Supplementary Information for

Bayesian Optimization and Molecular Dynamics Simulations to Guide Protic Ionic Liquid-Based Biorefining for Efficient Lignin Applications

Sharib Khan [a], Tormi Lillerand [a], Veerapandian Ponnuchamy [b], A. G. M. Zaman [a], Daniel Rauber [c, d], Udayakumar Veerabagu [a], Jüri Olt [a], Markus Gallei [c, d, *], Sabarathinam Shanmugam [a, *], Timo Kikas [a, *]

[a] Chair of Biosystems Engineering, Institute of Forestry and Engineering, Estonian University of Life Sciences, Tartu, Estonia

[b] InnoRenew CoE, Livade 6, 6310 Izola, Slovenia and University of Primorska, Andrej Marušič Institute; Koper, Slovenia

[c] Polymer Chemistry, Saarland University, Campus C4.2, 66123 Saarbrücken, Germany

[d] Saarene, Saarland Center for Energy Materials and Sustainability, Saarland University, 66123 Saarbrücken, Germany

*Correspondence: markus.gallei@uni-saarland.de, sabarathinam.shanmugam@emu.ee, timo.kikas@emu.ee

Table S1 Biomass composition of softwood before and after the PIL pretreatment.

Components	Untreated (%)	Treated (%)
Cellulose	47.9 ± 0.9	47 ± 1.3
Hemicellulose	13.4 ± 0.7	0.00
Lignin	27.3 ± 0.5	7.3 ± 0.6
Ash	3.4 ± 0.8	0.6 ± 0.2
Extractives	4.1 ± 0.7	0.00
Moisture	3.3 ± 0.4	0.00
Dissolved in PIL	0.00	44.5 ± 0.6

Table S2 Documented and experimentally quantified P-factor and lignin yields values.

Temperature (°*C*)	Temperature (*k*)	P-factor (P_f)	Lignin yield (%)
Literature collected P_f and lignin yields			
120	393	**500**	**13**
180	453	**500**	**40**
180	453	**1000**	**50**
195	468	**500**	**60**
195	468	**800**	**62**
195	468	**1500**	**65**
195	468	**2000**	**82**
195	468	**2500**	**55**
200	473	**2000**	**80**
210	483	**2500**	**100**
Quantitative P_f and lignin yields			
120	393	**12**	**11**
130	403	**30**	**15**
140	413	**75**	**26**
150	423	**179**	**46**
160	433	**408**	**52**
170	443	**896**	**62**
180	453	**1902**	**73**
190	463	**3903**	**78**
200	473	**7781**	**90**
210	483	**15055**	**120**

The P-Factor Assumption vs. Chemical Reality

The **P-factor** is a mathematical simplification used to quantify "reaction severity" by combining time and temperature into one number. Crucially, the formula for the P-factor assumes a **single activation energy of 125.6 kJ/mol**. In a real biorefinery, however, the process is not just one single reaction; it is a collection of competing chemical reactions happening simultaneously.

These include:

- The cleavage of lignin-carbohydrate linkages (which releases lignin).
- The fragmentation of lignin into smaller pieces.
- The degradation of carbohydrates (like xylan) into other products.
- The formation of "pseudolignin" (unwanted polyfurans that look like lignin during measurement).

Pseudolignin formation (from xylan degradation) can occur, which may cause measured lignin yields to exceed 100%. Lower P-factors (low severity) generally favor a higher number of linkages and higher carbohydrate (LCC) content, while higher P-factors increase lignin yield but decrease these specific structural moieties. The observation that the highest lignin yield occurs at different temperatures for different P-factors even though the P-factor is supposed to represent "total severity" reveals that the chemistry of a biorefinery is more complex than a single mathematical formula can capture.

How Different Activation Energies Change the Outcome

Each of these reactions has its own unique activation energy. In chemistry, a reaction with a high activation energy is very sensitive to temperature; it speeds up drastically as you get hotter. A reaction with a low activation energy is less sensitive to temperature changes.

Because these competing reactions respond differently to heat, a fixed "severity" (P-factor) will yield different results depending on whether toreached that severity through a long time at a low temperature or a short time at a high temperature.

Why Yield Shifts with Temperature

At Low Severity: If the highest yield is observed at a low temperature. This suggests that at this lower severity level, the primary reactions occurring are the "productive" ones the cleavage of native linkages that release lignin. If you were to raise the temperature to 210 °C at this same low severity, you might trigger different competing reactions that do not contribute as effectively to the measured yield.

At High Severity: If the highest yield occurs at the highest temperature. This is likely because high severity and high heat trigger the degradation of xylan (carbohydrates) into furfural, which then reacts to form "pseudolignin". Because this pseudolignin is counted as "lignin" during characterization, the measured yield appears to increase, sometimes even exceeding 100%.

Furthermore, the delignification percentages and lignin yield of pine wood biomass by [TEA][HSO_4] were determined by evaluating the residual lignin in the biomass before and after pretreatment on a dry weight basis as shown in **Equation S1** and **S2**.

$$\% D \left(\frac{w}{w}\right) = \left[\frac{(L_{rb)} - (L_{PIL\,pb)}}{L_{rb}}\right] \times 100 \tag{S1}$$

$$\% Lr \left(\frac{w}{w}\right) = \left[\frac{L_{r\,PIL}}{L_{rb}}\right] \times 100 \tag{S2}$$

D - Delignification
L_{rb} - Lignin remained in residual biomass
L_r - Lignin recovered
L_{PILpb} - Lignin recovery after pretreatment

Table S3 HSQC NMR Peak integrations of PIL extracted lignin.

Name	Peak assignments	References
Pinoresinol (β-β')	δ(13C), δ(1H) β: 54.21/3.08, γ: 71.45/4.23 and 3.85	[1]
Epiresinol (β-β')	δ(13C), δ(1H) γ: 70.8/3.75 and 4.12	[1]
Dihydro cinnamyl alcohol	δ(13C), δ(1H) α: 31.97/2.55, γ: 60.73/3.43	[2]
Aryl glycerol	δ(13C), δ(1H) γ: 63.4/3.47	[3]
Guaiacyl unit	δ(13C), δ(1H) G2: 110.9/6.91, G5: 114.9/6.86 G6: 119.2/6.77	[4]
Methoxyls (MeO)	δ(13C), δ(1H) 55.5/3.66	[4]
Phenylcoumaran (β-5)	δ(13C), δ(1H) 86.8/5.49 (α), 54.5/3.47 (β), 63.9/3.73/3.62 (γ)	[5–8]
Resinol (β-β)	δ(13C), δ(1H) 85.1/4.63 (α), 54.2/3.07 (β), 70.9/4.16/3.66 (γ)	[9,5–8]
Trans-stilbene (SB1) and SB5	δ(13C), δ(1H) 125.0/7.01 (α), δ(13C), δ(1H) 128.0/7.12 (α), 120.0/7.22 (β)	[5-10]
Carbohydrate, β-D-xylopyranoside	C1: 102.3/4.30, C2: 73.2/3.08 C3: 74.6/3.29, C4: 75.9/3.54 C5: 63.8/3.92 and 3.23	[1,10]

Table S4 Quantitative aliphatic, phenolic and carboxylic acid -OH in the kraft lignin.

aliphatic-OH[a]	**phenolic-OH from lignin**			**total phenolic-OH**	**carboxylic acid OH**[e]	**total OH**
	C_5 substituted OH[b]	**Guaiacyl-OH**[c]	**p-Hydroxyphenyl-OH**[d]			
2.17	0.73	1.13	0.16	2.02	0.74	4.93

[a] integral of 150 – 145 ppm; [b] integral of 145 – 140.8 ppm; [b] for both syringyl-OH and C5-condensed-OH; [c] integral of 140.6 – 138.8 ppm; [d] integral of 138.6 – 137.1 ppm; [e] integral of 136 – 133.6 ppm

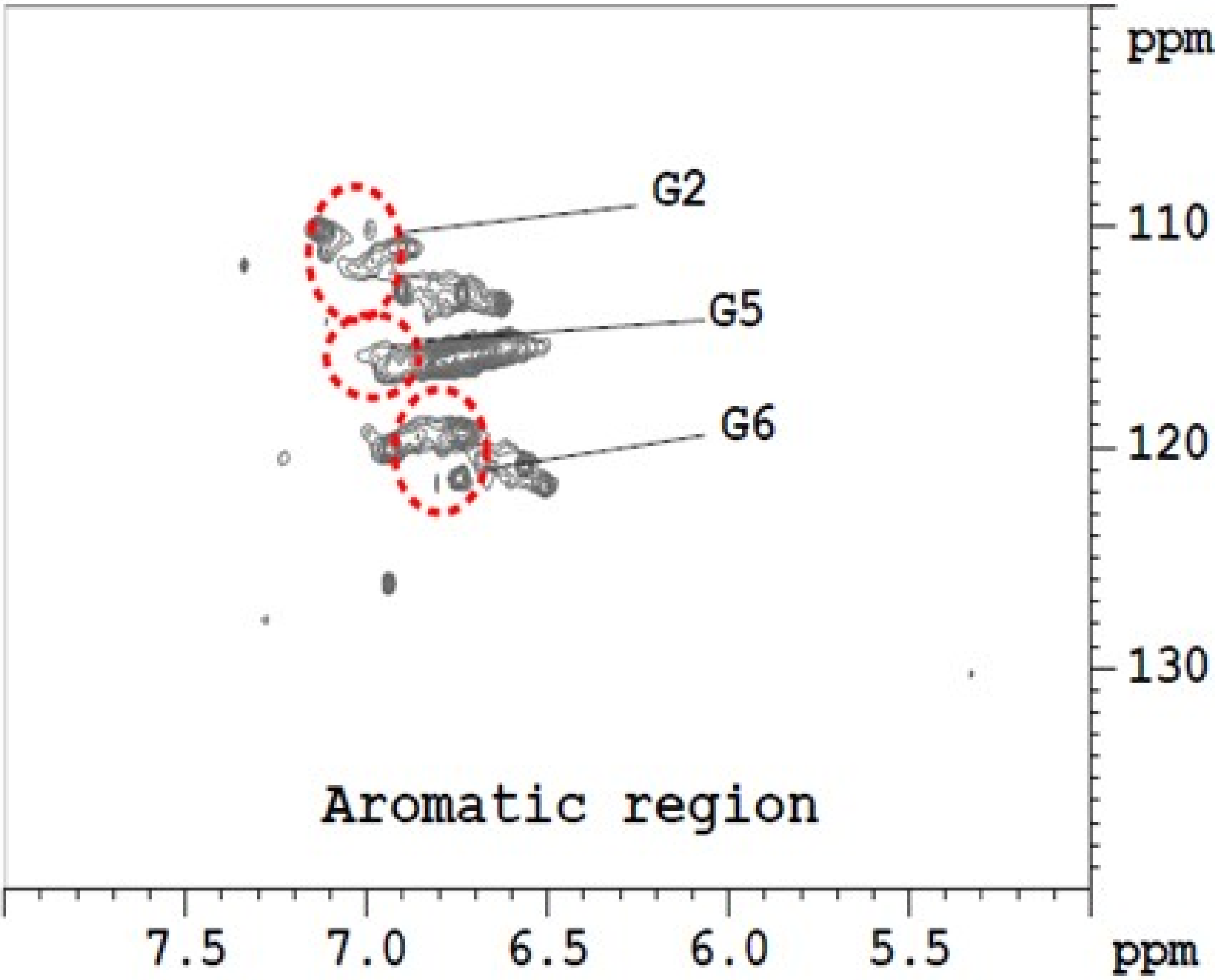

Figure S1 Aromatic HSQC NMR spectra of the kraft lignin.

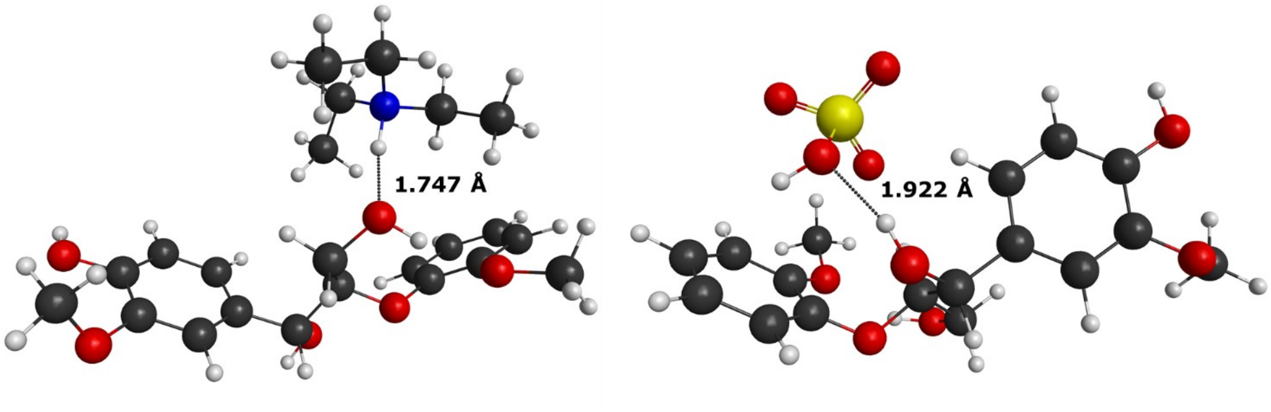

Figure S2 Optimized DFT structures of lignin model with PIL components (H-bond distances and interaction energy);
Lignin model, GGE – TEA = -33.201 kcal/mol
Lignin model, GGE – HSO_4 = -26.043 kcal/mol.

References:

1 R. Liu, A. Smeds, L. Wang, A. Pranovich, J. Hemming, S. Willför, H. Zhang and C. Xu, *ACS Sustain. Chem. Eng.*, 2021, **9**, 13862–13873.

2 D. Tarasov, P. Schlee, A. Pranovich, A. Moreno, L. Wang, D. Rigo, M. H. Sipponen, C. Xu and M. Balakshin, *Green Chem.*, 2022, **24**, 6639–6656.

3 I. Sapouna and M. Lawoko, *Green Chem.*, 2021, **23**, 3348–3364.

4 M. Y. Balakshin and E. A. Capanema, *RSC Adv.*, 2015, **5**, 87187–87199.

5 H. Ben and A. J. Ragauskas, *Energy and Fuels*, 2011, **25**, 2322–2332.

6 C. Zhao, S. Li, H. Zhang, F. Yue and F. Lu, *Int. J. Biol. Macromol.*, 2020, **152**, 411–417.

7 M. E. Eugenio, R. Martín-Sampedro, J. I. Santos, B. Wicklein, J. A. Martín and D. Ibarra, *Int. J. Biol. Macromol.*, 2021, **181**, 99–111.

8 T. Q. Yuan, S. N. Sun, F. Xu and R. C. Sun, *J. Agric. Food Chem.*, 2011, **59**, 10604–10614.

9 R. Liu, A. Smeds, T. Tirri, H. Zhang, S. Willför and C. Xu, *ACS Sustain. Chem. Eng.*, 2022, **10**, 14588–14599.

10 R. Liu, A. Smeds, S. Willför and C. Xu, *Ind. Crops Prod.*, 2024, **209**, 118055.