Unravelling lignin structure with mass spectrometry and generative modelling

Eliise Tammekivi[a], Avinash R. Pai[a], Joseph S. M. Samec[a], Anneli Kruve[a]
[a]Department of Chemistry, Stockholm University, Svante Arrhenius väg 16C, 114 18 Stockholm, Sweden
eliise.tammekivi@su.se

1 Background

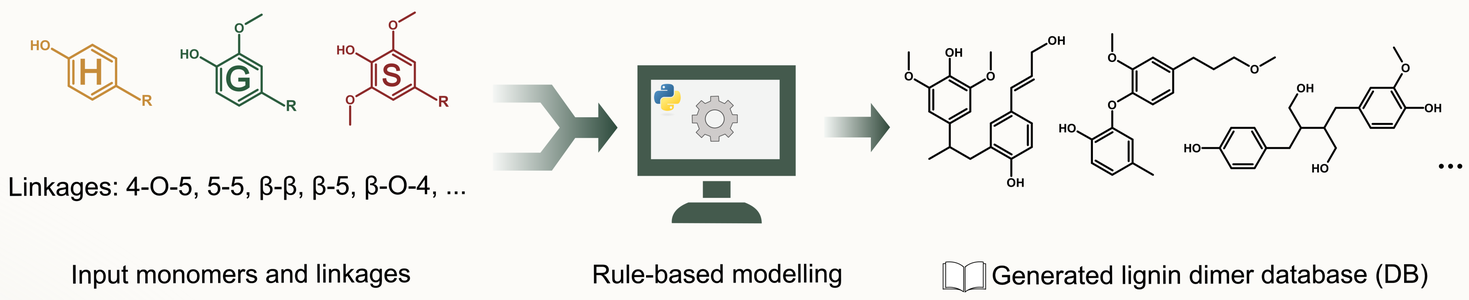
Lignin is a by-product of the paper and ethanol industries; however, it could be used as a resource for the production of novel biomaterials. Knowing the structure of lignin would aid with developing the valorisation processes.[1]

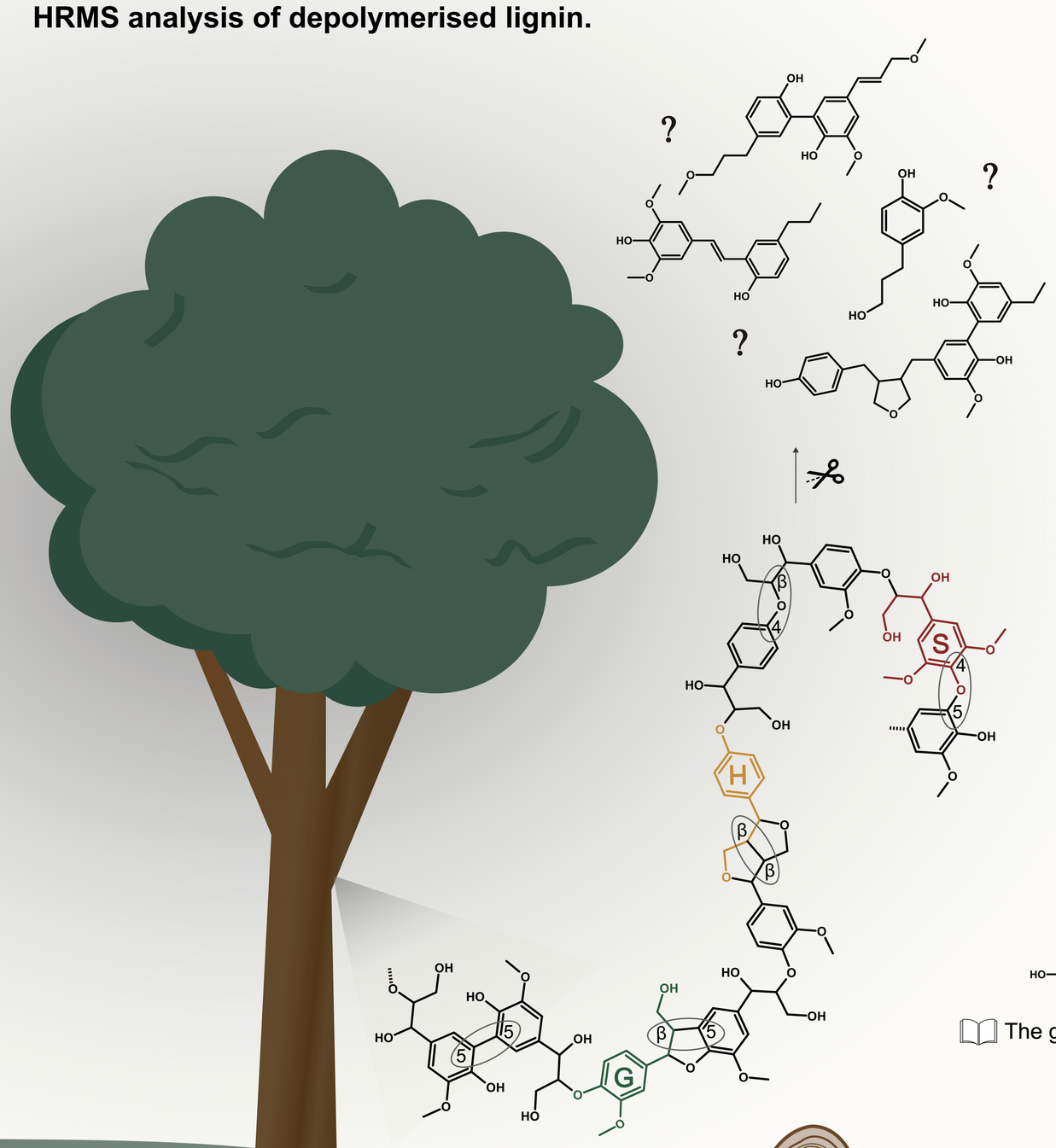
Structural characterisation of lignin with LC/HRMS is hampared by the absence of lignin dimers, trimers and larger oligomers from databases.[2]

In this project, we aim to expand the structural databases by generating lignin oligomer structures *in silico*. The database was used for annotating compounds detected during the LC/HRMS analysis of depolymerised lignin.

2 Structure generator

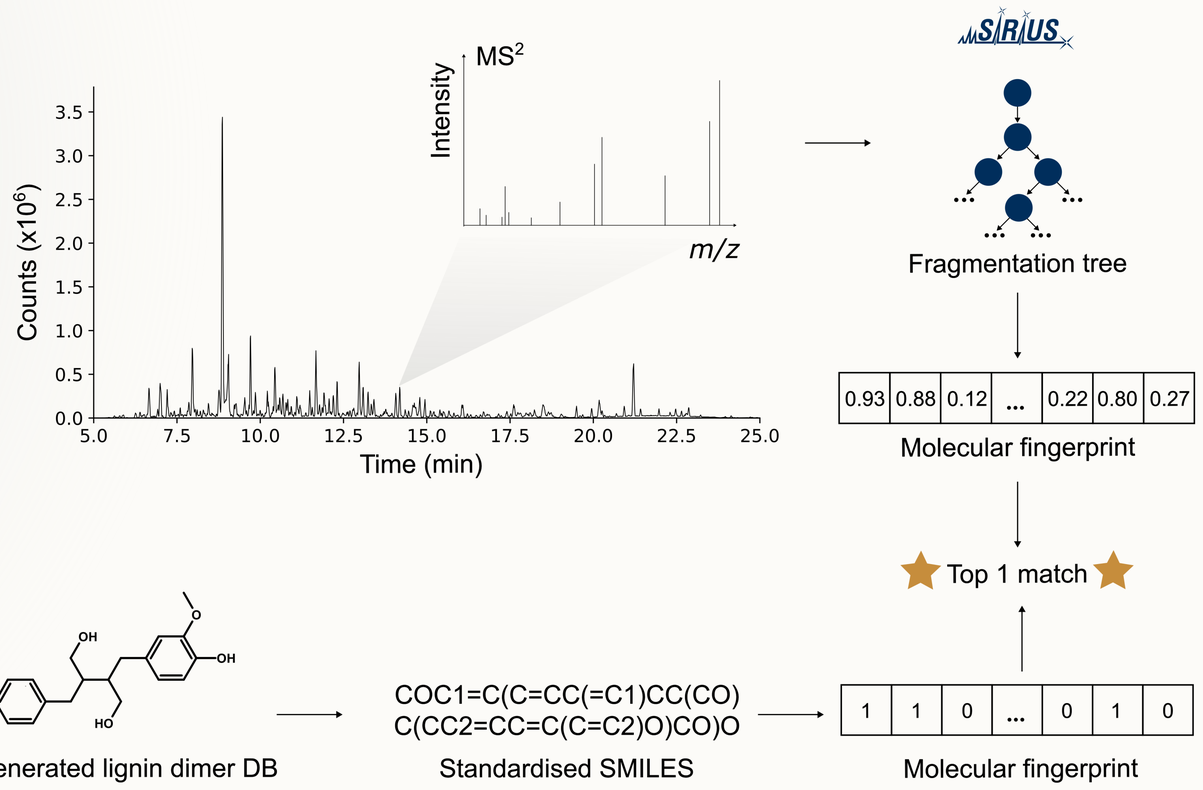
The reactions taking place during the reductive catalytic fractionation (RCF) of lignin were compiled.[3] Based on these reactions, a Python code utilising RDKit was written to generate lignin dimer structures.
ChatGPT (OpenAI, 2025) was used for coding assistance.





3 LC/HRMS^2 analysis

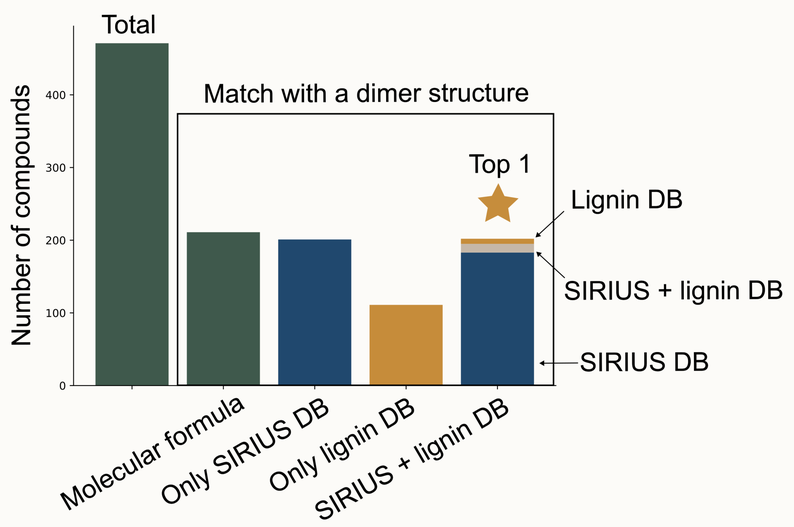
The MS^2 spectra were processed with SIRIUS+CSI:FingerID v. 6.0.4 for tentative structural annotation. Both existing databases, e.g. PubChem and lignin DB generated here were considered.



4 Preliminary results

The generated lignin dimer DB contains 777 compounds. From these, only 109 were found in the PubChem database.

For a depolymerised lignin sample, 470 compounds were detected. Although this is the first version of the lignin DB, already for some compounds (7) the top 1 match was obtained with the generated lignin DB.



5 Conclusions

Lignin structural DB was formed *in silico*, including structures absent from PubChem.

The lignin DB can be easily used in SIRIUS for tentative identification of lignin oligomers.

The next steps include expanding the lignin library even more by adding other monomers and linkages.

[1] S. Bertella, J. S. Luterbacker, Trends Chem. 2 (5) (**2020**) 440-453.
[2] E. Tammekivi et al., Anal. Chim. Acta. 1288 (**2024**) 342157.
[3] H. Dao Thi et al., Green Chem. 24 (**2022**) 191-206.

Acknowledgment
The project received funding from the Wenner-Gren Foundations project nr UPD2023-0118 and from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program grant agreement No. 101124488, LearningStructurE: Machine Learning and Mass Spectrometry for Structural Elucidation of Novel Toxic Chemicals.

Kruve lab