are the amide I (1600–1700 cm^{-1}) and amide II (1500–1600 cm^{-1}) band, respectively.¹⁰ Substantial light absorption by the HOH bending band of

Received: April 8, 2022

Accepted: July 25, 2022

Published: August 4, 2022



water at approximately 1645 cm^{-1} makes investigations of the overlapping amide I band with FT-IR instrumentation a cumbersome task. In order to avoid total absorption of IR radiation in this spectral region, optical path lengths of 6 to 8 μm are typically applied.^{11,12} Such limited path lengths are not suitable for LC-IR hyphenation as they lead to distinctly impaired robustness and sensitivity. For this purpose, complex schemes were developed that evaporate the solvent and deposit the protein almost simultaneously onto a substrate before FT-IR analysis.¹³ Even though these setups enabled protein secondary structure analysis from LC effluents,^{14,15} solvent evaporation interfaces can bear major challenges such as spatial heterogeneity and changes in analyte morphology over time.¹⁶ Moreover, in preparative LC runs, the effluent is usually fractionated after detection, making a preceding solvent evaporation step inapplicable. More recently, attenuated total reflection-FT-IR spectroscopy¹⁷ was coupled to an LC system for in-line monitoring of proteins.^{18–20} This configuration overcomes the limitations regarding ruggedness, but still requires high protein concentrations due to its limited sensitivity.

Significant progress in quantum cascade lasers (QCLs)²¹ has challenged conventional FT-IR spectrometers for biochemical sensing applications.²² Properties such as $\geq 10^4$ times higher brightness compared to thermal light sources and tunability over several hundred wavenumbers in external cavity (EC) configurations make QCLs highly beneficial for the analysis of proteins.²³ In this context, diverse academic setups were developed that employed EC-QCLs for protein investigations. Here, it has been demonstrated that the intense power outputs of QCLs allow to significantly increase the path length for transmission measurements and, thus, the ruggedness for protein amide I band analysis in aqueous solutions.²⁴ Due to the particular characteristics of water absorption in the protein amide I spectral region and emission properties of EC-QCLs,²⁵ it has turned out to be a challenging task to develop setups also covering the protein amide II region. However, simultaneous analyses of amide I + II bands were realized by different approaches, for example, by combining EC-QCLs with either mercury cadmium telluride detectors (MCTs) and optical filters,^{25,26} MCTs and acousto-optic modulators,²⁷ or quantum cascade detectors (QCDs).²⁸ Furthermore, the implementation of an advanced noise compensation strategy based on balanced detection led to robust protein measurements with limits of detection almost an order of magnitude lower than those from high-end FT-IR spectrometers.²⁶

Parallel to the rapid advances of laser-based optical setups in academic research, a commercially available QCL-IR spectrometer, the ChemDetect Analyzer (Daylight Solutions), was recently introduced.²⁹ This device covers a broad wavenumber range beyond protein amide I and II bands and offers robust and sensitive spectra acquisition with an optical transmission path of 25 μm . In a recent piece of work, the ChemDetect Analyzer was successfully applied for in-line monitoring of proteins from preparative LC.³⁰ Compared to conventionally used LC detectors, laser-based mid-IR spectroscopy offers the major advantage of providing near real-time information about protein quantity and secondary structure, which can otherwise merely be obtained by off-line measurements. LC-QCL-IR coupling, thus, bears a high potential for in-line analysis of CQAs, such as protein purity, which is further investigated in the present study.

For analysis of complex experimental data, chemometrics is typically applied to extract chemical information about individual analytes from spectroscopic data of multicomponent systems. Multivariate spectroscopic monitoring of dynamic processes, such as in LC-QCL-IR, generates two-way data matrices that comprise the information about the occurring spectral changes and the chemical perturbation profiles of the system. Multicomponent spectroscopic signals generally follow Beer–Lambert's Law and fulfill the concept of the so-called bilinear models.³¹ Among the most used chemometric techniques based on bilinear decomposition are self-modeling mixture analysis (SMMA)³² and multivariate curve resolution (MCR).³³ The advantage of these methods is that they do not require any a priori knowledge about the system, for example, the number or spectra of components, and all information can be deduced from the recorded data set. Even though the obtained pure variables do not represent a pure component, for systems with a reduced number of components, this approach can serve as a good and fast estimator of the chemical behavior of the system that can be readily compared to recorded spectra and, thus, allows straightforward interpretation by nonchemometricians.^{34,35}

In this work, LC-QCL-IR hyphenation was performed for in-line monitoring of proteins from coeluting chromatographic peaks. Two case studies involving three and four proteins, respectively, were performed based on SEC, and real-life conditions used in protein purification protocols were applied. The goal of this work is to employ in-line QCL-IR spectroscopy for obtaining qualitative and quantitative information, which can be conventionally only received by work- and time-intensive off-line high-performance LC (HPLC) analytics. For this purpose, chemometric analysis based on a bilinear decomposition model was performed to (i) extract IR absorption spectra of the individual proteins from the recorded multidimensional data set as well as to (ii) retrieve their concentration profiles over the chromatographic run. Achieved results were benchmarked against reference off-line IR spectra of pure protein solutions and HPLC measurements of the collected fractions, showing excellent agreement in both cases.

■ EXPERIMENTAL SECTION

Reagents and Samples. Ovalbumin (Ova, $\geq 90\%$), α -chymotrypsinogen A (α -CT) from bovine pancreas, myoglobin (Myo) from equine skeletal muscle ($\geq 95\%$), horseradish peroxidase type VI-A (HRP), and β -lactoglobulin (β -LG) from bovine milk ($\geq 90\%$) were obtained from Sigma-Aldrich (Steinheim, Germany). Appropriate amounts of protein powder were dissolved in SEC buffer. Ultrapure water (MQ) was obtained with a Milli-Q system from Merck Millipore (Darmstadt, Germany). Trifluoroacetic acid and acetonitrile, both HPLC-grade, were purchased from AppliChem (Darmstadt, Germany). All other chemicals used for the preparation of mobile phases were obtained from Carl Roth (Karlsruhe, Germany).

LC-QCL-IR Flow Path. The applied LC-QCL-IR setup is depicted in Figure 1. An ÄKTA pure preparative chromatographic system (Cytiva Life Sciences, MA, USA), equipped with a U9-M UV monitor and an F9-C fraction collector was used for all SEC runs. A HiLoad 16/600 Superdex 200 pg (Cytiva Life Sciences, MA, USA) was used as the SEC column for both Case study I and Case study II. A ChemDetect

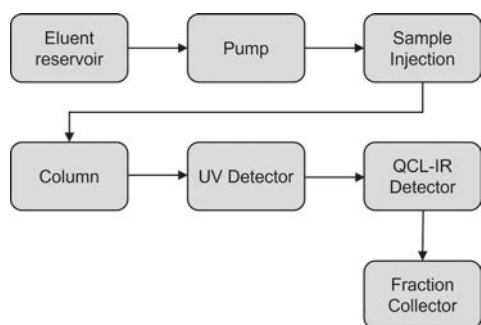


Figure 1. Scheme of the flow path in the LC-IR setup.

Analyzer (Daylight Solutions Inc., San Diego, USA) was used to record QCL-IR spectra.

Size Exclusion Chromatography Conditions. For preparative LC runs, the setup described in Figure 1 was used. Both runs were performed in the isocratic mode with a 50 mM phosphate buffer pH 7.4 (SEC buffer) with a constant flow of 7.5 cm/h (=0.25 mL/min). For Case study I, 0.5 mL of SEC buffer containing 10 mg/mL Ova, 10 mg/mL α -CT, and 10 mg/mL Myo were injected. For case study II, 0.5 mL of SEC buffer containing 10 mg/mL HRP, 10 mg/mL β -LG, 10 mg/mL α -CT, and 10 mg/mL Myo were injected. UV absorbance (280 nm) was recorded over the whole run and fractions with a volume of 1 mL were collected. For case study I, the protein concentration of the individual proteins in the collected fractions was measured using reversed-phase (RP) HPLC. For case study II, the concentration of HRP and Myo in the collected fractions were quantified using the described RP-HPLC method, while the concentrations of β -LG and α -CT were obtained using a cation exchange (CEX) HPLC method.

Laser-Based Mid-IR Measurements. All mid-IR measurements were acquired with a ChemDetect Analyzer. The equipped EC-QCL was operated between 1350 and 1750 cm^{-1} and thermally stabilized with an external water-cooling unit (set to 17 $^{\circ}\text{C}$). A custom-built, temperature-stabilized CaF_2 flow cell with an optical path length of 25 μm was used for all transmission measurements. The provided ChemDetect software package was used for spectra acquisition. For LC-QCL-IR in-line measurements, a background spectrum was acquired within 60 s by averaging 121 scans, followed by spectra acquisition every 10 s (averaging of 20 scans). Off-line reference measurements of pure protein solutions were performed by averaging 91 scans within 45 s. During spectra acquisition, the ChemDetect Analyzer was flushed with dry air to decrease the influence of water vapor from the atmosphere.

HPLC Reference Measurements. As an off-line analytical method to qualify and quantify proteins contained in the collected fractions, a previously published RP-HPLC method was used.³⁶ Because it was not possible to achieve satisfactory peak separation for β -LG and α -CT using the RP-HPLC method (case study II), additionally, CEX HPLC measurements were performed. For that purpose, an UltiMate 3000 HPLC system (Thermo Fisher, MA, USA) equipped with a quaternary pump module, a temperature-controlled autosampler, a column oven, and a UV/vis detector module was used. The method used a MabPac SCX-10 (250 mm) column (Thermo Fisher, MA, USA) with a constant column temperature of 35 $^{\circ}\text{C}$ and a constant flow rate of 1 mL/min. In total, three mobile phases (mobile phase A: 20 mM

phosphate citrate buffer pH 4; mobile phase B: 20 mM phosphate citrate buffer pH 4 with 1 M NaCl; and mobile phase C: 50 mM phosphate buffer pH 7.4 with 1 M NaCl) were used, and the exact gradient profile is shown in Figure S1. In order to achieve sufficient separation, the pH value of all samples was adjusted to pH 4 (10 M phosphoric acid) prior to the measurement. An injection volume of 20 μL was used for all samples and concentrations were calculated based on peak integration and comparison to measured standards with a known concentration. Standards were treated in the same way as samples, that is, dissolved to the desired concentration in SEC buffer and adjusted to pH 4 using 10 M phosphoric acid.

Data Analysis. In the present work, the separation of proteins by SEC was monitored with QCL-IR spectroscopy. For data analysis, the spectral range of the data matrix was cut to 1500–1700 cm^{-1} , corresponding to the protein amide I and amide II bands, and the temporal range was limited to cover periods of protein elution. Prior to chemometric resolution, aiming to improve the S/N ratio, averaging of two data points was performed in the spectral axis. For case study I, an additional averaging of two spectra was performed in the time axis. Finally, 273 \times 174 (case study I) and 271 \times 176 (case study II) matrices were obtained and subjected to chemometric analysis.

Multicomponent spectroscopic signals generally follow Beer–Lambert’s Law, hence they fulfill the concept of bilinear models described by

$$X = CS^T + E, \quad (1)$$

where X describes the two-way data matrix, and S and C contain the bilinear description of the data for both spectral profile and their relative concentrations, respectively; E contains the residuals of the model. In the applied workflow, spectral estimates for spectral profiles were obtained by SMMA. This group of techniques estimates the purest chemical factors and their contribution requiring any specific information about the data. In this regard, the pure variable-based methods, such as the simple-to-use interactive SSMA approach (SIMPLISMA), seek to obtain the selective spectral (or concentration) variables through the calculation of a purity value. The subsequently applied MCR, on the other hand, is a family of soft modeling techniques able to solve the bilinear description of the data for both spectral (S) profiles and their relative concentrations (C) through bilinear decomposition of the two-way data matrix X either by noniterative or iterative methods.

Data processing and chemometric analysis were performed in MATLAB R2020b (Mathworks, Inc., Natick, MA, 2020). MCR-ALS algorithms are available online at <http://www.mcrals.info/>.

Protein Quantitation of Reconstituted Individual LC-QCL-IR Chromatograms. Based on the concentration and spectral profiles obtained by the chemometric analysis, individual chromatogram matrices $X_n (=c_n s_n^T)$ were reconstituted for every protein. These reconstituted IR spectra were employed to calculate protein concentrations (c) across the chromatographic run according to the Beer–Lambert law

$$c = \frac{A}{\epsilon d} \quad (2)$$

Here, d is the path length of the transmission cell. The absorbance values (A) were obtained by integrating the amide II bands (1500–1600 cm^{-1}) of the reconstituted QCL-IR

spectra. Absorption coefficients (ϵ) of the selected proteins were obtained by integrating the same spectral region of off-line acquired QCL-IR spectra of reference solutions with known protein concentrations.

RESULTS AND DISCUSSION

To demonstrate the potential of the presented approach, monitoring of preparative SEC by a QCL-IR detector with subsequent chemometric analysis was performed with two-protein model systems exhibiting partial coelution of the proteins.

Case Study I: A Model System with Three Proteins.

For case study I, a SEC run with three different proteins was performed. For this purpose, Ova, α -CT, and Myo were identified to be proteins with different molecular weights and secondary structures. Figure 2A shows the results of in-line UV

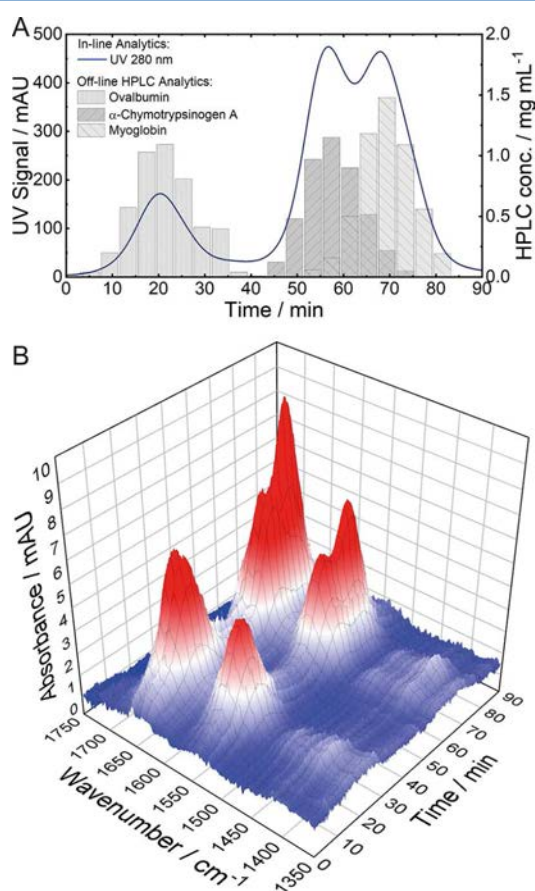


Figure 2. Experimental data obtained from the SEC run of case study I. (A) Results of in-line UV spectroscopy (left) and off-line HPLC analytics (right). (B) Spectral 3D plot recorded by the QCL-IR detector.

spectroscopy at 280 nm, indicating three chromatographic peaks. This signal is the most common for protein detection but does not provide any information regarding the secondary structure. Thus, in order to obtain qualitative and quantitative information about the eluting proteins, off-line HPLC analytics need to be performed (Figure 2A). The first chromatographic peak at 22 min can be related to Ova with a molecular weight of 44.5 kDa.³⁷ α -CT and Myo have more similar molecular masses of 25.6 kDa³⁸ and 17 kDa,³⁹ respectively, and show overlapping peaks at approximately 57 and 70 min. These

results agree with the separation principle of SEC, where large molecules elute first. The ChemDetect Analyzer was used to record mid-IR spectra of the LC effluent across the chromatographic run.

Figure 2B displays the 3D plot (wavenumber-time-absorbance) of the performed LC-QCL-IR measurements. The plot shows stable baseline and chromatographic peaks with the characteristic amide I and II bands at retention times corresponding to the three proteins. For rather basic qualitative interpretation and discrimination between the three eluting proteins, chromatograms at wavenumbers characteristic for individual secondary structures can be extracted from the 3D data set and compared.⁴⁰ However, to gain more insight into the qualitative and quantitative information, an in-depth chemometric analysis needed to be performed. For this purpose, first, the number of components was estimated by singular value decomposition. Then, the purest spectral profiles were received by using a SIMPLISMA-like approach.⁴¹ Subsequently, unconstrained MCR was applied to determine the corresponding time-dependent concentration profiles. At this point, it should be emphasized that for this analysis, no initial knowledge, for example, about the number and type of proteins, is required and all information can be derived from the recorded 3D QCL-IR data set. The obtained lack of fit (LOF, 2.5%) indicates a good description of the experimental data by the MCR model. By this approach, five components were determined to be needed to explain the experimental data, three of them were attributed to the eluting proteins in the chromatographic run, whereas the last two were associated with background signals. Figure 3A,B show the spectral and time-resolved concentration profiles, respectively, of the identified proteins. The retention times at the maximum of the concentration profiles agree very well with the peak maxima observed by reference techniques.

Analysis of the spectral profiles allows assigning secondary structures to the eluted proteins, even if no further reference information is available. The shapes of the amide I and amide II bands indicate that the first two eluting proteins are composed of mixed or β -sheet secondary structures, whereas the third protein mainly contains α -helices. In case reference spectra are available (Figure 3C), identification of the eluted proteins is also possible. Ova features both α -helical and β -sheet secondary structures, resulting in an amide I band maximum at 1656 cm⁻¹ with a shoulder at 1638 cm⁻¹ and a broad amide II band with the maximum at approximately 1545 cm⁻¹.^{42,43} α -CT also contains α -helices but is predominantly composed of β -sheets showing a characteristic broad amide I band with a maximum at 1635 cm⁻¹ and shoulders at 1650 and 1680 cm⁻¹. The amide II band features a broad shape with a maximum at approximately 1548 cm⁻¹.⁴⁴ Even though the spectral profiles of the first and second chromatographic peaks appear similar, evaluation of the different amide I band maxima and bandwidths allows the assignment of the first peak to Ova and the second peak to α -CT. Finally, the third resolved spectral profile can be assigned to Myo, which is mainly composed of an α -helical secondary structure.⁴⁵ The corresponding absorption spectrum shows a distinct amide I band at approximately 1656 cm⁻¹ and a narrow amide II band with a maximum at 1547 cm⁻¹.

Two additional components obtained from the chemometric analysis were identified in the recorded QCL-IR data set (Figure S2). One concentration profile featured negative dips at the same retention times as the chromatographic protein

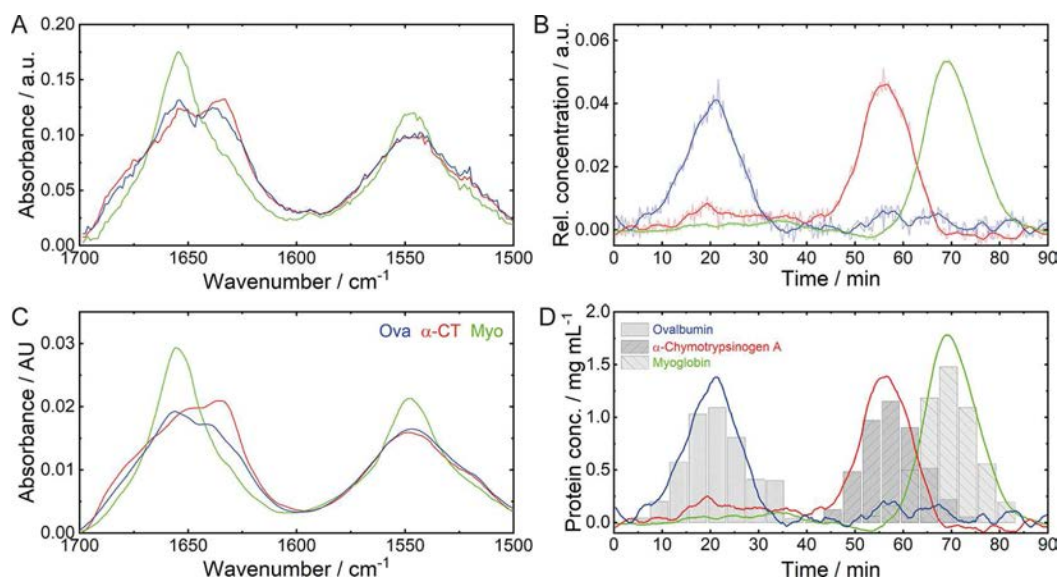


Figure 3. Results obtained by chemometric analysis of the bidimensional QCL-IR data set of case study I. (A) Spectral and (B) time-dependent concentration profiles retrieved by chemometric analysis. The obtained concentration profiles (thin) were smoothed by a Savitzky–Golay filter (thick). (C) Reference laser-based IR spectra of Ova, α -CT, and Myo. (D) Protein concentrations obtained by in-line QCL-IR analysis (lines) and off-line reference HPLC analytics (bars).

peaks. Thus, this profile was assigned to the dilution of the buffer during protein elution, as the presence of proteins reduces the relative water content as compared to the background spectrum (pure buffer). Thus, due to the high absorption coefficient of the HOH-bending band, even subtle variations of the water content can introduce an observable effect on the IR spectra.⁴⁶ One further component was assigned to varying baseline and instrumental responses.

Case Study II: A Model System with Four Proteins.

After the successful application of QCL-IR spectroscopy combined with chemometric analysis for protein structure identification and resolving of individual protein chromatograms with the model system comprising three proteins, a further, even more challenging case study was devised to validate the potential and versatility of the introduced method. To this end, a SEC run including HRP, β -LG, α -CT, and Myo was performed. Figure 4A shows the results of in-line UV spectroscopy at 280 nm as well as off-line HPLC analytics, indicating four chromatographic peaks with a severe overlap of the first two protein peaks at 25 and 32 min. The first of these chromatographic peaks can be attributed to HRP with the highest molecular weight of 44 kDa.⁴⁷ The second peak is related to β -LG, which is present at the employed pH conditions in its dimeric form with a molecular mass of 36.7 kDa.⁴⁸ Due to the similar masses of these two proteins, they show highly coeluting behavior. Finally, the two remaining peaks at 59 and 72 min are assigned to α -CT and Myo with molecular masses of 25.6³⁸ and 17 kDa,³⁹ respectively.

Figure 4B shows the 3D plot (wavenumber-time-absorbance) of the performed LC-QCL-IR measurement. This plot shows amide I and amide II maxima at retention times comparable with the reference techniques. Also, in this case, the data set was subjected to chemometric analysis to retrieve spectral and time-dependent concentration profiles of the eluting proteins. After obtaining initial estimates, MCR was performed obtaining an LOF of 2.3%, indicating a good fit of the MCR model to the experimental data. For this data set, six components were identified of which four could be assigned to

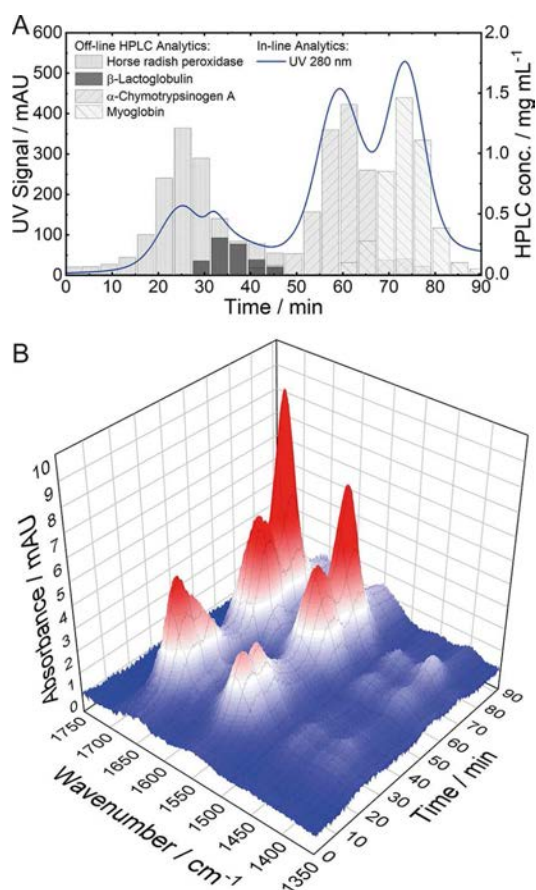


Figure 4. Experimental data obtained from the SEC run of case study II. (A) Results of in-line UV spectroscopy (left) and off-line HPLC analytics (right). (B) Spectral 3D plot recorded by the QCL-IR detector.

proteins in the chromatographic effluent. The spectral and time-resolved concentration profiles of the identified proteins

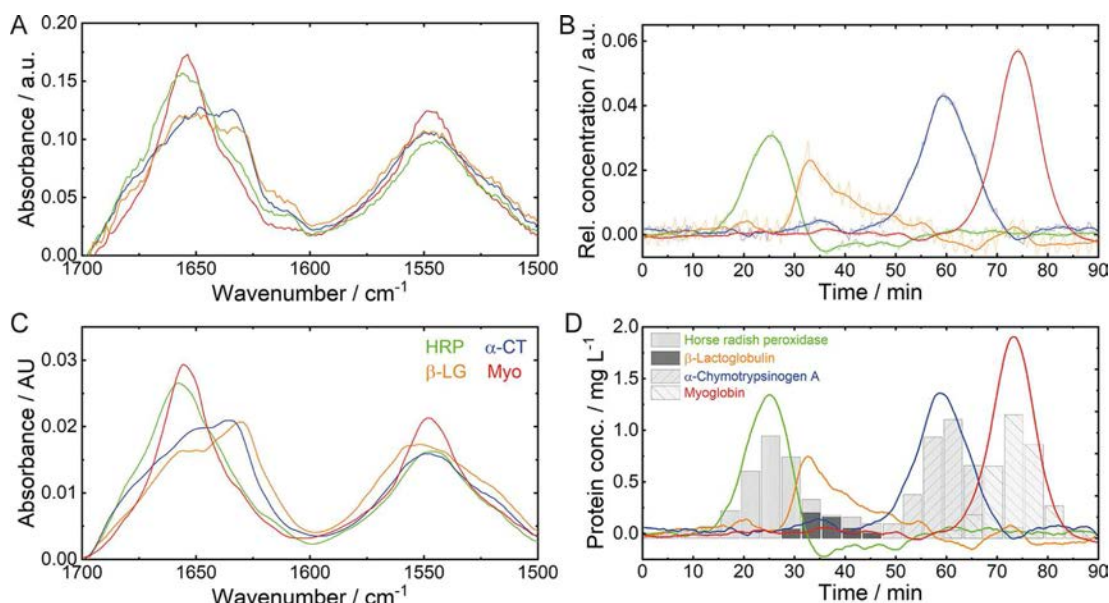


Figure 5. Results obtained by chemometric analysis of the bidimensional QCL-IR data set of case study II. (A) Spectral and (B) time-dependent concentration profiles retrieved by chemometric analysis. The obtained concentration profiles (thin) were smoothed by a Savitzky–Golay filter (thick). (C) Reference laser-based IR spectra of HRP, β -LG, α -CT, and Myo. (D) Protein concentrations obtained by in-line QCL-IR analysis (lines) and off-line reference HPLC analytics (bars).

are shown in Figure 5A,B. The maximum positions of the concentration profiles agree very well with the retention times observed by the reference techniques. Without any prior knowledge, the secondary structures of the first and fourth eluting proteins could be assigned to be mostly α -helical, while the second and third eluting proteins show spectral features of a mixed or β -sheet secondary structure. Comparison with the reference spectra allows identification of the first chromatographic peak as HRP with an amide I band maximum at 1656 cm^{-1} with shoulders at 1640 and 1680 cm^{-1} and an amide II band maximum at approximate 1545 cm^{-1} .^{49,50} It can be distinguished from Myo due to the slightly shifted amide I band maximum and the overall broader shape. The spectral profiles of the second and third identified chromatographic peaks have very similar shapes. However, comparison with reference spectra allows identifying the second peak as β -LG due to the narrower amide I band and the broader amide II band shapes. β -LG is predominantly composed of β -sheet secondary structures and shows a distinct amide I band with a maximum at 1632 cm^{-1} and a shoulder at 1660 cm^{-1} and a broad amide II band with a maximum at 1550 cm^{-1} .^{51,52} Finally, the spectral profile of the fourth chromatographic peak can be unambiguously attributed to Myo.

Evaluation of the concentration profile reveals negative relative concentrations of HRP at retention times between 30 and 50 min. Impaired resolution of the bidimensional data set by MCR in this time region is caused by the severe overlap in retention times of HRP and β -LG as also shown by the reference HPLC results. Furthermore, α -CT which also starts to elute in this period features similar IR spectral features as β -LG.

Chemometric analysis of this data set further identified two additional components (Figure S3), as occurred for case study I. Due to the shape of the concentration profiles, one was attributed to varying baseline and buffer dilution in the presence of proteins. The concentration profile of the other component shows a zigzag shape at rather low relative

concentrations. This periodic noise characteristic is partly also visible in the concentration profile of β -LG. It is present in the IR measurements, but not observable in the in-line UV measurements. The origin of these features was traced back to inconstant dry air supply throughout the QCL-IR measurements. The periodic behavior is introduced by a switching valve in the purge gas generator. Furthermore, the involvement of IR absorption of water vapor in this repetitive noise pattern is also supported by the narrow absorption bands in the related spectral profile.

Notwithstanding this instrumental issue, owing to the challenging model system and experimental difficulties, it was decided to present these results as proof to demonstrate the potency of the presented workflow to obtain reliable estimations by the combination of QCL-IR spectroscopy and chemometrics.

In-Line Protein Quantitation by QCL-IR Spectroscopy.

The reconstituted IR spectra obtained from chemometric analysis were employed for the calculation of the absolute protein concentrations. For this purpose, the area of the amide II band was integrated because this spectral region is less prone to water absorption-related intensity variations than the amide I band. Absorption coefficients of the proteins included in case studies I and II were obtained from QCL-IR off-line measurements with known protein concentrations. Figure 3D shows a comparison between the calculated concentrations based on QCL-IR spectroscopy and reference HPLC results of the collected fractions for case study I. The graph demonstrates highly overlapping concentration profiles between in-line and off-line reference measurements, indicating high validity of the presented approach based on QCL-IR spectroscopy and chemometrics. This evaluation was also performed for case study II (Figure 5D). Here, the elution profiles of all four proteins agree well between the two methods, even though absolute concentrations appear slightly shifted. These differences might be explained by the highly overlapping spectral features of β -LG and α -CT and the

pronounced overlap of the chromatographic peaks of HRP and β -LG. Those challenging circumstances may adversely affect chemometric analysis and lead to underestimation of the HRP content at retention times between 30 and 50 min while overestimating the β -LG content. Nevertheless, these good results, in spite of the complex data sets, indicate high flexibility and robustness of the presented LC-QCL-IR approach.

Case study II also was challenging to resolve for off-line RP-HPLC analytics. The insufficient peak resolution of β -LG and α -CT required the introduction of an additional CEX-HPLC method in order to accurately identify and quantify the proteins in the collected fractions. As two different off-line HPLC methods are required to analyze fractions in case study II, this further emphasizes the difficulties to establish straightforward reference analytics. Consequently, QCL-IR in-line detectors hold significant potential for providing near-real-time protein concentrations from chromatographic separation processes by achieving similar results as conventionally applied time- and cost-intensive off-line methods.

CONCLUSIONS AND OUTLOOK

In this work, an EC-QCL-based mid-IR spectrometer was successfully hyphenated to a preparative SEC system for in-line discrimination of proteins from highly overlapping chromatographic peaks. The advantages of QCL-IR detectors over conventional LC detectors were demonstrated in two case studies, involving mixtures of three and four different proteins, respectively. Due to similar molecular weights of the proteins, highly overlapping chromatographic peaks were obtained that could not be distinguished with a standard UV detector. In contrast, QCL-IR detection enabled the acquisition of multivariate data sets, containing mid-IR absorbance spectra across the chromatographic runs that provide information regarding protein secondary structure, thus allowing protein identification. These data sets were investigated by chemometrics to obtain spectral profiles of the individual proteins as well as their relative concentration profiles. The obtained spectra agree well with reference off-line spectra of pure protein solutions. Furthermore, absolute protein concentrations were calculated according to the Beer–Lambert law, showing high agreement with HPLC reference measurements of the collected effluent fractions. Consequently, QCL-IR in-line detectors can provide qualitative and quantitative information about proteins, comparable to time- and labor-intensive off-line methods that are inaccessible with conventional UV detectors. Hence, the in-line QCL-IR system has the capability to (i) accelerate process development and to (ii) monitor production processes on-line, allowing in-process control. Furthermore, the presented system enables QbD principles and concurs with the requirements of a PAT tool, providing information about protein secondary structure in real-time.

Finally, an important property of the presented in-line QCL-IR detection of preparative LC is its accordance with green analytical chemistry (GAC) principles.⁵³ For a comprehensive comparison of in-line QCL-IR spectroscopy and off-line HPLC analysis, the previously introduced Analytical GREENness (AGREE) metric approach,⁵⁴ a straightforward assessment approach based on the 12 principles of GAC (SIGNIFICANCE),⁵⁵ was applied. The results for both methods are shown in Figure S4. The scores of 0.84 for QCL-IR spectroscopy and 0.43 for off-line HPLC indicate a clear

superiority of the presented in-line QCL-IR method in terms of the greenness of the analytical procedure.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.analchem.2c01542>.

Gradient profile used in the CEX-HPLC method; additional spectra and time-resolved concentration profiles retrieved by chemometric analysis of the chromatographic run of case study I and case study II; results of evaluation according to the Analytical GREENness metric approach for in-line QCL-IR spectroscopy and off-line HPLC analysis; and selected options for the Analytical GREENness evaluation are presented here (PDF)

AUTHOR INFORMATION

Corresponding Authors

Andreas Schwaighofer – Institute of Chemical Technologies and Analytics, Technische Universität Wien, 1060 Vienna, Austria; orcid.org/0000-0003-2714-7056; Email: andreas.schwaighofer@tuwien.ac.at

Bernhard Lendl – Institute of Chemical Technologies and Analytics, Technische Universität Wien, 1060 Vienna, Austria; orcid.org/0000-0003-3838-5842; Email: bernhard.lendl@tuwien.ac.at

Authors

Christopher K. Akhgar – Institute of Chemical Technologies and Analytics, Technische Universität Wien, 1060 Vienna, Austria; orcid.org/0000-0001-8266-043X

Julian Ebner – Institute of Chemical, Environmental and Bioscience Engineering, Technische Universität Wien, 1060 Vienna, Austria

Mirta R. Alcaraz – Laboratorio de Desarrollo Analítico y Quimiometría (LADAQ), Cátedra de Química Analítica I, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral, S3000ZAA Santa Fe, Argentina; Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), C1425FQB CABA, Argentina

Julian Kopp – Institute of Chemical, Environmental and Bioscience Engineering, Technische Universität Wien, 1060 Vienna, Austria

Héctor Goicoechea – Laboratorio de Desarrollo Analítico y Quimiometría (LADAQ), Cátedra de Química Analítica I, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral, S3000ZAA Santa Fe, Argentina; Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), C1425FQB CABA, Argentina; orcid.org/0000-0001-7145-0082

Oliver Spadiut – Institute of Chemical, Environmental and Bioscience Engineering, Technische Universität Wien, 1060 Vienna, Austria

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acs.analchem.2c01542>

Author Contributions

The manuscript was written through the contributions of all authors. All authors have given approval to the final version of the manuscript. C.K.A. and J.E. contributed equally.

Funding

Open Access is funded by the Austrian Science Fund (FWF).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work has received funding from the COMET Center CHASE (project no. 868615), funded within the COMET—Competence Centers for Excellent Technologies program by the BMK, the BMDW, and the Federal Provinces of Upper Austria and Vienna. The COMET program is managed by the Austrian Research Promotion Agency (FFG). Additional funding was provided by the European Union's Horizon 2020 research and innovation program through the NUTRISHIELD project under grant agreement no. 818110. This research was further funded by the Austrian Research Promotion Agency (FFG) (project no. 874206) and by the Austrian Science Fund FWF (project no. P32644-N). M.R.A. and H.G. acknowledge CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina) for funding.

REFERENCES

- (1) Deutscher, M. P. Chapter 5 Setting Up a Laboratory. In *Methods in Enzymology*; Burgess, R. R., Deutscher, M. P., Eds.; Academic Press, 2009; pp 37–42.
- (2) Sun, Y.; Shi, Q. H.; Zhang, L.; Zhao, G. F.; Liu, F. F. 2.47—Adsorption and Chromatography. In *Comprehensive Biotechnology* (2nd ed.); Moo-Young, M., Ed.; Academic Press: Burlington, 2011; pp 665–679.
- (3) Barth, A. *Biochim. Biophys. Acta, Bioenerg.* **2007**, *1767*, 1073–1101.
- (4) Quintás, G.; Kuligowski, J.; Lendl, B. *Anal. Chem.* **2009**, *81*, 3746–3753.
- (5) Kuligowski, J.; Quintás, G.; Garrigues, S.; de la Guardia, M. *Talanta* **2010**, *80*, 1771–1776.
- (6) Kuligowski, J.; Quintás, G.; Garrigues, S.; de la Guardia, M. *Anal. Chim. Acta* **2008**, *624*, 278–285.
- (7) Kuligowski, J.; Quintás, G.; Garrigues, S.; de la Guardia, M. *J. Chromatogr. A* **2009**, *1216*, 3122–3130.
- (8) Quintás, G.; Lendl, B.; Garrigues, S.; de la Guardia, M. *J. Chromatogr. A* **2008**, *1190*, 102–109.
- (9) Quintás, G.; Kuligowski, J.; Lendl, B. *Appl. Spectrosc.* **2009**, *63*, 1363–1369.
- (10) Singh, B. R. Basic Aspects of the Technique and Applications of Infrared Spectroscopy of Peptides and Proteins. *Infrared Analysis of Peptides and Proteins*; American Chemical Society: Washington D.C., USA, 1999; pp 2–37.
- (11) Fabian, H.; Mantele, W. Infrared Spectroscopy of Proteins. In *Handbook of Vibrational Spectroscopy*; John Wiley & Sons: Hoboken, USA, 2006.
- (12) Yang, H.; Yang, S.; Kong, J.; Dong, A.; Yu, S. *Nat. Protoc.* **2015**, *10*, 382–396.
- (13) Turula, V.; de Haseth, J. *Appl. Spectrosc.* **1994**, *48*, 1255–1264.
- (14) Turula, V. E.; de Haseth, J. A. *Anal. Chem.* **1996**, *68*, 629–638.
- (15) Turula, V.; Bishop, R.; Ricker, R.; de Haseth, J. *J. Chromatogr. A* **1997**, *763*, 91–103.
- (16) Kuligowski, J.; Quintás, G.; Guardia, M.; Lendl, B. Liquid Chromatography—Liquid Chromatography—Fourier Transform Infrared. *Encyclopedia of Analytical Science*, 3rd ed.; Elsevier: Amsterdam, Netherlands, 2019; pp 75–85.
- (17) Ramer, G.; Lendl, B. Attenuated Total Reflection Fourier Transform Infrared Spectroscopy. In *Encyclopedia of Analytical Chemistry*; Meyers, R., Meyers, R., Eds.; John Wiley & Sons: Hoboken, USA, 2013.
- (18) Großhans, S.; Rüdts, M.; Sanden, A.; Brestrich, N.; Morgenstern, J.; Heissler, S.; Hubbuch, J. *J. Chromatogr. A* **2018**, *1547*, 37–44.
- (19) Sanden, A.; Suhm, S.; Rüdts, M.; Hubbuch, J. *J. Chromatogr. A* **2019**, *1608*, 460410.
- (20) Walch, N.; Scharl, T.; Felföldi, E.; Sauer, D. G.; Melcher, M.; Leisch, F.; Dürauer, A.; Jungbauer, A. *Biotechnol. J.* **2019**, *14*, No. e1800521.
- (21) Faist, J.; Capasso, F.; Sivco, D.; Sirtori, C.; Hutchinson, A.; Cho, A. *Science* **1994**, *264*, 553–556.
- (22) Schwaighofer, A.; Brandstetter, M.; Lendl, B. *Chem. Soc. Rev.* **2017**, *46*, 5903–5924.
- (23) Schwaighofer, A.; Lendl, B. Quantum cascade laser-based infrared transmission spectroscopy of proteins in solution. In *Vibrational Spectroscopy in Protein Research*; Ozaki, Y., Baranska, M., Lednev, I., Wood, B., Eds.; Academic Press: Cambridge, USA, 2020; pp 59–88.
- (24) Alcaráz, M. R.; Schwaighofer, A.; Kristament, C.; Ramer, G.; Brandstetter, M.; Goicoechea, H.; Lendl, B. *Anal. Chem.* **2015**, *87*, 6980–6987.
- (25) Schwaighofer, A.; Montemurro, M.; Freitag, S.; Kristament, C.; Culzoni, M. J.; Lendl, B. *Anal. Chem.* **2018**, *90*, 7072–7079.
- (26) Akhgar, C. K.; Ramer, G.; Žbik, M.; Trajnerowicz, A.; Pawluczyk, J.; Schwaighofer, A.; Lendl, B. *Anal. Chem.* **2020**, *92*, 9901–9907.
- (27) Chon, B.; Xu, S.; Lee, Y. J. *Anal. Chem.* **2021**, *93*, 2215–2225.
- (28) Dabrowska, A.; David, M.; Freitag, S.; Andrews, A. M.; Strasser, G.; Hinkov, B.; Schwaighofer, A.; Lendl, B. *Sens. Actuators, B* **2022**, *350*, 130873.
- (29) Schwaighofer, A.; Akhgar, C. K.; Lendl, B. *Spectrochim. Acta, Part A* **2021**, *253*, 119563.
- (30) Akhgar, C. K.; Ebner, J.; Spadiut, O.; Schwaighofer, A.; Lendl, B. *Anal. Chem.* **2022**, *94*, 5583–5590.
- (31) de Juan, A.; Jaumot, J.; Tauler, R. *Anal. Methods* **2014**, *6*, 4964.
- (32) Kucheryavskiy, S.; Windig, W.; Bogomolov, A. Chapter 3—Spectral Unmixing Using the Concept of Pure Variables. In *Data Handling in Science and Technology*; Ruckebusch, C., Ed.; Elsevier, 2016; pp 53–99.
- (33) Tauler, R. *Chemom. Intell. Lab. Syst.* **1995**, *30*, 133–146.
- (34) Hu, B.; Sun, D.-W.; Pu, H.; Wei, Q. *Talanta* **2020**, *217*, 120998.
- (35) Mansoldo, F. R. P.; Berrino, E.; Guglielmi, P.; Carradori, S.; Carta, F.; Secci, D.; Supuran, C. T.; Vermelho, A. B. *Spectrochim. Acta, Part A* **2022**, *267*, 120602.
- (36) Kopp, J.; Zauner, F. B.; Pell, A.; Hausjell, J.; Humer, D.; Ebner, J.; Herwig, C.; Spadiut, O.; Slouka, C.; Pell, R. *J. Pharm. Biomed. Anal.* **2020**, *188*, 113412.
- (37) Strixner, T.; Kulozik, U. 7—Egg proteins. In *Handbook of Food Proteins*; Phillips, G. O., Williams, P. A., Eds.; Woodhead Publishing, 2011; pp 150–209.
- (38) Wilcox, P. E. [5] Chymotrypsinogens—chymotrypsins. *Methods in Enzymology*; Academic Press, 1970; pp 64–108.
- (39) Zaia, J.; Annan, R. S.; Biemann, K. *Rapid Commun. Mass Spectrom.* **1992**, *6*, 32–36.
- (40) Akhgar, C.; Ebner, J.; Spadiut, O.; Schwaighofer, A.; Lendl, B. Laser-based mid-infrared spectroscopy enables in-line detection of protein secondary structure from preparative liquid chromatography. *Biomedical Vibrational Spectroscopy 2022: Advances in Research and Industry*; SPIE, 2022; Vol. 11957.
- (41) Windig, W.; Bogomolov, A.; Kucheryavskiy, S. Two-Way Data Analysis: Detection of Purest Variables. In *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*; Brown, S., Tauler, R., Walczak, B., Eds.; Elsevier, 2020; pp 275–307.
- (42) Stein, P. E.; Leslie, A. G. W.; Finch, J. T.; Turnell, W. G.; McLaughlin, P. J.; Carrell, R. W. *Nature* **1990**, *347*, 99–102.
- (43) Dong, A.; Meyer, J. D.; Brown, J. L.; Manning, M. C.; Carpenter, J. F. *Arch. Biochem. Biophys.* **2000**, *383*, 148–155.
- (44) Freer, S. T.; Kraut, J.; Robertus, J. D.; Wright, H. T.; Nguyen-Huu-Xuong, X. *Biochemistry* **1970**, *9*, 1997–2009.

- (45) Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C. *Nature* **1958**, *181*, 662–666.
- (46) Kuligowski, J.; Schwaighofer, A.; Alcaráz, M. R.; Quintás, G.; Mayer, H.; Vento, M.; Lendl, B. *Anal. Chim. Acta* **2017**, *963*, 99–105.
- (47) Wisdom, G. B. Horseradish Peroxidase Labeling of IgG Antibody. In *The Protein Protocols Handbook*; Walker, J. M., Ed.; Humana Press: Totowa, NJ, 2009; pp 681–683.
- (48) Madureira, A. R.; Pereira, C. I.; Gomes, A. M. P.; Pintado, M. E.; Xavier Malcata, F. *Food Res. Int.* **2007**, *40*, 1197–1211.
- (49) Gajhede, M.; Schuller, D. J.; Henriksen, A.; Smith, A. T.; Poulos, T. L. *Nat. Struct. Mol. Biol.* **1997**, *4*, 1032–1038.
- (50) Tavares, T. S.; da Rocha, E. P.; Esteves Nogueira, F. G.; Torres, J. A.; Silva, M. C.; Kuca, K.; Ramalho, T. C. *Molecules* **2020**, *25*, 259.
- (51) Dousseau, F.; Pezolet, M. *Biochemistry* **1990**, *29*, 8771–8779.
- (52) Monaco, H. L.; Zanotti, G.; Spadon, P.; Bolognesi, M.; Sawyer, L.; Eliopoulos, E. E. *J. Mol. Biol.* **1987**, *197*, 695–706.
- (53) Armenta, S.; Garrigues, S.; de la Guardia, M. *Trends Anal. Chem.* **2008**, *27*, 497–511.
- (54) Pena-Pereira, F.; Wojnowski, W.; Tobiszewski, M. *Anal. Chem.* **2020**, *92*, 10076–10082.
- (55) Gałuszka, A.; Migaszewski, Z.; Namieśnik, J. *Trends Anal. Chem.* **2013**, *50*, 78–84.

Recommended by ACS

Surface-Charged Hybrid Monolithic Column for MS-Compatible Peptide Separation with High Peak Capacity and Its Application in Proteomic Analysis

Chao Wang, Yukui Zhang, *et al.*

JUNE 28, 2022
ANALYTICAL CHEMISTRY

READ 

Pressure-Enhanced Liquid Chromatography, a Proof of Concept: Tuning Selectivity with Pressure Changes and Gradients

Szabolcs Fekete, Matthew A. Lauber, *et al.*

MAY 23, 2022
ANALYTICAL CHEMISTRY

READ 

Sense and Nonsense of Elevated Column Temperature in Proteomic Bottom-up LC–MS Analyses

Juraj Lenčo, František Švec, *et al.*

OCTOBER 21, 2020
JOURNAL OF PROTEOME RESEARCH

READ 

Sensitive and Reproducible Mass Spectrometry-Compatible RP-UHPLC Analysis of Tricarboxylic Acid Cycle and Related Metabolites in Biological Fluids: Application to...

Kerri M. Smith, Paul D. Rainville, *et al.*

DECEMBER 08, 2020
ANALYTICAL CHEMISTRY

READ 

[Get More Suggestions >](#)

B. Curriculum Vitae

Christopher Karim Akhgar

Contact	c.akhgar@gmx.net
Nationality	Austrian

Education

05 2019- 10 2022	PhD Candidate at Institute of Chemical Technologies and Analytics, TU Wien. Topic: <i>Novel Mid-infrared based Methods for Milk Analysis and Bioprocess Monitoring</i> . Supervisor: Prof. Bernhard Lendl.
05 2015- 02 2018	MSc in "Food Science and Technology", BOKU Wien Master Thesis: <i>Optimisation and validation of an LC-MS/MS based multi-target method for the determination of mycotoxins in food and feed matrices</i> . Supervisor: Prof. Franz Berthiller.
09 2015- 07 2016	Student exchange at Faculty for Bioscience Engineering, University of Ghent, Belgium, in the framework of Erasmus programme
11 2011- 04 2015	BSc in "Food Science and Biotechnology", BOKU Wien BSc Thesis: <i>Electrophoretic characterization of proteins from different categories of heat treated milk product samples</i> . Supervisor: Prof. Helmut Mayer.
09 2003- 06 2010	Bundesrealgymnasium, Salzburg with specialization in Informatics

Work Experience

05 2019- 10 2022	Project assistant at the Institute of Chemical Technologies and Analytics, TU Wien
03 2018- 04 2019	Laboratory Specialist at Romer Labs, Tulln an der Donau
01 2017- 08 2017	Master student at Romer Labs, Tulln an der Donau

Awards and Recognitions

06 2022	Front cover of Applied Spectroscopy
05 2022	Poster Award at 17 th International Symposium on Hyphenated Techniques in Chromatography and Separation Technology, Ghent, Belgium
10 2020	Poster Award at SciX Conference, Online

Peer-Reviewed Publications

ORCID: [0000-0001-8266-043X](https://orcid.org/0000-0001-8266-043X)

1. Akhgar, C. K.; Ebner, J.; Alcaraz, M.R.; Kopp, J.; Goicoechea, H.; Spadiut, O.; Schwaighofer, A.; Lendl, B. Application of Quantum Cascade Laser-Infrared Spectroscopy and Chemometrics for In-Line Discrimination of Coeluting Proteins from Preparative Size Exclusion Chromatography. *Anal. Chem.* **2022**, *94*, 11192-11200.
2. Akhgar, C. K.; Ebner, J.; Spadiut, O.; Schwaighofer, A.; Lendl, B. QCL-IR Spectroscopy for In-Line Monitoring of Proteins from Preparative Ion-Exchange Chromatography. *Anal. Chem.* **2022**, *94*, 5583-5590.
3. Akhgar, C. K.; Nürnberger, V.; Nadvornik, M.; Ramos-Garcia, V.; Ten-Doménech, I.;

- Kuligowski, J.; Schwaighofer, A.; Rosenberg, E.; Lendl, B. Fatty Acid Determination in Human Milk Using Attenuated Total Reflection Infrared Spectroscopy and Solvent-Free Lipid Separation. *Appl. Spec.* **2022**, *76*, 730-736.
4. Akhgar, C. K.; Nürnberger, V.; Nadvornik, M.; Velik, M.; Schwaighofer, A.; Rosenberg, E.; Lendl, B. Fatty Acid Prediction in Bovine Milk by Attenuated Total Reflection Infrared Spectroscopy after Solvent-Free Lipid Separation. *Foods* **2021**, *10*, 1054.
 5. Schwaighofer, A.; Akhgar, C. K.; Lendl, B. Broadband laser-based mid-IR spectroscopy for analysis of proteins and monitoring of enzyme activity. *Spectrochim. Acta A* **2021**, *253*, 119563.
 6. Akhgar, C. K.; Ramer, G.; Žbik, M.; Trajnerowicz, A.; Pawluczyk, J.; Schwaighofer, A.; Lendl, B. The Next Generation of IR Spectroscopy: EC-QCL-Based Mid-IR Transmission Spectroscopy of Proteins with Balanced Detection. *Anal. Chem.* **2020**, *92*, 9901-9907.

Conference Proceedings

7. Akhgar, C.K.; Ebner, J.; Spadiut, O.; Schwaighofer, A.; Lendl, B. Laser-based mid-infrared spectroscopy enables in-line detection of protein secondary structure from preparative liquid chromatography. *Proc. Spie.* **2022**, *11957*.

Conference Contributions

1. Akhgar, C. K.; Nürnberger, V.; Nadvornik, M.; Ramos-Garcia, V.; Ten-Doménech, I.; Kuligowski J.; Schwaighofer, A.; Rosenberg, E.; Lendl, *High-throughput Fatty Acid Profiling in Human Milk by Attenuated Total Reflection Infrared Spectroscopy and Solvent-free Lipid Separation*. Poster: Analytica, Munich. 21.06.2022 – 24.06.2022.
2. Akhgar, C.K.; Ebner, J.; Alcaraz M.R.; Kopp, J.; Goicoechea, H.; Spadiut, O.; Schwaighofer A.; Lendl B, Spadiut, O.; Schwaighofer, A.; Lendl, B. *Application of Laser-Based Mid-Infrared Spectroscopy for Inline Monitoring of Proteins from Preparative Liquid Chromatography*. Talk: Young Analytical Chemists Forum, Tulln. 12.05.2022 – 13.05.2022.
3. Akhgar, C.K.; Ebner, J.; Spadiut, O.; Schwaighofer, A.; Lendl, B. *Laser-Based Mid-Infrared Spectroscopy Enables In-line Detection of Protein Secondary Structure from Preparative Liquid Chromatography*. Poster: SPIE Photonics West, San Francisco. 22.01.2022 – 27.01.2022.
4. Akhgar, C. K.; Nürnberger, V.; Nadvornik, M.; Velik, M.; Schwaighofer, A.; Rosenberg, E.; Lendl, B. *Fatty Acid Profiling in Bovine Milk by Attenuated Total Reflection Infrared Spectroscopy and Rapid Lipid Separation*. Poster: SciX, Virtual Meeting. 26.09.2021 – 01.10.2021.
5. Akhgar, C. K.; Nürnberger, V.; Nadvornik, M.; Velik, M.; Schwaighofer, A.; Rosenberg, E.; Lendl, B. *Bovine Milk Fatty Acid Profiling by Attenuated Total Reflection Infrared Spectroscopy and Rapid Lipid Separation*. Flash Presentation: 11th International Conference on Advanced Vibrational Spectroscopy, Virtual Meeting. 23.08.2021 – 26.08.2021.
6. Akhgar, C. K.; Ramer, G.; Žbik, M.; Trajnerowicz, A.; Pawluczyk, J.; Schwaighofer, A.; Lendl, B. *Balanced Detection enables highly sensitive EC-QCL Based Mid-IR Transmission Spectroscopy of Proteins*. Poster: SciX, Virtual Meeting, 12.10.2020 – 15.10.2020.
7. Akhgar, C. K.; Schwaighofer, A.; Ramer, G.; Lendl, B. *The Next Generation of IR Spectroscopy: EC-QCL-Based Mid-IR Transmission Spectroscopy of Proteins with Balanced Detection*. Poster: Infrared and Raman Discussion Group (IRDG) Meeting, London, 20.12.2019.