Expanding the Foundations for Applied Lignin Analysis

Norberg, Mynta

2026

Expanding the Foundations for Applied Lignin Analysis

MYNTA NORBERG | CENTRE FOR ANALYSIS AND SYNTHESIS | LUND UNIVERSITY

Expanding the Foundations for Applied Lignin Analysis

Expanding the Foundations for Applied Lignin Analysis

by Mynta Norberg

LUND
UNIVERSITY

Thesis for the degree of Doctor of Philosophy
Thesis advisors: Peter Spégel, Margareta Sandahl, Charlotta Turner
Faculty opponent: Jeffrey Hawkes

To be presented, with the permission of the Faculty of Science of Lund University, for public criticism in Kemicentrum, Sal A on Monday, June 8 of 2026 at 09:00.

Organization	Document name
LUND UNIVERSITY Department of Chemistry Box 124 SE-221 00 LUND Sweden	**DOCTORAL DISSERTATION** Date of disputation 2026-06-08 Sponsoring organization
Author(s) Mynta Norberg	

Title and subtitle

Expanding the Foundations for Applied Lignin Analysis

Abstract

Lignins are aromatic polymers found in fibrous plants, and a major side-product of paper pulping. They are under exploited due to a combination of economic and technical challenges. Valorisation efforts aim to resolve this, but these rely on informative and reliable analytical feedback. The diverse and complex structures of compounds found in lignins makes analysis difficult. This thesis outlines two different aspects of lignin analysis. The first is the quantitative analysis of monomeric products from lignin depolymerisation, which is a mature application seen in the manufacture of i.e. vanillin. Here, the main analytical challenge is the reliability of comparisons across diverse feedstocks and depolymerisation methods. The second is the qualitative analysis of, in particular process-modified, oligomeric products. Here, more fundamental challenges are present, such as limited access to reference data, and even a lack of chemical nomenclature and taxonomy.

The articles included in this thesis addressed these analytical challenges. In the first article, a GC-FID method for the quantitative analysis of oxidatively depolymerised lignosulfonate was developed and validated, and then applied in a first-of-its-kind comparison of three fundamentally different oxidative depolymerisation methods. In the second article, a cheminformatic toolkit was developed to support the qualitative analysis of lignin oligomers. This included fundamental contributions such as a comprehensive nomenclature, alongside software written in R. This software, named Lignonaut, generates oligomer libraries through virtual combinatorial synthesis. A dictionary-based SMILES translation algorithm was also developed, along with various other features useful in cheminformatics. Lignonaut was applied to chemical space exploration, predictive model validation, and the annotation of high-resolution mass spectrometry data. In addition to the included articles, this thesis also comments on quality control in interdisciplinary studies, future challenges of lignin analysis, and negative results. It also includes more personal stories of instrument troubleshooting, programming, and lecturing.

Keywords

lignin, analysis, chemistry, cheminformatics, mass spectrometry, chromatography

Classification system and/or index terms (if any)

Supplementary bibliography information	Language English
ISSN and key title	ISBN 978-91-8104-937-4 (print) 978-91-8104-938-1 (pdf)
Recipient's notes	Number of pages 136 / Price
	Security classification

Expanding the Foundations for Applied Lignin Analysis

by Mynta Norberg

LUND
UNIVERSITY

A doctoral thesis at a university in Sweden takes either the form of a single, cohesive research study (monograph) or a summary of research papers (compilation thesis), which the doctoral student has written alone or together with one or several other author(s).

In the latter case the thesis consists of two parts. An introductory text puts the research work into context and summarizes the main points of the papers. Then, the research publications themselves are reproduced, together with a description of the individual contributions of the authors. The research papers may either have been already published or are manuscripts at various stages (in press, submitted, or in draft).

Cover illustration front: Tree in ink and charcoal, by Mynta Norberg.

Cover illustration back: Ensō in ink, by Mynta Norberg.

Funding information: This work was supported by the Research Council of Norway through the project L2BA – Lignin to BioAromatics (321427).

Faculty of Science, Department of Chemistry

ISBN: 978-91-8104-937-4 (print)
ISBN: 978-91-8104-938-1 (pdf)

Printed in Sweden by Media-Tryck, Lund University, 2026

Media-Tryck is a Nordic Swan Ecolabel certified provider of printed material. Read more about our environmental work at www.mediatryck.lu.se

MADE IN SWEDEN

It's the kind of thing that makes you glad you stopped
and smelled the pine trees along the way,
you know?

Gabbro

Contents

Abbreviations

ADHD Attention Deficit Hyperactivity Disorder
AS Autosampler
FID Flame Ionisation Detection
GAI Generative Artificial Intelligence
GC Gas Chromatography
GPC Gel Permeation Chromatography
GU Guaiacyl (monomer residue)
GUllol Coniferyl alcohol (monomer residue)
GUllic Ferulic acid (monomer residue)
HPLC High-Performance Liquid Chromatography
HRMS High-Resolution Mass Spectrometry
HYllol Paracoumaryl alcohol
IM(S) Ion Mobility (Spectrometry)
%RA Percent Residual Accuracy
RSD Relative Standard Deviation
SEC Size Exclusion Chromatography
SFC Supercritical Fluid Cromatography
SFE Supercritical Fluid Extraction
SMILES Simplified Molecular Input Line Entry System
SYllol Sinapyl alcohol (monomer residue)
timsTOF Trapped Ion Mobility Spectrometry [quadrupole] Time Of Flight [mass spectrometry]

List of publications

This thesis is based on the following publications, referred to by their Roman numerals:

I **Oxidative Depolymerization of Lignosulfonate to Low-Molecular Weight Aromatics: A Comparative Study**

M. Norberg, S. Bekirovska, J. Klein, F. Moeller, K. P. J. Gustafson, M. Sandahl, C. P. Hulteberg, C. Turner, O. Y. Abdelaziz, S. R. Waldvogel, P. Spégel, and O. Bengtsson
RSC Sustainability, 2025, **3**, pp. 4818-4824

II **Lignonaut: designing diverse combinatorial libraries for the exploration and annotation of lignin oligomer spaces**

M. Norberg, M. Sandahl, and P. Spégel
Manuscript in print with Journal of Cheminformatics

All papers are reproduced with permission of their respective publishers.

Acknowledgements

I remember seeing Jens Prothmann working on HRMS data analysis in “the cave” as a bachelor student, and how that inspired me. I could never have imagined that I would get to be his successor.

To my supervisors. Thank you to Lotta for being one of my early role-models. If it wasn't for you, I would probably be doing organic chemistry by now. The same goes for Maggan. You once gave me a spontaneous lecture on a napkin while helping me with instrument maintenance. That story says a lot about you. Thank you for being so generous. If I came for Lotta and Maggan, I stayed for Peter. You taught me the value of quality over quantity, and helped pull me out of cynicism. Thank you for making space for me, my wild ideas, and my recovery.

To my team. Dániel P., I walked home sobbing after hugging you goodbye. Having you around meant more than I had realised, both in and out of the lab. To Selda, thank you for seeing me, and making all those troubles bearable. Your bright glow is incredibly infectious. Ujala, thank you for your contributions to the lignin nomenclature. Azemina B., thanks for reminding me of the value of asking questions. Thank you Josep for being a warm beacon of humanity in a frigid land. A team is not just a handful of people. To Fiona, Simon, Mona, Stefan, Kuria, Srinivas, Andreia, Thamani, Oksana, Shandilya, Erika, Daniel M.-D., Veronika, Hamidreza, and to those I forgot. Thank you for the presentations, the lunchroom discussions, the wine nights, board game evenings, and numerous other adventures. Also thanks to Linnea, and all of my other students. Following your growth, and growing with you, was an absolute treat.

To my technical and administrative staff, and others often forgotten. Maria, Sara, Sofia E., Annette, Kornelije, Fatima, all three Ulfs, and others. Projects of this scope would never be possible without you. Also a huge thanks to my half-time opponent Daniel Globisch, to the reviewers, the opponent, and committee members for helping with improving our work, and holding us accountable to it. Science would not be possible without any of you.

Work would not be possible (or worthwhile) without life to balance it. Thanks to my mom for reminding me of putting happiness first, and my dad for not letting me give up. Thanks to my Zen teacher Dharman Sensei for helping me give up. Bless up Gabbro of Outer Wilds Ventures for staying so calm in the face of inevitable doom. Thanks to all the musicians, writers, and other artists out there for your tireless efforts. And if there is anyone who helped keep me in balance, that is Sofia. Every day I spend with you reminds me why all of this is worth it in the end.

Favours are borrowed. I return them to all beings in the ten directions, three worlds, all buddhas, bodhisattva-mahasattvas, and Maha prajna paramita.

Popular summary in english

Analytical chemists develop methods for gathering, separating, and measuring what chemicals that things contain, and how much of them there are. Today, it is common to use expensive and complicated machines to achieve this, which is why analytical chemists expertise is needed. These methods are eventually used by i.e. chemical engineers at factories, or lab technicians at hospitals or water purification plants. In this doctoral thesis, methods to study lignin were developed. Lignin is a kind of fiber that is found in large amounts in i.e. wood. Paper is made from paper pulp, and paper pulp is made from wood. When paper pulp is made the lignin is removed, because otherwise the paper will be bad. The lignin is then burned as a kind of fuel at the paper mill for cooking the paper pulp, to make some savings on the energy bill. Researchers want to use lignin in more valuable ways than this. For example, today much of the worlds vanillin (synthetic vanilla) is already made from lignin, instead of from oil.

A doctoral thesis in Sweden contains one or many scientific articles, including an overarching story which also includes stuff that didn't make the cut. In the first article a method was developed for measuring how much vanillin (and similar chemicals) that is around after engineers have attempted to make it from lignin. There are many different ways of making vanillin, so to determine what method is best for making vanillin, a measuring method that is reliable and fair is needed. This is not a given, because such a method has to take out the vanillin from the soup that is left over, separate it from other stuff with a complicated machine, and then carefully measure the amount of vanillin. However, we did succeed in developing such a method. Research groups from universities in Lund and Mainz (Germany), and a factory in Norway, developed their own methods for making vanillin. These could then be compared thanks to our measuring method. However, the three methods were quite similar, so it was not possible to determine a clear winner.

Vanillin is quite a small and simple molecule. Lignin, like other fibers, looks a bit like a chain. Here, vanillin is like a single link. Researchers want to be able to use longer chains (of several links) that are found in lignin, or made from it. One difficulty of this is that there are many different types of links, and these can also be put together in many different ways. So it is difficult to know what chains you have. So in the next article, a collection of tools to help with studying such chains were therefore developed. First of all a new language for better naming these were developed, which is crucial since it's otherwise not possible to talk about them. Next a computer program which can make long lists of up to millions of such chains were developed. These lists could be used as an aid in developing statistical models, or to interpret complicated data from expensive measuring equipment.

Other things discussed in the thesis are challenges with collaborations, things we tried but didnt work out like we had hoped, and various personal stories.

Populärvetenskaplig sammanfattning på svenska

Analytiska kemister utvecklar metoder för att samla in, separera, och mäta vilka kemiska ämnen som saker innehåller, och hur mycket som finns av dem. Idag används ofta dyra och komplicerade maskiner för att göra detta, vilket är varför analytiska kemisters expertis behövs. De här metoderna används så småningom av t.ex. kemiingenjörer på fabriker, eller labbtekniker på sjukhus eller vattenverk. I den här doktorsavhandlingen så utvecklades analytiska metoder för att studera lignin. Lignin är en slags fiber som finns i stor mängd i bl.a. träd. Papper görs av pappersmassa, som i sin tur görs av trä. Vid tillverkning av pappersmassa så tas ligninet bort, eftersom pappret annars blir dåligt. Ligninet bränns sen upp som ett bränsle på pappersbruken för att koka pappersmassan, för att spara på energikostnader. Forskare vill dock använda lignin på mer värdefulla sätt än så. Idag görs t.ex. redan mycket av världens vanillin (syntetisk vanilj) av lignin, istället för av olja.

En doktorsavhandling i Sverige innehåller en eller flera vetenskapliga artiklar, tillsammans med en överhängande berättelse som också inkluderar sånt som inte kom med i artiklarna. I den första artikeln så utvecklades en metod för att mäta hur mycket vanillin (och liknande ämnen) som finns efter att ingenjörer försökt göra det från lignin. Det finns många olika sätt att göra det på, och för att bestämma vilken metod som är bäst på att göra vanillin så krävs en mätmetod som är pålitlig och rättvis. Det här är inte självklart, eftersom en sån mätmetod måste plocka ut vanillinet ifrån den soppa som blir över, separera det från annat som följer med en komplicerad maskin, och sen noggrant mäta vanillinet. Vi lyckades dock utveckla en sån metod. Forskargrupper från universitet i Lund och Mainz (Tyskland), samt ett företag i Norge, utvecklade sina egna metoder för att göra vanillin, och de här kunde sen jämföras tack vare vår mätmetod. De tre metoderna var dock ganska lika, så det var svårt att utlysa en vinnare.

Vanillin är en ganska liten och enkel molekyl. Lignin, liksom andra fibrer, ser ut lite som en kedja. Vanillin är som en ensam länk. Forskare vill kunna använda längre kedjor (med flera länkar) som redan finns i lignin, eller som kan göras av det. En svårighet med det är att det finns massor med olika typer av länkar, och de kan sättas ihop med varandra på massa olika sätt. Så det är svårt att veta vad som är vad. I nästa artikel så utvecklades därför en samling verktyg för bättre kunna studera såna kedjor. Först och främst så utvecklades ett nytt språk för att bättre kunna namnge dem, vilket är viktigt eftersom det annars inte går att prata om dem. Sen så utvecklades även ett datorprogram som kan göra långa listor med upp till miljontals av den här sortens kedjor. De här listorna kunde sen användas som ett hjälpmedel för att utveckla statistiska modeller, eller tolka komplicerad data från dyra mätinstrument.

Annat som tas upp i avhandlingen är utmaningar med att samarbeten, sånt som inte funkade som vi hade hoppats, och olika personliga berättelser.

Expanding the Foundations for Applied Lignin Analysis

1 Preface

How did it come to this? There are an innumerable amount of stories to pick from. I could tell you about how I dreamt of becoming an inventor (a childs naive notion of a scientist) since pre-school, or how I as a young adult needed something new after realising the cabin-in-the-woods and farming dream wasn't going to work out. I could tell you about how research is amenable to my autistic neurology, and how a rejection for an industry position for being "too nerdy" helped me realise I wasn't done with academia. And I have already told the story about the many people that got me here.

All of these stories are of course true, but I believe the greater story that many PhD students will relate to is one of perseverance. I was actually the second pick for this position, but never let that get to my head. I chose to stick around after my midterm crisis and burnout. I allowed my love for teaching to grow, and was generously given the rare opportunity to lecture. I got to continue developing my programming skills. These would eventually blend with my love for writing, as I found myself taking a few days recreating the university's thesis template(s) in a more modern typesetting language. Some may call these distractions, but these little side-quests were what truly defined my PhD journey in the end. This was not what I had in mind, but if I knew where I would end up then it wouldn't have been research.

At the beginning of my PhD studies, fellow researchers would remark to me ad-nauseam how "lignin is difficult". While I definitely felt like a deer caught in the headlights at times, this was mostly because I was looking too far ahead. Being a young and smaller field, there is naturally much left to do. After taking a step back from the allure of applications, I realised that the fundamental research gaps that held me up were a rare and more exciting opportunity. And this excitement was enough to help me keep at it.

At the end of my PhD studies, I can confirm that lignin sure is difficult! This thesis describes my quest for how to make it a bit less so.

2 Introduction

2.1 LIGNIN AS A BIOMASS

Lignins are a class of aromatic polymers which form the bulk of plant dry matter, along with the more abundant cellulose. The inclusion of lignin in plant cell walls provides rigidity and decay resistance.[1] Due to its high natural abundance, it has potential as a renewable alternative to fossil-based feedstocks in the petrochemical industry.[2] Lignin rich liquor is acquired as a side-stream product from the delignification step in paper pulp production, which is burned in the pulping mills to recover energy. Only 2% of spent liquors produced globally are extracted to yield technical lignins, which are used for value-added applications such as vanillin synthesis. The limited commercialisation of lignin has been explained by a combination of techno- and socio-economical challenges. This underutilisation sets lignins apart from other renewable biomasses such as carbohydrates and oils, which is why it has seen a developing research interest.[3]

Through the phenylpropanoid biosynthesis, phenylalanine is converted into the three canonical lignin monomers, or monolignols; paracoumaryl, coniferyl and sinapyl alcohol. These then link to form lignin oligomers and polymers through a variety of well-documented oxidative coupling mechanisms. Different mechanisms result in a number of different linkage types.[1,4] The ratios between different types of monomer residues and linkages varies across different plant species.[4] Additionally, the structure of lignins are heavily altered during delignification, since covalent bonds have to be broken to extract lignin from the lignocellulosic biomass. This gives technical lignins unique properties that depend on both plant species and pulping process. Even soft, lab-scale extractions have been shown to modify the lignin structure, making the structures of native lignins difficult to study.[5] This lability and diversity of lignin structures is a fundamental and pervasive challenge of lignin research.

2.2 LIGNIN STRUCTURE, TAXONOMY, AND NOMENCLATURE

Symbolic naming is frequently used for lignin monomers and linkages, much like for other abundant polymers. The three most naturally abundant monomers – paracoumaryl, coniferyl and sinapyl alcohol – are often referred to by the symbolic names of H, G, and S, respectively. These consist of a phenolic ring with zero, one, and two methoxy-substituents, respectively; and an allyl hydroxide side-chain. The brevity and intelligibility offered by symbolic names allows lignin structures to be described as a sequence of units. This has been used to great effect in both experimental[6] and computational[7,8] sequencing of lignin.

While H, G, and S (the "canonical" monomers) have seen the most attention in scientific literature, many other monomers exist in nature.[9] Process modification, which occurs even

during extraction,[5] adds an even bigger level of complexity to the structural space of lignin. What constitutes lignin in principle is therefore only limited by how many features from the wider phenylpropanoid taxonomy (and beyond) that researchers are willing to include. A widely recognised addition is caffeyl alcohol (C), especially in the context of catechyl lignin.[10–13] However, many additions are only seen in single articles. For example, Thi et al. (2021) introduced new symbolic notation for reduced linkage variants, which were used to communicate identified oligomer sequences.[14]

While additions are being made, a nomenclature that can describe this ever-growing structural space of lignins is currently missing. However, that does not mean that no precedent exists. The International Union for Pure and Applied Chemistry (IUPAC) have published recommendations for the nomenclature and symbolic notation of many biopolymers. Examples include amino acids and peptides,[15] nucleic acids and polynucleotides,[16] and carbohydrates.[17] Surprisingly, in depth IUPAC recommendations for the related lignans and neolignans have also been available since 2000,[18] but not for lignin. Furthermore, initiatives such as the Hierarchical Editing Language for Macromolecules (HELM) highlights the need to denominate complex polymers at different structural levels,[19] which naturally follows from how polymers are analysed at different structural levels. This is also the case for lignins.

2.3 ANALYSIS OF LIGNINS

Numerous analytical methods have been used for the characterisation and quantification of lignins and lignin products, with each presenting with their own challenges[20]. Early methods were developed for the context of the paper and pulp industry, who needed to determine the total amounts of lignin in lignocellulosic biomasses. Two such quantities still in use today include Klason or acid soluble lignin. Despite remaining in use for over a century for the sake of comparability,[21] developments are still ongoing due inherent definitional uncertainty in total lignin as a measurand.[14,22] Comprehensive quantification of individual or classes of lignin monomers and (in particular) oligomers has instead been limited by separation technology. Developments have therefore centered around the introduction of new separation techniques, such as supercritical fluid chromatography (SFC)[23] or two-dimensional gas chromatography (GC)[14]. Others have looked even farther, and attempted to minimise the need for calibration standards through semiquantification[24] or even universal quantification approaches[25]. These attempts highlighted the need for more qualitative reference data, such as response factors. As such, while quantitative analysis is of more immediate significance to valorisation efforts, any major developments are still tied to advances in qualitative analysis and separation technology.

Recent research using size-exclusion chromatography (SEC), one of the more established techniques in polymer characterisation, illustrated the inherent challenges of qualitative lignin analysis through the scope of addressing bias in the molecular weight determination of lignins.[26,27] Process-modified lignins (such as depolymerised or technical lignins) have

been used as a challenging case to drive the development and application of new separation concepts. Sun et al. (2018) used a trapping column interface to couple reversed phase chromatography (RPLC) with SFC, and applied this method to rapidly separate low-molecular weight compounds in depolymerised lignin.[28] Musl et al. (2020) used hydrophobic interaction chromatography (HIC) to fractionate lignosulfonate, an amphiphilic technical lignin, and were able to achieve a high orthogonality in the offline combination with SEC.[29] Thi et al. (2022) used GCxGC with flame ionisation detection to quantify reductively depolymerised lignin, and were even able to elute lignin trimers at high temperatures.[14] Tammekivi et al. (2024) achieved a comprehensive separation and characterisation of a depolymerised lignin through an offline coupling of RPLC with SFC and high-resolution mass spectrometry (HRMS).[30]

As with other biological samples, HRMS has earned prominent role in the comprehensive, qualitative analysis of lignins.[31] In 2010, Morreel et al. pioneered important foundations, such as describing how lignin can be sequenced based on the characteristic fragmentation patterns of linkages.[32] They also introduced comprehensive lignin analysis, which they termed "lignomics".[6] The high polydispersity of native lignins makes their mass spectra difficult to interpret. However, Andrianova et al. (2018) used ion-mobility spectrometry to deconvolute and therefore confirm the presence of multiply charged species. They could also determine a well-defined correlation between mass and charge.[33] Many authors have since applied HRMS to native and technical lignin characterisation.[34–39] There have also been recent shifts towards studying oligomeric products in depolymerised lignins, to support the achievement of higher yields of monomeric products, or their use in value-added applications.[40,41] Because isomerism increases with degree of polymerisation, the concentrations of individual oligomers also decreases. Several authors have therefore studied improving ionisation efficiency, through better sample introduction,[42–44] or the use of dopants or additives.[45,46] Another recent development was introduced by Dong et al. (2023), where energy-resolved HRMS was demonstrated to be able to distinguish between different structural subunits. These may be particularly useful in de-novo annotation.[47] The aforementioned applications of more complex and multi-dimensional techniques has subsequently led to a growing amount and dimensionality of data. Because of this, there is an emerging need for high-throughput data processing and improved data annotation.[31]

2.4 DATA ANNOTATION AND CHEMINFORMATICS APPROACHES

Recent articles by authors such as Qi et al. (2020) and Sander et al. (2023) described the use of characteristic regions of Van Krevelen diagrams – determined with Fourier-transform ion cyclotron resonance (FT-ICR) MS – to annotate thousands of features from just a single sample as "lignin-like".[43,44] This illustrates two key developments. Firstly, researchers involved in comprehensive lignin analysis, or "lignomics", have turned to other -omics fields for data analysis approaches that can extract useful information from such large amounts

of data.[48] Secondly, in contrast to these fields, the annotation confidence levels[49,50] reached with nontargeted approaches are very low, here only producing a likeness-classification based on molecular formula. Prothmann et al. (2018) developed a Kendrick-mass defect based classification model for assigning degree of polymerisation,[34] and were able to use this to tentatively assign the degree of over 200 out of almost 400 features in a single technical lignin.[35] However, only 21 probable structures were assigned based on both MS^1 and MS^2 data, and 5 were confirmed by reference standard. Tammekivi et al. (2024) used an even simpler, ring-double bond equivalent basis for tentatively assigning the degree of polymerisation for almost 500 features from a depolymerised lignin sample.[30] Furthermore, they also used all 35 tentatively assigned structures from Prothmann et al. (2020) to investigate the coverage of popular databases such as PubChem, where only a single oligomer yielded an exact match. This illustrates how limited the size and scope of experimental lignin reference data currently is.

Computers have been used to simulate lignin polymer growth since the 1970s.[51] 40 years later, Yanez et al. (2016) reported the first in-silico libraries of lignins. These were generated stochastically from experimentally determined parameters.[52] They expanded the scope with another article in 2017.[53] This was eventually succeeded by two other groups, releasing the Lignin-KMC[54] and LigninGraphs[8] toolkits in 2019 and 2022, respectively. Terrell et al. (2020) drew from Yanez and Dellon et al. to generate in-silico libraries for the annotation of HRMS data.[55] They showed how 46 dimers could be combinatorially generated from just H, G, and S monomers, and β-O-4 linkages. Given the vast amount of possible structures, they argued for the use of previously developed stochastic approaches to generate managable but representative libraries. 100 lignin oligomers were therefore generated stochastically, followed by a heuristic expansion to 1200 structures, to improve coverage. Using this suspect list, they were able to assign tentative structures to 80 out of 478 detected features. Doran et al. (2021) proposed a comparable in-silico approach for the de-novo sequencing of linear, synthetic polymers.[56]

In this introduction, we have covered lignins incredibly rich biology and chemistry, its many frictions with both engineering and analysis, and the hopes of its many applications. It is clear that lignin research as a field is rapidly maturing. This thesis revisits some foundations left missing.

3 Commentary

3.1 SCOPE & EARLY AIMS

Typically a section like this would be called Results and Discussion, but this did not quite fit the intended scope. This section will not repeat content from included papers, but rather add further details, context, and metanarrative. This is similar to commentary sometimes added over movies. As such, reading the papers first is highly recommended to get the full story. Put a mark here and come back to this section later.

From the experience I have gathered across my PhD studies, it appears quite rare that PhD projects (or even individual papers for that matter) don't undergo fundamental changes. Equipment or experiments might fail, for whatever reason. Collaborations can fall apart. And in some cases, the aims might simply be too ambitious. This was the case here. Our (my) early aims for this PhD project were quite astonishing in hindsight:

1. The development of a standardised, inter-laboratory reproducible method for lignin monomer quantification, which will allow for confident comparisons of monomer yields of different lignin valorisation methods.
2. The use of different ionisation methods (ESI, APCI, APPI) in mass spectrometry analyses, which may be compared or even used together to improve characterisation through classification.
3. The novel application of IM-HRMS in lignin analysis, which will allow for the resolution of isobaric ions, and provide more information and higher interpretability per injection, and possibly even improve other performance characteristics. Especially for lignin oligomers.
4. The application of in-silico collision cross section (CCS) estimation, to raise identification confidence levels for compounds determined by IM-HRMS.
5. The further development of advanced data analysis methods, such as classification models, to improve interpretability, pre-selection, and identification confidence levels.
6. The development of oligomer quantification methods that minimises the burden of accessing chemical standards, such as the application of charged aerosol detection, the use of internal normalisation, or response factor prediction.

While aims 1 and 5 eventually produced Paper I and Paper II, respectively, all of these aims were explored to varying degrees. This highlights a major gap between work and output, which will be filled in through this commentary section.

3.2 QUALITY CONTROL IN INTERDISCIPLINARY STUDIES

During the planning for **Paper I** we came across the analytical work performed (mostly by chemical engineers) in various lignin-related articles, and realised that analytical norms can vary more than I had expected. Our collaboration would span multiple research fields and groups, three countries, and even the dreaded academia-industry divide. As such, one of our first roles as analytical chemists would be to establish agreed upon norms within the collaboration. A key quantity for **Paper I** was yield, and we realised across the course of several meetings that we did not even agree on how to best define this. While we eventually settled on a definition, this highlighted an essential lesson for collaborations: take nothing for granted.

For this reason, I set about writing a 13 page technical report. This was not published as a supplement to **Paper I**, but is included here in **Supplement S1.2**. The purpose of this report was to clearly define the bounds of the analytical method. This even included instructions on how to integrate chromatographic peaks, and accurately measure volumes and masses. This report highlights a major feature of the method which (for practical reasons) were not fully described in the published article: extensive quality control steps. This was expected to be necessary to achieve our strict targets of below 5% (as RSD) intralaboratory precision, and below 20% interlaboratory precision for monomeric yield. The main concern with respect to size of potential errors was expected to be the calibration, closely followed by the sample preparation. Contributions from i.e. peak integration or FID noise were expected to be very small by comparison. Given the reduced number of steps in the final sample preparation method, there was not much need for quality assurance (QA), other than e.g. the choice of internal standards/surrogates (IS), and prescribing certain times for vortexing. One compromise of note was the choice to recommend extracting the reaction liquors three times as opposed to just one, despite one extraction typically giving a near quantitative recovery. Our feeling was that this might improve robustness (unconfirmed), but also the samples would often need to be diluted anyway. However, the preparation of standards for calibration was a completely different story. Errors in the preparation of stock solutions could potentially add systematic errors, which might only become apparent during interlaboratory comparisons.

To avoid this, and to save time, we used a highly formulaic approach to sample preparation. This often made use of response factors estimated by single-point calibration. For example, IS peak areas should ideally be small enough to avoid interfering with analytes, but large enough to not be impacted by interferents. We aimed for a 1:1 ratio between analyte and IS. Analyte concentrations in a high-concentration sample would be estimated from the response factor estimates. These would then be used to calculate minimum concentrations for the IS mixture, followed by adjusting for IS interference (determined from unspiked sample). This formulaic approach was even more apparent during the calibration, where a formula was used to immediately find an appropriate calibration range. The samples were analysed first, and then the first calibration solution was prepared to be approximately twice

as concentrated as the largest peak for each analyte across the samples. The calibration solutions were then prepared through serial dilution. This ensured that all analytes would fall within the calibration range, as more standards at the lower end of the calibration range could always be added. A risk with serial dilution is that (potentially large) dilution errors in the first standard will carry over to all other standards. To address this, we kept response factor estimates for all analytes (determined in separate experiments), and used this to validate the concentration of the first calibration standard.

In addition to all of these quality assurance steps in the standard preparation, we also included quality control of the calibrations. Unfortunately, there is still no universally embraced approach to assessing the quality of calibrations. The squared correlation coefficient (R^2) is frequently used, despite it having little use beyond (re)establishing correlation. Residual plots are more informative (e.g. can actually assess linearity), but are difficult to apply consistently in routine analysis due to being visual. A compelling approach (which we ended up adopting) was the use of percent residual accuracy (%RA) by Logue and Manandhar (2018).[57] As described in full in the section 3.4.2 (Calibration Diagnostics) of **Supplement S1.2**, calibration models had to have relative residuals of less than 20% for all points and an average relative residual value below 5% (equivalent to a %RA above 95%) to pass. Calibration points would (if possible) be dropped, until reaching these targets. We frequently saw relative residuals of 100% or more at the extremes of the (often wide) calibration ranges, illustrating why this was necessary.

It quickly became obvious that all of these QA/QC calculations were a bit too much to manage. To improve ergonomics, but also to reduce the risk of calculation errors, a spreadsheet was distributed which automated and helped visualise these calculations. This was not published as a supplement to **Paper I**, but an example spreadsheet is included here in **Supplement S1.3**. A version of this was also used for the round-robin. Another important intervention was an excellent two week project-wide work exchange at Borregaard (one of the collaborators) in Norway. Meeting up allowed us to assess the method in person, and validate the analytical method across a range of samples. This led to a set of clarifications and improvements, which might not otherwise have been caught. We noticed poor repeatability when using non-inert liners, especially for vanillic acid, consistent with analyte adsorption and degradation in the injector. We also found that a sample volume of concentrated HCl was not sufficient to acidify the particularly basic reaction liquor, which allowed us to validate and approve H_2SO_4 as an alternative.

After all of these efforts, a severe deviation would still slip in at the very last step: the round-robin. While we were able to work around this, it did bring me back to the lesson that set all of this off. Compliance is expected, but not something we can take for granted. In the end, we are left to work with what we get.

3.3 MY FAIR SHARE OF INSTRUMENT ISSUES

As young researchers in analytical chemistry, we expect to be able to do exciting research using cutting-edge analytical instruments that can do essentially anything we want them to do. This was certainly the case for me with HRMS, which I (up until I got to work with it) believed to be a sort of analytical panacea. However, through a sort of right of passage, we are eventually forced to face the reality that most of our work in the lab will consist of troubleshooting. This was certainly the case for me too. Half of our early research aims (e.g. aims 2, 3, and 6 in Section 3.1) were instrument-driven. As such, many man-months of work went into acquiring, setting up, and troubleshooting instruments; some of which ended up playing no part whatsoever in the published papers.

As early as 2021, I managed to secure a research grant of 250 000 SEK from Kungliga Fysiografiska Sällskapet for the purchase of a Thermo Fisher Charged Aerosol Detector (CAD) Corona Veo RS. This acquisition was connected to aim 6 (Section 3.1), e.g. standard-free quantification of oligomers, since CADs have more uniform response factors than i.e. diode-array detectors (DADs). However, it soon became apparent through research by colleagues Papp et al. (2022) that achieving sufficient chromatographic separation was considerably more difficult with CADs than with DADs. This was difficult to the point of requiring peak-sharpening methods, even for just monomers and dimers.[25] While we had just acquired a new and improved CAD, the reality was that we were interested in more than dimers. As such, we quickly realised that these technical advances would not be enough.

At this point, we were also in the process of acquiring a Shimadzu Nexera UC system. This was a highly modular system, combining supercritical fluid extraction (SFE) with supercritical fluid chromatography (SFC) or high-performance liquid chromatograpy (HPLC). Given that this would be an MS-clean system, it was natural to include this in the project. As such I ended up taking on the role of instrument responsible for this system. The initial challenge of this system was the sheer complexity, given that it was built out of 13 modules. These included a CO_2 pump, two HPLC pumps and degassers, two backpressure regulators, and the SFE module along with a rack and a fraction collector. This did not include external detectors such as the CAD, or the Bruker timsTOF Pro 2 which we were also just about to acquire, and which made a brief feature in **Paper II** (with more details found in Section 3.4). I wanted to use this system for SFC, and to split the effluent between the CAD and the timsTOF. Our hope was that the ion mobility separation of the timsTOF would be able to provide the last bit of resolution needed to at least annotate the chromatographic peaks (including quantities) on a group level.

When we acquired the Nexera UC system, it was configured for online SFE-SFC, which (while interesting) was too difficult to bring into the mix. With a modular system like this, it was only natural that me and colleagues wanted to use this instrument in different configurations. Being the first research group in Sweden to attempt to use all of these modules together, there was little application support to receive. As such, we had to figure

this one out alone. While studying the instrument, I realised that there was no obvious way of shifting between different configurations. As such, every time a user would start their booked time on the instrument, they would have to spend a full day just going through every single connection. I therefore introduced the concept of standard hardware configurations (or "modes" for short). The intention behind introducing these were to allow for:

1. Strictly defined modes, where no other modes were permitted.
2. Step-by-step instructions for switching from one specific mode to another.
3. Minimising the number of disconnections when moving between the modes.

The first step when introducing these were to define which modes that were of interest. These ended up being SFE, LC-DAD, SFC-DAD, LC-E, and SFC-E, where E stands for an external detector such as the CAD or the timsTOF. These were then logged as a shorthand, to make identifying the current mode easy. Illustrations of all modes were made, such as the example shown in Figure 1, alongside other documentation. One major difference between the modes was the role of Pump C, or the second HPLC pump. In the LC modes, it was used as the cosolvent pump. In SFC-DAD, a second cosolvent pump. And in the other modes, it was used as a makeup pump. Few other disconnections were in fact necessary, which meant that moving between these modes was quite trivial in the end.

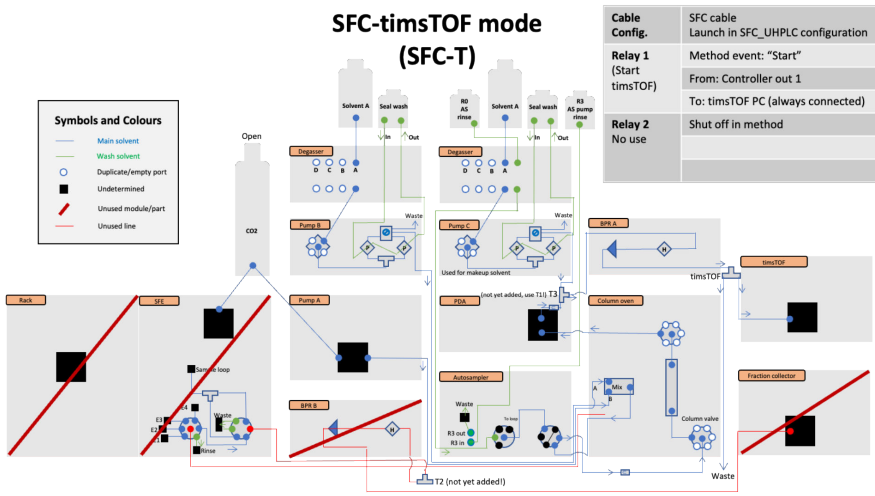


Figure 1: Illustration of a standard hardware configuration ("mode") for the Shimadzu Nexera UC system. The SFC-T mode combined SFC with the timsTOF, which involved adopting (HPLC) Pump C as a makeup pump.

With all that done, I was ready to move on to actually using the instrument. Except not, because the next years would be riddled with various issues, convoluted by us having to figure out our new timsTOF at the same time. These issues included (but were not limited to):

1. Leaking needles in two SFE positions.
2. Slow heating of SFE extraction vessels.
3. Misalignment of the SFE rack.
4. Leaking seals inside the CO_2 pump.
5. Pressure issues in SFC when changing between low and high amount of cosolvent.
6. Leaking (broken) flow cell in the DAD.
7. Systematic errors in injection volumes with partial loop injections in SFC.
8. Contamination issues with strange patterns, such as apparent seasonality.

The sheer number of simultaneous issues made the troubleshooting confusing and difficult. Fortunately, I was not alone. The contamination issue required an entire Masters thesis project to figure out, and was eventually published by Bajramova et al. (2026).[58] The injection volume issue was more pressing for me, since I needed small, concentrated sample bands to strike the right balance between detectability and column lifetime. To resolve the injection volume issue, I performed a deep dive into the (surprisingly complex) autosampler (AS). This had a module setting called the injection-port-to-loop (ITL) volume, which was used to ensure that the sample band was appropriately positioned relative to the sample loop. With our piping, the factory specified ITL was 6.1 µL. I hypothesised that the issue we observed was due to sample band displacement, caused by a smaller/larger than expected AS internal volume.

I therefore conducted an experiment where the ITL setting was varied in 1 µL steps for three different partial injection volumes (illustrated in Figure 2), and the peak areas were recorded for a 10% acetone sample. This showed *(i)* a maximum peak area at an ITL setting of 12 µL, twice the volume of the factory specification, and *(ii)* a band dispersion much broader than the 5 µL sample loop, independent of injection volume. The acquired peak areas could also be compared to the peak areas from full loop injection, which suggested that the maxima for all injection volumes were between 50-60% of what was expected. I would then explore various mitigation strategies, using an ITL setting of 9 µL as this was the point with the highest slope. This ensured that even small effects on the dispersion would have a large effect on the peak area. The air gap volume was varied, fittings were reseated, and the draw and dispensing volumes were reduced. None of these interventions had a significant effect on the dispersion. Given that the manufacturers technicians had even less experience than us with this system, my hope was that writing up and delivering these results to them would help them resolve the issue. Unfortunately, this was not the case.

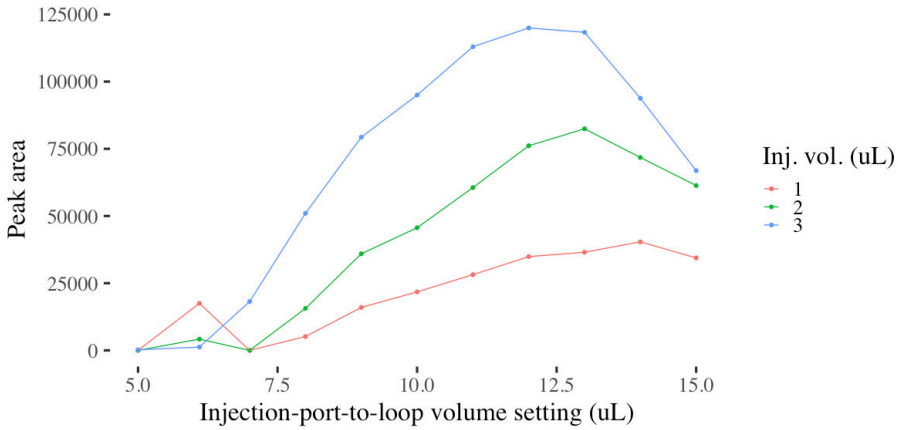


Figure 2: Peak areas for partial loop injections of 10% acetone sample across different injection-port-to-loop (ITL) volume settings. Peak area maxima were achieved at 12 µL, twice the factory specified volume. Areas were 50-60% relative to full loop injections. The broad curves suggest high dispersion.

Eventually I had spent at least six man-months across two years on troubleshooting this system, and failed to produce a single result. While a valuable learning experience, and one I do not regret, I eventually had to give up and move on to using the setup described in **Paper II**.

3.4 INTRODUCING LIGNIN SAMPLES TO A MASS SPECTROMETER

One of our early research aims (e.g. aim 2 in Section 3.1) involved sample introduction, or more specifically the use of different ionisation methods. While this had been done before by Kosyakov et al. as early as 2016,[42] we still had some interest in exploring this for ourselves. The reason for this was that previous research within the group, see Prothmann et al. (2018, 2020),[34,35] found a dramatic lack of pentamers and above. This was especially surprising since it was inconsistent with gel-permeation chromatography (GPC) experiments performed in our group on the very same samples by Papp et al. (2024).[27] We hypothesised this might be due to a decline in ionisation efficiency with increasing degree of polymerisation, as suggested by Andrianova et al. (2018). However, they could detect them after optimising ionisation conditions and applying charge deconvolution.[33] We were never able to reproduce these results with any of our mass spectrometers across multiple PhD projects. In the end, due to the lack of novelty, and because we had just purchased a Bruker timsTOF Pro 2, we did not end up exploring different ionisation methods.

My Favourite Negative Result: Dialysed Lignin for DI-timsTOF

Given the high amount of salts present in technical lignins (some of which are even salts themselves), large ionisation suppression is expected for direct infusion. The sodium lignosulfonate we used in i.e. **Paper I** contained a lot of sodium and other ions such as sulfate. If the salt is the problem, then removing it would help. We therefore attempted dialysis using a tube with a 500 Da cutoff, which would effectively remove excess salt, but keep at least most trimers and above.

This had no analytically significant effect on the sensitivity, regardless of if direction infusion or chromatography was used. However, it was still an interesting and typical negative result, highlighting how seemingly obvious and common-sense solutions might not work. Further studies into the ionisation dynamics of lignins is needed.

Despite dropping this line of investigation, we did consider possibilities for how we could make use of the timsTOF to resolve this sample introduction issue for larger oligomers. One alternate hypothesis to poor ionisation efficiency was that they failed to elute (with peak heights above the limit of detection). We had seen rapid filter and column clogging, which gave this hypothesis credibility. Direct infusion would avoid this, and was therefore appropriate for testing the hypothesis. On another mass spectrometer, we would not expect to see more than a charge convoluted mass distribution at best, as seen in e.g. Papp et al. (2024).[27] However, with the ion mobility separation of the timsTOF, such an experiment was worth investigating. Surprisingly, we did not see more than a few peaks in the mobilogram when injecting various lignin feedstocks, which went against our expectation of at least several tens or hundreds of oligomers. A third hypothesis that could explain the small number of peaks above the limit of detection is precisely this: that the diversity of oligomer structures means that the intensity of most individual oligomers are low. This would also be consistent with the growing number of possible masses with increasing degree of polymerisation (which we demonstrated in **Paper II**), and the lack of such an effect with GPC and non-MS detection, where the intensities of similar masses are combined. Most likely, all three of these hypotheses play some part in explaining the difficulties of lignin sample introduction. As such, it is a complex issue that still remains worthy of future investigation.

3.5 BRANCHING AND FUTURE OUTLOOK FOR LIGNIN NOMENCLATURES

The lignin oligomer taxonomy and nomenclature introduced with Lignonaut was orders of magnitude more extensive than what we have seen before, but in reality it only just scratched the surface. **Paper II** was not sufficient to cover all considerations for this nomenclature, and

particularly not the future outlook. This includes the introduction of new linkage notation, and considerations for branching.

The seven common lignin linkages have symbolic names which see frequent use. Two linkages of note are β-β and β-1, which actually consist of more than one bond between the monomeric units. For example, β-β also includes two γ-O-α linkages, which are ignored in the notation. However, such ignored linkages might become important to consider after chemical modification. For example, in a study on reductively fractionated pine lignin, Thi et al. (2022) identified two variants of β-β linkages. These have one or two unbound γ-hydroxyl groups, which they referred to as "β-β γ-OH", and "β-β 2x γ-OH", respectively.[14] In sequence notation, care must be taken to ensure clarity, especially when the linkage is not symmetrical. This is the case for "β-β γ-OH". The symmetrical "β-β 2x γ-OH" was in fact added to Lignonaut for **Paper II**, where it was given the symbol of [bOOb]. The capital letters indicate they are not oxygens (e.g. as in [bo4]). Instead, they stand for "open", e.g. the doubly open form of β-β. If added in the future, the two variants of "β-β γ-OH" could be referred to as [b.Ob] and [bO.b], respectively, where the point is arbitrarily added to help distinguish between the two.

When expanding the set of lignin linkage notations, it is also crucial to draw an appropriate line between monomer and linkage residues. This is highly arbitrary, but not particularly easy. For example, in the case of β-O-4 linkages, it is reasonable to allow the β-linked monomer unit to convey what the γ-substituent is. This is simpler than creating multiple β-O-4 variants, as was suggested for β-β above. Conversely, it is reasonable to consider the hydroxyl that is added at the α-carbon (e.g. the other side of the double bond) as a part of the β-O-4 residue. Given that this is taken for granted, modifying the symbol of the β-linked monomer would be redundant. In general, changing monomer symbols after linking should be avoided if the context of the sequence is sufficient to imply a particular structure. An exception to this arises with β-1 linkages, since the side-chain of 1-linked monomer is lost. This means that apparently different dimers, for example SYllol[β-1]GUllol and SYllol[β-1]GUllic, have the same structure. Therefore, the 1-linked monomer should be truncated (reflective of the chemistry) and the dimer written as SYllol[β-1]GU, to avoid confusion and synonyms.

The biggest unresolved challenge of lignin oligomer nomenclature is how to name branched oligomers. For lignin, branching can occur in two ways, depending on which residues that are involved. Either the branching occurs at a monomer node, or it occurs at a linkage node (see Figure 3). Each of these two cases have to be treated differently. Similar to SMILES, sequences with **branching at a monomer** can simply be represented with the help of a parenthesis, i.e. HYllol[5-5]HYllol([β-O-4]HYllol)[4-O-β]HYllol (for the structure illustrated in Figure 3a). Here, we have three synonymous sequences instead of two as with linear cases, as the [5-5]-linked unit can also be treated as the branch (e.g. placed in the parenthesis). Each branching nodes adds another such synonym.

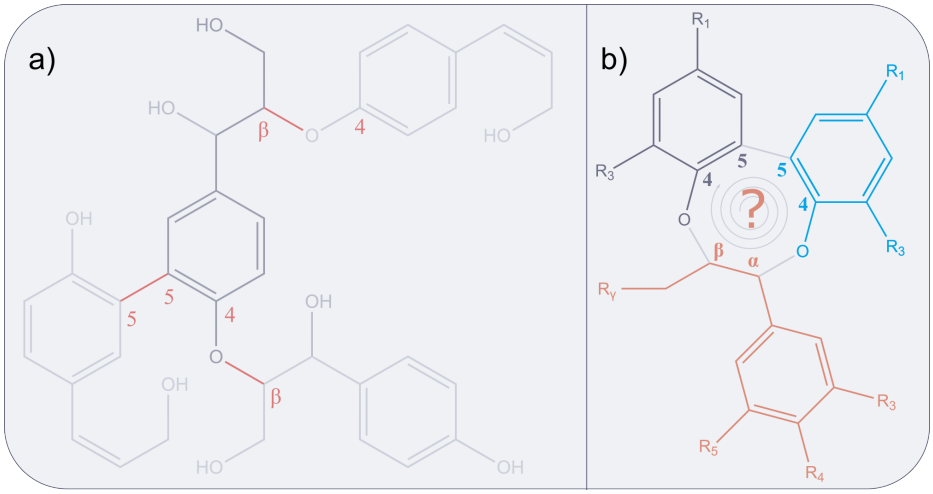


Figure 3: *(a)* Illustration of a tetramer branched at the central monomer node, which could be named HYllol[5-5]HYllol([β-O-4]HYllol)[4-O-β]HYllol. *(b)* Illustration of the dibenzodioxocin linkage, which forms a three-membered ring of monomer residues. Here, the linkage itself must be treated as the branching node.

Building linear oligomers is a straightforward task, in that only the reactivity of the end-groups and the reacting monomers need to be considered. For adding branches to a linear oligomer (hereafter just called "branching"), there are two further complications. The first complication is that the reactivity at all of the "inner units", e.g. all but the outermost two, now have to be considered. For each inner unit, four more columns of data, or 16 more bits per row, is required. All this brings a high memory cost to branching. There are potentially ways around this, such as reducing redundant information found at each $n - 1$ level, which is possible because there is a lot of copying going on. That is, a larger set of trimers is built from a smaller set of dimers, and so on. However, this will increase the computational cost instead, as it is more complex to vectorise the operations in this way. While it would make the code slower to run, it could still be worth looking into, since memory efficiency is more important as it limits library size.

The second complication is that there are a variety of ways that an oligomer can be branched, even when just considering branching at monomers. To find these, it is simplest to consider branching with respect to the longest linear backbone, similar to IUPAC nomenclature for branched, organic molecules. The smallest possible backbone that can be branched is a trimer. The only way to branch a trimer is by addition of one monomeric residue to the middle residue. If we add a dimeric branch to the trimer, the trimer is no longer the longest linear backbone, and we get an oligomer that could be equivalently described as a tetramer with a monomeric branch. To avoid this, we can derive a rule for singularly branched oligomers, where the maximum degree of a branch can be no longer than any of the other two "branches" on the backbone. In other words, the maximum degree of a branch is only determined by what node it is added to on the backbone, and how far away it is from the

end nodes. This pattern is extended up to a decamer backbone in Table 1. The maximum possible branching length is also equivalent to the number of possible branching lengths, since all shorter lengths are possible too. As such, Table 1 also illustrates how the number of possibilities grows with the length of the backbone. Given n monomers, the number of possible branchings with the trimer backbone is n, and for the tetramer backbone it is $2n + n^2$. For higher backbone lengths I am less sure of how to derive the formula (I am not a mathematician after all), but in any case this is sufficient to illustrate how branching at monomer nodes eventually (and quickly) reaches a computational limit. Given that we can form up to branched hexamers from a tetramer backbone, or up to branched nonamers from a pentamer backbone, it might not even be relevant to go beyond these. This is important to consider when developing nomenclature for this type of branched lignins, since it significantly reduces the difficulty of the task. Moving forward, other approaches (not starting from a linear backbone) should also be considered.

Table 1: The maximum (and number of) possible branching length(s) at different monomer nodes.

Backbone	**Node:**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Trimer		0	1	0							
Tetramer		0	1	1	0						
Pentamer		0	1	2	1	0					
Hexamer		0	1	2	2	1	0				
Heptamer		0	1	2	3	2	1	0			
Octamer		0	1	2	3	3	2	1	0		
Nonamer		0	1	2	3	4	3	2	1	0	
Decamer		0	1	2	3	4	4	3	2	1	0

The second type of branched lignins involve **branching at a linkage**, e.g. when the linkage acts as the branching node. At the time of writing, Lignonaut treats the reactive sites as belonging to the monomer units. Therefore, linkage units are currently not valid branching nodes. However, there are well-known special cases such as the dibenzodioxocin linkage (depicted in Figure 3b). This linkage involves three monomers, which form into a three-membered ring. This means that no one monomer can be considered as the branching node. For this reason it is necessary to consider linkages as potential branching nodes as well. Dibenzodioxocin linkages essentially consist of a combination of 5-5, α-O-4, and β-O-4 linkages. The difference from linear α-O-4 and β-O-4 linkages is that a 4-hydroxy has replaced the free hydroxy group on the β- or α-position, respectively. Representing the resulting oligomer sequence in a readable form is somewhat challenging. SMILES resolves rings by the use of numeric indices of where the ring has been "cut", that is placed to the right of each relevant atom. Perhaps something similar could be used here.

In conclusion, there are plenty of challenges left to solve, even for something as fundamental as nomenclature. Most likely, this will need to continue developing in a horizontal fashion, to ensure that the nomenclature remains in service of the lignin community.

3.6 TRACING NEUTRAL LOSSES WITH LIGNONAUT

Tracing viable neutral losses with Lignonaut is another exciting potential application. Mechanistically viable neutral losses can be traced back to the units by counting up the number of moieties that can produce each neutral loss. While these counts have a probabilistic effect on relative intensities of the corresponding fragments, the probabilities are not independent from other competing fragmentation pathways. As such, they are not reliable for predicting relative intensities, even for small compounds. However, they can be used to generate a list of fragment masses to look for during data analysis, and provide some aid in reducing candidate lists. They are also important for the determination of the molecular ion, as i.e. methyl loss is highly prevalent in lignin compounds.

One way of implementing this would be to include columns with neutral loss counts in the monomer database, based on e.g. Prothmann et al. (2017)[59]. This is quite straightforward, and was actually done at one point during the development of Lignonaut, according to Table 2. However, this was eventually postponed to restrict the scope of the first release. Neutral losses can also be traced to different linkages. Fragmentation patterns have been extensively characterised by various authors, along with proposed mechanisms.[6,60,61] Similar to the discussion in Section 3.5, connecting neutral losses to either monomeric or linkage residues is not always so straightforward. However, given enough engagement, and due to the usefulness of the application, I believe this will be one of the first challenges to be solved.

Table 2: Summary of which moieties are counted for each neutral loss column.

Column	Molecular ion	Fragment Explanation
loss.ch3	M - H^+ - CH_3^-	Methoxy and acetyl (ketone) moieties.
loss.cho	M - H^+ - CHO^-	Formyl (aldehyde) moieties.
loss.co2	M - H^+ - CO_2^-	Carboxyl moieties.
loss.h2o	M - H^+ - H_2O^-	sp^3 (alkane) bound hydroxide moieties.
loss.bc	M - H^+ - (R-CH_2)-	Side chain - CH_2 (benzyl cleavage).

3.7 PROGRAMMING AS AN ANALYTICAL CHEMIST

Outside of data analysis and visualisation, it seems rare for analytical chemists to engage much with programming. This is not very surprising, given that we receive little training in computer science. Despite this, I think few are better equipped for developing methods for i.e. cheminformatics or computational mass spectrometry. For this reason I think it is important to share stories of using programming in analytical chemistry, and of how it can be a worthwhile and relevant skill to develop. As has become a wide-spread compulsion of late, I too need to start out with a comment on generative AI (GAI). Vibe-coding (e.g. using GAI as a support in programming), much like other applications of GAI, blew up in the middle of my PhD project. In a way I am grateful it didn't happen earlier, since I otherwise

might not have gotten the opportunity to go as deep as I did. I eventually made a personal choice of abstaining from all use of GAI. This was due to my learning-oriented disposition, combined with concerns over novelty-induced ignorance (of costs) and perverse incentives. So in this section I would like to give a peek into what I learned, but also convey a sense of the joy of discovery that can be found through taking the time to do something well.

Lignonaut (featured in **Paper II**) was entirely written in R, which is a high-level (e.g. quite accessible) language used for computational programming. It is especially popular with statisticians, but also bioinformaticians, and (more importantly) -omics researchers and cheminformaticians. Computational programming is essentially just like an advanced graphing calculator. It is not used to write software, but to (reproducibly) perform calculations, data analysis, and data visualisation. As such, the choice of writing Lignonaut in R would be considered rather strange by those in the know. But I wanted a low barrier to contribution from other researchers. I also enjoy the language, especially in comparison to more natural choices such as Python. As I came to learn, R also has some compelling characteristics that make it quite well suited towards cheminformatic software. Because of how R was designed for working with data, most functions in R are vectorised. This means that they can operate on vectors as a whole, as opposed to looping over the individual elements. This not only makes the code more clean, but also potentially a lot more performant. As described in **Paper II**, one of the key advances of Lignonaut over previous approaches was performance. When optimising the various functions in Lignonaut, I would often find myself coming back to making good use of vectorisation. What follows in this section is a recollection of a chain of events from the process of optimising how the deduplication algorithm was implemented. I think that the stumbling blocks that came up offered some interesting insights into how R works.

Step 1: Restructure the data
Early on in the development of Lignonaut, the namestrings of the oligomer sequences would be built immediately, and stored as strings (e.g. text) in a single column. This created a performance bottleneck for an important function in Lignonaut that is used to identify mirror duplicates in the generated libraries. To achieve this, the function had to take the column vector with oligomer sequences and (element by element) split it up into a matrix of residues (e.g. multiple columns), so that the reverse sequence could be identified. To avoid these costly, element-wise string operations, the residues (monomers and linkages) would instead be kept in separate columns, and the full namestrings for the oligomer sequences would only be generated at the very end.

Step 2: Find a faster sorting algorithm
Next, to create the same identifier for both mirror duplicates, a sorting step would also be necessary. This was incredibly slow, since elements had to be individually moved between the columns of a matrix, which the general `sort()` function in R does not perform efficiently. Sorting operations can never be truly vectorised, so I looked towards using a more optimised (e.g. specialised) function such as `rowSort()` from the Rfast package. However, this was

not possible since it doesn't support character matrices (e.g. one filled with strings instead of numbers). The potential gains in speed here (several orders of magnitude) was enough to justify encoding all of the residue strings as integers instead. So I did.

Step 3: Encode the strings as integers

Encoding the residue strings as integers simply meant assigning them a number. For example, HYllol or paracoumaryl alcohol could be encoded as the integer 1, GUllol or coniferyl alcohol as 2, and so on. Translating back and forth between these two simply requires a dictionary (a type that stores pairs of keys and values). However, R does not have a dictionary type. Fortunately, a pair of vectors could achieve the same result, and would arguably even be more efficient. To translate from integer to string for the above example, the vector v = (HYllol, GUllol, …) would be enough. Why? Because vectors are an ordered type. This means that requesting the element at position 1 (by calling v[1]) would return HYllol, and so on. This achieves exactly what I needed. To translate from string to integer, a named vector of the integers would instead be used. A nice aspect of this solution is not just that it is fast, but that the resulting syntax is very reminiscent of a function call. All of this worked very well, and now the Rfast::rowSort() function could finally be used.

Step 4: Concatenate… the integers?

However, encoding everything as integers created a new issue to resolve, which was my favourite programming rabbit hole of all. I needed a single, unique identifier for the deduplication function. Before I had strings representing the residues, which could be concatenated (e.g. glued together). But now I had integers, which can't be concatenated. Or can they? I could of course just convert the integers to strings (e.g. the integer 1 to the string "1") and then concatenate them, but this didn't feel quite right. So as one would do before the advent of GAI chatbots, I desperately scoured Stack Exchange for tangentially related conversations in the hopes of finding some solution. And as often happens, I was able to find exactly the tool that I needed, which was a sort of Horner scheme (see Equation 1). This is a recursive algorithm which manages to "concatenate" integers through mathematics alone. E.g., it would take the keys 22 and 333, and return 22333. This discovery helped me avoid working with strings altogether. And best of all, it could be vectorised! In this algorithm, the key column vectors (k) are recursively concatenated by multiplying the previous key column (k_{n-1}) by the order of magnitude of ($10^{\lfloor 1+\log_{10}(k_n)\rfloor}$), and then adding, the next key column (k_n).

$$\begin{aligned} C_1 &= k_1 \\ C_n &= k_{n-1} \cdot 10^{\lfloor 1+\log_{10}(k_n)\rfloor} + k_n \end{aligned} \tag{1}$$

Step 5: Perform

Integer concatenation was the final piece that was missing. After all of this work, and other similar optimisation tangents, Lignonaut ended up becoming the incredibly performant piece of software that it is. To me, all of this was important. I personally could never be bothered with using a lot of the free open-source software for analytical chemistry out there,

due to them often being incredibly slow or even buggy (not that commercial software is that much better though). And because the virtual combinatorial synthesis approach used in Lignonaut was expected to push computers to their limits, cutting corners for the sake of reaching a minimum viable product faster was never really on the table.

In conclusion. While developing software as an analytical chemist can produce manuscripts that are publishable in relevant journals, as was the case with **Paper II**, I think there is more for us to find here. Given how much time researchers in analytical chemistry spend on data analysis, participation in the development of software gives us the opportunity to directly steer the direction of the tools that we rely on. And if nothing else, that feels great.

3.8 LESSONS FROM LECTURING AS A PHD STUDENT

Given that teaching is a primary task of Swedish universities, it is not surprising that it is also (more or less) a requirement for the PhD degree. PhD students in analytical chemistry typically act as teaching assistants for lab courses. Holding course lectures might be fairly normal in some fields, universities, or countries. But with us, it is rare. So I was very privileged to be asked to cover for a lecturer in my fifth and last year. Here I would like to tell the story of how I ended up there, and reflect on what I learned from it.

In hindsight, I always found teaching the lab courses to be a highlight of the year. It offered me a nice break from my research, provided a structure not otherwise found in academic work, and getting to meet new and returning students was always a treat. If I was told that this would have been the case when I started, I wouldn't have believed it. I taught my first lab course just a few month after starting my PhD in 2021. Just before then, I remember having a phone appointment with my work therapist, terrified at the idea of having to interact with a dozen students every day for weeks. I went into crisis management mode, and cancelled all other commitments, including telling loved ones that they might not hear from me for weeks. While I prepared for burning out, I instead found that I loved it. So in the years after, I would spend a lot of my personal time on developing lab manuals, report scoring rubrics, and even a writing an entire guide on how to write excellent lab reports. I also held a lecture in presentation skills between 2023-2025. In the midst of all of this, I never truly stopped to consider my development as a teacher, until I was eventually awarded with "Årets Labbhandledare 2024" (Teaching Assistant of 2024) from Kemi- och Bioteknikssektionen (the Guild of Chemical Engineering and Biotechnology) for my efforts in improving one of the course labs.

Authentic Learning

A multi-dimensional, constructivist approach emphasising the authenticity (e.g. relevance) of learning content, activities, and assessment. Schriebl et al. (2023) developed a model which deconstructs authenticity into personal (e.g. anything), real-world (e.g. professional), and disciplinary (e.g. academic) authenticity.[62]

So when I was asked to cover for the basic course lectures in chromatography, I was happy to do so within the scope of my PhD assuming I would be able to make a valuable learning experience out of it. I would first start with a 3 ECTS module studying education in chromatography, as a part of our PhD course in analytical chemistry. This gave me a good overview of modern approaches, which included the implementation of digital tools. I was also given a generous amount of time to prepare for teaching in the basic course, which involved preparing six hours of lecture material from scratch, along with two new exercises. While daunting, I had a very clear vision of what I wanted to do with these even before starting out. As was hinted at previously, I have personal experiences with dealing with disability (including as a student). Or to be more precise, I have personal experiences with being disabled by how teaching materials in general are built for a neurology different from my own. Because of this, I felt empowered by the opportunity to finally create the enabling teaching materials that I always needed as a student. More formally, I knew I wanted to create materials with a radical emphasis on authentic learning, but also drawing from cognitive load theory.

The Surprising Connection Between Falafel and HPLC

Developing contextualised intuitions is useful to avoid getting caught up in minor details. This student found one familiar to all of us in Lund.

Student: So what is the deal with high-performance liquid chromatography (HPLC) then, since we also have ultra-HPLC?

Me: HPLC used to be high-performance, but it's the new normal these days.

Student: Aah, so it's like Lundafalafel! You know how they only sell the falafels in large and extra-large?

Me: That's brilliant!

From teaching chromatography in the lab course, I knew that many students seemed to have (been) rushed past even the most basic fundamentals. I realised that my more pragmatic interest in chromatography would help here, as it would be easy to identify and emphasise content with high real-world authenticity. From my experiences as a student, I had also identified a common fallacy among lecturers of conflating horizontal development with

"depth", while neglecting vertical depth of content. In other words, a didactic quantity over quality. To address this, I chose to shift emphasis away from what I considered to be tangential topics with limited real-world authenticity, to make space for introducing more vertical layers to each topic. To achieve this I aimed for a scaffolded sequence, starting with personally contextualised intuition, followed by relational thinking (supported by models), and finally real-world authentic problem-solving (involving higher-order thinking). In fact, this was pretty much the exact structure I chose for the lecture series:

Lecture 1: A broad but superficial **introduction** to chromatography, emphasising accessibility and intuition over scientific rigor. The main objective here was to help with forming mental-organizational "boxes", in which the rest of the course could be sorted.

Lecture 2: A deepdive into chromatographic **dynamics**, emphasising models (rooted in familiar physical chemistry) for explaining and predicting chromatographic phenomena.

Lecture 3: A full lecture covering chromatographic **instrumentation**, all the way from syringe to detector. This would also introduce technical limitations such as backpressure, of high importance for understanding both instrument design and method development.

Lecture 4: A final lecture introducing chromatographic **method development** and data interpretation, to move from disciplinary to real-world authenticity. This is what we actually work with, after all.

Feedback on Scaffolding

"Mynta, she actually helped us with learning, instead of just recounting facts. She understands the importance of reviewing the basics and not getting lost in complicated tangents that leave us confused."

Another application of this scaffolded sequence was in the design of the exercises. We had been struggling with a low attendance rate during the exercise sessions for several years. While this was beyond just the chromatography module, I hypothesised that this partly was caused by the exercises being too focused on preparing the students for the exam. As such, students would only use them for preparing for the exam. Since the exercise sessions were fairly close to (and sometimes directly after) the lectures, I thought that it would be more fitting to focus on developing intuition and relational thinking. For this I would prepare a series of discussion questions rehearsing the lecture content, but also interspersed with head-scratchers such as:

- Is a high retention factor always good?

- Is it possible for an analyte to elute in GC if the column temperature is below its boiling point?
- A student attempted separating fatty acids on a C18 column, with buffer (pH $<$ 4) and heptane for the mobile phase. The chromatograms looked strange. Why was heptane a bad pick, despite its high elution strength?

From spending time with students in the lab, I knew that these are the kinds of questions that they actually needed to overcome in order to develop their confidence in chemical analysis. I even used a version of the last question above in the exam, where the students had to find more of this kind of severe technical mistakes in the chromatographic method hallucinated by a GAI chatbot. While discussion questions are all well and good to get students started, I wanted to offer something more. During the literature study I previously performed on education in chromatography, I ended up playing around with various web-based HPLC simulators for method development. These are highly useful as a learning tool for developing relational thinking. E.g., the student can see in real time how adjusting the amount of cosolvent affects retention. Another major benefit is that they provide immediate feedback (and gratification), in contrast to real instruments. For these reasons, I ended up developing a second exercise using the simulator hosted at www.multidlc.org. This appeared to be appreciated by students working alone, but also by students with ADHD, possibly due to the immediate gratification and increased perceptual load.[63]

Cognitive Load Theory

An instructional approach based in work by Sweller (1988), emphasising how limitations imposed by working memory impact learning, and that the cognitive load imposed by learning activities needs to be managed.[64]

With the prevalence of ADHD within the general population growing and approaching 10% according to some estimates,[65] it is now potentially one of the most disproportionally underserved student groups. One of the key characteristics of ADHD is an increased vulnerability to extraneous distraction.[63] Cognitive load theory is therefore a well suited framework for designing learning materials for this group, since it outright recommends the reduction of extraneous cognitive load for improved learning outcomes. I hypothesised that lectures most likely would be the most sensitive to these sorts of interventions. I therefore set out to limit extraneous cognitive load by reducing the amount of redundant text and visual elements (also to the benefit of dyslexics), and by making use of modern visual language (i.e. web design). While intended for the design of statistical graphics, the minimalist design principles by the famous Edward R. Tufte were highly useful here.[66] However, Forster et al. (2014) demonstrated reduced distractability in the presence of increased perceptive load, for participants with (and without) ADHD.[63] As such, stripping away too many elements might have a counterproductive effect. Fortunately, I was in need of a creative outlet at the

time. So I also ended up including a lot of eye-candy, such as custom designs, animations, and the consistent use of a palette that strikes a good balance between high-contrast, bold colors, and an overall soothing and coherent impression (Catppuccin Latté). An example of this can be seen in Figure 4. These efforts appeared to have been well received, as one student in the course evaluation explicitly wrote "[…] some lectures were well adapted for students with learning disabilities (Myntas lectures)".

During the COVID-19 pandemic, many lecturers turned to video recordings to fill the void left by stay-at-home notices, which would remain available even after the pandemic. From interacting with this (post-)COVID generation of students during the lab courses, I learned that many still enjoyed having access to these videos, despite them rapidly falling out of sync with the in-person lectures. So I committed myself to recording and uploading high-quality videos of all lectures in advance (see Figure 4). At the end of the first lecture, I pulled up the video of the lecture they had just seen, just to remind them that they didn't actually have to attend the lectures in person. While I'm sure this would outrage many lecturers, this simply connects back to authentic learning. I wanted students to attend the lectures because they found value in them, and not out of a fear of missing out. Waking up early, commuting, and cramming into a lecture hall with a hundred students comes at a great cost to some. I know because I was one of them when I was a student. While all this was a bit of a gambit, most students did in fact return for the next lecture.

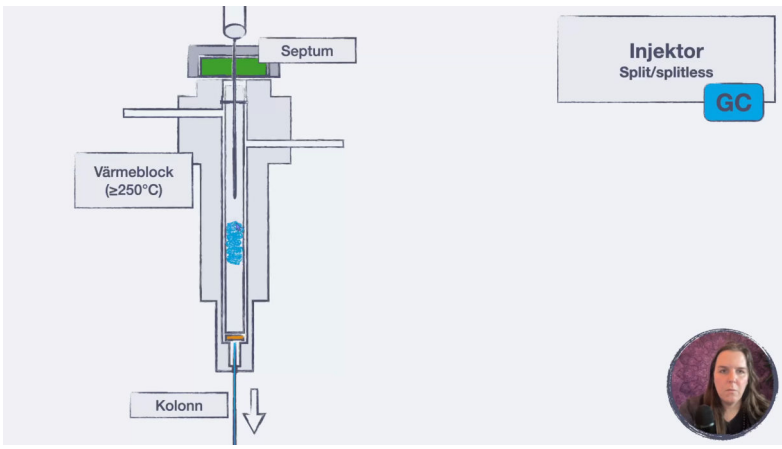


Figure 4: Drawing from my (limited) experience with game recording and streaming, I used OBS Studio to record high resolution and bitrate videos of the lectures. These are still available at https://www.youtube.com/playlist?list=PLHrbMWOAfulBlC3DdtDNiGF2cytygBJaV

Feedback on recorded lectures

"The recorded lectures were valuable to me. I couldn't attend some lectures, and I find it hard to learn by just learning from the book. I also appreciated that Mynta made recorded lectures too, even if she didn't have to."

While my intention was to better meet the needs of underserved groups, perhaps at the justified expense of the needs of the majority, the course evaluation would make it clear that this was a false dichotomy. The chromatography module serendipitously received universal praise from the students. Those airy slides that I created to minimise cognitive load? Students loved them because they were great for taking notes on with their tablets. All of those colorful designs intended for increasing perceptual load? Many simply appreciated them from an aesthetic point of view, with a student finding it "[…] very refreshing that she had made her own, new, modern slides". The recordings that I made for students not able to attend the lectures in person? Nearly everyone used them during the exercises, and while practicing for the exam. And even more importantly, pass rates for the course were significantly above average.

I think this might be the most profound insight of this whole endeavour. If proactive disability supports benefit everyone, then this suggests that disabled students play the role of canaries in a coal mine. And if that is truly so, then working towards their success might just be one of the most valuable pursuits we can undertake as teachers.

4 Final Notes

While much has already been said, some sort of happy ending to tie everything together would be nice. While intended to some degree, I think this thesis very much ended up mirroring my entire PhD project. I expected it to include mostly research, but it ended up including mostly instrument troubleshooting, programming, and teaching. I expected it to be incredibly stressful, but in the end it turned out to be quite chill. I didn't expect it to be so personal and informal, although I am not particularly surprised about it either given my disposition. I expected it to end up being a whole lot shorter than it was, and not that the time would just fly by. And at first I was incredibly excited to get going, but now all I want is to be done with it.

With that said, what has transpired here will stick with me for a long time. I was drawn to the position out of a love for learning and research, and found new passions in both programming and teaching. Lignonaut is now published on a code repository. While I will not continue active work on it, I will stay on as lead developer, given enough interest by the community. This includes overseeing, counseling, and approving new fixes and features. In the near future, I hope to be able to find some time for learning a low-level language as well. While analytical chemists can end up in a wide variety of positions, I feel like even more doors have been opened by this incredibly diverse PhD project. Only time will tell what I end up doing after this.

Very few actually take their time to read PhD theses. You are one of probably less than ten people who will ever read this. I wanted this to be a personal document of these last five years, but also something of use to you. I hope you enjoyed it!

5 References

1 R. Vanholme, B. Demedts, K. Morreel, J. Ralph and W. Boerjan, *Plant Physiology*, 2010, **153**, 895–905.

2 N. Zhou, W. P. D. W. Thilakarathna, Q. S. He and H. P. V. Rupasinghe, *Frontiers in Energy Research*.

3 J. Wenger, V. Haas and T. Stern, *Current Forestry Reports*, 2020, **6**, 294–308.

4 M. Balk, P. Sofia, A. T. Neffe and N. Tirelli, *International Journal of Molecular Sciences*, 2023, **24**, 11668.

5 J. Banoub, G.-H. Delmas, N. Joly, G. Mackenzie, N. Cachet, B. Benjelloun-Mlayah and M. Delmas, *Journal of Mass Spectrometry*, 2015, **50**, 5–48.

6 K. Morreel, O. Dima, H. Kim, F. Lu, C. Niculaes, R. Vanholme, R. Dauwe, G. Goeminne, D. Inzé, E. Messens, J. Ralph and W. Boerjan, *Plant Physiology*, 2010, **153**, 1464–1478.

7 S. c. d. Eswaran, S. Subramaniam, U. Sanyal, R. Rallo and X. Zhang, *Scientific Data*, 2022, **9**, 647.

8 Y. Wang, J. Kalscheur, E. Ebikade, Q. Li and D. G. Vlachos, *Journal of Cheminformatics*, 2022, **14**, 43.

9 J. C. del Río, J. Rencoret, A. Gutiérrez, T. Elder, H. Kim and J. Ralph, *ACS Sustainable Chemistry & Engineering*, 2020, **8**, 4997–5012.

10 F. Chen, Y. Tobimatsu, D. Havkin-Frenkel, R. A. Dixon and J. Ralph, *Proceedings of the National Academy of Sciences*, 2012, **109**, 1772–1777.

11 Y. Tobimatsu, F. Chen, J. Nakashima, L. L. Escamilla-Treviño, L. Jackson, R. A. Dixon and J. Ralph, *The Plant Cell*, 2013, **25**, 2587–2600.

12 D. Ando, F. Lu, H. Kim, A. Eugene, Y. Tobimatsu, R. Vanholme, T. J. Elder, W. Boerjan and J. Ralph, *Green Chemistry*, 2021, **23**, 8995–9013.

13 Y. Li, L. Shuai, H. Kim, A. H. Motagamwala, J. K. Mobley, F. Yue, Y. Tobimatsu, D. Havkin-Frenkel, F. Chen, R. A. Dixon, J. S. Luterbacher, J. A. Dumesic and J. Ralph, *Science Advances*, 2018, **4**, eaau2968.

14 H. D. Thi, K. V. Aelst, S. V. d. Bosch, R. Katahira, G. T. Beckham, B. F. Sels and K. M. V. Geem, *Green Chemistry*, 2022, **24**, 191–206.

15 *Biochemical Journal*, 1984, **219**, 345–373.

16 *Biochemical Journal*, 1970, **120**, 449–454.

17 A. D. McNaught, *Carbohydrate Research*, 1997, **297**, 1–92.

18 G. P. Moss, *Pure and Applied Chemistry*, 2000, **72**, 1493–1523.

19 T. Zhang, H. Li, H. Xi, R. V. Stanton and S. H. Rotstein, *Journal of Chemical Information and Modeling*, 2012, **52**, 2796–2806.

20 R. C. Rodrigues, B. Green Rodrigues, E. Vieira Canettieri, E. Acosta Martinez, F. Palladino, A. Wisniewski Jr and D. Rodrigues Jr, *Bioresource Technology*, 2022, **348**, 126627.

21 R. Hatfield and R. S. Fukushima, *Crop Science*, 2005, **45**, 832–839.

22 D. Díez, A. Urueña, R. Piñero, A. Barrio and T. Tamminen, *Processes*, 2020, **8**, 1048.

23 M. Sun, G. Lidén, M. Sandahl and C. Turner, *Journal of Separation Science*, 2016, **39**, 3123–3129.

24 E. Bartolomei, Y. Le Brech, A. Dufour, V. Carre, F. Aubriet, E. Terrell, M. Garcia-Perez and P. Arnoux, *ChemSusChem*, 2020, **13**, 4633–4648.

25 D. Papp, T. Rukkijakan, D. Lebedeva, T. Nylander, M. Sandahl, J. S. M. Samec and C. Turner, *Analytical Chemistry*, 2022, acs.analchem.2c04383.

26 A. A. Andrianova, N. A. Yeudakimenka, S. L. Lilak, E. I. Kozliak, A. Ugrinov, M. P. Sibi and A. Kubátová, *Journal of Chromatography A*, 2018, **1534**, 101–110.

27 D. Papp, G. Carlström, T. Nylander, M. Sandahl and C. Turner, *Analytical Chemistry*, 2024, **96**, 10612–10619.

28 M. Sun, M. Sandahl and C. Turner, *Journal of Chromatography A*, 2018, **1541**, 21–30.

29 O. Musl, I. Sulaeva, M. Bacher, A. K. Mahler, T. Rosenau and A. Potthast, *ChemSusChem*, 2020, **13**, 4595–4604.

30 E. Tammekivi, M. Batteau, D. Laurenti, H. Lilti and K. Faure, *Analytica Chimica Acta*, 2024, **1288**, 342157.

31 D. R. Letourneau and D. A. Volmer, *Mass Spectrometry Reviews*, 2021, 1–45.

32 K. Morreel, H. Kim, F. Lu, O. Dima, T. Akiyama, R. Vanholme, C. Niculaes, G. Goeminne, D. Inzé, E. Messens, J. Ralph and W. Boerjan, *Analytical Chemistry*, 2010, **82**, 8095–8105.

33 A. A. Andrianova, T. DiProspero, C. Geib, I. P. Smoliakova, E. I. Kozliak and A. Kubátová, *Journal of the American Society for Mass Spectrometry*, 2018, **29**, 1044–1059.

34 J. Prothmann, P. Spégel, M. Sandahl and C. Turner, *Analytical and Bioanalytical Chemistry*, 2018, **410**, 7803–7814.

35 J. Prothmann, K. Li, C. Hulteberg, P. Spégel, M. Sandahl and C. Turner, *ChemSusChem*, 2020, **13**, 4605–4612.

36 K. Li, J. Prothmann, M. Sandahl, S. Blomberg, C. Turner and C. Hulteberg, *Molecules*, 2021, **26**, 2887.

37 A. Mikhael, T. D. Fridgen, M. Delmas and J. Banoub, *Journal of Mass Spectrometry*, 2020, **56**, e4676.

38 A. Mikhael, T. D. Fridgen, M. Delmas and J. Banoub, *Rapid Communications in Mass Spectrometry*, 2020, **34**, e8910.

39 J. Zhang, Y. Jiang, L. F. Easterling, A. Anstner, W. Li, K. Z. Alzarieni, X. Dong, J. Bozell and H. I. Kenttämaa, *Green Chemistry*, 2021, **23**, 983–1000.

40 J. D. Guthrie, C. E. R. Rowell, R. O. Anyaeche, K. Z. Alzarieni and H. I. Kenttämaa, *Mass Spectrometry Reviews*, 2023, e21832.

41 Y. Han, B. A. Simmons and S. Singh, *Industrial Chemistry & Materials*, 2023, **1**, 207–223.

42 D. S. Kosyakov, N. V. Ul'yanovskii, E. A. Anikeenko and N. S. Gorbova, *Rapid Communications in Mass Spectrometry*, 2016, **30**, 2099–2108.

43 Y. Qi, P. Fu, S. Li, C. Ma, C. Liu and D. A. Volmer, *Science of The Total Environment*, 2020, **713**, 136573.

44 K. Sander, L. Dütsch, M. Bremer, S. Fischer, C. Vogt and J. Zuber, *Energy & Fuels*, 2023, **37**, 439–449.

45 D. S. Kosyakov, I. I. Pikovskoi and N. V. Ul'yanovskii, *Analytica Chimica Acta*, 2021, **1179**, 338836.

46 W.-Y. Song, H. Park and T.-Y. Kim, *Journal of Chromatography A*, 2022, **1685**, 463598.

47 X. Dong, H. B. Mayes, K. Morreel, R. Katahira, Y. Li, J. Ralph, B. A. Black and G. T. Beckham, *ChemSusChem*, 2023, **16**, e202201441.

48 R. Zhang, Y. Qi, C. Ma, J. Ge, Q. Hu, F.-J. Yue, S.-L. Li and D. A. Volmer, *Molecules*, 2021, **26**, 178.

49 E. L. Schymanski, J. Jeon, R. Gulde, K. Fenner, M. Ruff, H. P. Singer and J. Hollender, *Environmental Science & Technology*, 2014, **48**, 2097–2098.

50 A. C. Schrimpe-Rutledge, S. G. Codreanu, S. D. Sherrod and J. A. McLean, *Journal of the American Society for Mass Spectrometry*, 2016, **27**, 1897–1905.

51 W. G. Glasser and H. R. Glasser, *Macromolecules*, 1974, **7**, 17–27.

52 A. J. Yanez, W. Li, R. Mabon and L. J. Broadbelt, *Energy & Fuels*, 2016, **30**, 5835–5845.

53 L. D. Dellon, A. J. Yanez, W. Li, R. Mabon and L. J. Broadbelt, *Energy & Fuels*, 2017, **31**, 8263–8274.

54 M. J. Orella, T. Z. H. Gani, J. V. Vermaas, M. L. Stone, E. M. Anderson, G. T. Beckham, F. R. Brushett and Y. Román-Leshkov, *ACS Sustainable Chemistry & Engineering*, 2019, **7**, 18313–18322.

55 E. Terrell, V. Carré, A. Dufour, F. Aubriet, Y. Le Brech and M. Garcia-Pérez, *ChemSusChem*, 2020, **13**, 4428–4445.

56 D. Doran, E. Clarke, G. Keenan, E. Carrick, C. Mathis and L. Cronin, *Cell Reports Physical Science*, 2021, **2**, 100685.

57 B. A. Logue and E. Manandhar, *Talanta*, 2018, **189**, 527–533.

58 A. Bajramova, E. Synefakis, J. Larsen, J. Lindegaard Hjorth, C. Turner and M. Sandahl, Uncovering the origin of impurities in supercritical fluid chromatography coupled to high-resolution mass spectrometry, https://www.ssrn.com/abstract=6215043, (accessed April 1, 2026).

59 J. Prothmann, M. Sun, P. Spégel, M. Sandahl and C. Turner, *Analytical and Bioanalytical Chemistry*, 2017, **409**, 7049–7061.

60 S. O. Asare, P. Kamali, F. Huang and B. C. Lynn, *Energy & Fuels*, 2018, **32**, 5990–5998.

61 J. Zhang, E. Feng, W. Li, H. Sheng, J. R. Milton, L. F. Easterling, J. J. Nash and H. I. Kenttämaa, *Analytical Chemistry*, 2020, **92**, 11895–11903.

62 D. Schriebl, A. Müller and N. Robin, *Science & Education*, 2023, **32**, 1021–1048.

63 S. Forster, D. J. Robertson, A. Jennings, P. Asherson and N. Lavie, *Neuropsychology*, 2014, **28**, 91–97.

64 J. Sweller, *Cognitive Science*, 1988, **12**, 257–285.

65 R. Thomas, S. Sanders, J. Doust, E. Beller and P. Glasziou, *Pediatrics*, 2015, **135**, e994–e1001.

66 E. R. Tufte, *The visual display of quantitative information*, Graphics Press, 2nd edn., 1986.

Scientific Publications

Author contributions

PAPER I

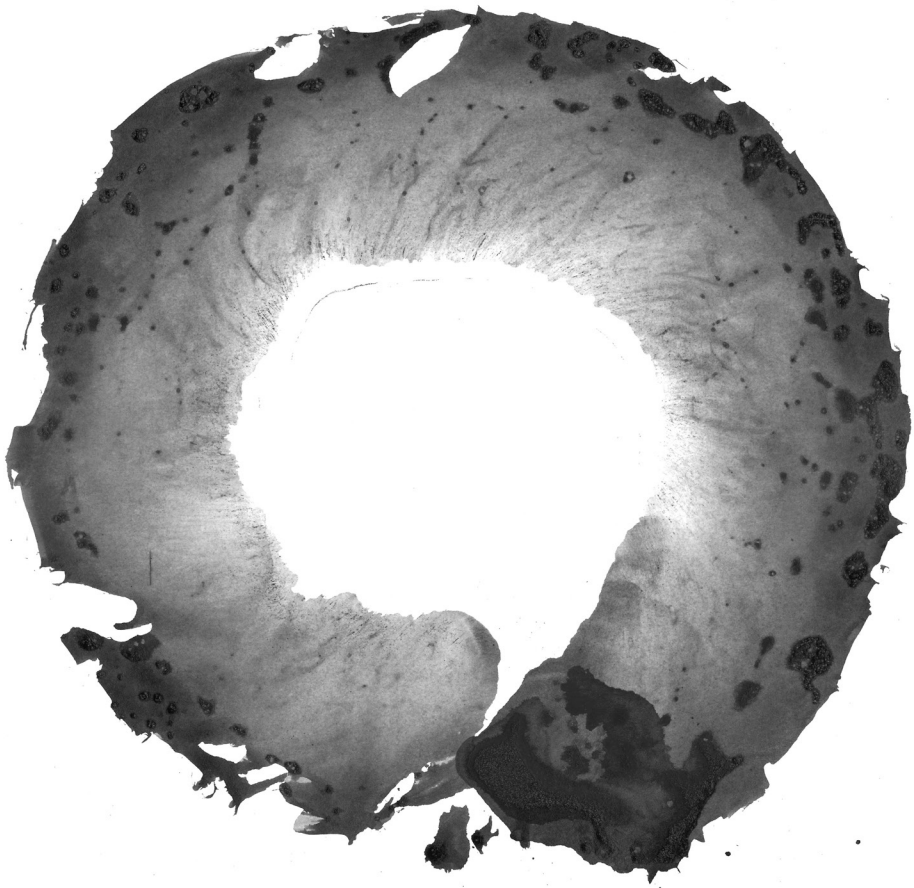
I contributed the following:

- Identification of analytical scope and purpose.
- Conception and development of sample preparation method.
- Conception and development of GC-FID method.
- Conception and development of data analysis and quality control methods.
- Conception and development of data analysis and quality control methods.
- Conception and conduction of round-robin study.
- Validation of analytical method.
- Analysis of data.
- Writing of technical documentation.
- Writing and revision of manuscript.
- Design and animation of figures.

PAPER II

I contributed the following:

- Development of lignin oligomer taxonomy and nomenclature.
- Design, validation, and characterisation of algorithms.
- Production, optimisation, and dissemination of code.
- Acquisition and analysis of data.
- Writing of technical documentation.
- Writing and revision of manuscript.
- Design and animation of figures.



Printed by Media-Tryck, Lund 2026 NORDIC SWAN ECOLABEL 3041 0903

LUND UNIVERSITY

ISBN 978-91-8104-937-4

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

9 789181 049374