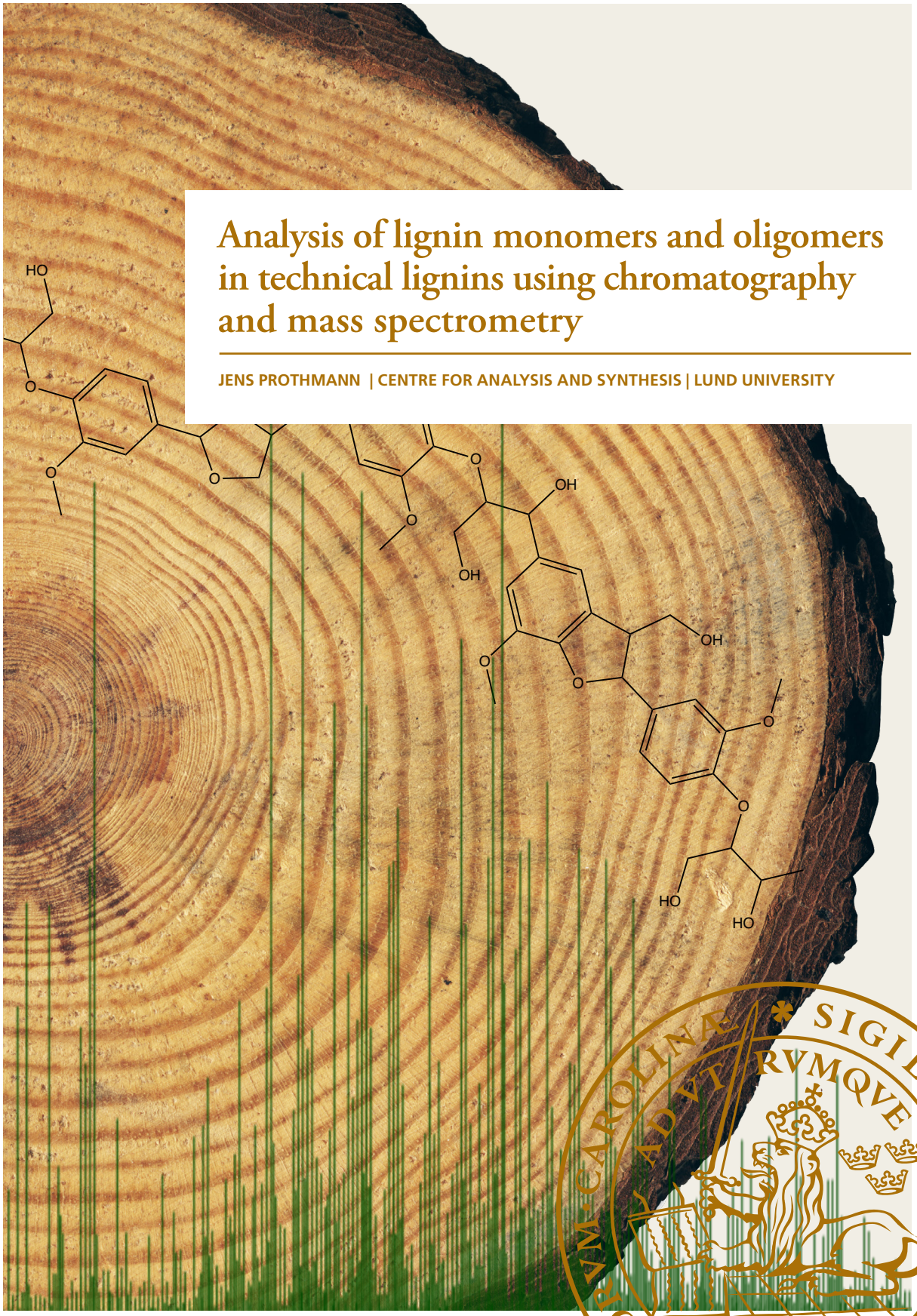
Analysis of lignin monomers and oligomers in technical lignins using chromatography and mass spectrometry

Prothmann, Jens

2020

Analysis of lignin monomers and oligomers in technical lignins using chromatography and mass spectrometry

JENS PROTHMANN | CENTRE FOR ANALYSIS AND SYNTHESIS | LUND UNIVERSITY



Analysis of lignin monomers and oligomers in technical lignins using chromatography and mass spectrometry

Analysis of lignin monomers and oligomers in technical lignins using chromatography and mass spectrometry

Jens Prothmann

LUND
UNIVERSITY

DOCTORAL DISSERTATION

by due permission of the Faculty of Science, Lund University, Sweden.
To be defended at lecture hall B, Kemicentrum, Naturvetarvägen 14, Lund, Sweden on Wednesday, December 9, 2020, at 14:00.

Faculty opponent
Prof. Hilkka I. Kenttämaa
Purdue University

Organization LUND UNIVERSITY Department of Chemistry Centre for Analysis and Synthesis Author(s) Jens Prothmann	**Document name** Doctoral Thesis in Analytical Chemistry **Date of issue** 2020-11-13 Sponsoring organization

Title and subtitle Analysis of lignin monomers and oligomers in technical lignins using chromatography and mass spectrometry

Abstract

Lignin is the most abundant aromatic biopolymer on earth and has the potential to play an important role in the transition from an oil-based refinery to a biorefinery-based industry. The isolation of lignin from the biomass can be achieved by several different technical processes and isolated lignins are referred to as technical lignins in literature. During the last decades the conversion of technical lignins into economically valuable aromatic chemicals became a growing research interest.

For the understanding of the chemical nature of technical lignins and for the understanding of the chemical reactions occurring during conversion processes, selective and sensitive analytical methods are crucial. Therefore, analytical methods using liquid chromatography (LC) and supercritical fluid chromatography (SFC) coupled to electrospray ionisation–mass spectrometry (ESI–MS) were developed in this thesis work. Main emphases were set on the optimisation of the ESI of lignin monomers (LMs), the identification of unknown LMs and lignin oligomers (LOs) in complex technical lignin samples and on the separation of LMs and LOs using SFC.

Several ESI parameter were studied for the ionisation of LMs to identify which ESI parameters have significant influences. First, the ESI of a wide range of LMs were studied using SFC/ESI–MS, then the influence of specific ESI parameters on different groups of LMs. Significant ESI parameters, such as the concentration of the makeup solvent additive or the desolvation gas temperature, were identified as significant parameters for the ESI of a wide range of LMs. Furthermore, it has been shown that for the different groups of LMs, different ESI parameters are of importance. For instance, compounds with two methoxy groups seem to need more desolvation energy compared to compounds with one or no methoxy groups.

The identification of unknown LMs and LOs in technical lignin samples using MS is very challenging due to the very complex sample mixtures and the lack of commercially available reference standards. Therefore, a non-targeted analysis strategy using SFC/ESI–high-resolution multiple stage tandem MS combined with Kendrick mass defect-based principal component analysis–quadratic discriminant analysis classification models were developed. The developed method assures an identification confidence of level 3 without the need of reference standards and without the study of MS fragmentation patterns. Furthermore, it was demonstrated that a higher identification confidence level and tentative chemical structures can be obtained with multiple stage MS.

For the characterisation of unknown compounds in complex samples, clean mass spectra of the investigated compound are useful. In this work, the application of SFC for the separation of LMs and LOs has been investigated using a stationary phase screening approach. It was found that a SFC stationary phase with both hydrogen bonding and π-π-interaction chemistries offers the highest overall resolution power combined with a selective separation of LMs and lignin dimers.

Key words Classification models, Design of experiments, Electrospray ionisation, High-resolution mass spectrometry, Supercritical fluid chromatography

Classification system and/or index terms (if any)

Supplementary bibliographical information		**Language** English
ISSN and key title		**ISBN** 978-91-7422-760-4
Recipient's notes	**Number of pages** 348	Price
	Security classification	

Analysis of lignin monomers and oligomers in technical lignins using chromatography and mass spectrometry

Jens Prothmann

LUND
UNIVERSITY

Cover designed by Jens Prothmann and Lisa Gehringer

Doctoral Thesis

Faculty of Science
Department of Chemistry
Centre for Analysis and Synthesis
P.O. Box 124
SE-22 100 Lund, Sweden

ISBN 978-91-7422-760-4 (printed version)
ISBN 978-91-7422-761-1 (online version)

Printed in Sweden by Media-Tryck, Lund University
Lund 2020

Abstract

Lignin is the most abundant aromatic biopolymer on earth and has the potential to play an important role in the transition from an oil-based refinery to a biorefinery-based industry. The isolation of lignin from the biomass can be achieved by several different technical processes and isolated lignins are referred to as technical lignins in literature. During the last decades, the conversion of technical lignins into economically valuable aromatic chemicals became a growing research interest.

For the understanding of the chemical nature of technical lignins and the understanding of the chemical reactions occurring during conversion processes, selective and sensitive analytical methods are crucial. Therefore, analytical methods using liquid chromatography (LC) and supercritical fluid chromatography (SFC) coupled to electrospray ionisation–mass spectrometry (ESI–MS) were developed in this thesis work. Main emphases were set on the optimisation of the ESI of lignin monomers (LMs), the identification of unknown LMs and lignin oligomers (LOs) in complex technical lignin samples and on the separation of LMs and LOs using SFC.

Several ESI parameters were studied for the ionisation of LMs to identify which ESI parameters have significant influences. First, the ESI of a wide range of LMs were studied using SFC/ESI–MS, then the influence of specific ESI parameters on different groups of LMs was analysed. Significant ESI parameters, such as the concentration of the makeup solvent additive or the desolvation gas temperature, were identified as significant parameters for the ESI of a wide range of LMs. Furthermore, it has been shown that for the different groups of LMs, different ESI parameters are of importance. For instance, compounds with two methoxy groups seem to need more desolvation energy compared to compounds with one or no methoxy groups.

The identification of unknown LMs and LOs in technical lignin samples using MS is very challenging due to the very complex sample mixtures and the lack of commercially available reference standards. Therefore, a non-targeted analysis strategy using SFC/ESI–high-resolution multiple stage tandem MS combined with Kendrick mass defect-based principal component analysis–quadratic discriminant analysis classification models was developed. The developed method assures an identification confidence of level 3 without the need of

reference standards and without the study of MS fragmentation patterns. Furthermore, it was demonstrated that a higher identification confidence level and tentative chemical structures can be obtained with multiple stage tandem MS.

For the characterisation of unknown compounds in complex samples, clean mass spectra of the investigated compound are useful. In this work, the application of SFC for the separation of LMs and LOs has been investigated using a stationary phase screening approach. It was found that a SFC stationary phase with both hydrogen bonding and π-π-interaction chemistries offers the highest overall resolution power combined with a selective separation of LMs and lignin dimers.

Popular Science Summary

This thesis is dedicated to the analysis of technical lignins using two different analytical techniques called chromatography and mass spectrometry.

Lignin is one of three major components of every plant besides cellulose and hemi-cellulose. For instance, a tree can consist of up to 30% lignin. In the paper industry only the cellulose and hemi-cellulose are used for paper production. The lignin on the other hand gets isolated from both and is then called technical lignin. Over the years many different isolation procedures have been developed resulting in different types of technical lignins with different chemical properties. So far technical lignins have no economic value for the paper industry and are often burned for energy production. However, since technical lignins show similar chemical structures to crude oil, many researchers believe that lignin has the potential to replace crude oil as the starting material for industrial products, like plastics or pharmaceutical drugs. Therefore, many studies have been performed to find a use for technical lignins either directly in their isolated form or through a chemical or biological conversion into specific chemicals of economic value.

Knowledge about the chemical structures of technical lignins both before the conversion as well as afterwards is essential for the different applications and possible process optimisations. Consequently, chemical analysis methods need to be developed to identify and quantify those chemical structures. The main problem is that technical lignins are very complex samples and can contain hundreds or thousands of different chemicals. In addition to that, depending on the different isolation processes, technical lignins are chemically different from each other and if a technical lignin is converted, its complexity will increase even further. A main challenge is therefore to identify the chemical structures of lignin in a highly reliable way.

Chromatography and mass spectrometry are suitable analytical instruments for this reliable identification of chemical structures in complex samples like technical lignins. In chromatography, chemicals dissolved in a liquid can be separated based on their chemical properties and in mass spectrometry chemicals can be separated and identified based on their mass. These two analytical techniques can be combined to use the strength of both systems at the same time.

In the first part of this thesis, studies were performed to improve the detectability of typical lignin compounds using mass spectrometry. A better detectability will

help to improve the identification reliability of unknown chemical structures in technical lignin samples. In the second part of this thesis, an analytical method using chromatography and mass spectrometry was used in combination with a developed and tailored data mining tool to identify unknown chemical structures in technical lignin samples. The developed method offers a solution to the mentioned problem and is a fast and reliable identification tool. In the last part of the thesis a study was conducted to investigate if the separation of chemical structures typical for lignin could be further improved by the use of so-called supercritical fluid chromatography.

List of publications

I. Ultra-high-performance supercritical fluid chromatography with quadrupole-time-of-flight mass spectrometry (UHPSFC/QTOF-MS) for analysis of lignin-derived monomeric compounds in processed lignin samples
Jens Prothmann*, Mingzhe Sun*, Peter Spégel, Margareta Sandahl, Charlotta Turner
Analytical and Bioanalytical Chemistry **2017**, *409*, 7049–7061.
*Shared first authorship

II. Functional group-based optimisation of the electrospray ionisation efficiency of lignin monomers using ultra-high-performance supercritical fluid chromatography coupled to electrospray ionisation–mass spectrometry (UHPSFC/ESI–MS)
Jens Prothmann, Daniel Molins-Delgado, Alexander Braune, Margareta Sandahl, Charlotta Turner
Manuscript

III. Identification of lignin oligomers in Kraft lignin using ultra-high-performance liquid chromatography/high-resolution multiple-stage tandem mass spectrometry (UHPLC/HRMS^n)
Jens Prothmann, Peter Spégel, Margareta Sandahl, Charlotta Turner
Analytical and Bioanalytical Chemistry **2018**, *410*, 7803–7814.

IV. Non-targeted analysis strategy for the identification of phenolic compounds in complex technical lignin samples
Jens Prothmann, Kena Li, Christian Hulteberg, Peter Spégel, Margareta Sandahl, Charlotta Turner
ChemSusChem **2020**, *13*, 4605–4612.

V. Separation of lignin-related phenolic compounds in lignosulphonate lignin using ultra-high-performance supercritical fluid chromatography
Jens Prothmann, Simon Palmer, Charlotta Turner, Margareta Sandahl
Manuscript

Author's contributions

I. I participated in developing the idea of the project, performed part of the lab work and wrote one part of the manuscript.

II. I synthesised the initial idea of the project, conducted all of the planning, most experiments, most data evaluation and wrote the manuscript.

III. I came up with the initial research idea, conducted all experiments, data evaluation and wrote the manuscript.

IV. I synthesised the initial research idea, conducted all of the planning, most experiments, all data evaluation and wrote the manuscript.

V. I synthesised the initial idea of the project, conducted all of the planning, performed all experiments, performed part of the data evaluation and wrote the manuscript.

Publications not included

I. Biological valorization of low molecular weight lignin
Omar Y. Abdelaziz, Daniel P. Brink, **Jens Prothmann**, Krithika Ravi, Mingzhe Sun, Javier García-Hidalgo, Margareta Sandahl, Christian P. Hulteberg, Charlotta Turner, Gunnar Lidén, Marie F. Gorwa-Grauslund
Biotechnology Advances **2016**, *34*, 1318–1346.

II. A sediment extraction and cleanup method for wide-scope multitarget screening by liquid chromatography–high-resolution mass spectrometry
Riccardo Massei, Harry Byers, Liza-Marie Beckers, **Jens Prothmann**, Werner Brack, Tobias Schulze, Martin Krauss
Analytical and Bioanalytical Chemistry **2017**, *410*, 177–188.

III. Oxidative Depolymerisation of Lignosulphonate Lignin into Low-Molecular-Weight Products with Cu-Mn/ δ-Al_2O_3
Omar Y. Abdelaziz, Sebastian Meier, **Jens Prothmann**, Charlotta Turner, Anders Riisager, Christian P. Hulteberg
Topics in Catalysis **2019**, *62*, 639–648.

IV. Identification of polyphenols in the brown algae, *Saccharina latissima* and *Ascophyllum nodosum*
Roya R. R. Sardari, **Jens Prothmann**, Olavur Gregersen, Charlotta Turner, Eva Nordberg Karlsson
Submitted

V. A study of the spatial distribution patterns of airborne PAHs in crowberry (*empetrum nigrum*) in Ilulissat, Greenland
Oskar Munk Kronik*, **Jens Prothmann***, Gaudry Troche*, Bo Svensmark, Nikoline J. Nielsen, Jan H. Christensen
*Shared first authorship
Submitted

VI. Production of muconic acid in engineered *Saccaromyces cerevisiae*
Daniel P. Brink, Nina Muratovska, **Jens Prothmann**, Viktor Persson, Charlotta Turner, Marie F. Gorwa-Grauslund
Manuscript

VII. Investigation of the effect of the depolymerisation of black liquor retentate on the production of monomers and oligomers
Kena Li, **Jens Prothmann**, Margareta Sandahl, Charlotta Turner, Christian Hulteberg
Manuscript

VIII. Optimisation of lignin extraction from woodchips using supercritical fluid extraction (SFE) and CO_2-expanded liquid extraction (CXLE)
Federica Nardella, **Jens Prothmann**, Margareta Sandahl, Erika Ribechini, Charlotta Turner
Manuscript

Abbreviations

1-AA	1-aminoanthracene
2-PIC	2-picolylamine
8-5	Phenylcoumaran-linkage
8-8	Resinol-linkage
8-O-4	8-aryl ether-linkage
APCI	Atmospheric pressure chemical ionisation
API	Atmospheric pressure ionisation
APPI	Atmospheric pressure photo ionisation
BEH	Ethylene bridge hybrid
CID	Collision induced dissociation
CSH	Charged surface hybrid
DEA	Diethylamine
DI	Direct infusion
DIOL	High-density diol
DoE	Design of experiments
EI	Electron ionisation
ESI	Electrospray ionisation
FP	Fluorophenyl
FT-ICR-MS	Fourier transform-ion cyclotron resonance-mass spectrometry
GC	Gas chromatography
HPLC	High-performance liquid chromatography
HRMS	High-resolution mass spectrometry
HRMS^n	High-resolution multiple stage tandem mass spectrometry
IR	Infrared spectroscopy
KMD	Kendrick mass defect
kNN	k-nearest neighbours
LC	Liquid chromatography
LD	Lignin dimers
LDA	Linear discriminant analysis
LLE	Liquid-liquid extraction
LQIT	Linear quadrupole ion trap
LM	Lignin monomer
LO	Lignin oligomer
LOD	Limit of detection
LTE	Lignin tetramer

LTR	Lignin trimer
MALDI	Matrix-assisted laser desorption/ionisation
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MS^n	Multiple stage tandem mass spectrometry
m/z	Mass-over-charge ratio
NMR	Nuclear magnetic resonance spectroscopy
PC	Principal component
PCA-QDA	Principal component analysis-quadratic discriminant analysis
QTOF-MS	Quadrupole–time-of-flight mass spectrometry
QQQ-MS	Triple quadrupole mass spectrometry
RDB	Ring double bond equivalent
RP	Reversed phase
$scCO_2$	Supercritical carbon dioxide
SEC	Size exclusion chromatography
SIMCA	Soft independent modelling of class analogy
TOF-MS	Time-of-flight mass spectrometry
UHPLC	Ultra-high-performance liquid chromatography
UHPSFC	Ultra-high-performance supercritical fluid chromatography

Table of Contents

1. Introduction

Lignin is the second most abundant biopolymer in the biomass on our planet. In the biosphere approximately 300 billion tons of lignin are available and approximately 20 billion tons are produced annually by biosynthesis[1]. With its aromatic nature, lignin is a promising renewable resource that could reduce our dependency on crude oil. Therefore, lignin might have an important role in the transition process from an oil refinery-based industry to a biorefinery-based industry[2]. Several different technical processes like the soda process, the organosolv process, the Kraft process or the lignosulphonate process were established over years to isolate lignin from the biomass[3,4]. In the literature, isolated lignins are usually referred to as technical lignins or industrial lignins. Using a technical lignin as a starting material, valuable aromatic compounds can be produced by tailor-made depolymerisation or biological conversion[3].

The chemical structure of native lignin is still not well known. It is however commonly accepted that the lignin biopolymer structure consists of the three main polymer subunits *p*-coumaryl alcohol (H-unit), coniferyl alcohol (G-unit) and sinapyl alcohol (S-unit). A tentative chemical structure of native lignin is shown in Figure 1. The monomeric subunits are linked to each other by different combinations of ether and carbon-carbon bonds. Several different types of linkages are reported in the literature[4]. The most dominant linkage type is the 8-aryl ether-linkage (8-O-4). Other dominant linkages are the resinol-linkage (8-8) and the phenylcoumaran-linkage (8-5). Also, a direct linkage between two aromatic rings is possible like in the 5-5' linkage. So far, no repeating sequence of monomeric subunits and linkages has been identified, which hints that the biosynthesis of the biopolymer occurs in a randomized manner. It also explains the complex heterogeneous structure of the lignin biopolymer. Furthermore, the chemical structure of lignin shows differences depending on the type of lignin, like for example softwood or hardwood lignin[4].

During the isolation process of lignin from the biomass, the native chemical structure of lignin is changed and the complexity of the structure increases. Lignin can be isolated from the biomass by different isolation processes. Torres *et al* recently grouped the isolation processes into four different groups according to their main operation conditions[5]. The isolation processes were grouped into sulfur-containing processes, e.g. the Kraft and lignosulphonate process, sulfur-free processes, e.g. the soda and organosolv process, new generation "greener"

processes, e.g. the ionic liquid and deep eutectic solvent processes, and in the last group other processes, e.g. pyrolysis and hydrolysis processes[5]. Technical details of specific lignin isolation processes can be found in review articles in literature, like by Torres *et al*[5]. Depending on the isolation process, different chemical reactions will take place leading to different chemical structures of technical lignins. For example, in the alkaline Kraft process, the changes are mainly due to hydrolysis and condensation reactions. Functional groups are eliminated and new functional groups, like phenolic hydroxyl groups or aliphatic hydroxyl groups, are formed. Mainly the ether groups in the different linkages are cleaved, leading to an increased amount of phenolic hydroxyl groups and a decreased average molecular mass. A more detailed description of main lignin reactions in alkaline processes is given by Berlin and Balakshin[4].

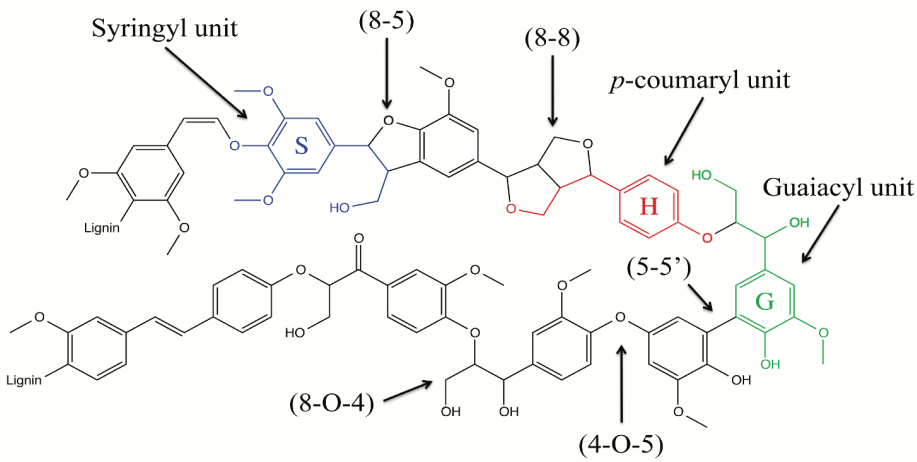


Figure 1. Tentative structure of native lignin with coloured subunits and different types of linkages. Figure from **Paper III,** reprinted with permission.

Many chemical and biological conversion processes of technical lignins have been proposed during the years for the production of small molecular weight compounds with a high value on the commercial market, such as vanillin or guaiacol[3]. With these conversion processes another chemical transformation process is introduced, leading possibly to an increased sample complexity. Moreover, the prediction of chemical reactions occurring during such conversion processes is challenging.

For a better understanding of the chemical processes during isolation, depolymerisation or biological conversion of technical lignins, knowledge about the chemical structures of its components before and after the treatment is crucial for further process improvements. Therefore, several different analytical techniques were used for the analysis of technical lignins during the years, such

as nuclear magnetic resonance spectroscopy (NMR)[6,7], infrared spectroscopy (IR)[7,8], size exclusion chromatography (SEC)[7,8], gas chromatography (GC)[9–11], liquid chromatography (LC)[12,13], supercritical fluid chromatography (SFC)[14] and mass spectrometry (MS)[15,16]. Each analytical technique has its own strengths and weaknesses for the analysis of complex chemical mixtures such as technical lignins.

With NMR and IR analysis functional groups within the complex mixture can be identified[7,8]. Furthermore, structural changes within the sample during depolymerisation processes can be monitored using the change of the peak area of for example the hydroxyl stretching region of 3700 cm^{-1} to 2700 cm^{-1} in IR[7]. However, the characterisation of single molecules in complex technical lignin samples using NMR or IR remains challenging due to interfering signals.

Size exclusion chromatography is widely used to determine the average molecular weight of technical lignins[17,18]. However, the molecular weight calibration is still challenging due to a lack of reference standards and also the complete dissolution of technical lignins in solvents suitable for SEC is still a problem[17,18]. But a recent study by Jacobs *et al* showed that a SEC system can also be calibrated for technical lignins using matrix-assisted-laser-desorption-ionisation MS[17].

A major interest in lignin research is the identification of lignin monomers (LMs) and lignin oligomers (LOs), which are produced from the native lignin biopolymer during isolation processed from the biomass and also during chemical and biological conversion of technical ligins[3]. NMR, IR and SEC are usually limited to the identification of certain functional groups within the sample or its average molecular weight. For the identification of LMs and LOs other techniques such as GC, LC, SFC or MS are often used. LMs and LOs are oxygen-containing aromatic compounds. Sometimes, depending on the used isolation and conversion process also other elements can be included in their chemical structure, like for example sulfur in lignosulphonate lignins[19]. LMs contain one aromatic benzene ring, while LOs contain at least two benzene rings. Several different functional groups, such as hydroxy groups, aldehydes or methoxy groups or also aliphatic chains containing several functional groups can be connected to the aromatic rings.

The analysis of technical lignins using GC is mainly focused on the analysis of LMs and often the GC system is coupled to a MS system[9–11]. The advantage of GC/MS for the analysis of technical lignins is the availability of databases, which are especially useful for the identification of LMs. However, the identification of LOs is limited due to a lack of commercially available reference standards and therefore missing entries in the databases. Furthermore, before the analysis by GC/MS, pyrolysis[9,10] or derivatisation of the sample is performed[11]. Pyrolysis or derivatisation methods cause chemical changes of the lignin samples and may

lead to biased results and results that are more difficult to interpret due to an increase of the sample complexity. Additionally, because of the complexity of technical lignin samples, the analysis times of reported GC/MS methods are up to 35 min[9,10].

An alternative separation method with no need of derivatisation or pyrolysis is LC, which is usually coupled to MS (LC/MS) using electrospray ionisation (ESI) in negative mode[12,20]. Reported LC/MS methods using reversed phase (RP) high-performance liquid chromatography (HPLC) columns have about 30 min analysis time, which is comparable to GC/MS methods[12,20]. A study presented by Kiyota *et al* using a RP ultra-high-performance liquid chromatography (UHPLC) column has a faster analysis time of 8 min[13]. Another alternative, offering high analysis speed and selectivity for LMs, is ultra-high-performance supercritical fluid chromatography (UHPSFC), as shown recently by Sun *et al,* presenting a method for the separation of 11 LMs in less than 6 min[14]. Same as for LC in SFC no derivatisation or pyrolysis is needed before the analysis.

Reported MS methods for the analysis of technical lignins are either using direct infusion (DI)-MS[15,21] or MS coupled to chromatographic techniques such as GC[9–11] or LC[12,13]. For the identification of unknown LMs or LOs in complex technical lignin samples high-resolution multiple stage tandem mass spectrometry ($HRMS^n$) is a powerful tool[22]. However, the large number of possible functional groups present in LMs and LOs make it difficult to establish a universal ionisation method for MS analysis using an atmospheric pressure ionization (API) technique like ESI. Particularly, lignin-related compounds with low molecular weight and only containing phenolic hydroxy groups and/or methoxy groups, like phenol, guaiacol or syringol, are difficult to ionize efficiently using ESI. The identification of LOs in complex technical lignin samples is challenging due to the lack of commercially available reference standards. In some studies own LO standards were synthesised[13,23], however, chemical synthesis is time-consuming and can only partly represent chemical structures of LOs present in complex lignin samples.

2. Aim of the thesis

This thesis aimed to extend the boundaries of analytical possibilities for the analysis of lignin monomers and oligomers in technical lignins with a focus on the use of mass spectrometry and multivariate analysis tools. Emphases were set on the electrospray ionisation efficiency of lignin monomers and on the development of a comprehensive non-targeted analysis strategy for the identification of lignin monomers and oligomers in complex technical lignin samples using high-resolution multiple stage tandem mass spectrometry in combination with multivariate classification models. Additionally, the separation of lignin monomers and oligomers using supercritical fluid chromatography was explored. The following research questions were intended to be addressed within the thesis project:

- Which electrospray ionisation parameters have significant influence on the ionisation efficiency for lignin monomers in MS and how does the type of functional group and the number of methoxy groups of a lignin monomer influences the electrospray ionisation?
- Can the combination of $HRMS^n$ and multivariate classification models improve the identification of unknown lignin monomers and oligomers in complex technical lignin samples and to what extent can the combination compensate the lack of commercially available reference standards?
- Which stationary phase chemistry available for SFC shows the best selectivity and best resolving power for lignin monomers and oligomers and why?

3. Mass spectrometry

Mass spectrometry is implemented nowadays in many research fields, such as chemistry, biochemistry or pharmacy, as an indispensable tool for the analysis of organic or inorganic compounds[24]. The main function of MS is the separation of generated ions of organic compounds or elements and their detection based on their mass-over-charge ratio (m/z). MS can be used for both qualitative and quantitative analysis and has been used for example for the identification of unknown compounds in the environment on earth[25], on Mars for the analysis of aeolian sediments and drilled sedimentary deposits[26], for sequencing biomolecules[27] and for the quality control of food[28].

The foundation of what today is known as MS was laid by Joseph J. Thomson in the late 19th and early 20th century[29]. Together with his student at that time, Francis W. Aston, Thomson built the first mass spectrometer in 1912 and obtained mass spectra from O_2, N_2, CO, CO_2 and $COCl_2$[29–31]. For the generation of ions, Thomson and Aston used gas discharge tubes. The generated ions then passed through parallel electric and magnetic fields and got then detected on a photographic plate[29]. The fundamental setup of modern mass spectrometers still follows the setup of the first build MS. Modern mass spectrometer consists mainly of three parts, an ionisation source, a mass analyser and a detector. Both Thomson and Aston received the Nobel Prize for their work related to MS, Thomson in Physics in 1906 and Aston in Chemistry in 1922. It took another thirty years until the first commercial mass spectrometer was introduced in 1942[30]. The foundation for the first modern ionisation source, an electron ionisation (EI) source, was published by Arthur J. Dempster in 1918[32]. The concept of a nowadays widely used mass analyser, the time-of-flight (TOF) mass analyser, was first described by William E. Stephens in 1946[29]. Another, nowadays widely used mass analyser, often coupled to GC, was the quadrupole mass analyser, which was first described by Paul and Steinwedel in 1953[33] and was commercially released in 1968[30]. In the 1980s, MS was established as a routine analytical technique for the analysis of small molecules, however, the analysis of large molecules such as proteins or complex carbohydrates remained challenging[29]. In the late 1980s electrospray ionisation (ESI)[34] and also matrix-assisted laser desorption/ionisation (MALDI)[35] was introduced by Fenn and Tanaka, respectively, which made the ionisation of larger molecules possible and extended the application range of MS enormously. For their developments, Fenn

and Tanaka shared the Nobel Prize in Chemistry in 2002. Melvin B. Comisarow and Alan G. Marshall laid the foundation for modern high-resolution mass spectrometry (HRMS) by implementing fourier transformation ion cyclotron resonance spectroscopy in 1974[36], leading to the development of fourier transform-ion cyclotron resonance (FT-ICR) mass analyser, which, until now, achieve the highest mass resolution of all mass spectrometers. Nowadays several different types of mass analyser exist, such as a quadrupole, TOF and ion traps. The newest type of mass analyser, the Orbitrap, was introduced by Alexander Makarov in 1999. The Orbitrap is an ion trap with a high performance using an electrostatic quadro-logarithmic field[37]. Also, different mass analysers can be combined to powerful analytical instruments offering many different options, such as the quadrupole–time-of-flight (QTOF), triple quadrupole (QQQ) or linear quadrupole ion trap (LQIT)–Orbitrap mass analyser.

Usually, mass analysers are classified into two different types, low-resolution mass analysers and high-resolution mass analysers. The mass resolution *R* in MS is defined as the smallest difference in m/z that can be achieved for a detected peak. While the resolving power is defined by the ability of the MS to distinguish two neighbouring peaks with the same height. HRMS are TOF-MS, FT-ICR-MS and Orbitrap-MS, which allow the determination of the exact mass of a compound, for what a separation of m/z of around 0.001 u or lower is needed. Mass analyser with a low-resolution are quadrupoles and ion traps, which allow the separation of a m/z around 1 u. While HRMS are powerful instruments for the identification of compounds, modern low-resolution MS, like a QQQ-MS, have very high detection sensitivity and therefore are often used for quantitative studies.

During the years several different types of ionisation sources have been developed, which have different application fields. Probably the most used ionisation sources today are EI, chemical ionisation (CI), MALDI and atmospheric pressure ionisation (API) sources, such as ESI, atmospheric pressure chemical ionisation (APCI) and atmospheric pressure photo ionisation (APPI). While EI is mainly used in GC/MS applications, API sources are used for the analysis in the liquid phase using direct infusion (DI)-MS or MS combined with LC or SFC. For solid samples, MALDI is often used. Ionisation sources are grouped in soft and harsh ionisation sources. With soft ionisation sources, like for example ESI, molecular ions of a compound are obtained, while with harsh ionisation techniques, such as EI, mainly fragment ions of the molecular ion are obtained. Besides EI, API and MALDI, many other ionisation sources have been developed. Currently, the development of ambient ionisation techniques for MS is a growing research field for the past two decades[38], for example desorption electrospray ionisation (DESI) or direct analysis in real time (DART) ionisation. DESI can be used for example for the analysis of tissue without any sample preparation and DART for the analysis of object surfaces[38].

3.1 Analysis of technical lignins using MS

For the analysis of technical lignins, both low-resolution MS and high-resolution MS, is applied in the literature. The most common used ionisation techniques for the analysis of LMs and LOs are ESI[13,39], APCI[40,41], APPI[21,42], EI[43,44] or MALDI[45,46]. Some articles present in the literature compared the different API sources for the ionisation of technical lignins and came to different conclusions[42,47,48]. Both, Kosyakov *et al*[42] and Qi *et al*[47], compared the use of ESI, APCI and APPI for the ionisation of lignin samples. While Kosyakov *et al*[42] concluded that APPI in negative ionisation mode shows the highest ionisation efficiency and ESI in negative mode the lowest for LMs and LDs, Qi *et al*[47] concluded in contradiction that ESI in negative mode generates the most ionised components and APCI in negative ionisation mode the fewest. Andrianova *et al*[49] compared in a method development the ionisation efficiency of ESI in negative and in positive ionisation mode for LM, LD and lignin trimer (LTR) model compound standards and also for a Kraft lignin sample. They concluded that the best ESI efficiency can be achieved by ESI in positive ionisation mode. Reasons for the different conclusions may be the use of different types of technical lignin samples or different mass analyser, but of course each API source also has its own ionisation selectivity. In this thesis work, ESI was applied in negative ionisation mode in all projects, therefore the section 3.2 is focused on the use of ESI for the analysis of technical lignins.

For qualitative studies, low-resolution MS is mainly performed using GC/quadrupole-MS due to the advantage of available MS databases, that can be used for the identification of lignin monomers[9,43]. Liquid chromatography coupled to low-resolution MS is rarely used in the literature for the analysis of technical lignin samples, probably because no databases like for GC/MS are available and also no exact mass can be determined, which is a valuable information for the identification of an unknown compound. However, Kiyota *et al* used UHPLC coupled to a QQQ-MS to study the MS fragmentation behaviour of a synthesised mixture including LMs and LOs and used the identified MS^2 fragmentation patterns for the identification of LMs and LOs in a technical lignin sample[13]. Yan *et al* also used a UHPLC/QQQ-MS system for the analysis of LMs. However, Yan *et al* used the high sensitivity of the QQQ mass analyser for the quantification of LMs and reported limits of detection (LODs) in the range of femtomoles[50].

High-resolution MS is widely used for the analysis of complex technical lignin samples due to its advantage of determining exact masses. For the analysis of LMs or LOs HRMS is used either with DI[21,51] with no prior sample separation or with different separation techniques such as LC[12,20] or GC[41]. The majority of reported HRMS applications for the analysis of technical lignins is mainly

performed using a FT-ICR-MS[15,39] or a TOF-MS[21,51], but Orbitrap MS[42,52] systems are also used.

3.2 Electrospray ionisation

Electrospray ionisation is nowadays probably the most applied API technique for the analysis of non-volatile and chargeable compounds[24,53]. The main reasons for the success of ESI is the good compatibility of ESI with LC and the large molecular weight ionisation range, ranging from low molecular weight molecules such as vanillin, with a molecular mass of 152 u, to high molecular weight molecules such as intact viruses with molecular weights exceeding 10^6 u[24,54]. Furthermore, as a soft ionisation technique, molecular ions can be obtained, which is very useful for the identification of unknowns in complex sample mixtures.

Ionisation in ESI is governed by either protonation of the analyte in positive ionisation mode or by deprotonation in negative ionisation mode. A strong electric field is applied under atmospheric pressure between a capillary and a counter electrode at the MS entrance[30]. Behind the MS entrance a vacuum is applied. The analyte solution is pumped through the capillary and at the tip of the capillary the electric field induces the formation of a mist of electrically charged droplets, which are attracted to the MS entrance by the electric field[53]. Desolvation processes, often supported by a heated gas, lead to the formation of smaller droplets. At a certain droplet size the Rayleigh limit is reached and electrostatic repulsion overcome the surface tension of the droplet and even smaller droplets are produced[24,30]. Two models have been proposed for the formation of ions from charged droplets, the charged-residue model (CRM) and the ion evaporation model (IEM)[24]. In the CRM, complete desolvated analyte-ions are formed by a successive loss of solvent molecules until only a charged analyte molecule remains. In the IEM, the formation of desolvated analyte-ions is described as a result of a direct evaporation of analyte-ions from the surface of highly charged microdroplets.

3.2.1 Electrospray ionisation of LMs and LOs

Electrospray ionisation in negative ionisation mode is applied in many MS studies for the analysis of technical lignins. The presence of phenolic groups, carboxylic acids and hydroxyl groups in LMs and LOs make ESI in negative ionisation mode a suitable ionisation technique for the ionisation of many compounds present in complex lignin samples[47]. For example, Haupert *et al* showed for LM and LD model compounds that ESI in negative ionisation mode

doped with NaOH provides molecular ions with little fragmentation, which is especially important for the identification of unknown compounds[55]. Owen *et al* combined the developed ESI method by Haupert *et al* with HPLC for the analysis of LM and LD model compounds and for the identification of unknown compounds in an organosolv lignin sample[12]. Recent studies by Dier *et al*[48] and Qi *et al*[47] show that also ESI in negative ionisation mode without a basic dopant, at least for DI-MS, is a suitable ionisation technique for the ionisation of LMs and LOs.

3.2.3 Optimisation of the ESI for LMs and LOs

In an ESI source, several ionisation parameters, such as the capillary voltage, cone voltage, source temperature and sheath gas flow rate, can be varied to improve the ionisation of targeted analytes. Due to different ionisation efficiencies in ESI of different compounds classes an optimisation of the ESI parameters is a crucial part of MS method development. While for one analyte the optimisation of the ionisation efficiency can be easily done, an optimisation of parameters for a simultaneous analysis of multiple analytes can be very challenging, especially if the chemical structures are very different. In this kind of cases multivariate optimisation strategies using design of experiments (DoE) are often used. DoE are systematic optimisation tools and are described in more detail in chapter 4, section 4.1.

Several studies conducted during the recent years have applied DoE approaches for the optimisation of ESI parameters[56–60]. For example, Reymond *et al* reported an optimisation approach using a DoE for the optimisation of ESI parameters for the analysis of a complex pyrolyzed lignocellulose biomass sample using LC/ESI–HRMS[59]. In their study, the effect of the nebulising gas flow rate, drying gas flow rate, sweep gas flow rate, transfer capillary temperature and capillary voltage for twelve selected model compounds were investigated. With the help of a desirability function, which combined all twelve models, a method with the best compromise in signal intensity for all model compounds was found[59]. A similar study by Grand-Guillaume Perrenoud *et al* for the ESI behaviour of pharmaceutical compounds showed that the MS peak intensity using UHPSFC/ESI–MS can be improved four to ten times compared to UHPLC/ESI–MS[60]. Grand-Guillaume Perrenoud *et al* conclude that the improved ESI sensitivity result from more efficient analyte desolvation processes in UHPSFC/ESI–MS due to the decompressed CO_2 in the SFC effluent entering the ESI source[60]. Additionally, if SFC is coupled to a MS, a post column solvent, called makeup solvent, which is commonly used in SFC, offers opportunities to improve the MS ionization efficiency of the analyte with less negative dilution effects compared to LC/ESI–MS[60,61]. In LC/ESI–MS a post column solvent is rarely used because ESI is a concentration dependent ionisation technique.

In **Paper I** the ESI efficiency in negative ionisation mode was optimised for in total 39 LMs using UHPSFC/ESI–QTOF-MS. For the investigation a reference standard mixture including all 39 LMs was used. The influences of nine parameters, including the type of makeup solvent, type of makeup solvent additive, concentration of the makeup solvent additive, makeup solvent flow rate, ion source temperature, desolvation gas temperature, desolvation gas flow, capillary voltage and cone voltage, were investigated using a DoE approach. As response the number of detected peaks with a base peak intensity of equal or higher than 10^5 was used. In this way only one DoE for all analytes needed to be performed, instead of one DoE for each investigated analyte. Furthermore, the response was used to ensure that a good MS^2 spectrum of each analyte can be obtained, which is important for compound identifications. A coefficient plot showing the influence of each investigated parameter is shown in Figure 2. From the investigated parameters the type of makeup solvent, the type of makeup solvent additive, the desolvation gas temperature and the cone voltage showed a significant influence on the ESI efficiency for the investigated LMs. Furthermore, five different two-factor interactions were determined. The concentration of makeup solvent additive did not show a significant influence, but it interacts with the desolvation temperature and cone voltage and might therefore have an important role in the ionisation process. Two different makeup solvents, isopropanol and methanol were investigated. While for isopropanol a significant negative influence was determined, methanol showed a significant positive influence. This shows that the makeup solvent in UHPSFC/ESI–MS should be considered during method developments. The influences of the makeup solvent may be related to its ability to solubilize the analytes, but also due to physicochemical properties of the solvent such as surface tension, volatility or solvation properties for ions[62]. Furthermore, it was shown that for the ionisation of LMs in ESI negative mode ammonia is a suitable additive and has a positive influence on the ionisation efficiency.

In **Paper II** the ESI efficiency of 24 selected LMs was studied separately using reference standards and a UHPSFC/ESI–QQQ-MS system. The analytes were selected to cover a wide range of functional groups. The LMs were categorized in eight groups based on their functional group on position 1 at the benzene ring and into three groups based on their number of methoxy groups (0, 1 or 2). An overview of the selected analytes and their group membership is illustrated in Figure 3.

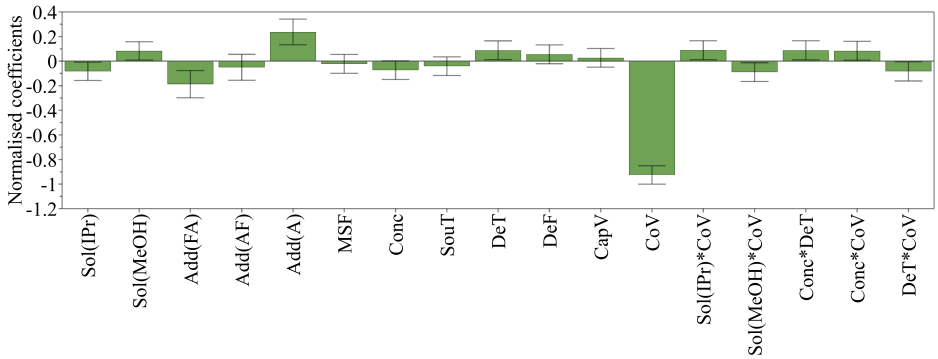


Figure 2. Normalized coefficient plot showing the influence of nine ESI parameters and their interactions on the ionisation of 39 LMs using a UHPSFC/ESI–QTOF-MS system. Sol: solvent, IPr: isopropanol, MeOH: methanol, Add: additive, FA: formic acid, AF: ammonium formate, A: ammonia, MSF: makeup solvent flow rate, Conc: concentration of makeup solvent additive, SouT: ion source temperature, DeT: ion source desolvation gas temperature, DeF: ion source desolvation gas flow rate, CapV: ion source capillary voltage, CoV: ion source cone voltage. Figure from **Paper I**, reprinted with permission.

A DoE approach was applied to investigate the influence of seven ESI parameters on every studied LM. The studied ESI parameters were the capillary voltage, gas temperature, gas flow rate, sheath gas temperature and sheath gas flow rate. In addition to this, the feed speed and the overfeed volume were studied, which are injection parameters but might influence ESI sensitivity. In this study, makeup solvent parameters were not included, because **Paper I** showed that methanol with a low concentration of ammonia is a good makeup solvent for the analysis of LMs. Furthermore, in **Paper I**, the makeup solvent flowrate showed no significant influence on the ionisation of LMs and was therefore kept constant. The numbers of positive or negative significant influences for each investigated ESI parameter for compounds within the acid and alcohol group and for compounds with two, one or no methoxy group are illustrated in Figure 4. Different trends within the different groups of analytes can be seen. For example, for the carboxylic acids the sheath gas temperature (significant for four out of six compounds) and the sheath gas flow rate (significant for five out of six compounds) appear to have a significant positive influence on the ionisation efficiency. For the alcohols, however, for only two out of six compounds a significant positive influence and for one a significant negative influence of the sheath gas temperature was found and only for one compound a significant positive effect of the sheath gas flow rate was found. Some trends can also be identified by comparing the three groups with different numbers of methoxy groups.

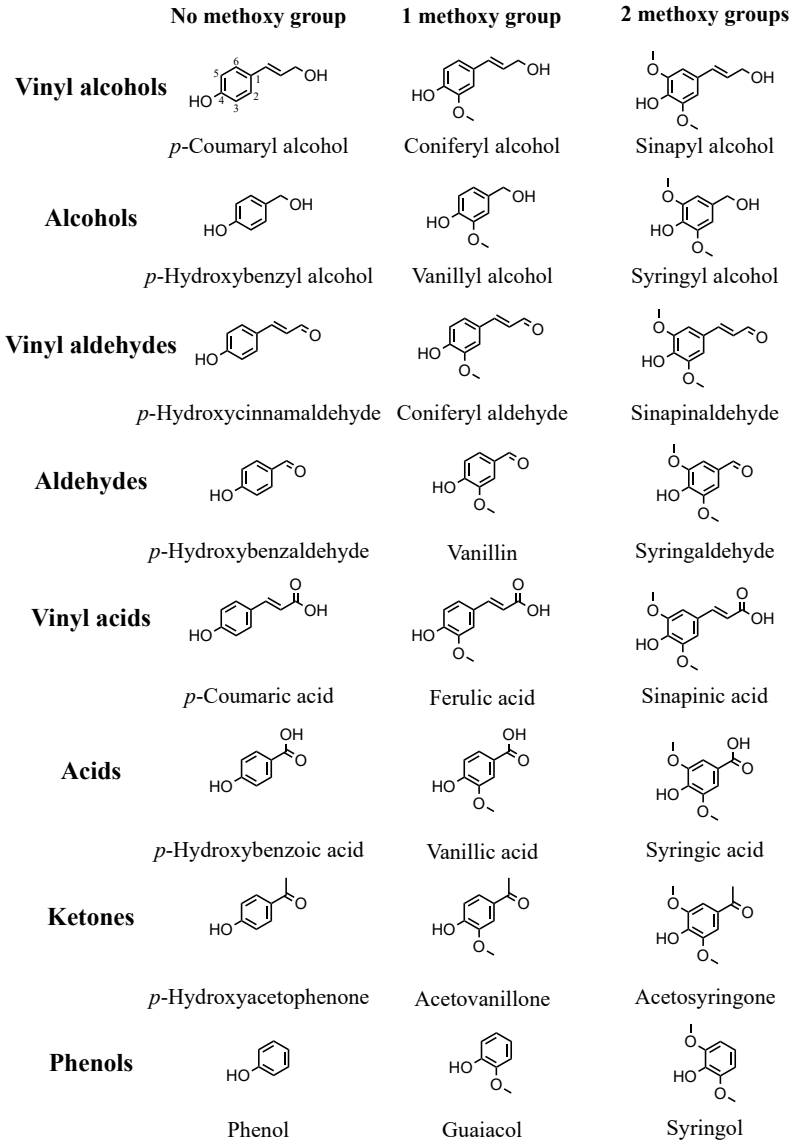


Figure 3. Selected analytes investigated in **Paper II** and their group membership. Figure taken from **Paper II**.

This is exemplified by the group with two methoxy groups, where a significant positive influence of the gas temperature was found for all eight compounds. In contrast to that for the group with one methoxy group, a significant influence of the gas temperature was found for three compounds, two positive and one negative. In addition to that, for the compounds with no methoxy groups the gas temperature shows for two compounds a significant positive influence and for

three a significant negative influence. Both trends may be explained by a relatively better solubility of compounds with two methoxy groups and with a carboxylic acid functional group in methanol, which was used as co-solvent and makeup solvent. As described by Kebarle and Verkerk a better solubility of a compound in the used solvent leads to a lower surface activity of the compound on the surface of the ESI droplets, meaning higher solvation energy needs to be exceeded by desolvation energy[53]. The better solubility in methanol may be explained by different hydrogen bonding acceptor and donator properties of the compared compounds. Carboxylic acids can potentially create more hydrogen bondings with methanol compared to alcohols and therefore need potentially more support by the sheath gas for the desolvation process. Due to potentially more hydrogen bonds between methanol and the oxygen atoms of the methoxy groups, a similar process may explain the significant positive influence of the gas temperature for the compound group with two methoxy groups.

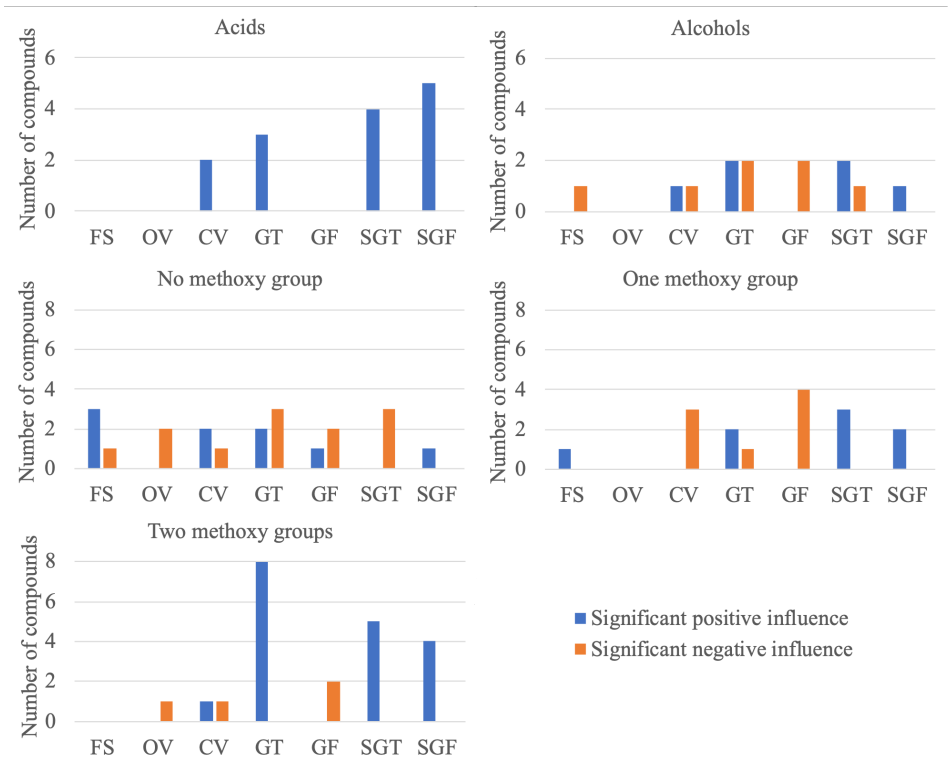


Figure 4. Number of analytes for the acid group, the alcohol group and for compounds having 2, 1 or no methoxy group showing significant positive or significant negative influence on the ESI efficiency of the feed speed (FE), the overfeed volume (OV), the capillary voltage (CV), gas temperature (GT), gas flow rate (GF), the sheath gas temperature (SGT) and the sheath gas flow rate (SGF). Compounds belonging to each group can be seen in Figure 3. Figure from **Paper II**.

In **Paper I** and **II**, the ESI efficiency was optimised for LM model compounds. However, since for LOs almost no commercially produced reference standards are available, the ESI efficiency of LOs was studied in **Paper III** for LDs, LTRs and lignin tetramers (LTEs) identified in a Kraft lignin sample using LC/ESI–MS. In **Paper III** a DoE approach was used too. The influence of the capillary voltage, sheath gas flow rate and auxiliary gas flow rate on the peak intensity of 36 identified LOs was studied. Obtained coefficient plots of all compounds were compared and the best setting with a compromise for all investigated analytes was chosen. Significant positive influences for most of the analytes were identified for the capillary voltage and the ion source sheath gas flow rate, while for the ion source auxiliary gas flow rate a significant negative effect was found. Furthermore, significant negative interactions between the ion source capillary voltage and the ion source auxiliary gas flow rate, and between the ion source sheath gas flow rate and the ion source auxiliary gas flow rate were identified. A second DoE was applied to investigate different ranges of the ion source sheath gas flow rate and the ion source auxiliary gas flow rate. With the optimised method the peak intensity was significantly improved for 25 out of 36 LOs.

3.3 Identification of LMs and LOs by MS^n fragmentation pathways

Multiple stage tandem mass spectrometry is a powerful tool for the identification of unknown compounds[63]. Especially in complex matrices, such as technical lignin samples, valuable information about the chemical structure of unknown compounds can be obtained by specific neutral losses. For example, if ESI operates in negative ionisation mode, the loss of 44 u is well known to be correlated to the loss of CO_2 from a carboxylic acid group.

3.3.1 Identification of LMs using MS^n fragmentation pathways

In a comprehensive study, Marcum *et al* studied the fragmentation patterns of LMs triggered by collision induced dissociation (CID) in MS^n experiments up to MS^6 using a LQIT-MS with ESI in negative ionisation mode[63]. In **Paper IV**, MS^n using a LQIT–Orbitrap-MS was used up to MS^3 for the identification of LMs and LOs in three different technical lignin samples. For example, in a Lignoboost Kraft lignin sample, vanillin was identified, which showed in the MS^2 a neutral loss of a methyl radical and in the MS^3 stage two fragment ions were identified, one with a loss of CO and one with a loss of CO_2. The same MS^3 fragmentation pattern was reported for vanillin by Marcum *et al*[63]. Especially in combination with other information, such as the chemical formula obtained by exact mass

measurements and ring double bond equivalents (RDB), MS^n fragmentation patterns can be used to propose chemical structures of unknown compounds. In **Paper IV**, MS^3 fragmentation was also obtained for some unknown compounds with the characteristics of LMs. For instance, for the m/z 231.0661, a neutral loss of a methyl radical was obtained in the MS^2 stage, followed by a neutral loss of CHO and a neutral loss of CO in the MS^3 stage. A proposed tentative structure and MS^3 fragmentation pathway is illustrated in Figure 5.

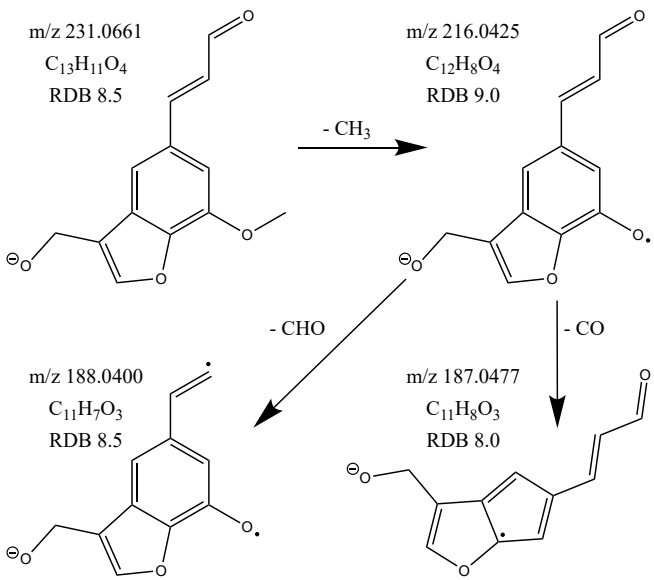


Figure 5. Proposed structure and MS^3 fragmentation pathway of the LMs with m/z 231.0661. Figure from **Paper IV**, reprinted with permission.

3.3.2 Identification of LOs using MS^n fragmentation pathways

The identification of LOs is a very challenging task due to a lack of commercially available reference standards. In some reported studies in literature oligomeric model compounds were synthesised[13,15]. However, the synthesis of oligomeric compounds is time-consuming and often costly. Furthermore, since many different types of technical lignins exist, the number and diversity of LOs is enormous, especially if chemically or biologically converted technical lignins are taken into account. Therefore, a tool like MS^n is indispensable for the identification and structural elucidation of LOs.

Nevertheless, reported MS^n fragmentation pathway studies using synthesised LO standards, like by Morreel *et al*[15] or Kiyota *et al*[13], made an important contribution to the understanding of MS^n fragmentation patterns of LOs. Morreel *et al* first studied the fragmentation pathways of four different lignin linkages using

synthesised LDs[16]. For example, they found out that the first major neural losses for the 8-5 linkage are a water and a formaldehyde molecule, followed by a neutral loss including a benzene ring[16]. The proposed fragmentation pathways were then used for the identification of unknown LOs extracted from wild-type poplar[15].

In **Paper III**, MS^n fragmentation pathways were used for the structural elucidation of identified LOs in a Kraft lignin sample. MS^n studies are often performed with high MS resolution in the MS^1 stage, while for higher MS^n low MS resolution is used. In **Paper III**, using an LQIT–Orbitrap system, a reduced MS resolution of 30.000 was used at each MS stage, which still produces high-resolution MS data, without a high duty cycle of the MS. An example of a proposed tentative structure and fragmentation pathway of a LTR is illustrated in Figure 6. With the high-resolution at every MS stage molecular formulas based on the detected exact masses and RDB equivalents were obtained, which are useful information for the structural elucidation of unknown compounds. An important identification parameter for LOs is the neutral loss of a benzene ring unit. Like in the given example in Figure 6, the neutral loss of a benzene ring unit is often not detected in the MS^2, but in the MS^3 or MS^4 stage. Another benefit of using high mass resolution in every MS stage is the possibility to differentiate neutral losses with very similar masses. For instance, the neutral loss of 30 u is often reported as a loss of formaldehyde (CH_2O). However, in **Paper III** it was shown that tentative LOs can lose both, a formaldehyde molecule and two methyl radicals. The losses of 30.0106 u for CH_2O and 30.0470 u for C_2H_6 are very similar, only differing by 0.0364 u. It would have not been possible to obtain this information with a low resolution at MS stage 2 or higher and it might have led to false conclusions.

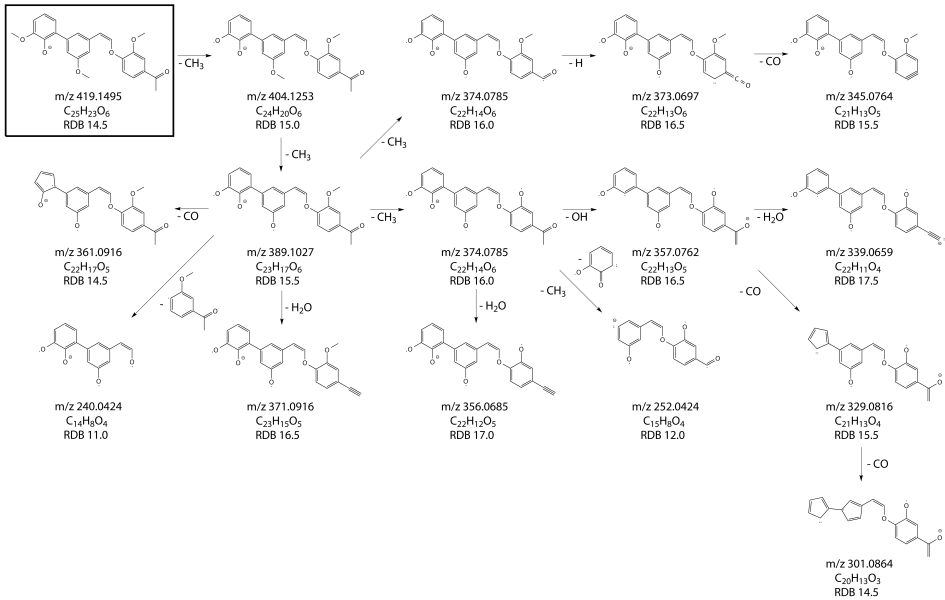


Figure 6. Proposed tentative structure and MS^7 fragmentation pathway of an identified lignin trimer with m/z 419.1495 ($[M-H]^-$). Figure from **Paper III**, reprinted with permission.

In this chapter, the ESI of LMs and LOs was discussed and it was shown that DoE are useful chemometric tools for the optimisation of the ESI efficiency. Therefore, DoE will be discussed in more detail in the next chapter. It was also shown, that MS^n fragmentation pathways can be used for the identification of unknown LMs and LOs in technical lignin samples. However, the identification of LMs and LOs based on the analysis of MS^n fragmentation patterns is time consuming, since the MS^n fragmentation patterns for every detected compound in the analysed sample need to be evaluated. With a pre-selection data mining tool for tentative LMs and LOs only MS^n fragmentation patterns of promising compounds would need to be evaluated. A developed pre-selection data mining tool using classification models will be presented in the following chapters.

4. Chemometric tools

Chemometrics is a chemical discipline that was established by Svante Wold and Bruce Kowalski in the early 1970s and uses statistical and mathematical methods to either design or optimize measurement procedures or experiments, or to extract maximum chemical information from chemical data[64]. Multivariate approaches are used in chemometrics for the design of experiments or the analysis of chemical data. Multivariate approaches allow the simultaneous investigation of effects of multiple parameters and furthermore the investigation of possible parameter interactions. Chemometrics has gained importance during the years due to the rise of complex analysis techniques such as ESI-MS, where several parameters can be changed during a method optimisation and very complex data can be produced. Chemometric tools, such as design of experiments (DoE), principal component analysis (PCA) and classification models are nowadays well integrated tools within analytical chemistry and used for many different applications[65].

4.1 Design of experiments

A good experimental design is crucial for the success of any kind of study and should be well planned before the first experiment is performed. DoE are logically formulated procedures for the optimisation of any kind of process with several process parameters. DoE can also be used to optimise a certain process in a time and cost-saving way[65]. In DoE a multivariate approach is used, in which several experimental parameters are changed at the same time, which also assures a multivariate interpretation of the results.

Classic DoE are based on symmetric geometrical forms, like for instance a cube. Examples of two classical designs, a two-level full factorial design and a three-level face-centred central composite design, are shown in Figure 7. A full factorial design was used in **Paper III** to study the influence of three quantitative ESI parameters on ESI efficiency of identified LOs in a Kraft lignin sample. As it can be seen in Figure 7, the experimental points are distributed around a reference point in the geometric centre of the design space, which is called the centre point. The centre point is usually replicated to determine the reproducibility of the measurements.

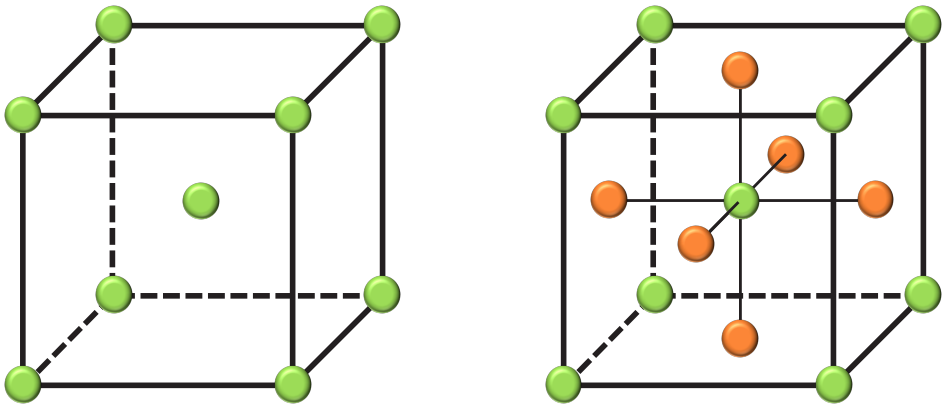


Figure 7. Geometric forms of a full factorial design (left) and a face-centred central composite design (right).

DoE are evaluated mathematically by linear or polynomial models[66]. For example, a full factorial design with three investigated parameters can be described by the following linear model,

$$y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_{12}X_1X_2 + b_{13}X_1X_3 + b_{23}X_2X_3 + b_{123}X_1X_2X_3 + \varepsilon$$

with y being the response, X_i the model parameters, b_0 the constant term, b_i the coefficients of the model parameters and ε the error term. In DoE parameters are categorized into two types of parameters, quantitative parameters and qualitative parameters. Quantitative parameters need to have a quantifiable nature, like for example the temperature of a ESI source, while qualitative parameters do not have a quantifiable nature, like for example different types of makeup solvent additives. Looking at the model above, an advantage of the multivariate approach can be seen, which is that possible interaction between the investigated parameters are also considered. If the parameters would be tested separately by keeping one parameter constant, while the other gets changed, possible interactions between them would remain undiscovered. An interaction design was e.g. used in **Paper I** and several two-factor interactions of ESI parameters were identified. For three-level designs, such as the face-centred central composite design, a quadratic term of each parameter would need to be included into the model.

If many parameters need to be investigated, the number of experiments can increase rapidly. The number of experiments for a full factorial design is equal to 2^k, with k being the number of parameters[66]. For example, if 6 parameters are investigated, 64 experiments plus the centre points need to be performed. If two-parameter interactions are not of main interest, the number of experiments can be reduced using fractional factorial designs[66]. In fractional factorial designs, only a fraction of the design space is investigated, which reduces the number of needed

experiments. However, this goes in hand with an increased confoundedness, meaning that effects of parameters or two-parameter interactions cannot be distinguished from other parameters or two-parameter interactions. Fractional factorial designs are categorized in different resolution levels. If a fractional factorial design has a resolution of III or lower, parameters are confounded with two-parameter interactions and two-parameter interactions are confounded with each other. Therefore, a minimum resolution of IV is recommended when using fractional factorial designs, because then the parameters are not confounded with the two-parameter interactions, but the two-parameter interactions are still confounded with each other[66]. In **Paper II** a fractional factorial design (resolution IV) was used for the investigation of seven quantitative parameters for the ESI efficiency of LMs using UHPSFC/ESI–MS. With the used design the number of experiments was reduced from 128 for a full factorial design to 16 (without centre points).

Besides the classical and symmetric forms for DoE, geometrically non-symmetric DoE exist too, which can be used for specific problems where a classic DoE cannot be applied. A common model design for such cases is the determinant-optimal (D-optimal) design. D-optimal designs are often used when qualitative factors are to be included in the DoE or if parameter boundaries of the studied system need to be considered. In **Paper I** a D-optimal design was applied because with the type of makeup solvent and type of makeup solvent additive, two qualitative parameters were included in the DoE.

The main evaluation tools of performed DoE are coefficient plots and response contour plots. In a coefficient plot, as shown in Figure 2 in chapter 3, section 3.2.3, the effects of the coefficients on the model of each parameter is shown. For the estimation of the coefficients, the multiple linear regression or the partial least squares methods can be used. The included confidence intervals for each coefficient in the coefficient plots describe their uncertainty and their size depends on the noise of the model[66]. A coefficient is considered to make a significant contribution to the model if the confidence interval is not crossing zero. Response contour plots, like the one shown in Figure 8, show the behaviour of the response depending on two parameters. Response contour plots can visualise interactions between parameters and can help to find the parameter settings for the best response.

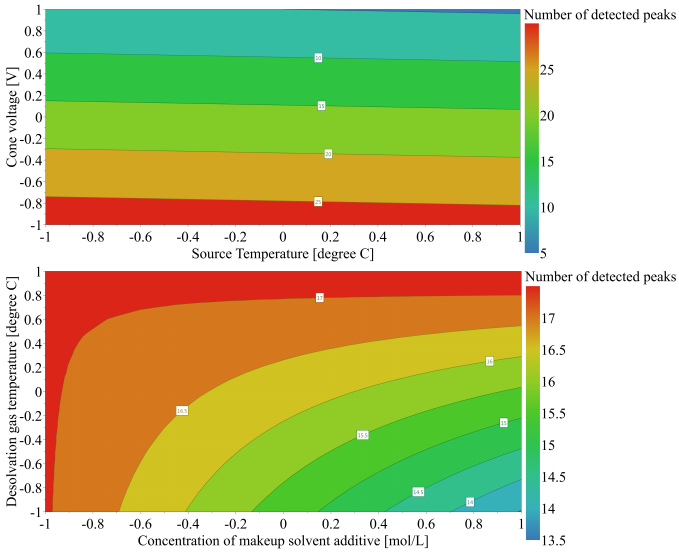


Figure 8. Examples of contour plots showing the effect on the response of the source temperature and cone voltage showing no significant two-factor interaction (top) and the effect on the response on the concentration of makeup solvent additive and desolvation gas temperature showing a significant two-factor interaction (bottom). The contour plots are based on the D-optimal design applied in **Paper I** and correspond to the coefficient plot shown in Figure 2.

4.2 Principal component analysis

Principal component analysis is a multivariate projection method used for the extraction and visualisation of systematic variation in a dataset[67]. Applying PCA to a dataset can help to identify for example relations between different samples or parameters, also new patterns can be found, which might have otherwise remained undiscovered when parameters are examined in isolation, as in a univariate analysis. PCA searches for linear relationships within the explored original parameters and creates a new parameter if linear relationships are detected[68]. Such new parameters are called principal components (PCs). Since a PC is a combination of all original parameters, a PC can explain more variation in the data compared to one original parameter[68]. Identified PCs are ordered according to their explained variation of the dataset, PC 1 explains the most variation in the data, PC 2 the second most, and so on. With the use of two principal components a model plane can be defined and used for visualisation of the data in a score plot[67]. Figure 9 shows a score plot from a PCA-based classification model described in **Paper III**, which is based on LOs described in the literature. The used parameters are the number of carbon atoms (#C), the number of hydrogen atoms (#H), the number oxygen atoms (#O) and the RDB

equivalent. As it can be seen in Figure 9, PC 1 describes more than 90% of the variation in the data, while PC 2 describes around 9%. The score plot shows that with a multivariate approach the difference of LOs based on their #C, #H, #O and RDB equivalents can be visualised and clusters of LDs, LTRs and LTEs can be identified. Furthermore, the potential positions of clusters for lignin pentamers, hexamer and heptamers can be identified. To understand the observed patterns in the score plot and which influence each parameter has on the data separation, a loading plot, as shown in Figure 10, can be used. A loading plot is created by the loadings of the PCs and the loading vectors are called latent variables (LVs). In a loading plot the parameters with a similar contribution to the information of the dataset are grouped together[67]. In Figure 10 it can be seen that for example the #C, #H and #O are relatively close to each other compared to the RDB equivalent, and therefore contribute to a similar extent to the information of the dataset. Moreover, the distance of the parameter loadings towards the point of origin of the plot represents the effect of the parameter on the PCA model, meaning the larger the distance the stronger the effect on the model[67]. For instance, the RDB equivalent and the #C show stronger effects on the model compared to #H and #O. The position of a score in the score plot is directly correlated to the position of a parameter in the loading plot. For example, the position of the LD labelled as 'D12' in Figure 9 is more influenced by the RDB equivalent as the three other parameters, while the position of the LTR labelled as 'TR34' depends more on the #O than on the RDB equivalent.

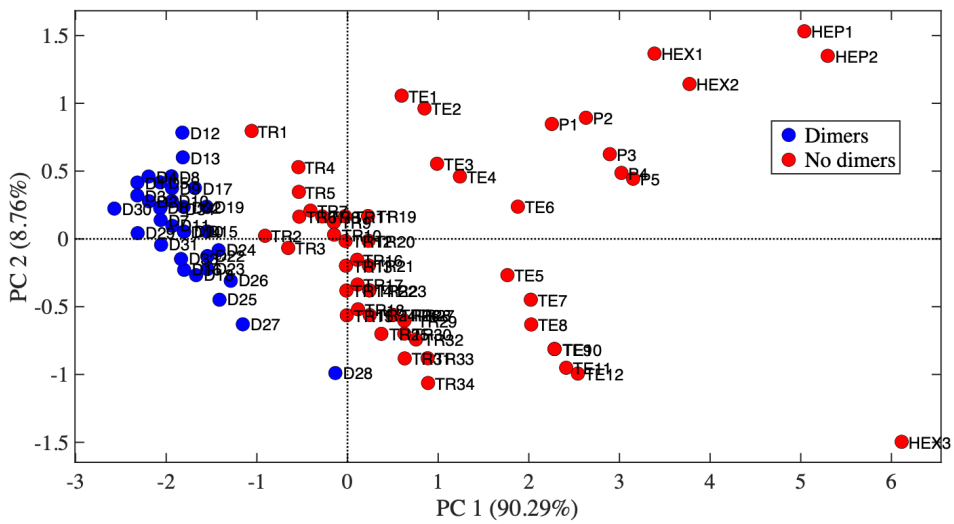


Figure 9. PCA score plot. PC: principal component, D: dimers, TR: trimers, TE: tetramers, P: pentamers, HEX: hexamers, HEP: heptamers. Figure from **Paper III**, reprinted with permission.

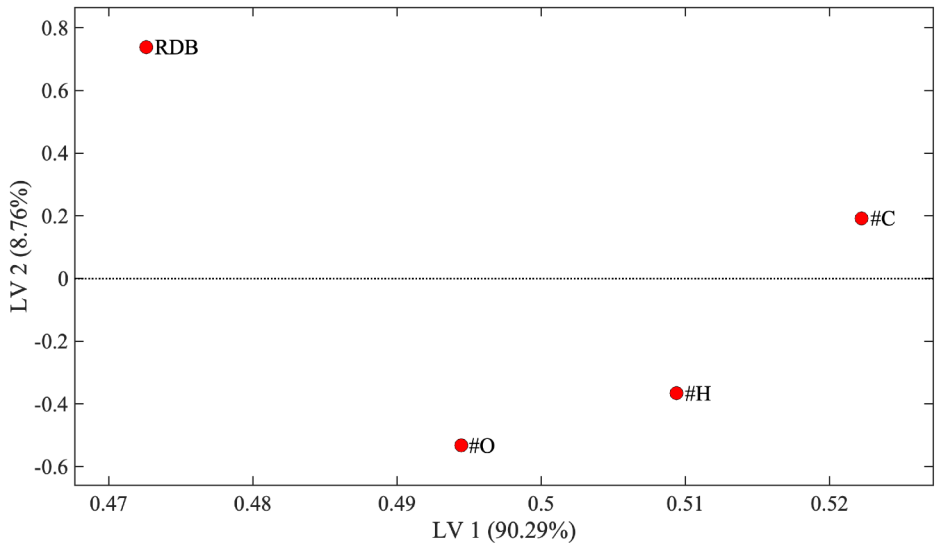


Figure 10. PCA loading plot. LV: latent variable, #C: number of carbon atoms, #H: number of hydrogen atoms, #O: number of oxygen atoms, RDB: ring double bond equivalent. Figure from **Paper III**, reprinted with permission.

4.3 Multivariate classification models

Multivariate classification models are used to recognize the membership of a sample or a compound to a specific class, like LMs or LDs. For the development of a classification model, data of samples or compounds from a class of interest are needed, for example the RDB, #C, #H and #O of LMs or LDs. The PCA score plot illustrated in Figure 9 is the basis for a classification model and it can be seen, that the shown score plot includes the information about the membership of the projected compounds, either dimer or no dimer. Classification models are often used in combination with PCA, like for example linear discriminant analysis (LDA), quadratic discriminant analysis (QDA) and soft independent modelling of class analogy (SIMCA)[67,69]. The k-nearest neighbour (kNN) method is another model that is often used because of its simplicity[69]. The kNN classification method is based directly on the class memberships of the sample towards the nearest samples in for example a PCA model. The number k defines the number of the nearest neighbours used for classification purposes. If for example k = 3, then a sample is recognized as part of a class if at least two out of three of the nearest neighbours are from the same class. In LDA or QDA, linear or quadratic functions are used to distinguish different classes within the dataset. In SIMCA a PCA model is created for each defined class. The class based PCA

models are used for the classification of a sample and the classification is based on the tolerance intervals of the class-based PCA models[67].

4.4 Multivariate classification models for LMs and LOs

In the literature, different approaches for the classification of detected m/z values in complex lignin samples are reported[20,48]. One pre-selection strategy for LMs, LDs and lignin-carbohydrate complexes based on the definition of C:O ratio, RDB equivalent ranges and molecular weight ranges are described in a study by Jarrell *et al*[20]. Other reported approaches use Van Krevelen plots, which plot the O/C ratio against the H/C ratio, and regions for specific compound types can be assigned[70]. Furthermore, mass defect filtering techniques based on Kendrick mass defects (KMD) are used. KMDs are based on the rescaling of the mass scale by using the molecular structure of CH_2 as the base of the mass scale instead of carbon[71]. As reported by Dier *et al,* KMD plots allow the identification of homologous series of detected LMs and LOs, for example compounds with the same RDB equivalent, but with an increasing number of oxygen-atoms[48]. Any other molecular structure, like phenol with C_6H_4O or methoxy OCH_2 as shown by Crawford *et al* can be used as the base for KMDs[52]. Furthermore, Crawford *et al* showed, that if the KMD defect of for instance OCH_2 is plotted against the RDB equivalent, more homologous series based on the present monomeric subunit or the type of linkage can be identified in lignin MS data and also regions within the plot for LMs, LDs and LTRs[52].

In **Paper III**, the first multivariate classification models for LDs and LTR were created based on LOs previously reported in literature, which were identified in different types of technical lignin samples and detected with different types of ionisation sources coupled to HRMS. The first multivariate classification models were based on a PCA model using the reported masses and four different parameters, which were the #C, #H, #O and the RDB equivalents. The score plot and the loading plot of the created PCA model are shown in Figure 9 and 10, respectively and were already discussed in section 5.2. Based on the PCA model in Figure 9, QDA classification models can be created for LDs and LTRs and can be used for the classification of detected LOs in a Kraft lignin sample.

The established PCA-QDA classification models were further improved in **Paper IV**. New parameters based on the KMD of six different functional groups typical for LMs and LOs were included. The selected functional groups were aldehydes, carboxylic acids, phenols, secondary alcohols, primary alcohols and methoxy groups. However, the two latter, primary alcohols and methoxy groups, have the same molecular formula and were therefore counted as one parameter. The RDB

equivalent parameter was removed, since some technical lignin samples also include sulfur in their chemical structure, due to the used lignin isolation process. Furthermore, LMs and recently in literature identified LOs were added to the PCA dataset.

The score and loading plots of the KMD-based PCA-QDA classification model for LTRs are exemplarily shown in Figure 11. Compared to the score plot from the first PCA model in Figure 9, the improved PCA model includes a cluster of LMs projected, as expected, next to the LDs and also the clusters for LDs, LTRs and LTEs are more defined due to more available data. The data separation on PC 1 and PC 3 is mainly influenced by the #C, #H, #O, the KMDs with a methoxy/primary alcohol base on the positive side of PC 1 and by the KMDs with a carboxylic acids and aldehydes base on the negative side of PC 1, while on PC 3 mainly by #C and #O. The KMDs with a secondary alcohol and a phenol base seem to have only minor influences on the PC 1 and PC 3. Based on the KMD-based PCA model QDA classification models were created for LMs, LDs, LTRs and LTEs. Since the KMD-based PCA-QDA classification models are based on literature data from different types of technical lignin samples and different used MS ionisation sources, the classification models can be applied to any type of technical lignin samples and HRMS data. Furthermore, multivariate and data specific class ranges are implemented with the multivariate approach. This means that no monovariate classification limits are needed, such as molecular weight ranges. Additionally, no reference standards or the study of MS^n fragmentation pattern are needed for the classification of LMs and LOs using the developed classification models.

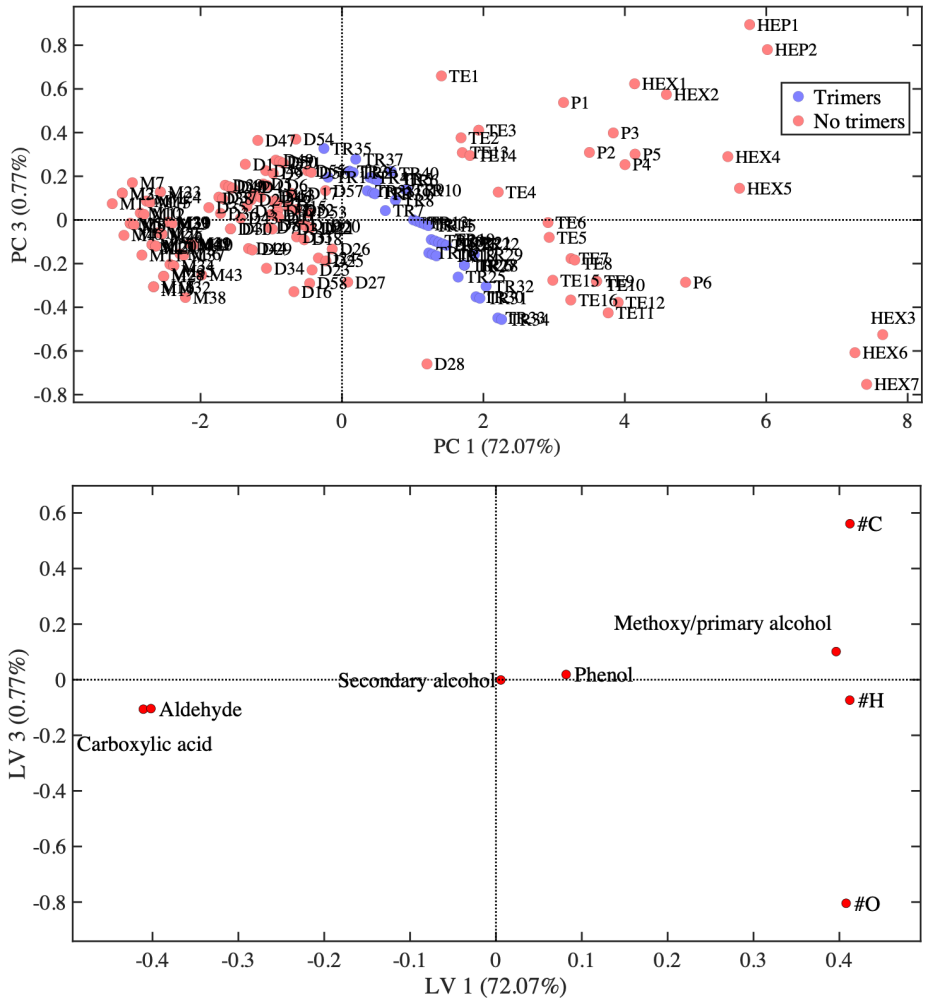


Figure 11. PCA score plot (top) and loading plot (bottom) from a KMD-based PCA-QDA classification model for lignin trimers. PC: principal component, LV: latent variable, #C: number of carbon atoms, #H: number of hydrogen atoms, #O: number of oxygen atoms. Figure from **Paper IV**, reprinted with permission.

In **Paper IV** the established KMD-based PCA-QDA classification models were applied for the classification of detected masses in a Lignoboost Kraft lignin sample, a Lignosulphonate lignin sample and a depolymerised Kraft lignin sample. Figure 12 shows the projection of 371 different masses detected in the Lignoboost Kraft lignin sample. It can be seen that the majority of the detected masses are projected around the LM, LD and LTR clusters.

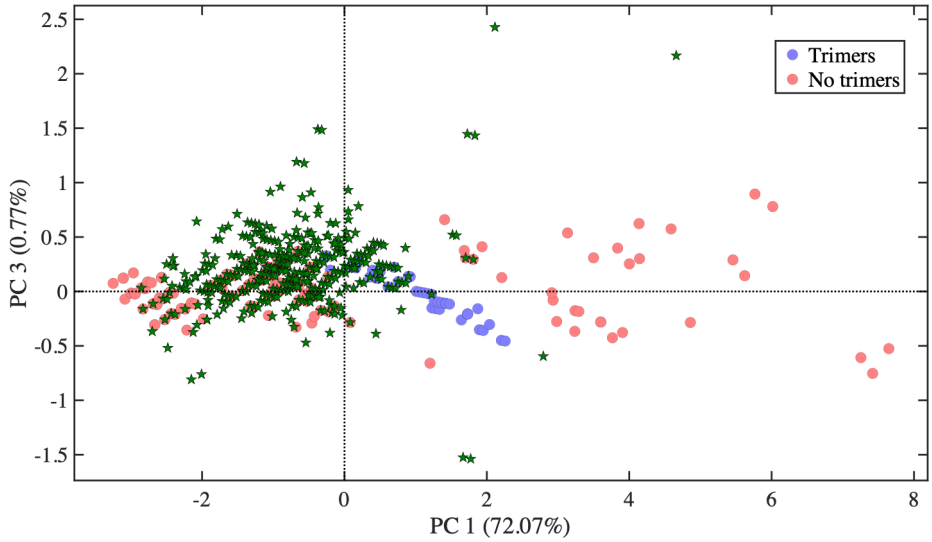


Figure 12. PCA score plot from a KMD-based PCA-QDA classification model for lignin trimers based on literature data (blue and red points) including projected positions from 371 different masses (green stars), which were identified in a Lignoboost Kraft lignin sample. PC: principal component. Figure from **Paper IV**, reprinted with permission.

In this chapter, DoE and the development of KMD-based PCA-QDA classification models for LMs, LDs, LTRs and LTEs were discussed. The developed classification models were embedded in an identification strategy for the identification of unknown LMs and LOs in technical lignin samples, which will be presented in the next chapter.

5. Identification strategies for technical lignins using MS

Generally, three conceptually different analysis strategies are applied for the analysis of all kinds of samples using MS: targeted analysis, suspect analysis and non-targeted analysis. The main differences between targeted, suspect and non-targeted analysis are the available information of the analytes and the availability of reference standards[72]. In targeted analysis, detailed information of the analytes, such as chemical formulas or chemical structures, are known and reference standards are available, which are usually used for method development. The only difference between targeted analysis and suspect analysis is that for suspect analysis no reference standards are available, but detailed information about the analytes are known. In non-targeted analysis, however, no information about the analytes are present before the analysis and therefore naturally no reference standards are available.

MS, often in combination with a separation technique such as LC, GC or SFC, has been proven to be a very useful tool for the identification of unknown compounds in complex matrices[25,72]. However, the identification of unknown compounds in complex matrices without any prior information about possible analytes is very challenging. As previously demonstrated by Schymanski *et al*, the identification of unknowns using non-targeted analysis approaches is still very time-consuming and also still needs well-experienced researchers, who can focus mainly only on the identification of compounds from compound classes within their expertise[25]. Therefore, non-targeted analysis strategies are often tailored for certain compound classes[73–75]. A characteristic of such compound classes searched for in tailored non-targeted analysis strategies is often a large number of different compounds and a large variety of chemical functionalities, like industrially produced pesticides or pharmaceuticals[75]. Also compounds from compound classes already known to be present in for instance biological or natural samples, like blood plasma or technical lignins, are searched for using tailored non-targeted analysis strategies. In such complex samples, the aim is often to identify new derivatives or degradation products of certain compound classes, like in metabolomics or lipidomics. In such tailored non-targeted methods specific characteristics of the compound class are often used, like characteristic neutral losses in MS/MS fragmentation, which can be used for

neutral loss scans[73], or characteristic chemical structure functionalities, which can be used for example for Kendrick mass defect filtering[48].

Schymanski *et al* introduced identification confidence levels for non-targeted analysis to establish a general system to communicate the identification of compounds[76]. The authors suggest five different levels of identification confidence, with level 5 being the lowest and level 1 the highest identification confidence level. At level 5 only the exact mass of a compound is known, while at level 4 also a chemical formula and MS isotope patterns are known. Tentative candidates are known at level 3 and structural functionalities are known based on MS/MS data or other experimental data, such as a classification by different data mining techniques. At level 2 a probable structure is known based on different evidence such as a library spectrum match or by diagnostic evidence, such as a characteristic neutral loss. Level 1 represents the ideal case and the compound can be confirmed using a reference standard[76].

5.1 Targeted and suspect analysis of LMs and LOs using MS

Targeted analysis of lignin-related phenolic compounds using MS is applied in literature mainly for LMs[50,77], due to a lack of commercially available reference standards of LOs. Suspect analysis of LMs and LOs has only rarely been used. However, for example Huis *et al* used reported identified LMs and LOs in literature for the identification of LMs and LOs in a lignin sample extracted from flax[78].

A targeted analysis method for LMs was developed in **Paper I** and several LMs were identified in depolymerised Kraft lignin samples. In **Paper III** and **IV**, suspect lists, including LMs and LOs identified in literature, were used to establish PCA-QDA classification models and for the identification of LOs in different technical lignin samples.

5.2 Non-targeted analysis of unknown LMs and LOs using MS

In principle, non-targeted analysis has been already widely used in the past for LMs and LOs, although the term *non-targeted* was simply not used. For the identification of unknown LMs in technical lignin samples GC/MS has been mainly used, because of available comprehensive EI databases[79,80]. But also LC/MS/MS or LC/MS^n have been used for the identification of unknown LMs in

technical lignin samples[12,63,81]. However, since in LC/MS, due to non-standardised ionisation sources and settings, no comprehensive databases are available as in GC/MS and different strategies for the identification of LMs needed to be developed. As mentioned in chapter 3, section 3.3.1, Marcum *et al* studied the fragmentation pathways in MS^n experiments[63]. The reported MS^n fragmentation pathways were then used for the identification of LMs in a technical lignin sample using LC/MS^n[63]. Another interesting approach, reported by Zhu *et al*, used ion-molecule reaction between phenolic compounds and diethylmethoxyborane (DEMB), which was introduced into a LQIT[81]. Compounds with a phenolic functionality were identified in a technical lignin sample by the formation of DEMB adducts or DEMB adducts including the loss of a methanol molecule using LC/MS[81].

In the literature, the identification of LOs using MS is mainly based on the evaluation of MS/MS or MS^n fragmentation pathways. However, some studies also used synthesised LOs for identification[13,15]. If an assumed LO is not confirmed by a synthesised reference standard, the maximum identification confidence level, according to Schymanski *et al*[76], that can be achieved is level 2. To increase identification confidence reported studies in literature use different classification approaches as previously discussed in chapter 5, section 5.4.

In **Paper III** a non-targeted analysis strategy was developed for the identification of LDs, LTRs and LTEs using UHPLC/HR data-dependent neutral loss MS^3 experiments and the developed PCA-QDA classification models. The developed non-targeted analysis strategy from **Paper III** was then further improved in **Paper IV**. A schematic overview of the non-targeted analysis strategy used in **Paper IV** is illustrated in Figure 13.

The strategy starts with dissolving a technical lignin sample, which is often already a challenge by itself. Sample preparation is a crucial part of the analytical chain and especially for non-targeted analysis, a sample preparation with as less as possible sample loss and bias should be prioritized, because no validation of the sample preparation method is possible. In this thesis work, a mixture of acetone/water (70/30, v/v) was used to dissolve solid technical lignin samples, as this solvent combination has been shown by Boeriu *et al* to be a suitable solvent for a wide spectrum of different technical lignins[82]. For liquid samples, like e.g. the depolymerised Kraft lignin sample in **Paper IV**, which was dissolved in an aqueous solution, LLE using ethyl acetate as extraction solvent was performed.

In the second step, one UHPSFC/fullscan-HRMS and five UHPSFC/HR-data dependent neutral loss MS^3 experiments, screening for five characteristic neutral losses of LMs and LOs previously reported in literature[16,63], were performed separately. In the next step, a peak list was created using the fullscan-HRMS data and a first reduction of the peak list was done by for example removing all chemical formulas with less than six carbon atoms. The developed KMD-PCA-

QDA classification models for LMs, LDs, LTRs and LTEs, which were described in chapter 5, section 5.4, were applied to recognize classes of remaining detected m/z values in the list. All classified compounds were then validated using multiple criteria, such as $^{13}C/^{12}C$ ratios. In the last step, the identification confidence level according to Schymanski *et al*[76] was determined.

Table 1 shows an overview of the obtained results of the analysis of three different technical lignin samples using the developed non-targeted strategy from **Paper IV**. It can be seen that most compounds were identified with identification confidence level 3, which were obtained by exact mass, chemical formula and a classification using the KMD-PCA-QDA classification models. Identification confidence level 2 was given if a classified LM showed typical neutral losses and if a classified LO showed a neutral loss including a benzene ring unit. For some LMs identification confidence level 1 was achieved by comparison with reference standards.

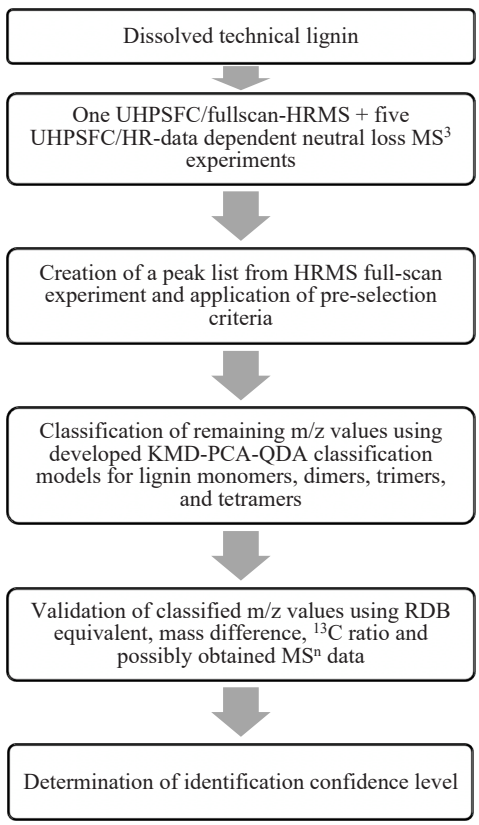


Figure 13. Schematic overview of a developed non-targeted analysis strategy for the identification of LMs and LOs in technical lignin samples. Figure from **Paper IV**, reprinted with permission.

Table 1. Overview of identified lignin monomers, dimers, trimers and tetramers in three different technical lignin samples using the developed non-targeted analysis strategy from **Paper IV**. LKL: Lignoboost Kraft lignin, SLS: sodium lignosulphonate lignin, DKL: depolymerised Kraft lignin, [a]: using peak list creation workflow and pre-selection criteria (see **Paper IV**), identification confidence level according to Schymanski *et al*[76]. Table from **Paper IV**, reprinted with permission.

Sample	LKL	SLS	DKL
Number of obtained m/z values[a]	372	132	113
Number of classified m/z values	211	67	58
Number of validated m/z values	120	29	44
Lignin monomers	15	14	10
Lignin dimers	78	13	34
Lignin trimers	23	2	0
Lignin tetramers	4	0	0
Number of m/z values matching reported data	51	11	22
Number of classified and validated m/z values with more than one retention time	73	12	20
Identified phenolic compounds	260	50	77
Lignin monomers	23	25	63
Lignin dimers	194	23	63
Lignin trimers	38	2	0
Lignin tetramers	5	0	0
Phenolic compounds with identification confidence level 1	5	6	7
Phenolic compounds with identification confidence level 2	21	7	9
Phenolic compounds with identification confidence level 3	234	37	61

The analysis of three different technical lignin samples demonstrates the wide applicability of the developed non-targeted analysis strategy using UHPSFC/HRMS^n in combination with a KMD-PCA-QDA classification model. Furthermore, neither reference standards nor investigation of MS^n fragmentation patterns are needed to classify the unknown LMs and LOs. The implemented multivariate classification models assure that at least an identification confidence level 3 is achieved. As it can be seen in Table 1, for several of the classified and validated m/z values more than one retention time was observed, which indicates the presence of structural isomers. Hence, the use of a separation technique prior the MS analysis is inevitable for the identification of isobaric structural isomers. The chromatographic separation of technical lignin samples will be discussed in more detail in the next chapter.

6. Liquid and supercritical fluid chromatography

Liquid chromatography and supercritical fluid chromatography are both used for the separation of non-volatile compounds. LC was commercially introduced in the early 1970s[83] and the first commercially available SFC system was launched around ten years later in the early 1980s[84]. While in LC the mobile phase consists of liquid organic solvents, like e.g. methanol or acetonitrile, and water, in SFC the mobile phase is usually a compressed liquid, supercritical carbon dioxide ($scCO_2$) or a combination of both. Compared to commonly used mobile phases in LC, $scCO_2$ has a much lower viscosity. This can be used to improve the separation speed and efficiency due to higher obtainable flow rates and reduced mass transfer resistance[85].

The separation efficiency of a chromatographic method can be calculated with the Van Deemter equation, which includes the terms that can cause peak broadening effects in chromatographic separations[86]:

$$H = A + \frac{B}{u} + C \times u$$

With H being the separation efficiency, also referred to as plate height, A describing multiple flow paths, B the longitudinal diffusion and C the mass transfer resistance as a function of the flow rate u. The particle size affects both, the A-term and the C-term and smaller particle sizes will lead to higher separation efficiency. However, smaller particles will also increase the column backpressure in the same time, resulting in technical limitations of chromatographic systems. The longitudinal diffusion term describes the diffusion along the flow path axis and is correlated proportionally to the diffusion coefficient of the analyte in the mobile phase, Dm. Since B is correlated anti-proportional to u, its peak broadening effect can be reduced at high flow rates. A major advancement in modern LC in terms of separation efficiency was introduced in 2004 with UHPLC columns, which are packed with particles smaller than 2 µm, and a LC system that can provide the high pressure that is needed to pump the mobile phase through the more dense packed columns[87]. In modern SFC also packed sub-2 µm particle columns are standard nowadays[85].

An important parameter to describe the separation between two compounds is the chromatographic resolution R_S:

$$R_S = \frac{2 \times (t_{R2} - t_{R1})}{w_{b1} + w_{b2}}$$

With t_{R1} being the retention time of the first eluting peak and w_{b1} the peak width at the base of the peak, t_{R2} the retention time of the second eluting peak with its corresponding peak width w_{b2}. A complete separation, also called baseline separation, is achieved if R_S is ≥1.5. Since the peak width is an important parameter to achieve high R_S, R_S depends directly on the overall separation efficiency of the chromatographic method. In this thesis work, R_S is used to compare the separation efficiency of different LC or SFC stationary phases.

6.1 Separation of LMs and LOs using LC and SFC

For the separation of LMs and LOs, RP-LC is usually chosen, mostly with typical RP stationary phases like C18 or phenyl[12,13]. Probably one of the first HPLC method developed for the separation of LMs was published by Steinberg *et al* in 1984[88]. Steinberg *et al* used a packed C18 column with a particle size of 5 µm[88]. A modern sub-2 µm particle UHPLC column with an ethylene bridge hybrid (BEH) C18 stationary phase was used by Kiyota *et al* for the separation of synthesised LMs, dimers and LTRs and for the separation of a lignin sample extracted from sugarcane[13].

The analysis of LMs and LOs using SFC has rarely been reported in the literature, to my knowledge, only two reported studies by Sun *et al*[14,89], one by Abdelaziz *et al* (Publication III under 'Publications not included') and **Paper I** and **IV** included in this thesis. All five studies were performed within our research group during the past five years. Probably the first work describing a separation of LMs using SFC was reported by Sun *et al* in 2016[14]. Sun *et al* reported a method optimisation for the separation of 11 selected LMs using a UHPSFC system coupled to a diode array detector. The method optimisation included a UHPSFC column screening and the investigation of the effect on the separation of the column temperature, the system backpressure and three different organic modifier additives. The optimised method was then used for the identification of LMs in an oxidized humic acid sample[14]. In another work, Sun *et al* presented a developed comprehensive two-dimensional method using RP-LC x UHPSFC with diode array detection for the separation of LMs[89]. In **Paper I** a UHPSFC/QTOF-MS method was developed for the separation of 40 lignin model compounds including 39 LMs and one LD. The developed UHPSFC/QTOF-MS method was then applied for the analysis of a depolymerised lignosulphonate lignin sample (Publication III under 'Publications not included'). In **Paper V** the

developed UHPSFC/QTOF-MS method was used as a starting method for a column screening for the separation of LMs and LDs identified in a lignosulphonate sample. The developed UHPSFC method was embedded in **Paper IV** in a non-targeted analysis method for the identification of LMs and LOs in technical lignin samples using UHPSFC coupled to a LQIT–Orbitrap-MS. In **Paper II** UHPSFC coupled to a QQQ-MS was used for the investigation of the ESI behaviour of selected LMs.

6.2 Selection of stationary phase in RP-LC and SFC

The selection of the stationary phase is a key aspect to achieve a good separation of the targeted analytes. Due to the broad variety of possible functional groups present in LMs and LOs and the possibility of several functional groups being present in one molecule many different interactions are possible between them and the stationary phase. Furthermore, the difference in the number of benzene rings will also affect the retention behaviour. Therefore, a selection of a suitable stationary phase is challenging. Column screening of promising stationary phases can be a useful experimental approach to find the best performing stationary phase for the targeted analytes.

Column screening for the separation of LMs and LDs was e.g. reported by Owen *et al*[12]. In their study they investigated three different HPLC columns, a C18, a phenyl and a core-shell pentafluorophenyl column, for the separation of 11 lignin model compounds, of which nine were LMs and two LDs[12]. The core-shell pentafluorophenyl column showed the poorest chromatographic resolution and the eluted peaks were significantly widened compared to the C18 and the phenyl column, which showed similar separation efficiency[12].

In **Paper III**, the separation efficiency of a BEH C18, a BEH phenyl and a charged surface hybrid (CSH) phenyl-hexyl column was investigated for the separation of identified LDs, LTRs and LTEs in a Kraft lignin sample using RP-LC. Resolution-level graphs were used to compare the separation performances of the different columns. A resolution-level graph is showing the cumulative number of peak pairs at a certain resolution. Hence, the column with the best overall resolution will show the highest cumulative number of peak pairs over the full resolution range. A cumulative-resolution level graph comparing the three investigated columns is shown in Figure 14. As it can be seen in Figure 14, no significant difference in terms of chromatographic resolution in UHPLC was achieved by the three tested RP-columns. Looking at the BEH and CSH phenyl columns, also the difference in particle technology shows no significant difference in terms of resolution.

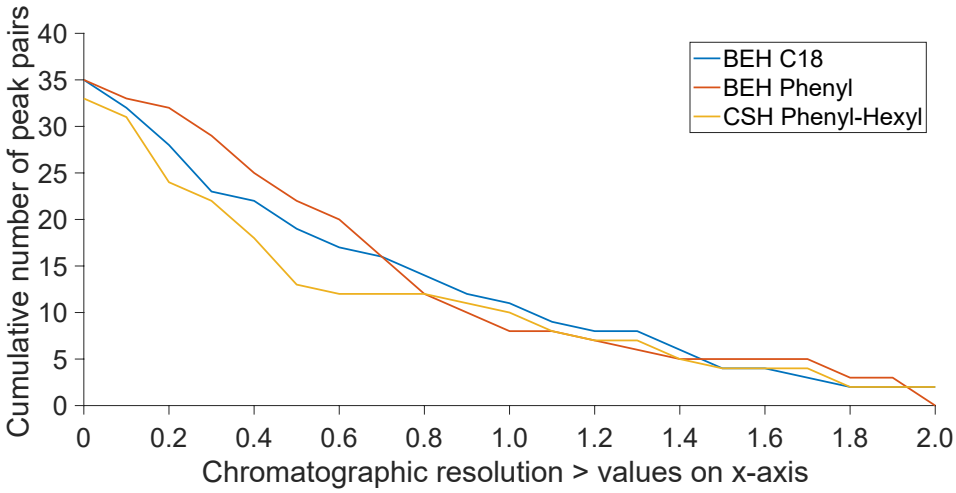


Figure 14. Resolution-level graph comparing a BEH C18, a BEH phenyl and a CSH phenyl column for the separation of lignin dimers, trimers and tetramers in a Kraft lignin sample using RP-LC. Mobile phase: water with 10 mM ammonium formate (A) and acetonitrile/water (95/5; v/v) with 10 mM ammonium formate; linear elution gradient: 30% B at 0 min to 70% B at 67 min; flow rate: 250 µL/min; injection volume: 5 µL: column temperature: 50 °C. Figure from **Paper III**, reprinted with permission.

The large variety of available stationary phases in SFC from polar, such as a DIOL stationary phase, to non-polar, such as a fluorophenyl stationary phase, offers many opportunities in terms of separation selectivity. An overview of different stationary phases available for SFC is illustrated in Figure 15. The large variety of SFC stationary phases available today is because columns developed for either RP-LC, normal phase-LC or hydrophilic interaction LC can also be used in SFC and during the recent years several column manufacturers introduced new stationary phases tailor-made for SFC[90,91]. The availability of many different stationary phase chemistries in SFC makes the selection of a suitable stationary phase for a specific separation problem a challenge. In SFC, separation mechanisms can be based on a combination of several different interaction chemistries, such as dipole-dipole interactions, hydrogen bonding, electrostatic interactions or π-π-interactions[91]. Therefore, West *et al* introduced a linear solvation energy relationships based classification model for UHPSFC stationary phases considering seven possible interaction chemistry parameters: dipole-dipole interactions, dispersive interactions, hydrogen bonding acceptor, hydrogen bonding donator, charge transfer interaction and possible anionic or cationic states of ionisable analytes[91]. If the analytes of interest have very different functional groups within their compound class and several interactions between analytes and stationary phase are possible, column screening, same as in LC, can also be used SFC to find the best performing stationary phase.

In **Paper I** UHPSFC column screening was performed for the separation of 39 different LMs. In total seven different UHPSFC columns were tested. The best overall resolving power was obtained with the diethylamine (DEA), the 1-aminoanthracene (1-AA), the 2-picolylamine (2-PIC) and the high-density diol (DIOL) stationary phases. Compared to those stationary phases, poor overall resolving power was obtained with the fluorophenyl (FP), ethylene bridged hybrid (BEH) and C18 stationary phases. The DIOL stationary phase was selected for the final method, due to relatively short analysis time and good overall resolution for the investigated LMs. In **Paper V**, five different UHPSFC columns were screened for the separation of 11 LMs and 14 LDs identified in a lignosulphonate lignin sample. Figure 16 shows a resolution-level graph comparing the obtained overall resolution of the five screened stationary phases. Comparing the obtained resolutions for all peak pairs, the best separation was obtained using the DEA stationary phase, followed by the 1-AA. The 2-PIC and the DIOL stationary phases show very similar resolving power for the investigated compounds, while the FP stationary phase showed the lowest of all screened stationary phases. A comparison of Figures 14 and 16 demonstrates the advantage of the wide spectrum of stationary phase chemistries for SFC compared to RP-LC. While the resolution-level graphs for the three different investigated RP-LC columns in Figure 14 show very similar progressions, the progressions of the resolution-level graphs obtained for the investigated SFC columns in Figure 16 are very different. The graph progressions for the SFC columns are different due to the very different separation selectivities of the tested SFC stationary phases.

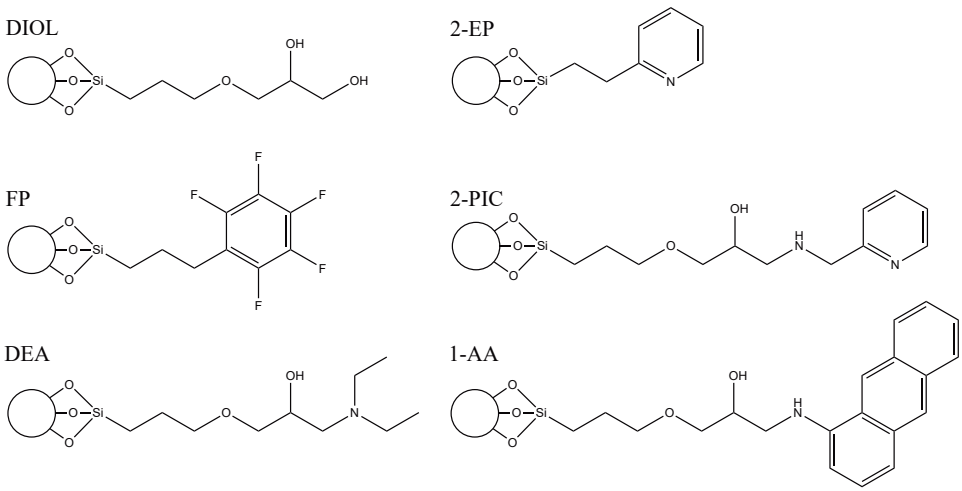


Figure 15. Chemical structures of different SFC stationary phases. DIOL: high-density diol, 2-EP: 2-ethylpyridine, FP: fluorophenyl, 2-PIC: 2-picolylamine, DEA: diethylamine and 1-AA: 1-aminoanthracene.

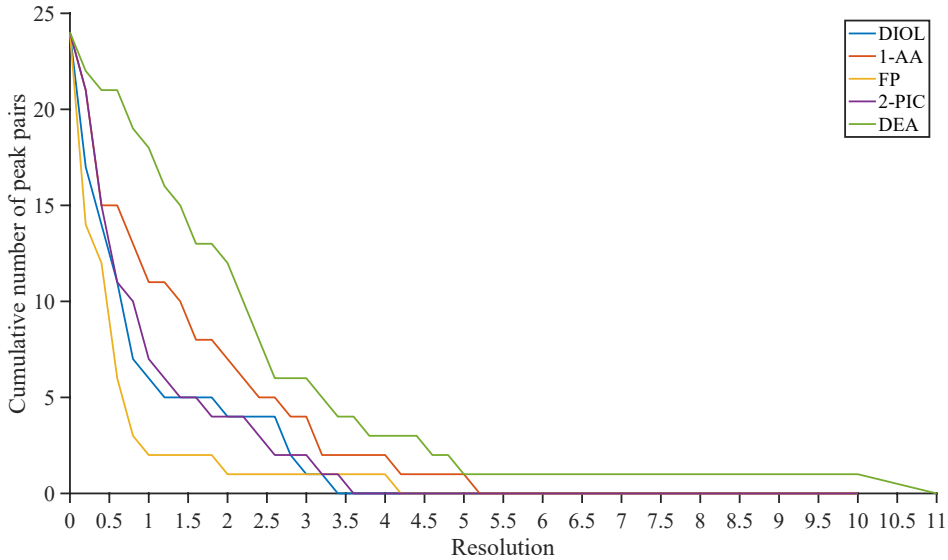


Figure 16. Resolution-level graph comparing the separation efficiency of five different UHPSFC stationary phases for the separation of LMs and LDs in a lignosulphonate lignin sample. All experiments were performed using the same UHPSFC settings. Figure from **Paper V**.

6.3 Selection of mobile phase and elution gradient in RP-LC and SFC

Besides the stationary phase, also the selection of the mobile phase and the elution gradient are important factors to consider in LC. Commonly used solvents in RP-LC are for instance water, methanol and acetonitrile. Important mobile phase parameters to be considered by the selection of a mobile phase composition are their elution strength, selectivity and viscosity. The elution strength depends on the interactions between the mobile phase and stationary phase, while the selectivity of a solvent is described by analyte – mobile phase interactions[92]. A high elution strength will lead to a faster elution of analytes and a shorter analysis time can be achieved. Solvents with a low viscosity create lower column backpressure and higher flow rates can be used, leading to shorter analysis times and also less peak broadening effects caused by mass transfer resistance. Therefore, a acetonitrile/water mixture was chosen as the mobile phase in **Paper III**, even though acetonitrile/water mixtures can suppress π-π-interactions and dipole-dipole interactions when using a phenyl column[93].

Carbon dioxide is the main mobile phase component in modern packed column SFC and usually an organic modifier is added to extend the application range towards more polar analytes[94]. The most commonly used modifier is methanol,

but also other solvents such as ethanol, isopropanol or acetonitrile are used[95]. Same as for HPLC, important parameters to consider for the selection of a mobile phase composition in SFC are the solvent strength, selectivity and viscosity. Compared to RP-LC, in SFC often polar stationary phases are used, such as a DIOL or DEA stationary phase, wherefore interactions between the organic modifier and the stationary phases need to be considered more carefully. Modifiers can be adsorbed on the stationary phase and may change its interaction chemistry, volume or three-dimensional structure[95]. Also, the addition of small amounts of water to the organic modifier can improve the separation efficiency, especially when polar stationary phases are used[96].

Elution gradients are a useful tool to improve the separation in terms of time and resolution. In an elution gradient, the mobile phase composition is changed over time, leading to a change in elution strength. Same as in LC, the use of an elution gradient in SFC can improve the separation efficiency. In **Paper V**, seven different gradients were used for each investigated stationary phase. Figure 17 shows the overall resolution obtained with the best performing gradient on each stationary phase. Except for the DEA stationary phase, where the first performed gradient gave already the best chromatographic resolution, for each column an improvement of the resolving power was achieved. However, the gradient optimisation did not lead to a change in the order of columns from the best to the worst-performing column. Therefore, it is advantageous to start the method development in SFC with column screening followed by gradient optimisation of the best performing column.

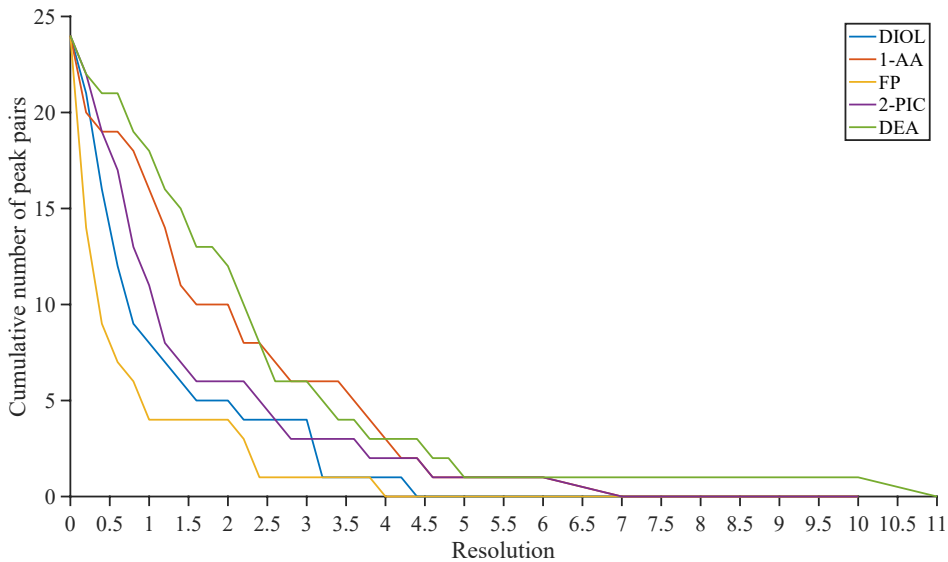


Figure 17. Resolution-level graph comparing the separation efficiency of the best performing elution gradient of each investigated UHPSFC stationary phase for the separation of LMs and LDs in a lignosulphonate lignin sample. Except for the elution gradients, all UHPSFC settings were kept constant. Figure from **Paper V**.

6.4 Separation of lignin monomers and dimers using SFC

As mentioned above, one strength of SFC is the availability of many different stationary phases, offering different analyte – stationary phase interaction chemistries. This strength was used in **Paper V** to investigate if a separation of LMs from LDs can be achieved using SFC. A chromatographic separation of LMs and LDs could potentially be used as a classification parameter in non-targeted analysis approaches for the identification of unknown LMs and LDs in complex technical lignin samples. The separation of LMs and LDs using different UHPSFC stationary phases is illustrated in Figure 18 by combined extracted ion chromatograms from each investigated analyte. It can be seen that no complete separation of LMs and LDs was achieved with any of the five stationary phases. However, the 1-AA stationary phase shows a high separation efficiency for LMs and LDs, since only one LM elutes in the LD region. Due to the incorporated anthracene ring, the 1-AA stationary phase offers enhanced π-π-interactions between analyte and stationary phase, which is beneficial for the separation of LMs and LDs. The latest eluted LM, which co-elutes with some LDs, is 4-hydroxybenzoic acid. The polar carboxylic acid group might interact by ion-dipole interaction or hydrogen bonding with the secondary amine or the hydroxyl group next to the anthracene ring in the 1-AA stationary phase, leading to higher retention. This hypothesis can be supported by the retention of vanillic acid, which is the second last eluting LM. The separation of LMs and LDs using the FP stationary is also dominated by π-π-interactions. That is why the FP column shows the second-best separation efficiency for the two compound classes of the five investigated stationary phases, however, with a lower resolving power and more co-elution of LMs and LDs compared to the 1-AA stationary phase. With the DIOL, 2-PIC and the DEA stationary phases, a separation between LMs and LDs was rarely achieved. The DIOL stationary phase has a hydrogen bonding-based separation mechanism, which might explain why the elution ranges of LMs and LDs overlap to a large extend. For both, the 2-PIC and the DEA, the separation is mainly based on ion-dipole interaction and hydrogen bonding, which both seem not to offer enough selectivity for the separation of LMs and LDs.

As discussed above the DEA column shows the best overall resolution power for the investigated analytes of the investigated SFC columns, however, looking at the separation of LMs and LDs the DEA column shows the poorest separation selectivity for the two compound classes. The 1-AA column shows a good overall resolving power and the best separation selectivity for the compound classes of LMs and LDs. The DIOL and the 2-PIC show no good separation of LMs and LDs and also not the best overall resolving power. Interestingly, the FP column

shows the poorest overall resolving power of all investigated columns, but on the other hand, the FP column shows the second-best separation of LMs and LDs.

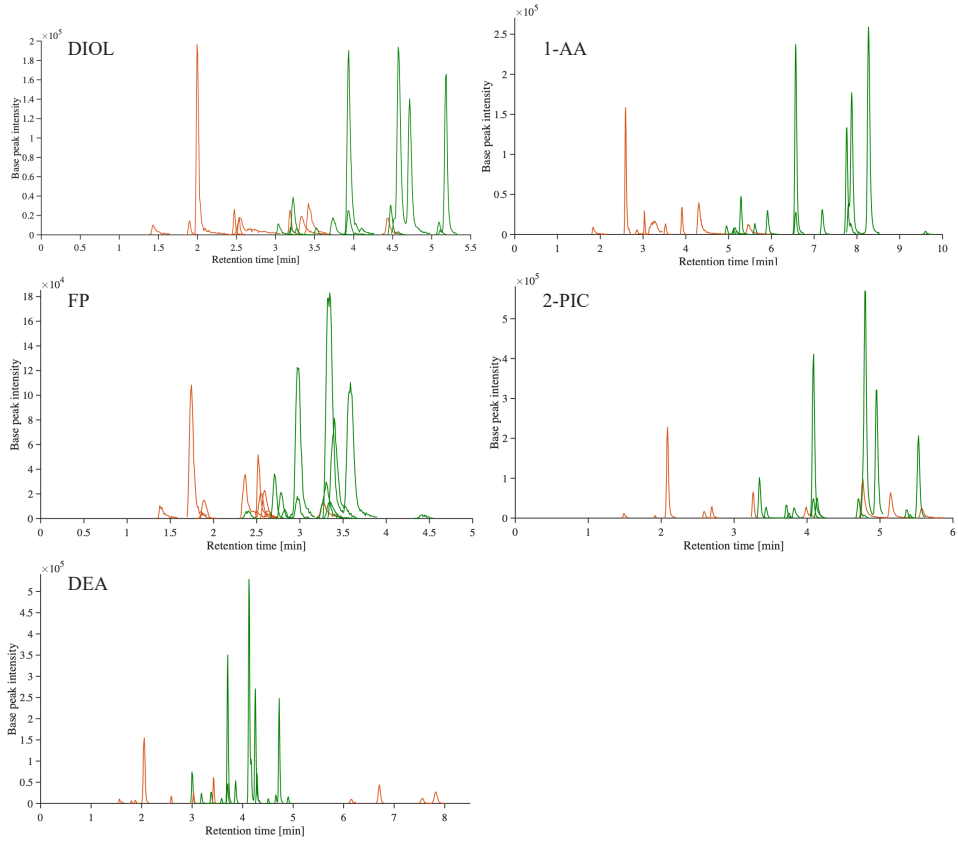


Figure 18. Combined extracted ion chromatograms showing the separation obtained with the best gradient of each SFC column for the analytes investigated in **Paper V**. Orange: lignin monomers, green: lignin dimers.

In this chapter, it was demonstrated that for the separation of LMs the DEA, 1-AA, 2-PIC and DIOL stationary phases can be used to achieve a good overall resolving power. In **Paper V** it was shown that the DEA and 1-AA stationary phases offer the best overall resolving power for the separation of LMs and LDs. If the separation of the two compound classes of LMs and LDs is aimed, then the 1-AA stationary phase should be selected, which shows a good overall resolving power and an almost complete separation of the investigated LMs an LDs. Also, the FP stationary phase shows a good separation of the two compound classes, however, with a relatively poor resolving power compared to the other investigated stationary phases.

7. Conclusions

7.1 Concluding remarks

This thesis work was dedicated to the analysis of lignin monomers and oligomers in technical lignins using SFC and MS.

The electrospray ionisation efficiency of LMs in negative ionisation mode using SFC coupled to ESI–MS was investigated from two different viewpoints. First, the overall ionisation of a wide spectrum of LMs with a wide spectrum of different chemical functionalities was investigated, followed by an investigation of the ionisation efficiencies of selected groups within LMs categorized by their chemical functionalities and number of methoxy groups. In both cases, a systematic optimisation of the electrospray ionisation efficiency was performed using design of experiments. For the electrospray ionisation of a wide range of LMs, the type of makeup solvent, the type of makeup solvent additive, the cone voltage and the desolvation gas temperature were identified as significant parameters. Furthermore, the concentration of the makeup solvent additive showed significant interactions with the cone voltage and the desolvation gas temperature and should therefore be considered as an important parameter for the electrospray ionisation of LMs. The investigation of the electrospray ionisation of LMs based on their functional group and number of methoxy groups showed that different electrospray ionisation parameters are significant for the different selected groups. For example, the gas temperature shows a significant positive influence on the electrospray ionisation for all compounds with two methoxy groups, but for compounds with one or no methoxy groups no strong influence of the gas temperature was found. Also, differences between the different functional groups were identified. For example, the sheath gas flow rate was found to be important for the ionisation of LMs with a carboxylic acid group, but not for LMs with an alcohol group.

A non-targeted analysis strategy tailored for the identification of unknown LMs and LOs in technical lignin samples has been developed within this thesis work. The combination of UHPSFC/ESI–HRMS^n and the developed KMD-PCA-QDA classification models based on identified LMs and LOs from literature studies assures an identification confidence level 3 for unknown LMs and LOs. Identification confidence level 2 can be obtained for LMs and LOs, if besides the

classification and validation typical neutral losses in MS^n experiments are found. Especially, the detection of a neutral loss including a benzene unit can be used for the identification of LOs. It was also shown that the use of a high resolution at every MS^n stage is beneficial for the structural elucidation of unknown LMs and LOs, since some neutral losses could be misinterpreted. For example, the neutral loss of 30 u can be a loss of formaldehyde, with an exact mass of 30.0106 u, or of two methyl radicals, with an exact mass of 30.0470 u. The classification models can be applied to any type of technical lignin samples since they are based on LMs and LOs identified in different technical lignin samples. Additionally, it was demonstrated that the developed classification models can be used as a selective pre-selection tool to recognize classes of detected masses without the need of reference standards and without the time-consuming evaluation of MS^n data to search for characteristic fragmentation patterns. Furthermore, with the multivariate classification approach, no monovariate limits of classification parameters are needed.

The full potential of SFC for the separation of LMs and LOs has not been fully explored yet, but a first impression is given within the presented work. The available wide spectrum of chromatographic selectivities in SFC was demonstrated by column screening for the separation of LMs and LDs. Column screenings are usually performed using reference standards. However, in this work, the lack of commercially available reference standards for LDs needed for the column screening was compensated by using the developed KMD-PCA-QDA classification model. The best overall chromatographic resolving power was obtained for the investigated LMs and LDs for the DEA stationary phase. However, a class-based separation of LMs and LDs was not achieved using the DEA column. The class-based separation was most likely not obtained using the DEA column, because of too strong hydrogen bonding interactions, which apparently do not offer a good separation selectivity for LMs and LDs. The best separation selectivity with an almost complete separation of the investigated LMs and LDs was achieved with the 1-AA column. The second-best separation selectivity was found for the FP column. Both columns offer π-π-interactions, which most likely are the reason for the good separation of the compound classes of LMs and LDs.

7.2 Outlook on future research

Many new research questions emerged along the way during this thesis project and, as common in PhD student projects, could not be addressed by now due to time limitations.

The used fractional factorial design for the functional group-based optimisation of the electrospray ionisation for LMs could not reveal two-parameter interactions and maybe the importance of some parameters might have been missed. The injection parameters, feed speed and overfeed volume were suspected to have a significant influence on the electrospray ionisation, but both parameters showed significant influences only for a few compounds. If an interaction model would be applied the importance of the two parameters might be discovered by significant interactions with other parameters.

For some reason, no oligomers higher than tetramers have been detected so far in the different analysed technical lignin samples with the developed HRMS method. One reason may be that they were simply not present in the investigated samples. This, however, seems to be unlikely when looking at size exclusion chromatograms of similar samples. Another reason may be that the used ESI settings are not suitable for their ionisation. Therefore, different ESI settings or the use of different ionisation sources, such as APCI or APPI, could be explored. Moreover, higher molecular weight oligomers may have multiple charges and the produced MS data should be screened for multiple charged ions. If somehow more oligomers could be identified, potentially KMD-PCA-QDA classification models could also be created for lignin pentamers, hexamers and higher molecular weight oligomers.

The obtained score plot of the KMD-PCA model shows a pattern, especially clear within the LM cluster, that reminds to a typical pattern of KMD plots. The sub-cluster could contain useful information about identified LMs and LOs, e.g. the presence of specific functional groups and should therefore definitely be investigated.

Another issue that is worth being addressed are the structural isomers of LMs and LOs, which presence could be shown within this work, but have not be investigated in more detail. The structural elucidation of isobaric isomers of LMs and LOs in technical lignins using mass spectrometry is probably one of the most challenging tasks within this research field. Some isobaric isomers could already be separated using UHPSFC and if the ionisation intensities are good enough first MS^n experiments could be performed, potentially revealing different fragmentation pathways. However, the ionisation efficiency is often not good enough for even an MS^2 experiment and an optimisation of the ionisation settings might be needed. Potentially, more isobaric isomers may be present in the samples, which have been undiscovered by now. Modern ion mobility techniques may be able to discover more isobaric isomers and their application for the analysis of technical lignins should therefore be explored. Maybe, with the combination of high-resolution ion mobility instruments isobaric isomers can be separated and investigated separately by $HRMS^n$ for structural elucidation.

The separation efficiency and selectivity of SFC have been only investigated so far for the separation of LMs and LOs. However, also class-based separations of LMs, LDs, LTRs and LTEs could be studied. For example, all four compound classes were identified in the Lignoboost Kraft lignin sample, which could be used for a column screening. Additionally, more SFC stationary phases offering π-π-interactions could be investigated, such as the polar-reversed phase stationary phase with ether-linked phenyl groups.

8. Acknowledgments

During the last five years many people have supported me on my way to finish this thesis and I am very grateful to every single one of them.

First of all, I like to express my deepest gratitude to my supervisors **Lotta** and **Maggan**. Without your outstanding guidance and unlimited support, I might have not been able to finish this thesis. I am especially grateful for your support during the times when things did not work out nicely and everything seemed to be hopeless. I can truly say that I cannot imagine better supervisors than the two of you!

I also like to thank my co-supervisor **Christian** for all the helpful advices about lignin, especially at the beginning of my PhD journey. Furthermore, I like to thank **Ulf** for being my department representative.

I like to thank **Peter Sp.** for all the nice and nerdy discussions about chemometrics and mass spectrometry. It sometimes felt like having another co-supervisor.

I owe a deep gratitude to **Sofia** for all the help in the mass spectrometry lab and for showing me all the little tricks to get the instruments running again.

Maria and **Katarina** have been enormously helpful when it came to administrative issues and I am very grateful for all your help during all the years.

I like to thank all present and former members of the **Green Technology Group (GTG)** I worked with: Thamani, Fiona, Federica, Veronica, Daniel P., Daniel M. D., Mahdi, Epi, Cecilia, Jingwen, Hafiz, Said, Bob, Nabomo, Victor, Firas, Barbara, Linda, Larissa, Tea, Merichel, Emil, Alexander, Simon, Yu, Mynta, Sujata, Andrés, Mar, Petter, Teshome, Tura and Yannick. I especially like to thank **Hafiz** for all the great times at work and also beside work, and for becoming a great friend. I will especially remember our little adventure during and besides the Analitika conference in South Africa and I like to thank thereby Lotta and Maggan to made it happen and **Yannick**, **Somandla**, **Prof Tutu** and **Luke** for their warm welcome and great time in South Africa. **Daniel M. D.**, you really deserve my deepest gratitude for your support in both working and private life and for being a great friend. I like to thank **Said** for all the nice research and non-research discussions when having lunch together at the design centre. I like to thank **Bob** for introducing me to lignin research during the begin of my PhD life

and for our little road trip adventure in the US after a HPLC conference. I like to thank **Victor** for many great concert evenings in Skåne and Copenhagen and **Firas** for answering all my questions during my first year as a PhD student. I also like to thank **Alexander** and **Simon** for their contribution to this thesis within their diploma thesis projects.

Furthermore, I like to thank current and former members of the analytical chemistry section within CAS: Oksana, Maider, Alicia, Romy, Marta, Claudia, Kathi and Irene. I like to thank especially **Alicia** for becoming a wonderful friend! I owe you a lot and I hope we can meet again very soon!

I like to thank all current and former members of the **Lignin Project** at Lund University: Krithika, Kena, Omar, Daniel B., Daniel P., Per, Magnus, Christian, Marie, Henrik, Gunnar, Lotta, Maggan and Javier.

I also like to thank my Team-Moss-mates **Oskar** and **Gaudry** for the great time during our little Greenland expedition and for becoming real moss experts. Thereby I also like to thank **Jan** and **Nikoline** for organizing the expedition to Greenland and such a great course!

Besides work, many good friends supported me on my way to the end of this thesis. I am especially thankful to **Moritz**, who continuously cheered me up all the years! I also like to thank **Seba**, **Pitti**, **Julia**, **Zoppe**, **René**, **Peter St.** and **Karin** for being such amazing friends and for always letting me feel like being home immediately when I came around. I am also thankful to **Sebastian L.**, without your advice back in somewhere around 2007, this thesis would have never been even started. I like to thank my PhD-life-struggling-mates **Geert** and **Roman** for all the great evening in Malmö during the last years. I also like to thank my old climbing crew: Hafiz, Jakov, Marta, Karl-Erik and Geert. I like to thank **Daniel L.-G.** and **Claire** for all the great concert evenings and for becoming very good friends over the years. I like to thank my band-mates from **Ho Chi Minh Love Affair**, **Edwin**, **Alex** and **Sebastian M.**, let's keep on rockin'! I also like to thank all the great musicians I played together with during the last years: **Evelina** (Ho Chi Minh Love Affair), **Geert** and **Mariela** (**The Wretched Kids**), **Moritz**, **Sebastian L.** and **René** (**Artsem**), **Niels** and **Pablo** for some Iron Maiden jamming and **Daniel O.** for some nice metal jamming.

I am grateful to my **Parents**, my **Sister** and my **Brother** for all your support and believe in me during all the years!

I am very thankful to **Lisa**. You gave me all the help and patience I needed during the last stressful months of this thesis. The day we met at Malmö C was the best day in my life!

9. References

1. Ganewatta, M. S., Lokupitiya, H. N. & Tang, C. Lignin biopolymers in the age of controlled polymerization. *Polymers (Basel).* **11**, (2019).
2. Bruijnincx, P. C. A., Rinaldi, R. & Weckhuysen, B. M. Unlocking the potential of a sleeping giant: lignins as sustainable raw materials for renewable fuels, chemicals and materials. *Green Chem.* **17**, 4860–4861 (2015).
3. Abdelaziz, O. Y. *et al.* Biological valorization of low molecular weight lignin. *Biotechnol. Adv.* **34**, 1318–1346 (2016).
4. Berlin, A. & Balakshin, M. Industrial Lignins. in *Bioenergy Research: Advances and Applications* (eds. Gupta, V. K., Tuohy, M. G., Kubicek, C. P., Saddler, J. & Xu, F.) (Elsevier, 2014).
5. Zevallos Torres, L. A. *et al.* Lignin as a potential source of high-added value compounds: A review. *J. Clean. Prod.* **263**, (2020).
6. Jiang, X. *et al.* Fractionation and Characterization of Kraft Lignin by Sequential Precipitation with Various Organic Solvents. *ACS Sustain. Chem. Eng.* **5**, 835–842 (2017).
7. Rönnols, J. *et al.* Structural changes in softwood kraft lignin during nonoxidative thermal treatment. *Nord. Pulp Pap. Res. J.* **30**, 550–561 (2015).
8. Martín-Sampedro, R. *et al.* Chemical and thermal analysis of lignin streams from Robinia pseudoacacia L. generated during organosolv and acid hydrolysis pre-treatments and subsequent enzymatic hydrolysis. *Int. J. Biol. Macromol.* **140**, 311–322 (2019).
9. Del Río, J. C., Gutiérrez, A., Romero, J., Martínez, M. J. & Martínez, A. T. Identification of residual lignin markers in eucalypt kraft pulps by Py-GC/MS. *J. Anal. Appl. Pyrolysis* **58**–**59**, 425–439 (2001).
10. Ohra-Aho, T., Tenkanen, M. & Tamminen, T. Direct analysis of lignin and lignin-like components from softwood kraft pulp by Py-GC/MS techniques. *J. Anal. Appl. Pyrolysis* **74**, 123–128 (2005).

11. Raj, A., Krishna Reddy, M. M. & Chandra, R. Identification of low molecular weight aromatic compounds by gas chromatography-mass spectrometry (GC-MS) from kraft lignin degradation by three Bacillus sp. *Int. Biodeterior. Biodegrad.* **59**, 292–296 (2007).

12. Owen, B. C. *et al.* High-performance liquid chromatography/high-resolution multiple stage tandem mass spectrometry using negative-ion-mode hydroxide-doped electrospray ionization for the characterization of lignin degradation products. *Anal. Chem.* **84**, 6000–6007 (2012).

13. Kiyota, E., Mazzafera, P. & Sawaya, A. C. H. F. Analysis of soluble lignin in sugarcane by ultrahigh performance liquid chromatography-Tandem mass spectrometry with a do-it-yourself oligomer database. *Anal. Chem.* **84**, 7015–7020 (2012).

14. Sun, M., Lidén, G., Sandahl, M. & Turner, C. Ultra-high performance supercritical fluid chromatography of lignin-derived phenols from alkaline cupric oxide oxidation. *J. Sep. Sci.* **39**, 3123–3129 (2016).

15. Morreel, K. *et al.* Mass Spectrometry-Based Sequencing of Lignin Oligomers. *Plant Physiol.* **153**, 1464–1478 (2010).

16. Morreel, K. *et al.* Mass spectrometry-based fragmentation as an identification tool in lignomics. *Anal. Chem.* **82**, 8095–8105 (2010).

17. Jacobs, A. & Dahlman, O. Absolute molar mass of lignins by size exclusion chromatography and MALDI-TOF mass spectroscopy. *Nord. Pulp Pap. Res. J.* **15**, 120–127 (2000).

18. Asikkala, J., Tamminen, T. & Argyropoulos, D. S. Accurate and reproducible determination of lignin molar mass by acetobromination. *J. Agric. Food Chem.* **60**, 8968–8973 (2012).

19. Aro, T. & Fatehi, P. Production and Application of Lignosulfonates and Sulfonated Lignin. *ChemSusChem* **10**, 1861–1877 (2017).

20. Jarrell, T. M. *et al.* Characterization of organosolv switchgrass lignin by using high performance liquid chromatography/high resolution tandem mass spectrometry using hydroxide-doped negative-ion mode electrospray ionization. *Green Chem.* **16**, 2713–2727 (2014).

21. Banoub, J., Benjelloun-Mlayah, B., Ziarelli, F., Joly, N. & Delmas, M. Elucidation of the complex molecular structure of wheat straw lignin polymer by atmospheric pressure photoionization quadrupole time-of-flight tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **21**, 2867–2888 (2007).

22. Banoub, J. *et al.* A critique on the structural analysis of lignins and application of novel tandem mass spectrometric strategies to determine lignin sequencing. *J. Mass Spectrom.* **50**, 5–48 (2015).

23. Morreel, K. *et al.* Profiling of Oligolignols reveals monolignol coupling conditions in lignifying Poplar xylem. *Plant Physiol.* **136**, 3537–3549 (2004).

24. Gross, J. H. *Mass Spectrometry - A Textbook*. (Springer Berlin Heidelberg, 2011).

25. Schymanski, E. L. *et al.* Non-target screening with high-resolution mass spectrometry: Critical review using a collaborative trial on water analysis. *Anal. Bioanal. Chem.* **407**, 6237–6255 (2015).

26. Stern, J. C. *et al.* Evidence for indigenous nitrogen in sedimentary and aeolian deposits from the Curiosity rover investigations at Gale crater, Mars. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 4245–4250 (2015).

27. Gallagher, K. J. *et al.* Isotope Depletion Mass Spectrometry (ID-MS) for Accurate Mass Determination and Improved Top-Down Sequence Coverage of Intact Proteins. *J. Am. Soc. Mass Spectrom.* **31**, 700–710 (2020).

28. Chen, G., Cao, P. & Liu, R. A multi-residue method for fast determination of pesticides in tea by ultra performance liquid chromatography-electrospray tandem mass spectrometry combined with modified QuEChERS sample preparation procedure. *Food Chem.* **125**, 1406–1411 (2011).

29. Griffiths, J. A brief history of mass spectrometry. *Anal. Chem.* **80**, 5678–5683 (2008).

30. De Hoffmann, E. & Stroobant, V. *Mass Spectrometry: Principles and Applications*. (John Wiley & Sons Ltd, 2007).

31. R., W. Application of Positive Rays. *Nature* **92**, (1914).

32. Dempster, A. J. A New Method of Positive Ray Analysis. *Phys. Rev.* **11**, 316 (1918).

33. Paul, W. & Steinwedel, H. Ein neues Massenspektrometer ohne Magnetfeld. *Zeitung für Naturforsch.* 448–450 (1953).

34. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science (80-.).* **246**, 64–71 (1989).

35. Tanaka, K. *et al.* Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **2**, 151–153 (1988).

36. Comisarow, M. B. & Marshall, A. G. Fourier Transform Ion Cyclotron Resonance Spectroscopy. *Chem. Phys. Lett.* **25**, 282–283 (1974).

37. Makarov, A. Electrostatic axially harmonic orbital trapping: A high-performance technique of mass analysis. *Anal. Chem.* **72**, 1156–1162 (2000).

38. Feider, C. L., Krieger, A., Dehoog, R. J. & Eberlin, L. S. Ambient Ionization Mass Spectrometry: Recent Developments and Applications. *Anal. Chem.* **91**, 4266–4290 (2019).

39. Owen, B. C. *et al.* High-performance liquid chromatography/high-resolution multiple stage tandem mass spectrometry using negative-ion-mode hydroxide-doped electrospray ionization for the characterization of lignin degradation products. *Anal. Chem.* **84**, 6000–6007 (2012).

40. Reymond, C. *et al.* Characterization of liquid–liquid extraction fractions from lignocellulosic biomass by high performance liquid chromatography hyphenated to tandem high-resolution mass spectrometry. *J. Chromatogr. A* **1610**, 460569 (2020).

41. Larson, E. A., Hutchinson, C. P. & Lee, Y. J. Gas Chromatography-Tandem Mass Spectrometry of Lignin with CSI : FingerID. *J. Am. Soc. Mass Spectrom.* 1908–1918 (2018).

42. Kosyakov, D. S., Ul'yanovskii, N. V., Anikeenko, E. A. & Gorbova, N. S. Negative ion mode atmospheric pressure ionization methods in lignin mass spectrometry: A comparative study. *Rapid Commun. Mass Spectrom.* **30**, 2099–2108 (2016).

43. Del Río, J. C. *et al.* Determining the influence of eucalypt lignin composition in paper pulp yield using Py-GC/MS. *J. Anal. Appl. Pyrolysis* **74**, 110–115 (2005).

44. Liu, Q., Li, P., Liu, N. & Shen, D. Lignin depolymerization to aromatic monomers and oligomers in isopropanol assisted by microwave heating. *Polym. Degrad. Stab.* **135**, 54–60 (2017).

45. Albishi, T. *et al.* Top-down lignomic matrix-assisted laser desorption/ionization time-of-flight tandem mass spectrometry analysis of lignin oligomers extracted from date palm wood. *Rapid Commun. Mass Spectrom.* **33**, 539–560 (2019).

46. Qi, Y. & Volmer, D. A. Chemical diversity of lignin degradation products revealed by matrix-optimized MALDI mass spectrometry. *Anal. Bioanal. Chem.* **411**, 6031–6037 (2019).

47. Qi, Y. *et al.* Assessment of molecular diversity of lignin products by various ionization techniques and high-resolution mass spectrometry. *Sci. Total Environ.* **713**, 136573 (2020).

48. Dier, T. K. F., Egele, K., Fossog, V., Hempelmann, R. & Volmer, D. A. Enhanced Mass Defect Filtering to Simplify and Classify Complex Mixtures of Lignin Degradation Products. *Anal. Chem.* **88**, 1328–1335 (2016).

49. Andrianova, A. A. *et al.* Electrospray Ionization with High-Resolution Mass Spectrometry as a Tool for Lignomics: Lignin Mass Spectrum Deconvolution. *J. Am. Soc. Mass Spectrom.* **29**, 1044–1059 (2018).

50. Yan, G. & Kaiser, K. A rapid and sensitive method for the analysis of lignin phenols in environmental samples using ultra-high performance liquid chromatography-electrospray ionization-tandem mass spectrometry with multiple reaction monitoring. *Anal. Chim. Acta* **1023**, 74–80 (2018).

51. Evtuguin, D. V. & Amado, F. M. L. Application of electrospray ionization mass spectrometry to the elucidation of the primary structure of lignin. *Macromol. Biosci.* **3**, 339–343 (2003).

52. Crawford, E. A., Gerbig, S., Spengler, B. & Volmer, D. A. Rapid fingerprinting of lignin by ambient ionization high resolution mass spectrometry and simplified data mining. *Anal. Chim. Acta* **994**, 38–48 (2017).

53. Kebarle, P. & Verkcerk, U. H. Electrospray: From Ions in solution to Ions in the gas phase, what we know now. *Mass Spectrom. Rev.* **28**, 898–917 (2009).

54. Fuerstenau, S. D. *et al.* Mass Spectrometry of an Intact Virus. *Angew. Chemie Int. Ed.* **40**, 982–982 (2001).

55. Haupert, L. J. *et al.* Characterization of model compounds of processed lignin and the lignome by using atmospheric pressure ionization tandem mass spectrometry. *Fuel* **95**, 634–641 (2012).

56. Klara, K., Brianna, G., Fisher, S. & Kubátová, A. Optimization of Electrospray Ionization for Liquid Chromatography Time-of-Flight Mass Spectrometry Analysis of Preservatives in Wood Leachate Matrix. *Chromatographia* **82**, 1677–1685 (2019).

57. Pedro, L., Van Voorhis, W. C. & Quinn, R. J. Optimization of Electrospray Ionization by Statistical Design of Experiments and Response Surface Methodology: Protein–Ligand Equilibrium Dissociation Constant Determinations. *J. Am. Soc. Mass Spectrom.* **27**, 1520–1530 (2016).

58. Moreiras, G., Leão, J. M. & Gago-Martínez, A. Design of experiments for the optimization of electrospray ionization in the LC-MS/MS analysis of ciguatoxins. *J. Mass Spectrom.* **53**, 1059–1069 (2018).

59. Reymond, C., Le Masle, A., Colas, C. & Charon, N. A rational strategy based on experimental designs to optimize parameters of a liquid chromatography-mass spectrometry analysis of complex matrices. *Talanta* **205**, 120063 (2019).

60. Grand-Guillaume Perrenoud, A., Veuthey, J. L. & Guillarme, D. Coupling state-of-the-art supercritical fluid chromatography and mass spectrometry: From hyphenation interface optimization to high-sensitivity analysis of pharmaceutical compounds. *J. Chromatogr. A* **1339**, 174–184 (2014).

61. Akbal, L. & Hopfgartner, G. Effects of liquid post-column addition in electrospray ionization performance in supercritical fluid chromatography–mass spectrometry. *J. Chromatogr. A* **1517**, 176–184 (2017).

62. Kostiainen, R. & Kauppila, T. J. Effect of eluent on the ionization process in liquid chromatography-mass spectrometry. *J. Chromatogr. A* **1216**, 685–699 (2009).

63. Marcum, C. L. *et al.* A Fundamental Tandem Mass Spectrometry Study of the Collision-Activated Dissociation of Small Deprotonated Molecules Related to Lignin. *ChemSusChem* **9**, 3513–3526 (2016).

64. Héberger, K. Chemoinformatics — multivariate mathematical – statistical methods. in *Medical applications of mass spectrometry* (eds. Vekey, K., Telekes, A. & Vertes, A.) 141–169 (Elsevier Science, 2008).

65. Brereton, R. G. *et al.* Chemometrics in analytical chemistry—part I: history, experimental design and data analysis tools. *Anal. Bioanal. Chem.* **409**, 5891–5899 (2017).

66. Eriksson, L., Johansson, E., Kettaneh-Wold, N., Wikström, C. & Wold, S. *Design of Experiments*. (Umetrics, 2008).

67. Eriksson, L., Byrne, T., Johansson, E., Trygg, J. & Vikström, C. *Multi- and Megavariate Data Analysis*. (Umetrics, 2013).

68. Bro, R. & Smilde, A. K. Principal component analysis. *Anal. Methods* **6**, 2812–2831 (2014).

69. Morais, C. L. M., Lima, K. M. G., Singh, M. & Martin, F. L. Tutorial: multivariate classification for vibrational spectroscopy in biological samples. *Nat. Protoc.* **15**, 2143–2162 (2020).

70. Qi, Y. & Volmer, D. A. Rapid mass spectral fingerprinting of complex mixtures of decomposed lignin: Data-processing methods for high-resolution full-scan mass spectra. *Rapid Commun. Mass Spectrom.* **33**, 2–10 (2019).

71. Kendrick, E. A Mass Scale Based on CH2 = 14.0000 for High Resolution Mass Spectrometry of Organic Compounds. *Anal. Chem.* **35**, 2146–2154 (1963).

72. Krauss, M., Singer, H. & Hollender, J. LC-high resolution MS in environmental analysis: From target screening to the identification of unknowns. *Anal. Bioanal. Chem.* **397**, 943–951 (2010).

73. Bloch, R. *et al.* Non-targeted mercapturic acid screening in urine using LC-MS/MS with matrix effect compensation by postcolumn infusion of internal standard (PCI-IS). *Anal. Bioanal. Chem.* **411**, 7771–7781 (2019).

74. Tian, L., Zheng, J., Goodyer, C. G. & Bayen, S. Non-targeted screening of plastic-related chemicals in food collected in Montreal, Canada. *Food Chem.* **326**, 126942 (2020).

75. Wang, X. *et al.* Suspect and non-target screening of pesticides and pharmaceuticals transformation products in wastewater using QTOF-MS. *Environ. Int.* **137**, 105599 (2020).

76. Schymanski, E. L. *et al.* Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. 2097–2098 (2014).

77. Zheng, M. *et al.* Development and validation of a sensitive UPLC–MS/MS instrumentation and alkaline nitrobenzene oxidation method for the determination of lignin monomers in wheat straw. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **1055**–**1056**, 178–184 (2017).

78. Huis, R. *et al.* Natural Hypolignification Is Associated with Extensive Oligolignol Accumulation in Flax Stems. *Plant Physiol.* **158**, 1893–1915 (2012).

79. Rana, M. *et al.* Production of Phenol-Rich Monomers from Kraft Lignin Hydrothermolysates in Basic-Subcritical Water over MoO3/SBA-15 Catalyst. *Energy and Fuels* **32**, 11564–11575 (2018).

80. Kang, J., Irmak, S. & Wilkins, M. Conversion of lignin into renewable carboxylic acid compounds by advanced oxidation processes. *Renew. Energy* **135**, 951–962 (2019).

81. Zhu, H. *et al.* Identification of the Phenol Functionality in Deprotonated Monomeric and Dimeric Lignin Degradation Products via Tandem Mass Spectrometry Based on Ion–Molecule Reactions with Diethylmethoxyborane. *J. Am. Soc. Mass Spectrom.* **27**, 1813–1823 (2016).

82. Boeriu, C. G. *et al.* Fractionation of five technical lignins by selective extraction in green solvents and characterisation of isolated fractions. *Ind. Crops Prod.* **62**, 481–490 (2014).

83. Dores-Sousa, J. L., De Vos, J. & Eeltink, S. Resolving power in liquid chromatography: A trade-off between efficiency and analysis time. *J. Sep. Sci.* **42**, 38–50 (2019).

84. Gere, D. R., Board, R. & McManlgill, D. Supercritical Fluid Chromatography with Small Particle Diameter Packed Columns. *Anal. Chem.* **54**, 736–740 (1982).

85. Nováková, L. *et al.* Modern analytical supercritical fluid chromatography using columns packed with sub-2μm particles: A tutorial. *Anal. Chim. Acta* **824**, 18–35 (2014).

86. van Deemter, J. J., Klinkenberg, A. & Zuiderweg, F. J. Longitudinal diffusion and resistance to mass transfer as causes of nonideality in chromatography. *Chem. Eng. Sci.* **5**, 271–289 (1956).

87. Mazzeo, J. R., Neue, U. D., Kele, M. & Plumb, R. S. Advancing LC performance with smaller particles and higher pressure. *Anal. Chem.* **77**, 460–467 (2005).

88. Steinberg, S., Venkatesan, M. I. & Kaplan, I. R. Analysis of the products of the oxidation of lignin by CuO in biological and geological samples by reversed-phase high-performance liquid chromatography. *J. Chromatogr. A* **298**, 427–434 (1984).

89. Sun, M., Sandahl, M. & Turner, C. Comprehensive on-line two-dimensional liquid chromatography × supercritical fluid chromatography with trapping column-assisted modulation for depolymerised lignin analysis. *J. Chromatogr. A* **1541**, 21–30 (2018).

90. West, C. & Lesellier, E. A unified classification of stationary phases for packed column supercritical fluid chromatography. *J. Chromatogr. A* **1191**, 21–39 (2008).

91. West, C., Lemasson, E., Bertin, S., Hennig, P. & Lesellier, E. An improved classification of stationary phases for ultra-high performance supercritical fluid chromatography. *J. Chromatogr. A* **1440**, 212–228 (2016).

92. Snyder, L. R., Glajch, J. L. & Kirkland, J. J. Theoretical basis for systematic optimization of mobile phase selectivity in liquid-solid chromatography. Solvent-solute localization effects. *J. Chromatogr. A* **218**, 299–326 (1981).

93. Croes, K., Steffens, A., Marchand, D. H. & Snyder, L. R. Relevance of π-π and dipole-dipole interactions for retention on cyano and phenyl columns in reversed-phase liquid chromatography. *J. Chromatogr. A* **1098**, 123–130 (2005).

94. Berger, T. A. *Supercritical Fluid Chromatography*. (Agilent Technologies, 2015).

95. West, C. & Lesellier, E. Effects of mobile phase composition on retention and selectivity in achiral supercritical fluid chromatography. *J. Chromatogr. A* **1302**, 152–162 (2013).

96. Khvalbota, L. *et al.* Enhancing supercritical fluid chromatographic efficiency: Predicting effects of small aqueous additives. *Anal. Chim. Acta* **1120**, 75–84 (2020).

Printed by Media-Tryck, Lund 2020 NORDIC SWAN ECOLABEL 3041 0903

ISBN 978-91-7422-760-4

Centre for Analysis and Synthesis
Department of Chemistry
Faculty of Science
Lund University