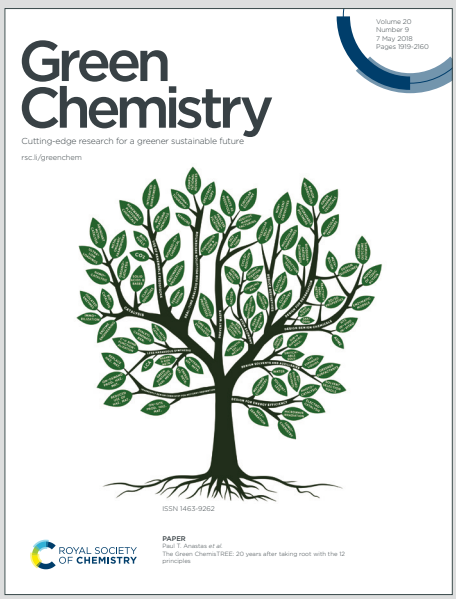
Green Chemistry

Cutting-edge research for a greener sustainable future

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

rsc.li/greenchem

View Article Online
DOI: 10.1039/D6GC02613C

Green Foundation

1. We advanced a protic ionic liquid (PIL)-based biomass fractionation process with a strong focus on lignin recovery by integrating Bayesian optimization and molecular dynamics simulations. This integrated strategy enables computationally driven efficiency, reduces the need for extensive experimentation, and clarifies the role of carbohydrate impurities in lignin extraction.
2. The key achievements were the production of low-molecular-weight lignin and the identification of an optimal balance between lignin yield and its quality.
3. To further enhance the sustainability of this approach, future research should focus on full life cycle assessments, PIL recycling, and the development of fully bio-derived PILs, thereby enabling a closed-loop, zero-waste biorefining process.

View Article Online
DOI: 10.1039/D6GC02613C

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

Abstract

The transition to a circular bioeconomy depends on the sustainable use of renewable resources, with forest-based biomass playing a critical role. Lignin, in particular, is a promising feedstock for replacing fossil-derived aromatics due to its unique chemical structure. However, lignin obtained from conventional pulp mills often contains impurities and structural modifications that limit its potential for producing high-value chemicals and materials. In contrast, emerging biorefinery pretreatment strategies enable a "lignin-focused" approach, where every major biomass component cellulose, hemicelluloses, and lignin are valorized for intended applications, allowing the tailored production of bio-based products with improved structural quality. Traditionally, the development of biorefining processes requires extensive experimentation guided largely by empirical advances. In this study, we demonstrate how Bayesian optimization can accelerate the development of a sustainable biomass fractionation process based on the protic ionic liquid (PIL)-triethylammonium hydrogen sulfate ([TEA][HSO_4]). Open-source Python-based tools were employed to optimize lignin extraction from softwood and evaluate the influence of key processing

View Article Online
DOI: 10.1039/D6GC02613C

parameters, including temperature and process severity, on lignin yield. Molecular dynamics simulations and in-depth literature analyses were adopted to identify the optimal trade-offs between lignin recovery and structural properties. Notably, the optimized PIL fractionation process achieved 82% delignification with lignin yields of up to 73%, while consistently producing lignin with low molecular weights because of in situ depolymerization during pretreatment. Furthermore, advanced qualitative characterization was conducted to assess the lignin structure and evaluate the relationship between lignin yield, structural quality, and potential downstream applications.

Keywords: Biorefining; Protic Ionic Liquid; Bayesian Optimization; Molecular Dynamics Simulations; Softwood Biomass; Lignin

View Article Online
DOI: 10.1039/D6GC02613C

1. Introduction

To achieve a truly sustainable energy system, carbon emissions from chemical and industrial production must be eliminated. This transformation requires shifting from traditional petrochemicals derived from fossil fuels to renewable feedstocks [1,2]. Moreover, this transition is accelerating the move from linear production models toward circular, bio-based economic frameworks [3]. Thus, by adopting circular economic strategies, nations currently dependent on fossil resources can reduce their reliance on linear production pathways while promoting resilient and sustainable growth. In this context, a forest-based economy has emerged as a promising solution for maximizing the value of renewable biomass for the continuous production of materials and chemicals [4]. Renewable biomass is generally classified into four generations based on its origin: first (food-derived biomass), second (lignocellulosic biomass), third (single-cell organisms), and fourth (engineered single-cell organisms) [5]. Among these, second-generation lignocellulosic biomass (LCB) represents a particularly attractive resource owing to its abundance and composition of cellulose, hemicellulose, and lignin [6]. However, due to its complex and recalcitrant structure, the efficient separation of LCB's components poses a significant challenge during fractionation processes. For instance, pulp industry focused on converting LCB into cellulose for paper production, with the residual black liquor (comprising lignin and hemicellulose) serving mainly as an energy source [7,8]. Conversely, the emerging biorefinery concept seeks to achieve the comprehensive valorization of all LCB components by employing advanced fractionation techniques [9]. To foster a robust forest-based economy, it is essential for both conventional and novel systems to evolve, placing greater emphasis on the utilization of biomass side-streams [10]. A major barrier to progress is that pulp mills, supported by their secured cellulose market, have limited incentives to invest in advanced fractionation technologies [8,11]. Simultaneously, biorefineries face challenges in establishing viable markets for hemicellulose and lignin derivatives [12,13]. Consequently, the development of advanced fractionation technologies capable of efficiently separating the three primary LCB components into high-purity, sustainable raw materials remain a critical bottleneck for enabling green industrial transitions.

Ionic liquid (IL)-based biomass fractionation, commonly known as the IonoSolv process, has recently gained significant attention as a promising strategy for LCB valorization [14]. The appeal of IonoSolv process lies in the ability of ILs, particularly protic ionic liquids (PILs) to selectively

View Article Online
DOI: 10.1039/D6GC02613C

dissolve and recover the primary components of LCB through several key features: intrinsic hydrogen-bonding capability essential for biomass processing; facile, cost-effective, and atom-efficient synthesis via simple acid–base neutralization, which is well-suited for large-scale applications; catalytic properties at elevated temperatures, including reverse reactions that reform the acid; and ease of handling and processing due to the absence of pressure build-up at higher temperatures [15]. On top of this, certain PILs can effectively extract and enrich cellulose, thereby enhancing its subsequent saccharification [16,17]. In addition, PILs can be designed to exhibit high selectivity toward lignin, improving extraction efficiency and facilitating potential in situ structural modification during pretreatment [17,18]. Compared to conventional methods, including mechanical [19], chemical [20], thermochemical [21], and kraft pulping processes [22], IL-based fractionation offers several advantages. Depending on the feedstock, the IonoSolv process can also provide improved component separation in a potentially more cost-effective and sustainable manner [23]. Furthermore, the tuning properties of PILs towards different biomass components, particularly cellulose and lignin, make them relevant for several high-value applications, such as emulsions, antioxidants, and thermoplastic formulations [16,24]. Nevertheless, tailoring the properties of LCB components or renewable materials traditionally requires extensive experimental screening, substantial resources, and detailed characterization. These challenges can be overcome by emerging machine learning (ML) methods, which can significantly speed up the process and make the IonoSolv approach more sustainable and economically relevant.

Recent studies have increasingly explored the application of ML to advance biomass fractionation, particularly for optimizing pretreatment conditions [25], modeling representative compounds [26], and tailoring the LCB side-stream properties for valorization. For instance, Löfgren et al. [27] demonstrated the potential of ML to accelerate the development of sustainable chemical processing strategies for targeted lignin extraction and its properties. Similarly, Diment et al. [28] and Chrząstowska et al. [29] applied ML approaches to improve the scalable production of lignin–carbohydrate complexes (LCCs) and pulp extractive analysis. Rummukainen et al. [30] introduced an ML-based framework for pilot-scale comparison of wood delignification processes. In addition, Gisperg et al. [31] highlighted the applicability of Bayesian optimization (BO) in bioprocess engineering, whereas Bertelsen et al. [32] developed an open-source Python package to facilitate real-world optimization using BO. Despite these advances, current research lacks a consolidated focus on lignin yield and its corresponding molecular weight (M_w), which is a key parameter

View Article Online
DOI: 10.1039/D6GC02613C

influencing downstream valorization. To the best of our knowledge, no study has investigated the integration of IL-based fractionation, particularly PILs, within a BO-driven ML framework to optimize lignin extraction. The absence of a general-purpose optimization strategy capable of controlling the M_w of lignin limits the development of standardized lignin valorization pathways. This challenge is further compounded by the fact that different fractionation methods produce lignin with varying M_w [33,34], complicating process modeling, comparison, and refinement [35–37]. Additionally, process optimization is often hindered by variability in biomass feedstocks and solvent systems. Conventional optimization approaches, such as linear programming, dynamic programming, and search algorithms, often struggle with the highly nonlinear and discontinuous nature of LCB fractionation, resulting in high computational costs [38,39].

Molecular dynamics (MD) simulations provide a powerful computational approach for understanding the dynamic behavior of materials at the molecular scale. These methods have been widely applied to investigate the interactions between lignin and various solvents to optimize the dissolution process and nanoparticle synthesis. In particular, MD simulations provide valuable insights into the movement of particles and molecules, which can further elucidate the complex interactions between lignin and ILs during fractionation [40–42]. Previous studies have also indicated that basic anions, such as acetate and glycinate, function as strong hydrogen-bond acceptors capable of disrupting both inter- and intra-molecular interactions within the lignin matrix [40,42]. MD Simulations further demonstrated that selective lignin dissolution relative to cellulose can be optimized by tuning the anion basicity. By quantifying the binding energies and diffusion coefficients, MD simulations provide a thermodynamic framework for the rational design of task-specific ILs.

Keeping these considerations and challenges in mind, our study developed a BO-guided PIL-based lignin extraction methodology that integrates MD simulations with lignin yield predictions to evaluate the feasibility of targeted extraction conditions at the early stages of the process design, thereby reducing reliance on trial-and-error experimentation. In addition, the study proposes integrating BO with extensive literature datasets to generate surrogate models that correlate key PIL properties with lignin characteristics while minimizing processing constraints and satisfying multiple experimental objectives. This approach enables the simultaneous maximization of lignin yield and optimization of structural properties, facilitating the prediction of suitable downstream

View Article Online
DOI: 10.1039/D6GC02613C

applications of lignin. By standardizing the lignin yield and structural properties within an ML framework that incorporates PIL selectivity and optimized extraction conditions, this study aims to advance predictive process design and significantly enhance lignin valorization strategies.

View Article Online
DOI: 10.1039/D6GC02613C

2. Materials and Methods

2.1. Materials and Reagents

Raw Scots pine (*Pinus sylvestris*) was collected from the southern part of Tartu, Estonia. Commercially available alkali kraft lignin was purchased from Sigma-Aldrich (USA). Moisture content of the biomass was determined using a Kern MLS-50-3D moisture analyzer (Kern & Sohn GmbH, Balingen, Germany). Fiber content of the pine wood biomass was analyzed using an Ankom 200 fiber analyzer (ANKOM Technology, Fairport, NY, USA) [43]. The estimated and optimized compositional analysis of pretreated pine wood biomass (expressed as % dry mass) is provided in the *Supplementary Table S1*. All chemicals were purchased commercially, and heating experiments were conducted in a conventional drying oven.

2.2. Selection and synthesis of Triethylammonium hydrogen sulfate ([TEA][HSO_4])

Based on a comprehensive review of the scholarly literature (**Table 1**), triethylammonium hydrogen sulfate ([TEA][HSO_4]) was selected for biomass fractionation, simulations study and lignin extraction. In-depth literature analysis identifies [TEA][HSO_4] as one of the most prominent protic ionic liquids (PILs) candidates for industrial applications, as it has been specifically investigated for scale-up and production viability. This conclusion is supported by the PIL's straightforward synthesis of commodity chemicals and its effective fractionation performance.

The synthesis was carried out as follows. Triethylamine (75.9 g, 750 mmol) was cooled in an ice bath within a round-bottom flask. Sulfuric acid (5 M solution) was added dropwise with stirring. Water was removed under reduced pressure using a rotary evaporator, and the product was dried overnight at 40 °C on a Schlenk line. The resulting PIL, [TEA][HSO_4], was obtained as a white hygroscopic solid.

Table 1 Detailed summaries of PIL pretreatment processes using [TEA][HSO_4] for various feedstocks.

Summary	Ref.
Rapid [TEA][HSO_4] 15-min pretreatment at 180°C enabled highly efficient enzymatic saccharification. A saccharification yield exceeding 75% of the theoretical maximum was reported, demonstrating the effectiveness and rapidity of the ionoSolv pretreatment.	44
[TEA][HSO_4] to remove heavy metals from contaminated biomass (HMCBs) before subjecting it to thermochemical or biological conversion processes.	45
[TEA][HSO_4] was evaluated using a combined methodology of detailed process simulation and life cycle assessment (LCA) to determine its economic viability and environmental footprint.	46
A hybrid pretreatment process was developed by combining ionoSolv and organosolv fractionation. Compared to the standard ionoSolv process, this hybrid method demonstrated superior performance. It produced a cellulose-rich pulp with higher enzymatic accessibility and achieved more extensive lignin removal.	47
A complete process was developed to transform oil palm empty fruit bunches (OPEFBs) using [TEA][HSO_4] as a pretreatment agent. Systematic optimization of key pretreatment parameters, including the PIL composition and temperature. Finally, a method for the recovery and recycling of PIL was established, forming a comprehensive and sustainable processing loop for biomass.	48
The biofuel potential of the perennial grass *Pennisetum polystachion* was evaluated using a pretreatment process with [TEA][HSO_4]. An optimal pretreatment condition was established using an 80% concentration of PIL at 140 °C for 45 min with a 10% solid load. This resulted in a high delignification rate of 65.8%, effectively breaking down the biomass structure for subsequent conversion. Furthermore, the process demonstrated strong sustainability through PIL recycling, with recovery rates reaching 90% while maintaining significant delignification efficiency across multiple cycles.	49
A microwave-assisted fractionation process using [TEA][HSO_4] was developed for corn stover. This strategy effectively deconstructed the biomass to facilitate the production of monomeric sugars and support downstream acetone–butanol–ethanol (ABE) fermentation. A mass balance analysis of the integrated process demonstrated that from 100 g of raw corn stover, this approach could generate 8.1 g of ABE solvents and 16.61 g of technical lignin.	50
This study designed and commissioned a versatile semi-batch reactor with inherent flexibility for reconfiguration into continuous-flow operation. The reactor was successfully implemented for production, demonstrating the capacity to generate ([TEA][HSO_4]) at a rate of 0.8 kg per hour and a concentration of 80% (w/w).	51
Molecular simulations of ionic liquids for lignin solvation and in situ depolymerization.	52

View Article Online
DOI: 10.1039/D6GC02613C

2.3. Bayesian optimization

Bayesian optimization (BO) is the cornerstone of scientific research and industrial applications. It is a strategic approach for fine-tuning various parameters to achieve the best possible outcome in terms of efficiency, yield, or cost-effectiveness. The application of structured process optimization often produces significant cost savings, improved product quality, enhanced performance, and increased productivity. In this context, data collection was guided by BO, building on the approach described by Bertelsen et al.[32], and the theory behind BO can be acknowledged elsewhere [32,53,54]. BO was performed on a lignin yield dataset comprising *literature-gathered data* and *experimentally concluded observations*, as shown in **Fig. 1**. A Gaussian process (GP) surrogate model was used to guide batch-wise BO of lignin yield as a function of temperature and P-factor (discussed later in the text). Temperature and P-factor were min–max normalized before model fitting, and lignin yield was used as the response variable. The GP model was implemented in Python using scikit-learn with an RBF-based covariance kernel and an added white-noise term. The optimization was initialized with four sequential observations, which served as the Batch 0 training set. At each subsequent BO iteration, the GP was refitted using all observations available up to that point. The remaining unrevealed experimental conditions were then ranked using an Upper Confidence Bound (UCB) acquisition function, $UCB = \mu + k\sigma$, where μ is the GP-predicted mean yield, σ is the predictive standard deviation, and $k = 2.0$ controls the balance between exploitation and exploration. The two candidates with the highest UCB scores were selected, their measured lignin yields were revealed, and these observations were added to the training set for the next iteration.

View Article Online
DOI: 10.1039/D6GC02613C

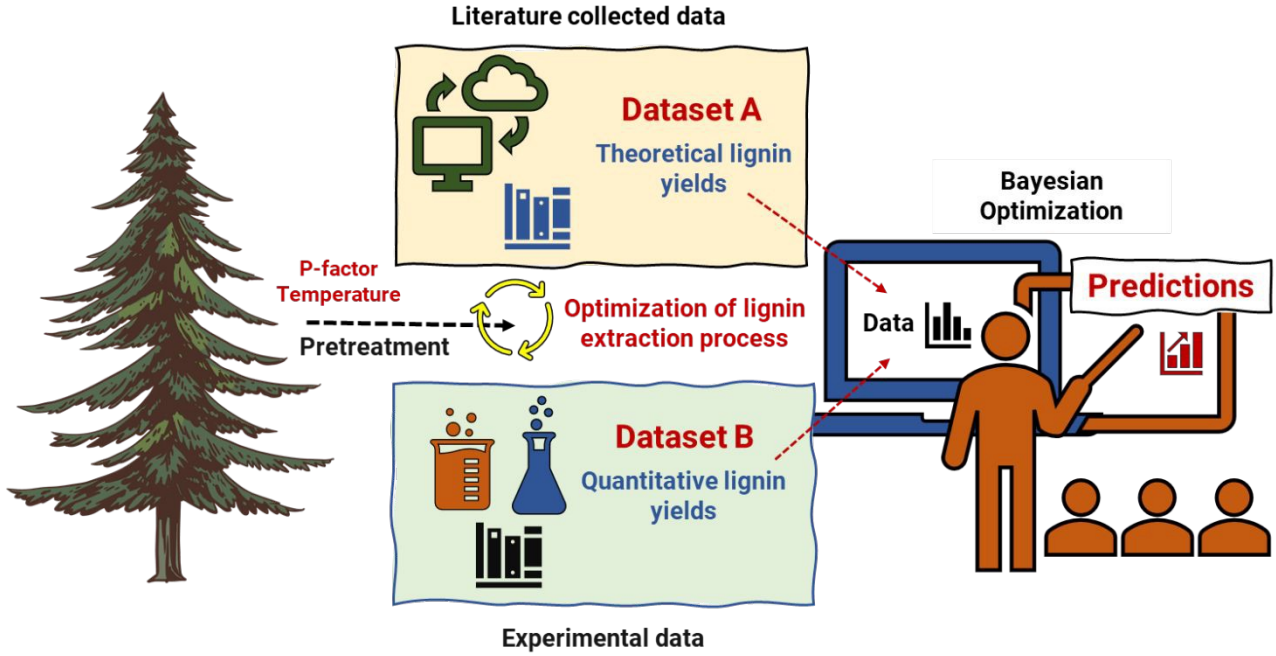


Fig. 1 Machine learning-driven workflow that utilizes a dual-dataset system to optimize lignin yield via Bayesian optimization. **Dataset A**, trained in literature data, proposes initial extraction parameters (e.g., P-factor and temperature), which are then validated through experimental PIL-based lignin extraction in **Dataset B**. The results from these experiments are fed back to sequentially updating the model, creating a closed-loop framework that continuously improves predictive accuracy and refines the optimal PIL-based lignin extraction conditions.

2.4. Molecular dynamics (MD) simulations

Lignin is a structurally heterogeneous biopolymer composed of *p*-coumaryl (H), coniferyl (G), and sinapyl (S) alcohol units, whose relative abundance and linkage patterns depend on biomass origin. Because no universal lignin structure exists for modelling, this study adopts a representative softwood lignin model (MW ≈ 8.5 kDa). The chosen model contains 46 G units and 1 H unit, connected primarily through β-O-4, β-5, 5-5, β-1, α-O-4, 4-O-5, and β-β linkages, with β-O-4 bonds accounting for ~46% of the structure. The lignin structure was generated using the Lignin-builder tool [55] and minimized with TopoGromacs [56] and NAMD [57], then converted to the GROMACS [58] format using VMD/TopoTools [59]. All simulations were performed using GROMACS with the CHARMM [60] force field to describe intra- and intermolecular interactions. Triethylammonium hydrogen sulfate ([TEA][HSO_4]) was used as the PIL solvent. CHARMM-GUI [61,62] was utilized to prepare both inter- and intra- molecular force-field parameters for the TEA^+ cation and HSO_4^- anion.

View Article Online
DOI: 10.1039/D6GC02613C

A simulation box containing one lignin molecule dissolved in 2792 molecules of [TEA][HSO_4] at approximately 1.5 wt% lignin was constructed as shown in **Fig. 2**. After energy minimization by the steepest descent method, the system was equilibrated for 5 ns at 298.15 K. Production simulations were then carried out for 100 ns in the NPT ensemble at 1 bar. The temperature was controlled using a velocity-rescaling thermostat, and the pressure was controlled using a Berendsen barostat. The leapfrog algorithm was used with a 2 fs time step, and hydrogen-containing bonds were constrained using the LINCS method [63]. Lennard-Jones interactions were switched off between 1.0 and 1.2 nm with a dispersion correction applied, and long-range electrostatics were treated using the Particle Mesh Ewald method. Structural and thermodynamic analyses were performed over the final 70 ns of the trajectory and saved every 2 ps. GROMACS in-built compiled codes were used for all analyses performed in this study.

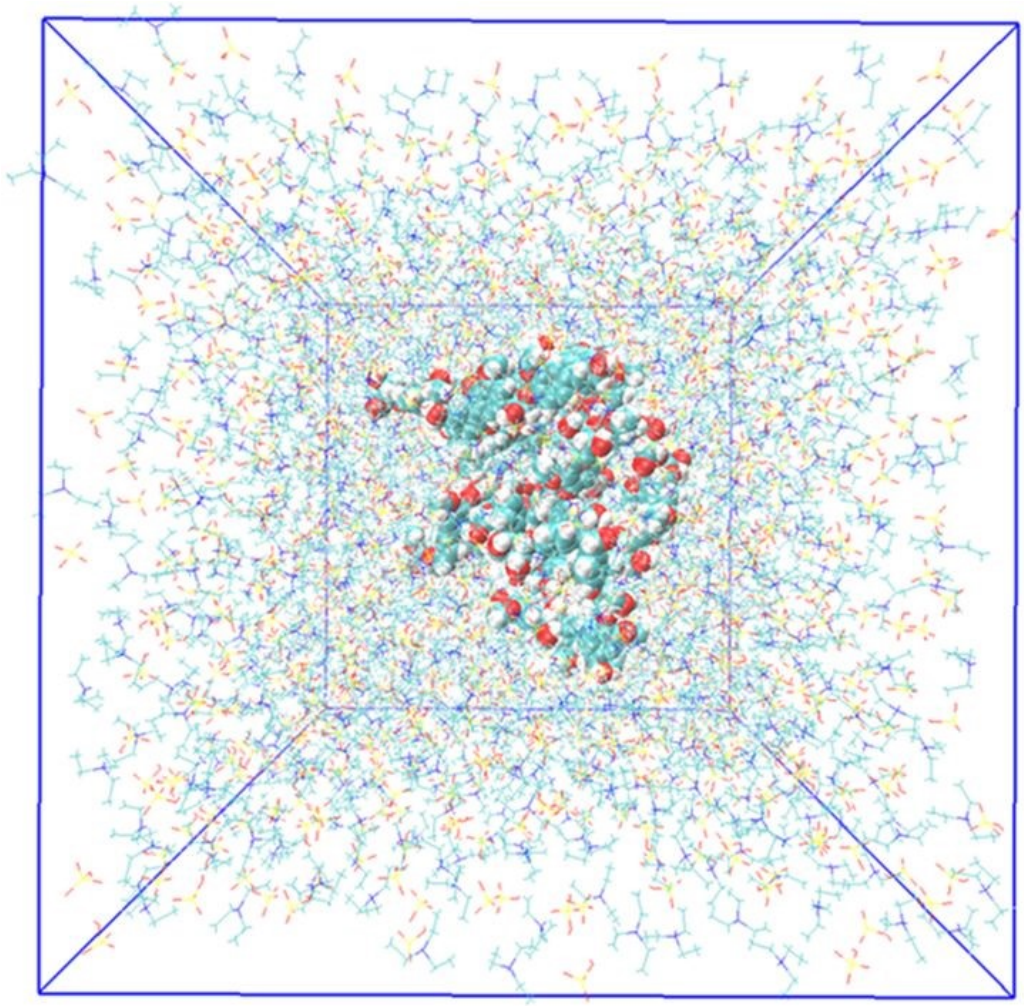


Fig. 2 Lignin molecule dissolved in 2792 molecules of [TEA][HSO_4].

View Article Online
DOI: 10.1039/D6GC02613C

2.5. PIL-Mediated Biomass Fractionation

The collected pinewood was debarked, air-dried, ground to 1-2 mm, maintained at <10% moisture, and stored in an airtight container. Fractionation and extraction of lignin from biomass using [TEA][HSO_4] were adapted according to our previous methodology [43]. Briefly, in a 100 mL ACE pressure tube with a silicone ring (front), PIL with a biomass-to-solvent ratio of 1:3 was added, vortexed, and incubated in a preheated oven. The samples were treated in triplicate at different temperatures (120-210 ℃) in batches iteratively (0, 1, 2 and 3) for 1.5 h as the incubation time. Following pretreatment, ethanol was added to the pressure tube to separate the cellulosic pulp from the PIL-lignin mixture via centrifugation (4000 rpm for 10 min). Later, PIL lignin dissolved in ethanol was fractionated using a pressure filtration system equipped with Whatman™ nylon membrane filters (pore size 0.8 mm) to extract lignin rich fractions. Finally, ethanol was recycled using a rotary evaporator (Buchi Rotavapor R-200, Buchi, Switzerland) and after evaporation, Milli Q water was added to the concentrated ionic liquid solubilized lignin as an antisolvent to precipitate lignin fractions from PIL; the precipitated lignin was subsequently washed with the Milli Q water (3×) to ensure complete removal of PIL, which was then recovered through centrifugation, as shown in **Fig. 3**. Furthermore, this study does not evaluate the enzymatic digestibility or fermentability of the cellulosic residue, nor does it provide PIL recycling data. These aspects remain the subjects of ongoing studies and are necessary for the development of a fully integrated biorefining process.

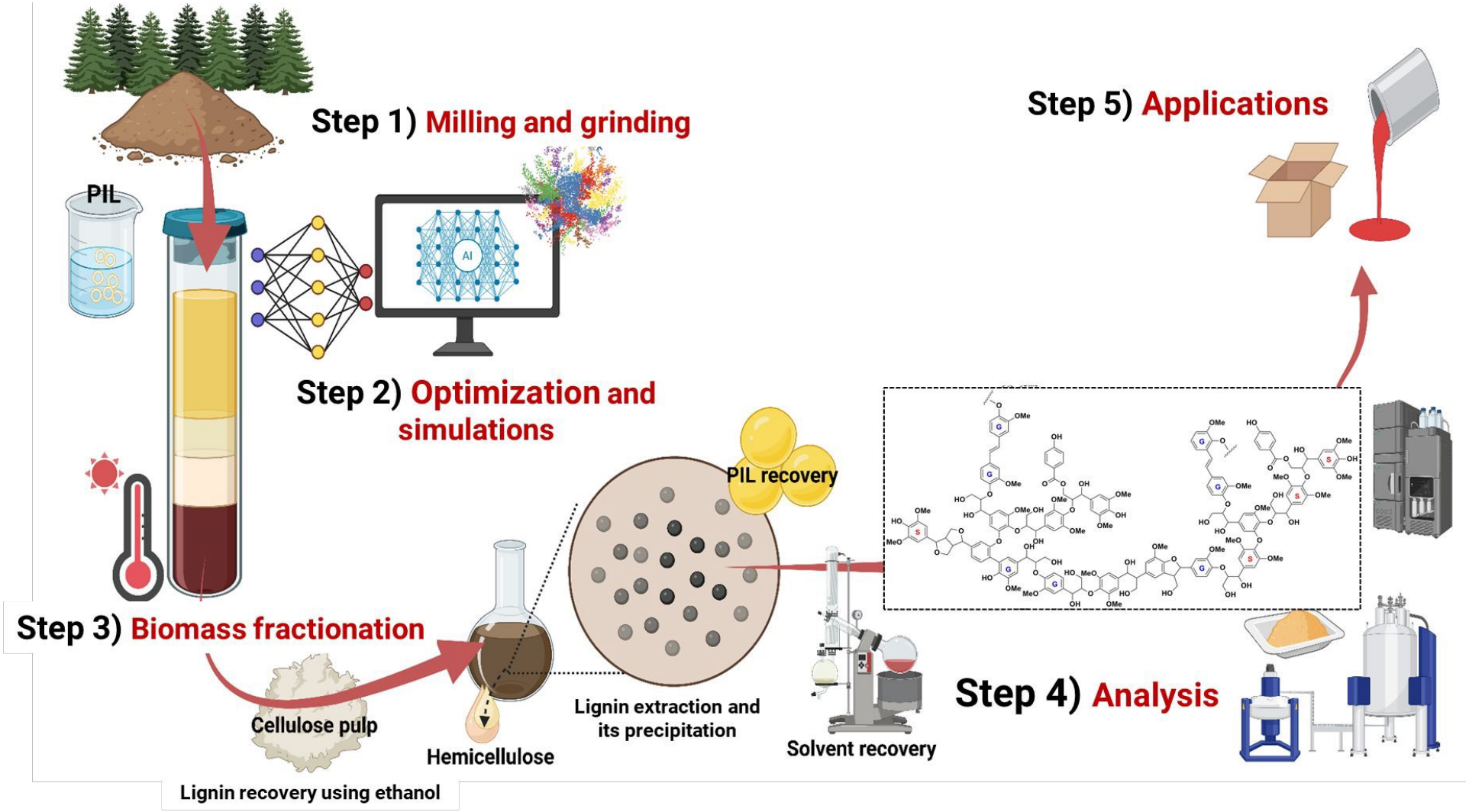


Fig. 3 Optimization of the lignin extraction process (steps 1-5) using combined Bayesian optimization, molecular dynamics simulations, PIL-based biomass fractionation, and quantitative and qualitative analyses for targeted lignin applications.

View Article Online
DOI: 10.1039/D6GC02613C

2.6. Data collection

Several previous studies have described the P factor [64,65] as a parameter that quantifies the effects of time and temperature. To quantify and control the severity of the PIL reaction, we used the same concept as that used for the pre-hydrolysis factor (P-factor). It converts time and temperature into a single variable that depends on the heating curve of the reactor. The relative reaction rate expresses the reaction rate at a given temperature relative to the reference reaction rate at 100 °C, as previously described [28]. Based on the reaction rate, the P-factor can be described as follows:

$$P - factor = \int_0^t \frac{K(T)}{k_{100℃}} dt = \int_0^t e^{40.48 - \frac{15106}{T}}$$

where *t* is the reaction time (h), *T* is the temperature (K), and *k* is the rate constant. The activation energy used in Equation (1) is 125.6 kJ/mol (or 30 kcal/mol), which is slightly smaller than the kraft cooking activation energy of 134 kJ/mol (or 32 kcal/mol) used in the H-factor [66]. The delignification of softwood by PIL was determined by estimating the residual lignin in the biomass before and after pretreatment on a dry weight basis (*Supplementary Information*). To construct a reliable initial dataset for BO, we therefore applied stringent selection criteria when mining the literature. First, we retained only studies that employed IonoSolv and Organosolv pretreatment of wood (both softwood and hardwood). Second, we further narrowed the selection to those studies that implemented a p-factor, which allowed us to extract a consistent severity index as a model input. The steps, including biomass fractionation, BO, molecular dynamics simulations, qualitative lignin analysis, and their applications, are shown in **Fig. 3**.

2.7. Characterization of Recovered Lignin

2.7.1. Attenuated Total Reflectance Fourier Transform Infrared (ATR-FTIR)

The surface functionalities of the lignin's obtained from PIL pretreatment and commercially available alkaline lignin were measured via FTIR spectroscopy (Spectrum BXII, Perkin Elmer Inc., Waltham, MA, USA) using the universal attenuated total reflection method. The spectra were recorded with an average accumulation of 16 scans in the range of 4000-600 cm^{-1} at a resolution of 4 cm^{-1}.

2.7.2. Thermogravimetric and elemental analysis

View Article Online
DOI: 10.1039/D6GC02613C

The thermal degradation properties of lignin were determined using a NETZSCH STA 449 F3 Jupiter simultaneous (TGA and DSC/DTG) thermal analyzer (NETZSCH, Selb, Germany). Approximately 5 mg of the lignin sample was heated to 900 ℃ under N_2 gas with a flow rate of 100 mL/min at a heating rate of 10 ℃/min.

For elemental analysis, approximately 100 mg of lignin was encapsulated in tin foil. The carbon, hydrogen, nitrogen, and sulfur contents were determined using an Elementar Vario Macro Cube based on EVS-EN ISO 16948:2015. The oxygen content was calculated as the difference (O = 100 − (C + H + N + S)).

2.7.3. Gel permeation chromatography and acid methanolysis

The average molecular weight distribution of lignin's was analyzed using gel permeation chromatography (GPC) equipped with LabSolutions GPC software. A high-performance liquid chromatography (HPLC) system (Shimadzu Prominence-i, LC-2030C 3D Plus, Shimadzu Corporation, Kyoto, Japan) contained a pump, an autosampler, a set of two MCX columns (1000 Å and 100,000 Å), and a pre-column (8 mm × 50 mm) (Polymer Standards Service (PSS), GmbH, Mainz, Germany) with a UV detector (280 nm). Lignin samples were dissolved in 0.1 M NaOH (5 mg/mL), and an isocratic flow was maintained with 0.1 M NaOH solution at a flow rate of 0.5 mL/min with an injection volume of 20 µL. The relative molecular weight of lignin was determined using polystyrene sulfonate sodium salt standards (PSS, GmbH, Mainz, Germany) ranging in size from 1100 to 100,000 Da.

The sugar content was quantified using 10 mg of lignin samples via acid methanolysis. Following acid methanolysis and silylation, carbohydrate analysis was performed by gas chromatography using an HP-1 GC column (Shimadzu GC-2010 AF) equipped with a flame ionization detector (Shimadzu, Kyoto, Japan), as previously described[67]. After neutralization with pyridine, the sample was derivatized overnight with HMDS/THMS, followed by the addition of 1.0 mL of internal standard (0.1 mg/mL resorcinol in methanol). The silylated analytes were separated using hydrogen as the carrier gas at a flow rate of 1 mL/min.

2.7.4. Heteronuclear Single Quantum Coherence (HSQC) and Phosphorus-31 Nuclear Magnetic Resonance spectroscopy (^{31}P NMR)

View Article Online
DOI: 10.1039/D6GC02613C

The functional groups and interunit linkages present in lignin were characterized using ^{31}P NMR and HSQC spectroscopy. Detailed methods are described in our previous report [43]. All NMR measurements including HSQC, ^{1}H NMR, ^{13}C NMR, and ^{31}P NMR spectra were done using an Ascend Neo 500 MHz NMR spectrometer (Billerica, MA, USA) equipped with a TCI Prodigy cryoprobe head.

View Article Online
DOI: 10.1039/D6GC02613C

3. Results and Discussion

3.1. Bayesian optimization (BO) of PIL extracted lignin

To implement BO in the optimization of PIL-based lignin extraction, a predictive model was first developed and refined using data compiled from literature. Among the parameters influencing biomass fractionation, temperature and pretreatment time are widely recognized as the most influential. Their combined effect is commonly expressed through a severity factor, a mathematical construct that quantifies the integrated impact of temperature and time during pretreatment. Originally introduced in 1987 [68], the severity factor has become a standard metric for characterizing biomass pretreatment processes. Subsequent studies have extended this concept to various lignocellulosic fractionation strategies, including autohydrolysis and alkaline lignin extraction [69,70,71]. More recently, the severity factor often referred to as the P-factor in the context of organic solvent-based processes that has been employed to describe the extraction of lignin and lignin–carbohydrate complexes (LCCs) from biomass using different organic solvents [64,27,72]. Within this framework, the P-factor serves as a key variable that integrates the combined effects of time and temperature, thereby enabling systematic comparison and optimization of fractionation conditions.

Notably, the P-factor was originally developed to model hemicellulose removal, and its application to direct PIL-based lignin extraction remains largely unexplored. To the best of our knowledge, this study represents the first attempt to correlate the P-factor with lignin yield in a PIL solvent system. Given the limited availability of prior work in this area, an iterative batch-based validation framework was adopted (detailed analysis and datasets are provided in the *Supplementary Information*). The Bayesian optimization progression is shown in **Fig. 4A** and **B** for the two lignin-yield datasets. Batch 0 represents the initial GP model trained using the first four observations, which are shown as white circles to distinguish them from BO-selected experiments. In subsequent batches, all observations available before the current acquisition step are shown as black circles, including the initial observations after batch 0, while the newly selected BO acquisitions are shown as lime-green circles. For each batch, the top row presents the GP-predicted lignin yield across the temperature P-factor space, with warmer colors indicating higher predicted yield and cooler colors indicating lower predicted yield. The bottom row presents the corresponding predictive uncertainty, expressed as σ in lignin-yield units. Regions with higher σ indicate areas where the

View Article Online
DOI: 10.1039/D6GC02613C

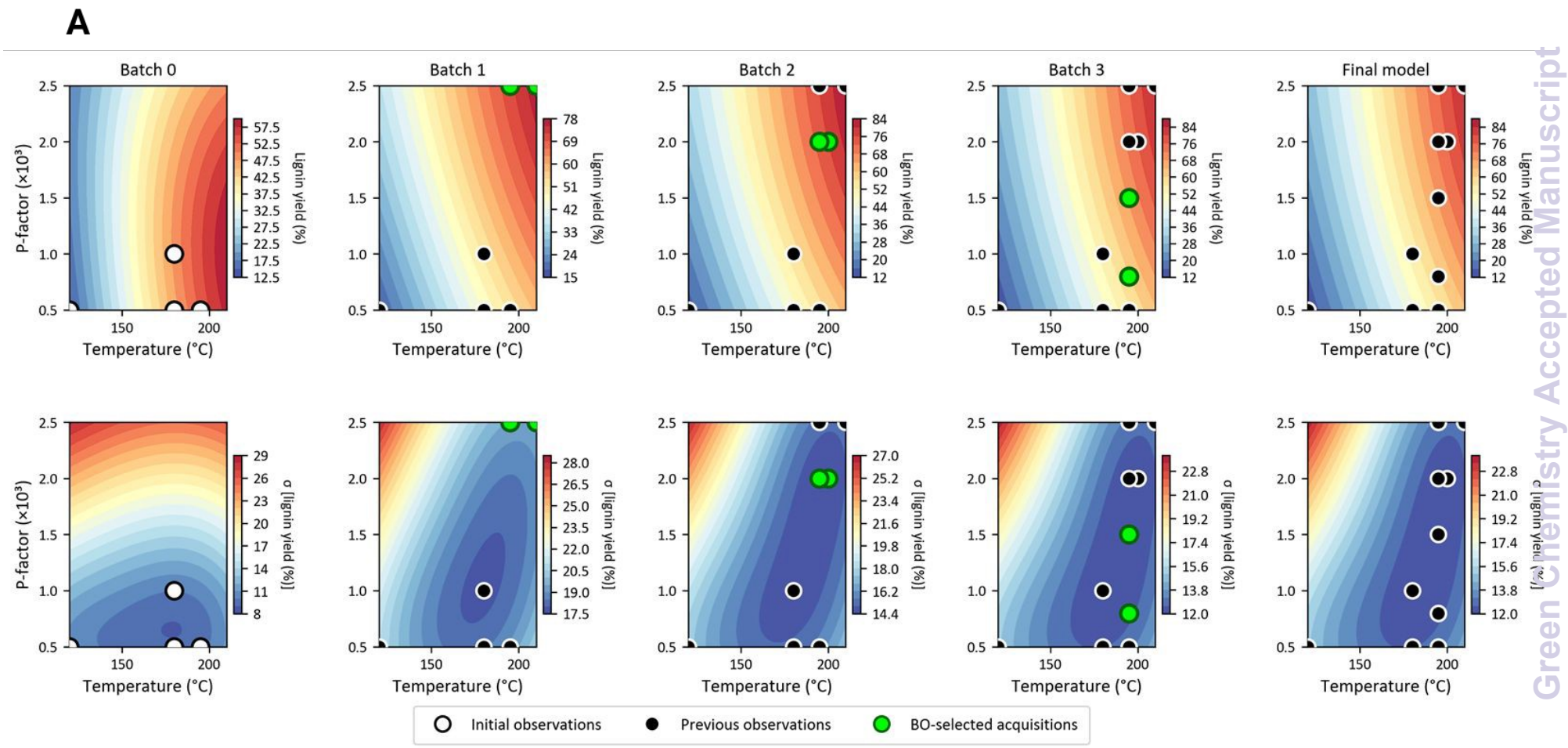
model has less information, whereas lower σ regions occur near sampled experimental conditions. As additional BO-selected observations are incorporated, the GP model is updated, uncertainty decreases around sampled regions, and the predicted yield landscape becomes progressively refined. The final column represents the final model after all observations have been incorporated, rather than an additional BO acquisition batch. This sequential process demonstrates how BO can guide the selection of informative process conditions while reducing the number of experiments required to identify promising lignin-yield regions.

The theoretical data presented in **Fig. 4A** were compiled from recent literature and illustrate the effects of the P-factor and temperature on lignin recovery. According to the results, the optimal pretreatment temperature for maximizing lignin yield during subsequent organosolv extraction using solvents such as ethanol, acetone, or aqueous solvent mixtures was 200 °C, with a P-factor of 2000.

However, these results are not directly comparable to the results obtained with a PIL, as shown in **Fig. 4B**. In the PIL process, the ionic liquid serves as a catalyst in the pretreatment, thus the optimal pretreatment temperature range for lignin extraction is between 170 and 180 °C. Moreover, when higher temperatures are employed with PIL, the reported lignin yield exceeds 100%, which is attributable to the co-extraction and condensation of hemicellulose and cellulose with lignin, as previously documented [17].

A major challenge associated with literature data is heterogeneity, as illustrated in **Fig. 4A**. The collected P-factor and temperature values, along with their corresponding lignin yields, vary considerably due to differences in solvent systems, biomass types, and extraction conditions across studies. This variability complicates the establishment of direct correlations. Nonetheless, the literature data clearly indicate the importance of temperature and its role in lignin extraction.

To address this variability and better understanding of the temperature dependence, the model simplifies the analysis by focusing on the fundamental relationship between lignin yield and the two primary controllable variables P-factor and temperature within the PIL extraction framework.



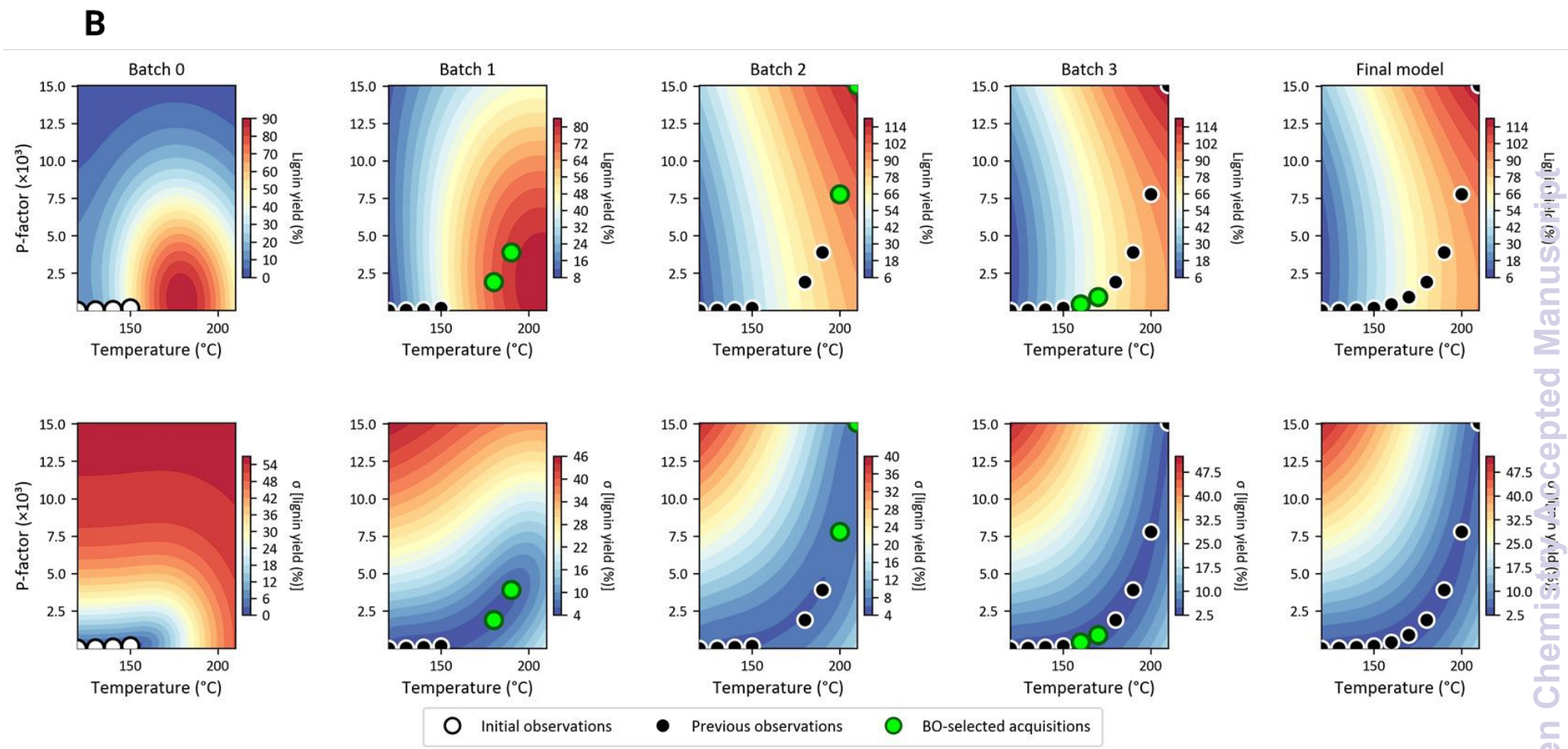


Fig. 4 Bayesian optimization of lignin yield in four batches (0, 1, 2 and 3), where 0 corresponds to the initial batch (white dots) and the predicted values (green dots). (**A**) Data gathered from previous studies to compare the effects of temperature and the P-factor, and (**B**) PIL-based lignin extraction process.

View Article Online
DOI: 10.1039/D6GC02613C

An in-depth iterative experimental series was subsequently conducted to focus on quantitative lignin yield across ten experimental observations (see *Supplementary Information* **Table S2**), in which time was held constant while temperature was increased iteratively up to 210 °C. The input variables examined were temperature (120-210 °C) and P-factor (12-15,055), with lignin yield (11-120%) serving as the response variable. The relationship between these process parameters and lignin yield, along with the corresponding model predictions derived from the literature dataset, is presented in **Fig. 4B**. The results identified a critical processing region near a P-factor of 2,000. A notable lignin yield (approaching 73%) with minimal carbohydrate impurities (discussed later) was achieved at a P-factor of approximately 1,902.

Although higher lignin yields (up to 78%) were also observed at elevated P-factors, these conditions were not employed in this work for two primary reasons. First, experimental results indicated that increased recovery levels led to higher residual PIL and saccharide contamination in the extracted lignin, resulting in greater filter residue and reduced lignin purity as shown in **Fig. 5**. Specifically, at a lignin yield of 73%, a notable increase in filter residue was observed during ethanol-based lignin recovery, a trend that became more pronounced at higher yields. This issue is further addressed later in the text. Second, consistent with previous studies, elevated temperatures are known to promote lignin condensation reactions, which can alter the native lignin structure and limit its potential for downstream valorization [17,19,35,73]. In particular, elevated temperatures during biomass fractionation can induce lignin condensation through the formation of LCCs, where lignin forms covalent linkages with hemicellulose especially xylan generating recalcitrant structures [74–76]. Within the PIL fractionation process, which solubilizes both lignin and hemicelluloses during the lignin recovery, this presents a specific challenge. Higher temperatures enhance delignification but also promote LCC formation and extensive cleavage of β-O-4 linkages in native lignin [67]. Consequently, increased delignification results in the co-extraction of lignin and LCCs. As this work focuses on extracting lignin with targeted properties, a balance must be struck between the degree of delignification and the purity and recoverability of lignin in subsequent stages. Henceforth, a delignification level of 82% with a lignin recovery of 73% was selected for detailed characterization of pretreatment-related impurities and lignin properties relevant to its efficient applications.

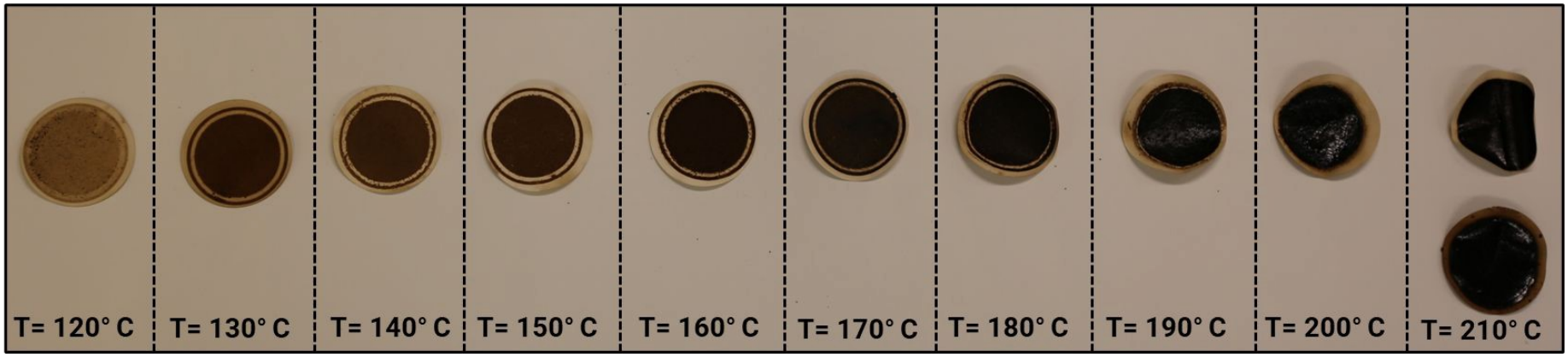


Fig. 5 Filter residues after ethanol-based lignin recovery from the PIL fractionation process. Increasing the pretreatment temperature from 120 °C to 210 °C significantly raises the p-factor, resulting in a larger amount of filter residue.

View Article Online
DOI: 10.1039/D6GC02613C

In summary, the influence of P-factor and temperature on the structural properties of lignin can be explained, both quantitatively and qualitatively, by well-established principles of lignin chemistry and by relevant literature data. However, the experimentally optimized results reported in this study require advanced characterization to enable detailed qualitative analysis, particularly regarding the role of the PIL in lignin solubilization during fractionation. Similarly, as a next step, optimization of time and PIL/biomass ratio as independent variables will be required to establish correlations between the lignin yield, cellulose retention and sidestreams valorization. Nevertheless, challenges remain in defining an optimal lignin yield, as the preferred outcome "whether lower molecular weight lignin with minimal carbohydrate impurities or higher molecular weight lignin with higher yield" depends on the specific application (discussed later in the text). In this context, our surrogate model landscapes provide the quantitative predictions necessary for large-scale lignin valorization, particularly for producing lower molecular weight lignin with reduced carbohydrate content.

Accordingly, this study represents a significant step in applying BO to elucidate the influence of temperature on PIL-based fractionation of pine wood biomass. BO enabled systematic exploration of the trade-off between maximizing lignin yield and preserving lignin quality. The results further confirm the applicability of the P-factor concept to PIL-based lignin extraction.

3.2. Influence of BO optimized condition on lignin structure and its properties

The BO-optimized process achieved 82% delignification at 180 °C with a P-factor of 1,902, corresponding to a lignin recovery of 73%, as shown in **Fig. 6** (Mass balance of the optimize lignin recovery process). Lignin removal exhibited a strong correlation with temperature, as discussed previously, while the reaction time was held constant at 1.5 h. Gschwend et al. reported a correlation between delignification and enzymatic saccharification yield for softwood biomass when using [TEA][HSO_4] [17]. In addition to the results reported by Gschwend et al. in this study, the increase in lignin content observed at higher temperatures (above 180 °C) during the later stages of pretreatment has been attributed to the formation of pseudo-lignin, which exhibits increased in situ depolymerization of recovered lignin, during which β-O-4 bonds are cleaved and lignin conjugates with hemicellulose-derived components [77]. These processes contribute to the generation of high-molecular-weight lignin, which is indeed difficult to characterize and valorize for targeted applications, as discussed in our previous report [67]. Similarly, at higher pretreatment

View Article Online
DOI: 10.1039/D6GC02613C

temperatures, condensed polymers may precipitate onto the pulp surface a phenomenon commonly referred to as pseudo-lignin formation, which predominantly originates from lignin–hemicellulose conjugation.

However, conventional compositional analysis techniques do not distinguish between native lignin and pseudo-lignin. Therefore, pseudo-lignin formation is strongly suspected under conditions where the amount of precipitated lignin exceeds the measured delignification. At this condition, the apparent yield exceeded 100% (see *Supplementary Information*), particularly for PIL-based pretreatments conducted above 180 °C [17,44,78]. This reprecipitation of pseudo-lignin onto the cellulose surface is known to negatively affect saccharification yields and kinetics by reducing the accessibility of the cellulose substrate [19,73]. Altogether, to better understand the effects of temperature and pretreatment associated impurities, several analytical techniques including GPC, ATR-FTIR, HSQC, ^{13}C and ^{31}P NMR, and TGA were employed to characterize lignin properties and its tailored applications [17,44].

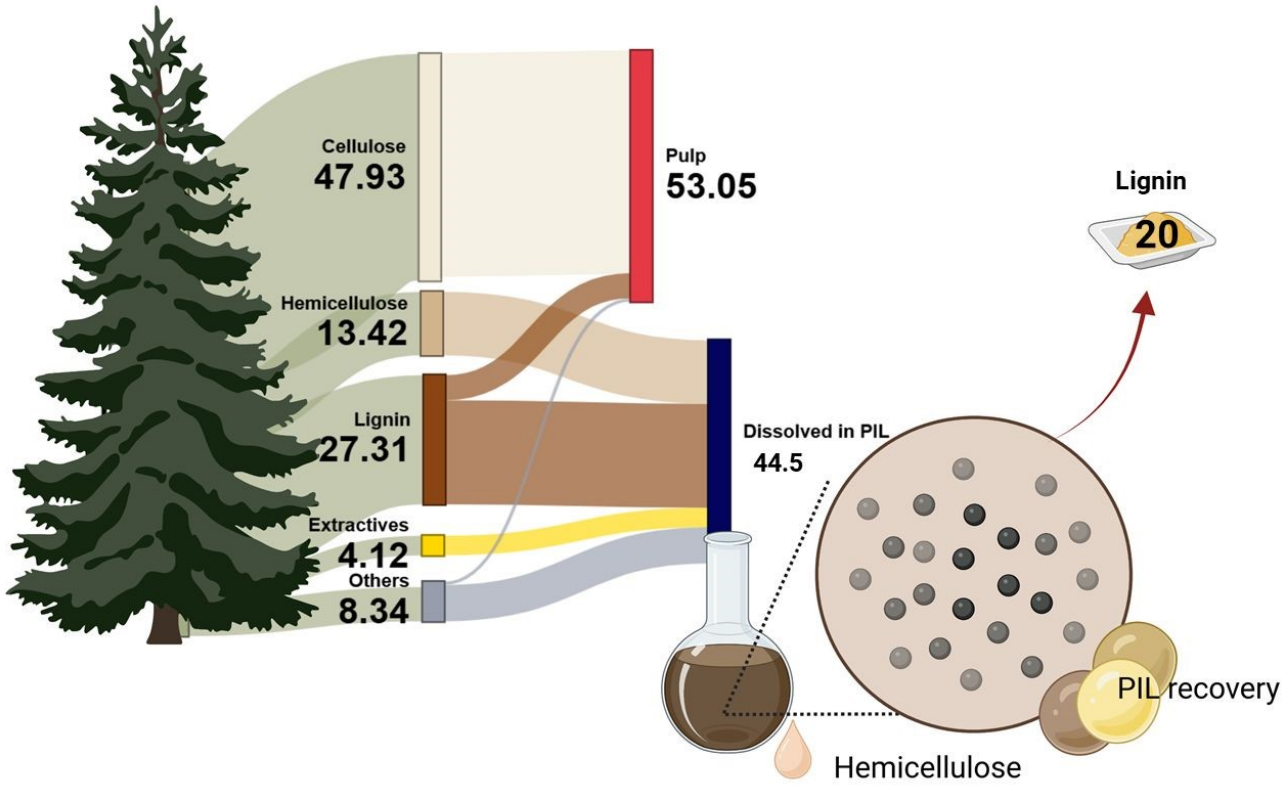


Fig. 6 Mass balance of PIL fractionated softwood pine.

View Article Online
DOI: 10.1039/D6GC02613C

GPC analysis of the optimized PIL-extracted lignin revealed lower average molecular weight (Mw) and polydispersity index (PDI) values compared to commercial kraft lignin, as summarized in **Table 2** and **Fig. 7A**. These differences are likely attributable to the PIL fractionation and lignin recovery process, particularly the ethanol washing step, which may bias the recovered lignin fraction toward lower molecular weight components rather than directly altering the lignin structure under higher pretreatment temperatures [79].

Table 2 Average molecular weights of lignin's.

Sample	M_n (g mol^{-1})	M_w (g mol^{-1})	Đ
kraft lignin	601	4585	7.63
PIL extracted lignin	850	2760	3.25

M_n - number average molecular weight, M_w - weight average molecular weight, *Đ* - polydispersity index (M_w/M_n)

The GPC results exhibited trends like those described in the ML optimization and were consistent with the delignification selection criteria discussed previously. This interpretation is further supported by studies demonstrating that sulfate-based PILs including the PIL used in the present study (**Table 1**) can dissolve up to 70 wt% of lignin prior to ethanol extraction. Moreover, ethanol has considerably lower lignin solubility compared to acetone and solvent-water mixtures, which may also contribute to the lower molecular weight observed in lignin from the PIL process [79]. Similarly, pseudo-lignin formation may be associated with the washing step, which promotes the precipitation of larger, ethanol-insoluble fragments, leading to pseudo-lignin deposition on the cellulose surface solution. This observation aligns with the apparent trend of increased ether cleavage at higher pretreatment temperatures. Overall, the decrease in Đ and M_w is attributed to the increase in hydrogen bonding between ethanol and lignin, combined with optimum PIL pretreatment conditions [80]. Therefore, lower molecular weight lignin fractions are more soluble in ethanol while higher molecular weight remained in the filtrate, as observed in this study (**Fig. 5**), resulting in a narrower molecular weight distribution in the isolated lignin. Moreover, the sugar content in the optimized lignin was found below 2 %, whereas the filter residue contained over 7 % sugars, with xylose as the predominant component, accounting for almost 40 % of the total sugars. Additionally, approximately 3 % of the impurities in the lignin were found to originate directly from the PIL insertion process due to the higher pretreatment temperature, specifically

View Article Online
DOI: 10.1039/D6GC02613C

from triethylamine[67]. Taking together, these findings suggest that while delignification efficiency and the intrinsic structure of the extracted lignin are largely affected at higher temperatures, both lignin yield and the molecular weight of the isolated lignin are also influenced by the recovery process.

ATR-FTIR spectroscopy was employed to identify the functional groups present in optimized lignin extracted using PIL and to compare them with those in commercial kraft lignin (KL) as shown in **Fig. 7C**. Absorption bands in the range of 3550–3250 cm^{-1} were attributed to the O-H stretching vibrations of aromatic and aliphatic groups in lignin [81]. Shoulder peaks were observed at approximately 2920 and 1453 cm^{-1}, corresponding to the symmetric and asymmetric C–H vibrations of the methylene and methyl groups, respectively [48]. The intense peak at 1604 cm^{-1} in the PIL-extracted lignin indicates carbonyl (C=O) vibrations conjugated with the aromatic skeletal structure [23]. The broad peak at 1250 cm^{-1} was assigned to the C–O stretching of the guaiacyl unit, whereas in KL, the presence of syringyl units was confirmed at 1220 cm^{-1} [23]. The peaks at 1080 cm^{-1} and 916 cm^{-1} are attributed to C-O-C stretching with aromatic C-H in-plane deformation and aromatic C-H out-of-plane bending vibrations, respectively [82]. These observations indicated that the extracted lignin primarily contained guaiacyl units, whereas the absence of condensed guaiacyl at 1362 cm^{-1} suggested that the extracted lignin was free from condensation.

TGA revealed an initial weight loss (stage I) between 100 and 150 °C (**Fig. 7B**), corresponding to approximately 5 % for PIL lignin, which is attributed to the evaporation of physically bound water molecules [83]. In the second pyrolysis stage, PIL lignin exhibited more significant mass loss and reduced thermal stability. Similarly, PIL lignin underwent gradual pyrolysis between 200 and 435 °C (stage II and III), reaching a maximum weight loss of 76 %. At temperatures above 450 °C (stage IV), no significant loss was found. The greater degradation observed at stages II and III for PIL lignin resulted in a small solid residue, indicating fewer unreacted by-products [84,85], which is advantageous for reducing waste and improving process efficiency. Similarly, the lower stability of PIL lignin can be attributed to its lower molecular weight. TGA results therefore suggest that the PIL fractionation produces lignin with lower molecular weights and higher volatility, as reflected by its lower decomposition temperature 288°C and greater overall mass loss (~76 %). Consistent with GPC, ATR-FTIR, TGA observations, elemental analysis showed a carbon content exceeding 70 % in PIL lignin compared with KL, as presented in **Fig. 7D.**

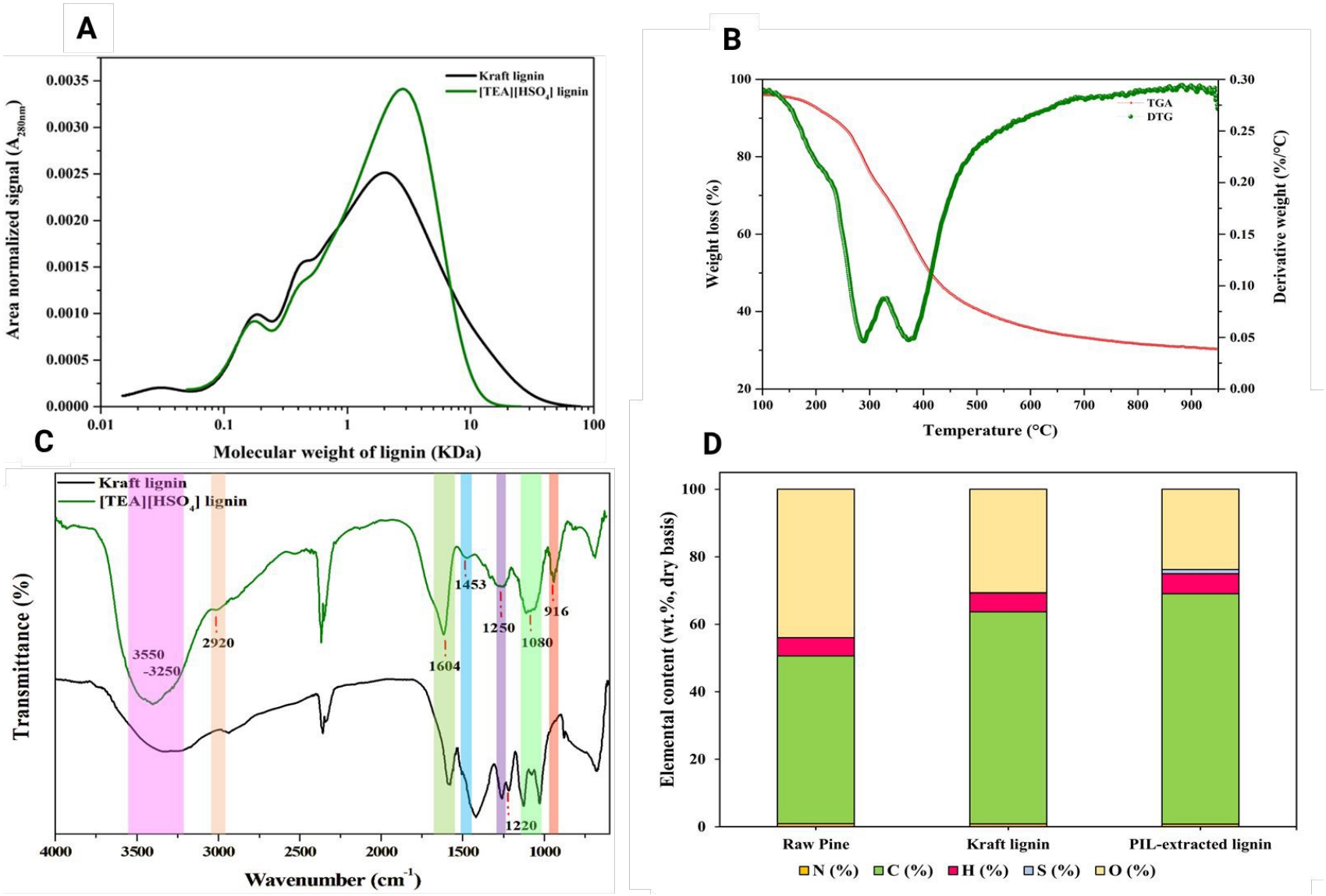


Fig. 7 (**A**) GPC, (**B**) TGA, (**C**) ATR-FTIR spectra, and (**D**) elemental analysis of different lignin's.

View Article Online
DOI: 10.1039/D6GC02613C

HSQC NMR analysis revealed minor changes in the chemical functionality of recovered lignin compared with the conventional lignin's (detailed **Table S3**, **S4** and **Fig. S1** in the *Supplementary Information*). In the aromatic region of HSQC spectra (**Fig. 8B**), the inter-unit linkages at δ(13C), δ(1H) 110.9/6.91, 114.9/6.86, and 115/6.8 ppm correspond to C_2–H_2, C_5–H_5, and C_6–H_6 of guaiacyl (G) units, respectively. Correlations at δ(13C), δ(1H) 110.5/7.2 and 124/7.1 ppm were assigned to C_2–H_2 and C_6–H_6 of oxidized guaiacyl (G′) units. The β-5 (**SB5**) linked trans-stilbene exhibited an α signal at δ(13C), δ(1H) 128.2/7.20 ppm and a β signal at δ(13C), δ(1H) 120.1/7.22 ppm, in contrast, the trans-stilbene with a β-1 linkage (**SB1**) exhibits only its α signal at δ(13C), δ(1H) 125.6/6.97 ppm. Similarly, in the **aliphatic region**, the methoxy groups of guaiacyl units were observed at δ(13C), δ(1H) 56/3.8 ppm, whereas signals at δ(13C), δ(1H) 8-52.5/0.5-4.5 correspond to alkyl side chains in the extracted lignin [86]. A decrease in ether linkage content (**Table 3**) was observed in the isolated lignin, accompanied by an increase in carbohydrate (**C**) peaks assigned to **C1** (δ(13C), δ(1H) 102.3/4.30), **C2** (δ(13C), δ(1H) 73.2/3.08), **C3** (δ(13C), δ(1H) 74.6/3.29), **C4** (δ(13C), δ(1H) 75.9/3.54), and **C5** (δ(13C), δ(1H) 63.8/3.92 and 3.23) [28,87]. Comparison of the assigned peaks showed that **C2**, **C3**, and **C5** were clearly visible in the HSQC spectra, suggesting that pseudo-lignin formation began at this yield, a finding consistent with experimental quantification. The appearance of carbohydrate peaks may result from an increased concentration of dissolved hemicellulose fragments in the PIL-lignin solution.

To further understand the reactions occurring during extraction, the isolated lignin was analyzed for subunit composition and interlinkage distribution, where ^{31}P NMR analysis identified aliphatic and phenolic OH regions (**Table 3**). Signals in the range δ(31P) 145.8-149 ppm correspond to aliphatic hydroxyl (R-OH) groups, whereas phenolic OH groups appeared within δ(31P) 137.5-145 ppm, indicating β-O-4 bond cleavage and limited condensation through Cα-Caryl linkages. These reactions lead to the formation of free phenolic structures, particularly guaiacyl OH at δ(31P) 137.5-142 ppm, which showed high signal intensity in the PIL-extracted lignin. In addition, ^{31}P NMR (**Fig. 8A**) revealed carboxylic acid groups (-COOH) within δ(31P) 134-136.6 ppm, likely formed through oxidation reactions of alkyl side chains [35,37].

Altogether, the ^{31}P and ^{13}C NMR results show that a substantial fraction of β-O-4 ether linkages, the most abundant linkages in native pine wood was cleaved after pretreatment. In contrast, ether bonds associated with the less abundant β-β′ and β-5′ linkages described in **Table 3** degraded more

View Article Online
DOI: 10.1039/D6GC02613C

slowly, consistent with previous observations. Higher pretreatment temperatures increased the in situ depolymerization of lignin, as evidenced by reduced G6 and G2 signal intensities. Furthermore, when comparing the ^{31}P and ^{13}C NMR results with the HSQC spectra, it becomes clear that increasing lignin yield beyond this point leads to extensive lignin modifications, which in turn complicates lignin characterization and the understanding of its properties.

The combination of HSQC and ^{31}P NMR analyses therefore provides detailed structural insight into the lignin units and inter-unit linkages present in lignin extracted from pine biomass. This critical analysis advances lignin-carbohydrate characterization and provides valid information on how PIL-soluble lignin depolymerizes in situ (via β-O-4 bond cleavage) during pretreatment. Henceforth, to further advance the understanding of PIL-lignin interactions and support these findings, molecular dynamics simulations were performed to relate the results to lignin applications.

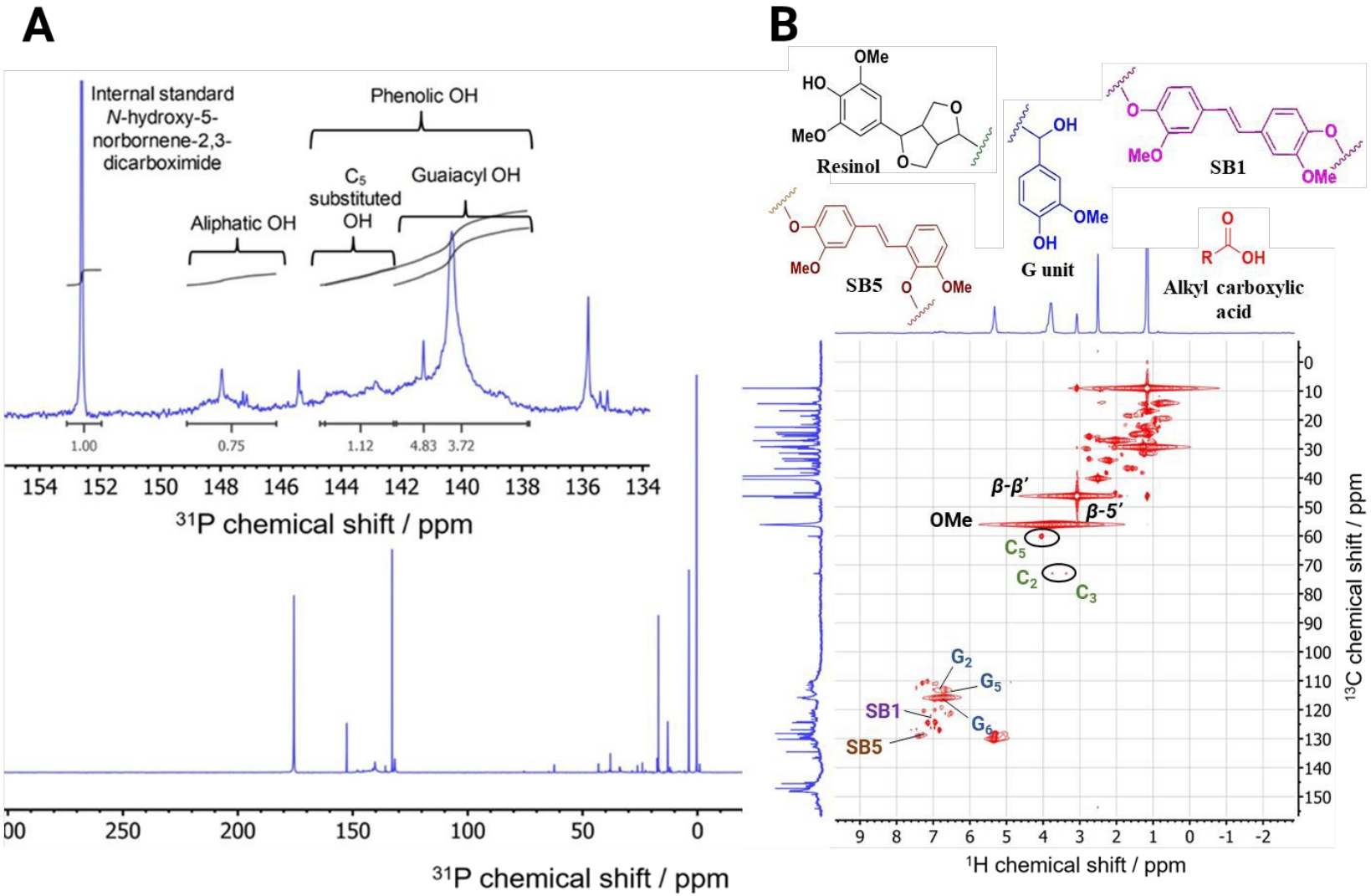


Fig. 8 (**A**) ^{31}P and (**B**) HSQC NMR spectra of PIL-extracted lignin.

Table 3 Quantitative ^{31}P (total hydroxyl groups) and ^{13}C NMR of PIL-extracted lignin.

Properties	**Lignin backbone composition**				
Hydroxyl content [a]	Aliphatic OH (mmol g^{-1})	Phenolic OH (mmol g^{-1})			Total OH (mmol g^{-1})
		C_5 substituted	Guaiacyl	Total Phenolics	
	0.46	0.62	1.95	2.37	3.03
Interunit linkages [b] (Abundance/100 aromatic units)		Phenylcoumaran (β-5')	Pinoresinol $^{\alpha}$ (*β-β′*)	Pinoresinol $^{\alpha'}$ (*β-β′*)	
		0.03	0.01	0.01	
Aryl-vinyl moieties [b] (Abundance/100 aromatic units)		Stilbene $^{\alpha}$ (*β-5′*)	Stilbene $^{\beta}$ (*β-5′*)	Stilbene $^{\alpha}$ (*β-1′*)	Cis-enol ether $^{\alpha}$	Cis-enol ether $^{\beta}$
		0.02	6.41	1.81	2.16	0.05
Side chain structure in end groups [b] (Abundance/100 aromatic units)		Dihydrocinnamyl alcohol $^{\alpha}$		Dihydrocinnamyl alcohol $^{\gamma}$	Aryl-glycerol $^{\gamma}$	
		0.02		0.08	0.02	
Lignin polysaccharide complex linkage [b] (Abundance/100 aromatic units)				Benzyl ether $^{\alpha}$		
				0.01		

[a] Quantification using ^{31}P-NMR; [b] Calculated using a quantitative combination of HSQC and ^{13}C-NMR; $^{\alpha/\beta/\gamma}$ represent assignments using ^{13}C and ^{1}H.

View Article Online
DOI: 10.1039/D6GC02613C

3.3. Molecular dynamic simulations of PIL and lignin

Molecular modeling and predictive screening have become essential approaches for addressing the inherent recalcitrance of lignocellulosic matrices, which are primarily composed of cellulose, hemicellulose, and lignin, a complex three-dimensional aromatic polymer [52]. An important advancement in these simulations is the transition from simple monomeric models, such as coniferyl alcohol, to a representative polymeric lignin structure (~5 kDa) incorporating major linkages, such as β-O-4, β-β, 4-O-5, α-O-4, and β-5. Monomeric models do not adequately reflect the structural heterogeneity and extensive hydrogen-bond networks characteristics of native lignin (**Table 4**). The lignin model structure used for MD simulations in this study is presented in **Fig. 9**. Benchmarking studies demonstrate that longer polymeric models provide significantly improved correlations with experimental solubility data, yielding R^2 values of up to 0.88 for activity coefficients and 0.78 for excess enthalpy, compared with substantially lower accuracy for monomeric or dimeric structures [52,88–90].

Table 4 Lignin and its bonds-attributed model.

Model	Softwood lignin
H/G/S composition	0-5/95-100/0
Molecular weight	~ 8.5 kDa
Linkage composition	
β-O-4	21 (~46 %)
β-5	7
5-5’	6
β-1	5
α-O-4	4
4-O-5	2
β-β	1

Effective anions typically exhibit negative σ-potentials in regions of positive charge density, facilitating strong interactions with the lignin surface. A key principle derived from these simulations is that the interaction energy between the anion and cation of the IL must be weaker

View Article Online
DOI: 10.1039/D6GC02613C

than the interaction energy between the anion and lignin, thereby enabling the ions to solvate the polymer [52].

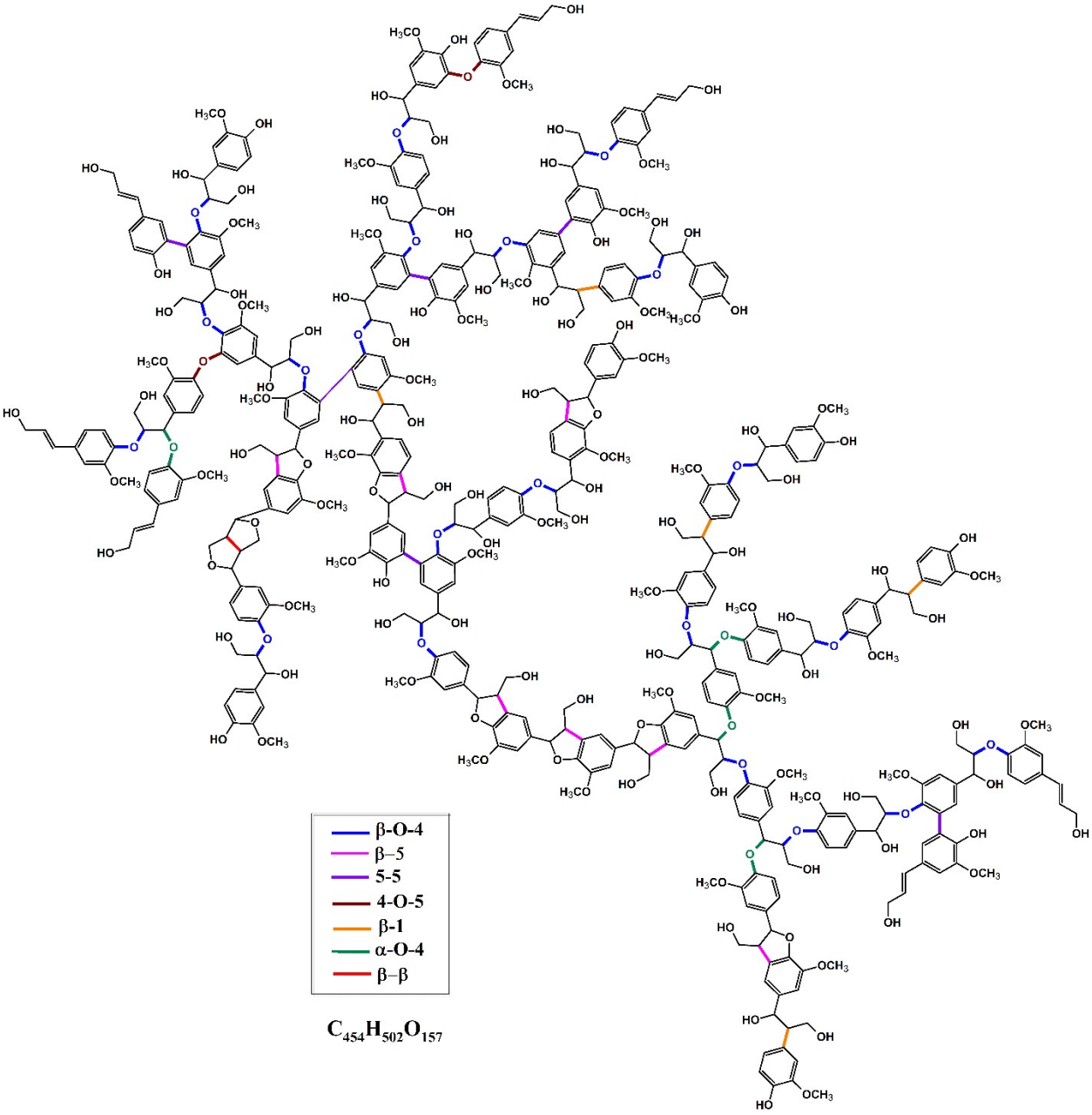


Fig. 9 Lignin model structure used for MD simulations.

To elucidate lignin behavior in a PIL environment, we analyzed time-dependent structural stability, conformational sampling, hydrogen-bonding patterns, and energy contributions. These descriptors link molecular interactions with polymer expansion and solvation behavior. The

View Article Online
DOI: 10.1039/D6GC02613C

RMSD profile (**Fig. 10A**) gradually increases and stabilizes at approximately 0.55-0.65 nm, indicating substantial conformational rearrangement of lignin before reaching a dynamically stable configuration in the PIL. This behavior reflects adaptation of the polymer to the ionic environment rather than its structural instability. Concurrently, the radius of gyration (Rg) increases steadily from 1.45 to approximately 1.60-1.65 nm throughout the trajectory, indicating that lignin adopts a more expanded conformation in [TEA][HSO_4]. These values are consistent with those observed in effective solvents, where lignin typically exhibits Rg values between 15.75 Å and 18.85 Å, whereas in poorer solvents such as water Rg values are generally below 13 Å [80,91].

The 2D projection obtained from principal component analysis (PCA) revealed a broad U-shaped conformational landscape (**Fig. 10B**). This indicates that the lignin polymer explores multiple conformational states during the simulation rather than remaining trapped in a single basin. Such extensive sampling suggests that lignin remains flexible in the PIL and is not kinetically trapped into a rigid configuration. Continuous migration across the phase space indicates stabilization through solvent interactions rather than by intrachain hydrogen bonding, consistent with the increasing RMSD and Rg values shown in **Fig. 10A**.

Hydrogen-bond analysis revealed a clear predominance of intermolecular hydrogen bonds between lignin and PIL ions, fluctuating between approximately 55-70 throughout the simulation trajectory. In contrast, the intramolecular lignin-lignin hydrogen bonds remained significantly weaker (~12-20). This imbalance demonstrates that lignin preferentially forms hydrogen bonds with the ionic liquid rather than with itself. Consequently, the dissolution process is characterized by disruption of intramolecular (intrachain) hydrogen bonds and their replacement with lignin-PIL interactions. These transitions explain the polymer expansion observed in the RMSD and Rg profiles, as internal hydrogen bonds are disrupted, the polymer structure opens and becomes more accessible to the solvent.

Energy decomposition analysis shows that Coulombic interactions dominate the lignin-PIL system, particularly interactions between lignin and HSO_4^- (HS) anions. The total interaction energy for lignin-HS is substantially more negative than that for lignin-TEA^+, indicating that the anion plays a primary role in stabilizing lignin in solution. Although van der Waals interactions contribute to the overall stabilization, their effect is secondary compared with electrostatic and

View Article Online
DOI: 10.1039/D6GC02613C

hydrogen-bonding interactions. The strongly negative total interaction energy for the combined Lignin-TEA-HS system confirms that solvation is energetically favorable, consistent with the hydrogen-bonding trends observed in **Fig. 10C**.

The RDF profiles (**Fig. 10E** and **10F**) for lignin-TEA hydrogen bonding, a small but distinct first peak appears near ~2 Å, corresponding to interactions between lignin oxygen atoms and the ammonium hydrogen of TEA. However, this peak is less pronounced than that observed for anionic interactions; the aliphatic Oxygen (O) of lignin shows an intensity of approximately 0.28, while interactions between aromatic O in lignin with H- in TEA are negligible. A second maximum at approximately 5.5 Å, indicating that solvation is distributed along the lignin backbone rather than localized at specific sites. In contrast, RDFs for lignin-HSO_4^- interactions display sharper and stronger maxima at shorter distances (~1.9-2.0 Å), indicating the presence of a well-defined coordination shell around lignin oxygen atoms. These results are consistent with previous observations that PIL anions form a dense hydrogen-bond network around lignin functional groups. The presence of this short-range, highly ordered solvation shell explains the dominance of the Coulombic interactions observed in **Fig. 10D** and the high contact probability between lignin and PIL ions. In effective PIL system, this contact probability can be up to one order of magnitude greater than that observed in weaker solvents, thereby driving lignin solubility.

In summary, MD simulations primarily represent the early-stage PIL-lignin interaction, and limitations of using a static native model for the full extraction process would be needed to precisely acknowledge the lignin modifications at dynamic environments. Altogether, the findings propose that the anion plays a key role in lignin solubilization, making its selection a cornerstone of the process. Similarly, concluding these results is challenging because, under dynamic and optimized pretreatment conditions, lignin undergoes simultaneous reactions i.e., with ethanol during the recovery process, associated impurities from hemicellulose and extractives, severity of the process and complex biomass components condensation reactions, which obscures specific lignin-PIL interactions. Therefore, future studies should eliminate impurities such as those from pre-hydrolysis (which removes hemicellulose) and residual extractives since simulating the full complexity of biomass with all impurities is theoretically prohibitive. Our findings further indicate that PIL-based biomass fractionation extracts lignin that resembles that obtained from

View Article Online
DOI: 10.1039/D6GC02613C

conventional Organosolv-based recovery methods. In this context, extensive literature supports the applications of PIL-extracted, lower-molecular-weight lignin, as discussed later in the text.

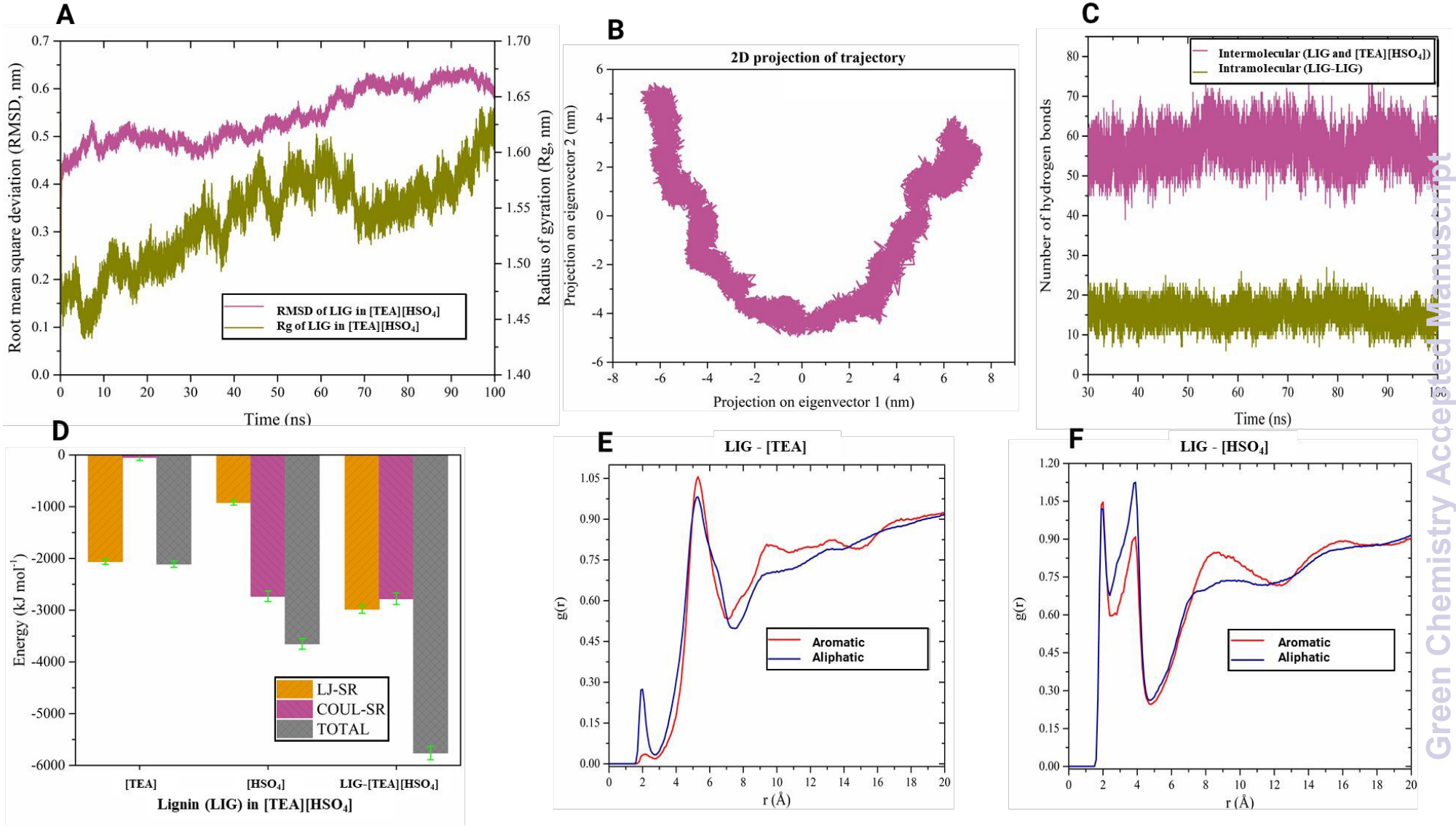


Fig. 10 (A) Root mean square deviation (RMSD), **(B)** 2D projection of trajectory, **(C)** inter- and intra-molecular interactions, **(D)** Coulombic energy in panel, and **(E)** and **(F)** are the RDF profiles for lignin-PIL hydrogen bonding.

View Article Online
DOI: 10.1039/D6GC02613C

3.4. Fate of lignin after PIL separation

The chemical structures of lignin's recovered from different pretreatments are characterized by a complex three-dimensional macromolecular architecture formed through the radical coupling of three primary phenylpropane monolignols: p-coumaryl (H), coniferyl (G), and sinapyl (S) alcohols. The distribution of these units depends on the botanical source. Where the softwood lignin is predominantly composed of G units, hardwood lignin contains both G and S units, and grass lignins incorporate a mixture of all three [6]. To convert lignin into value-added products, it must first be separated from cellulose and hemicellulose through extraction processes that generate "technical lignins".

PIL-separated lignin (IonoSolv lignin), such as that obtained in this study, often undergoes in situ depolymerization during pretreatment, thereby reducing the need for additional chemical modification [43]. This structural transformation during extraction positions PIL-derived lignin as a promising candidate for integrated downstream applications, a concept recently highlighted by Yang et al. [92]. Similarly, Zhang et al. [93] proposed a multiple-freezing strategy to incorporate lignin/ionic liquid into a dual-cross-linked network of polyvinyl alcohol and gelatin, enabling the preparation of a multifunctional green composite hydrogel. Furthermore, technical lignin's can serve as feedstock for three principal categories of chemical modification: fragmentation (depolymerization), synthesis of new chemically active sites, and functionalization of existing hydroxyl groups for targeted applications [94]. However, economic feasibility of modified-lignins[95] and its applications is a concerning fact which need market acceptability [96,97].

Therefore, we compile and compare different lignin applications in **Table 5**. Interestingly, most technical lignin-based applications require modification strategies. For example, fragmentation aims to depolymerize lignin into low-molecular-weight aromatics such as vanillin, which is currently the only lignin-derived chemical produced at significant industrial scale [98]. Another strategy involves the introduction of new reactive sites that impart functionality not present in native lignin [99], One such example is hydroxyalkylation, enhancing lignin reactivity and enabling partial substitution for phenol in phenol-formaldehyde resins [100]. Lastly, the functionalization of existing hydroxyl groups represents one of the most versatile routes for lignin valorization [94,101].

Table 5. Modifications required for optimal lignin valorization based on recent literature.

Lignins	Modifications	Applications	References
Kraft lignin	Lignin nanoparticles (LNPs)	Digital light processing (DLP) 3D printing of lignin-polymer nanocomposites	102
	Epoxidized lignin, Colloidal lignin particles	Fire and water-resistant composite adhesives	103
	Oxypropylation	Biocomposites for packaging applications	104
	Esterification and Nanoparticles	Targeted Chemical Delivery in Plant Protection	105
	Hydroxymethylation and Nanoparticles	Organic Battery Electrodes	106
	Hydroxyethyl modification, Esterification and Fractionation	Polyethylene terephthalate	107
	Aminopropyl/ methyl silsesquioxane (WAPMSS)	Wood coating	108
Soda, Kraft, Organosolv and Aldehyde-protected lignin	Acid phenolation	Antiviral Coating Material	109
Alkaline lignin	Preoxidation	Lignin-derived hard carbon anode	110
	Polymerized lignin formulation	Barrier coating in packaging	111
Organosolv and Milledwood lignin	Formulation	Natural sunscreens	112

Depolymerized lignin	Lignin hydrogenolysis oil Chlorohydrin, Glycidyl Ether, Cyclic Carbonate	Nonisocyanate Polyurethane/Epoxy Thermoset Materials	113
RCF lignin oil	Molybdenum carbide to deoxygenate lignin	Jet-range aromatic hydrocarbons	114
Hydrolysis lignin	Depolymerization and glycidylation of lignin	Epoxy resin for metal coating applications	115
Ionosolv lignin	Spinning dope	Carbon fibres	92
Sodium lignosulfonate and kraft lignin	Emulsions	Hair conditioning	116
Aldehyde-assisted fractionated lignin	Formulation	Thermal paper	117
Aminated lignin	Amino-nanoscale or micron-scale solid particles	CO_2 capture	118

View Article Online
DOI: 10.1039/D6GC02613C

Based on the applications described in **Table 5**, for PIL-separated lignin, integrating lower-molecular-weight lignin and esterification with a modifier as described by Nelis et al., [117] and Babaeipour et al.[119], where the resulting formulations are directly applicable as coatings on fiber-based substrates for packaging applications is promising. Entirely, proposed applications require further research not only on minimum modifications but also on the effects of such modifications on biodegradability. Similarly, based on a comparative analysis of the literature and the results reported in this study, future research would focus on minimizing modification steps as shown in **Fig. 11** and on applying green chemistry strategies such as esterification for lignin-based industrial coatings, thereby advancing sustainable IonoSolv lignin valorization.

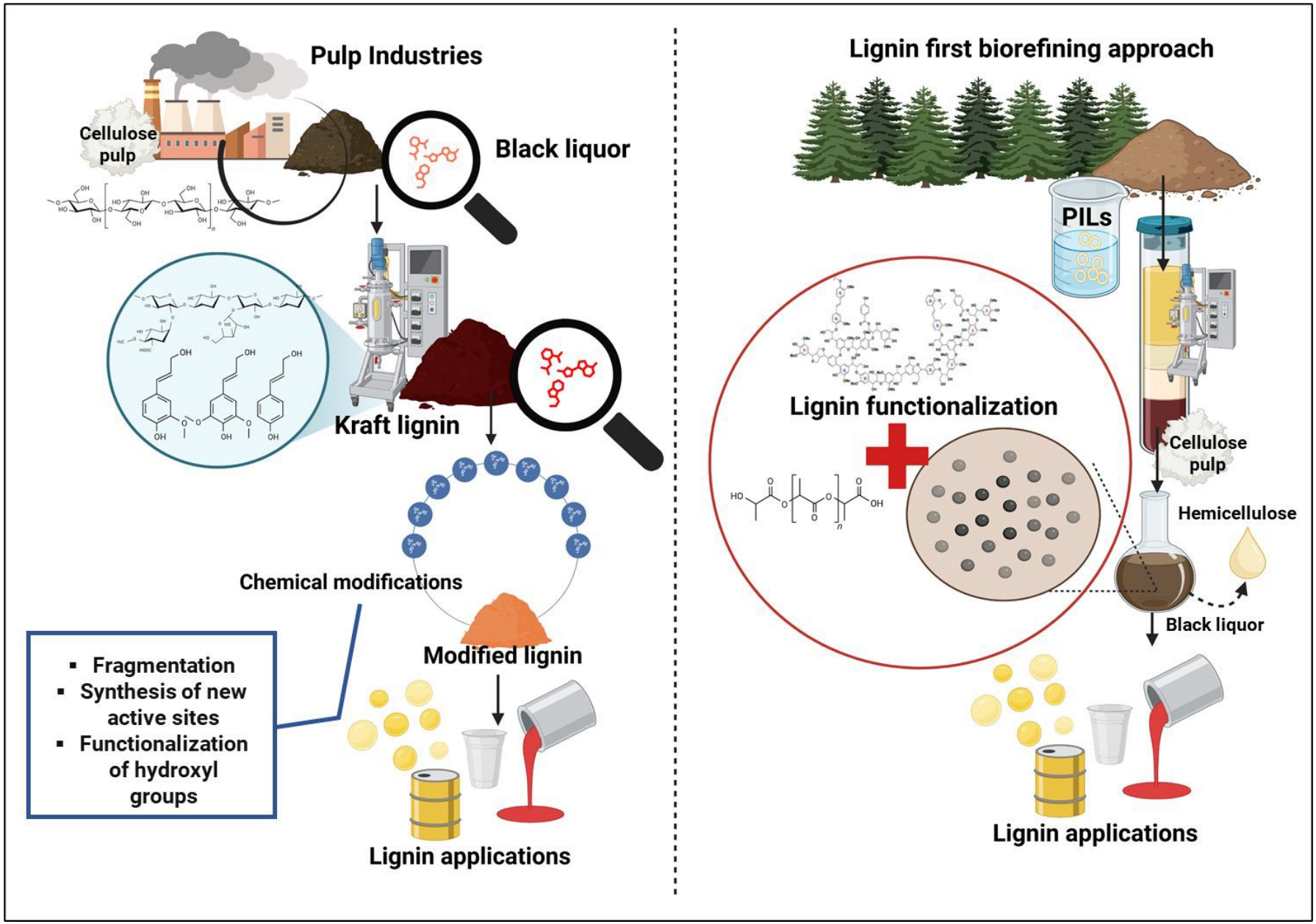


Fig. 11 Traditional kraft lignin extraction and valorization vs. lignin-focused protic ionic liquid (PIL)-biorefining approach.

View Article Online
DOI: 10.1039/D6GC02613C

4. Conclusions

This study successfully demonstrates the application of a machine learning-guided "lignin-focused" biorefinery strategy using PIL to address limitations associated with conventional lignin. By reducing reliance on intuition-driven experimentation, this approach enabled the identification of optimal processing conditions that balance the critical trade-off between lignin yield and structural quality. The optimized PIL fractionation process achieved high delignification and delivered substantial lignin recovery and consistently producing a high-quality product characterized by a low molecular weight, resulting from in situ depolymerization during fractionation. However, PIL recyclability and LCA is major limitation of the present study, which need further development in future studies. Furthermore, these findings establish a robust, data-driven framework for accelerating the development of sustainable biomass fractionation processes designed for high-value lignin valorization. Future works will also focus on expanding the computational framework by extending Python-based optimization tools to incorporate a broader range of lignin descriptors, including phenolic hydroxyl content, monomeric composition (S/G ratio), and specific functional groups, alongside yield. Integrating these parameters with application-oriented performance data obtained through collaborative research will enable the model to more effectively address the complex relationship between yield, structural quality, and application performance. Ultimately, this approach will facilitate the predictive and tailored production of lignin optimized for downstream applications, including polymers, resins, or coatings.

Author contributions

S.K.: investigation, methodology, conceptualization, data analysis, and writing the original draft. **T.L.**: investigation, methodology, formal analysis, reviewing and editing. **V.P.**: investigation, methodology, data analysis, and writing the original draft. **A.Z.**: investigation, reviewing and editing. **D.R.**: investigation, formal analysis, reviewing and editing. **U.V.**: investigation, formal analysis, reviewing and editing. **J.O.**: methodology and supervision. **M.G.**: methodology, supervision, reviewing and editing. **S.S.**: supervision, formal analysis, data curation, conceptualization, project administration, reviewing and editing. **T.K.**: supervision, data curation, conceptualization, funding acquisition, project administration, reviewing and editing.

Funding

View Article Online
DOI: 10.1039/D6GC02613C

This work was supported by the Estonian Research Council Grant **PRG2730**.

Acknowledgements

D. R. and **M. G.** would like to thank the state government of Saarland for supporting this study as part of the **EnFoSaar project**, which is funded by the Saarland Transformation Program for Research and Knowledge Transfer (Transformations programm Forschung und Wissenstransfer Saar). **S. K.** would like to thank Prof. **Jason Hallett** (Imperial College London) for the discussions about PILs and their role in lignin extraction. To Dr. **Antoine Mialon** (Neste) for discussing the role of lignin recovery in modern biorefinery technologies and to late Prof. **Mikhail Balakshin** (Aalto University) for discussing the lignin purity, its solubilization and advanced characterization using different solvents.

Availability of data and materials

The data supporting this article have been included as part of the *Supplementary Information*.

References:

1 R. Meys, A. Kätelhön, M. Bachmann, B. Winter, C. Zibunas, S. Suh and A. Bardow, *Science (80-.).*, 2021, **374**, 71–76.

2 P. Gabrielli, L. Rosa, M. Gazzani, R. Meys, A. Bardow, M. Mazzotti and G. Sansavini, *One Earth*, 2023, **6**, 682–704.

3 R. Meys, F. Frick, S. Westhues, A. Sternberg, J. Klankermayer and A. Bardow, *Resour. Conserv. Recycl.*, 2020, **162**, 105010.

4 M. Y. Balakshin, E. A. Capanema, I. Sulaeva, P. Schlee, Z. Huang, M. Feng, M. Borghei, O. J. Rojas, A. Potthast and T. Rosenau, *ChemSusChem*, 2021, **14**, 1016–1036.

5 P. Verdía Barbará, H. Choudhary, P. S. Nakasu, A. Al-Ghatta, Y. Han, C. Hopson, R. I. Aravena, D. K. Mishra, A. Ovejero-Pérez, B. A. Simmons and J. P. Hallett, *Chem. Rev.*, 2025, **125**, 5461–5583.

6 W. Schutyser, T. Renders, S. Van Den Bosch, S. F. Koelewijn, G. T. Beckham and B. F. Sels, *Chem. Soc. Rev.*, 2018, **47**, 852–908.

7 D. Di Francesco, C. Dahlstrand, J. Löfstedt, A. Orebom, J. Verendel, C. Carrick, Å. Håkansson, S. Eriksson, H. Rådberg, H. Wallmo, M. Wimby, F. Huber, C. Federsel, M. Backmark and J. S. M. Samec, *ChemSusChem*, 2021, **14**, 2414–2425.

8 D. D. S. Argyropoulos, C. Crestini, C. Dahlstrand, E. Furusjö, C. Gioia, K. Jedvert, G. Henriksson, C. Hulteberg, M. Lawoko, C. Pierrou, J. S. M. Samec, E. Subbotina, H. Wallmo and M. Wimby, *ChemSusChem*, DOI:10.1002/cssc.202300492.

View Article Online
DOI: 10.1039/D6GC02613C

9 D. M. Alonso, S. H. Hakim, S. Zhou, W. Won, O. Hosseinaei, J. Tao, V. Garcia-Negron, A. H. Motagamwala, M. A. Mellmer, K. Huang, C. J. Houtman, N. Labbé, D. P. Harper, C. T. Maravelias, T. Runge and J. A. Dumesic, *Sci. Adv.*, DOI:10.1126/sciadv.1603301.

10 M. M. Abu-Omar, K. Barta, G. T. Beckham, J. S. Luterbacher, J. Ralph, R. Rinaldi, Y. Román-Leshkov, J. S. M. Samec, B. F. Sels and F. Wang, *Energy Environ. Sci.*, 2021, **14**, 262–292.

11 L. Dessbesell, M. Paleologou, M. Leitch, R. Pulkki and C. (Charles) Xu, *Renew. Sustain. Energy Rev.*, 2020, **123**, 109768.

12 K. Melin, E. Sermyagina, J. Saari, J. Lahti, E. Vakkilainen, M. Mänttäri and M. Kallioinen-Mänttäri, *Energy*, DOI:10.1016/j.energy.2025.134886.

13 G. Semaan, A. B. Öztürk and G. Kumar, *Biochem. Eng. J.*, 2025, 109892.

14 C. L. Chambon, P. Verdía, P. S. Fennell and J. P. Hallett, *Sci. Rep.*, 2021, **11**, 1–15.

15 A. Brandt, J. Gräsvik, J. P. Hallett and T. Welton, *Green Chem.*, 2013, **15**, 550–583.

16 P. Y. S. Nakasu, T. C. Pin, J. P. Hallett, S. C. Rabelo and A. C. Costa, *Renew. Energy*, 2021, **172**, 816–828.

17 F. J. V. Gschwend, C. L. Chambon, M. Biedka, A. Brandt-Talbot, P. S. Fennell and J. P. Hallett, *Green Chem.*, 2019, **21**, 692–703.

18 A. Brandt, L. Chen, B. E. Van Dongen, T. Welton and J. P. Hallett, *Green Chem.*, 2015, **17**, 5019–5034.

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

20 L. Shuai, M. T. Amiri, Y. M. Questell-Santiago, F. Héroguel, Y. Li, H. Kim, R. Meilan, C. Chapple, J. Ralph and J. S. Luterbacher, *Science (80-.).*, 2016, **354**, 329–333.

21 F. Momayez, M. Hedenström, S. Stagge, L. J. Jönsson and C. Martín, *Bioresour. Technol.*, DOI:10.1016/j.biortech.2022.127466.

22 A. Kalliola, P. Kangas, I. Winberg, T. Vehmas, H. Kyllönen, J. Heikkinen, O. Poukka, K. Kemppainen, P. Sjögård, L. Pehu-Lehtonen and T. Liitiä, *Nord. Pulp Pap. Res. J.*, 2022, **37**, 394–404.

23 F. L. Wang, S. Li, Y. X. Sun, H. Y. Han, B. X. Zhang, B. Z. Hu, Y. F. Gao and X. M. Hu, *RSC Adv.*, 2017, **7**, 47990–47998.

24 I. Hasanov, M. Raud and T. Kikas, *Energies*, 2020, **13**, 4864.

25 M. Madadi, E. Kargaran, S. S. Hashemi, C. Sun, J. F. M. Denayer, K. Karimi, F. Sun and V. K. Gupta, *Adv. Sci.*, 2025, **12**, 1–15.

26 L. Leng, J. Gao, W. Zhang, W. Liu, L. Jiang, H. Zhan, H. Huang and H. Li, *Biomass and Bioenergy*, 2026, **205**, 108531.

27 J. Löfgren, D. Tarasov, T. Koitto, P. Rinke, M. Balakshin and M. Todorović, *ACS Sustain. Chem. Eng.*, 2022, **10**, 9469–9479.

View Article Online
DOI: 10.1039/D6GC02613C

28 D. Diment, J. Löfgren, M. Alopaeus, M. Stosiek, M. Cho, C. Xu, M. Hummel, D. Rigo, P. Rinke and M. Balakshin, *ChemSusChem*, DOI:10.1002/cssc.202401711.

29 J. Chrząstowska, T. Sokalski, E. Korotkova, J. Bobacka, A. Sundberg and C. Xu, *Holzforschung*, 2025, **79**, 453–464.

30 H. Rummukainen, H. Hörhammer, P. Kuusela, J. Kilpi, J. Sirviö and M. Mäkelä, *Heliyon*, DOI:10.1016/j.heliyon.2024.e24484.

31 F. Gisperg, R. Klausser, M. Elshazly, J. Kopp, E. P. Brichtová and O. Spadiut, *Biotechnol. Bioeng.*, 2025, **122**, 1313–1325.

32 S. Bertelsen, S. Carlsen, S. Furbo, M. B. Nielsen, A. Obdrup and R. Taaning, *J. Chem. Inf. Model.*, 2025, **65**, 1702–1707.

33 I. Sulaeva, G. Zinovyev, J. Plankeele, I. Sumerskii, T. Rosenau and A. Potthast, *ChemSusChem*, 2017, **10**, 629–635.

34 G. Zinovyev, I. Sulaeva, S. Podzimek, D. Rössner, I. Kilpeläinen, I. Sumerskii, T. Rosenau and A. Potthast, *ChemSusChem*, 2018, **11**, 3259–3268.

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

36 S. H. M. van Leuken, D. J. G. P. van Osch, P. D. Kouris, Y. Yao, M. A. Jedrzejczyk, G. J. W. Cremers, K. V. Bernaerts, R. A. T. M. van Benthem, R. Tuinier, M. D. Boot, E. J. M. Hensen and M. Vis, *Green Chem.*, 2023, **25**, 7534–7540.

37 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.

38 N. Emami, L. A. Gomez-Moreno, A. Klemettinen, R. Serna-Guerrero and M. Todorović, *Chem. Eng. J.*, 2025, **510**, 161128.

39 Z. Ba, Y. Shi, H. Zhang, A. Robiños, M. S. Zafar, Y. Li, F. Yu, J. Bobacka, D. Liang, Y. Wang, Y. Xie and C. Xu, *Appl. Energy*, 2025, **401**, 126845.

40 R. Radhakrishnan, B. Manna and A. Ghosh, *Renew. Energy*, 2023, **206**, 47–59.

41 P. Griffin and J. Kostal, *Green Chem.*, 2023, **25**, 7283–7291.

42 J. Zhang, G. Zhou, T. Fang, Z. Ding and X. Liu, *Ind. Eng. Chem. Res.*, 2023, **62**, 14700–14711.

43 S. Khan, D. Rauber, L. Wang, U. Veerabagu, C. W. M. Kay, C. Xu, S. Shanmugam and T. Kikas, *RSC Sustain.*, 2025, **3**, 4466–4477.

44 F. J. V. Gschwend, F. Malaret, S. Shinde, A. Brandt-Talbot and J. P. Hallett, *Green Chem.*, 2018, **20**, 3486–3498.

45 W. Dastyar, M. Zhao, W. Yuan, H. Li, Z. J. Ting, H. Ghaedi, H. Yuan, X. Li and W. Wang, *ACS Sustain. Chem. Eng.*, 2019, **7**, 11571–11581.

46 H. Baaqel, I. Díaz, V. Tulus, B. Chachuat, G. Guillén-Gosálbez and J. P. Hallett, *Green*

View Article Online
DOI: 10.1039/D6GC02613C

Chem., 2020, **22**, 3132–3140.

47 M. Chen, F. Malaret, A. E. J. Firth, P. Verdía, A. R. Abouelela, Y. Chen and J. P. Hallett, *Green Chem.*, 2020, **22**, 5161–5178.

48 M. Nurdin, H. Abimanyu, H. Putriani, L. O. M. I. Setiawan, M. Maulidiyah, D. Wibowo, A. Ansharullah, M. Natsir, L. O. A. Salim, Z. Arham and F. Mustapa, *Sci. Rep.*, 2021, **11**, 1–11.

49 N. J. Poolakkalody, K. Ramesh, S. Palliprath, S. N. Nittoor, R. Santiago, S. P. Kabekkodu and C. Manisseri, *Renew. Energy*, 2023, **209**, 420–430.

50 Y. Wang, D. Cai, Y. Jiang, X. Mei, W. Ren, M. Sun, C. Su, H. Cao, C. Zhang and P. Qin, *Biotechnol. Biofuels Bioprod.*, 2024, **17**, 1–14.

51 A. H. M. Dezashibi, J. P. Hallett and P. S. Fennell, *Chem. Eng. Process. - Process Intensif.*, 2025, **213**, 110312.

52 M. Mohan, B. A. Simmons, K. L. Sale and S. Singh, *Sci. Rep.*, DOI:10.1038/s41598-022-25372-2.

53 Q. Lu, H. Zhang, R. Fan, Y. Wan and J. Luo, *Sep. Purif. Technol.*, 2025, **363**, 132122.

54 J. A. Paulson and C. Tsay, *Curr. Opin. Green Sustain. Chem.*, 2025, **51**, 100983.

55 J. V. Vermaas, L. D. Dellon, L. J. Broadbelt, G. T. Beckham and M. F. Crowley, *ACS Sustain. Chem. Eng.*, 2019, **7**, 3443–3453.

56 J. V. Vermaas, D. J. Hardy, J. E. Stone, E. Tajkhorshid and A. Kohlmeyer, *J. Chem. Inf. Model.*, 2016, **56**, 1112–1116.

57 J. C. Phillips, D. J. Hardy, J. D. C. Maia, J. E. Stone, J. V. Ribeiro, R. C. Bernardi, R. Buch, G. Fiorin, J. Hénin, W. Jiang, R. McGreevy, M. C. R. Melo, B. K. Radak, R. D. Skeel, A. Singharoy, Y. Wang, B. Roux, A. Aksimentiev, Z. Luthey-Schulten, L. V. Kalé, K. Schulten, C. Chipot and E. Tajkhorshid, *J. Chem. Phys.*, DOI:10.1063/5.0014475.

58 D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, *J. Comput. Chem.*, 2005, **26**, 1701–1718.

59 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graph.*, 1996, **14**, 33–38.

60 K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov and A. D. Mackerell, *J. Comput. Chem.*, 2010, **31**, 671–690.

61 S. Jo, T. Kim, V. G. Iyer and W. Im, *J. Comput. Chem.*, 2008, **29**, 1859–1865.

62 S. Kim, J. Lee, S. Jo, C. L. Brooks, H. S. Lee and W. Im, *J. Comput. Chem.*, 2017, **38**, 1879–1886.

63 B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J. Comput. Chem.*, 1997, **18**, 1463–1472.

64 T. V. Lourencon, L. G. Greca, D. Tarasov, M. Borrega, T. Tamminen, O. J. Rojas and M.

View Article Online
DOI: 10.1039/D6GC02613C

Y. Balakshin, *ACS Sustain. Chem. Eng.*, 2020, **8**, 1230–1239.

65 M. S. Tunc and A. R. P. van Heiningen, *Nord. Pulp Pap. Res. J.*, 2009, **24**, 46–51.

66 T. E. S. Segura, M. A. Vivian, E. A. Bonfatti Júnior, C. Sarto, F. Schmidt and F. G. da Silva Júnior, *Sci. For.*, 2019, **47**, 791–798.

67 S. Khan, D. Rauber, U. Veerabagu, R. Wu, C. W. M. Kay, C. Xu, S. Shanmugam and T. Kikas, *Molecules*, 2025, **30**, 2630.

68 R. P. Overend and E. Chornet, *Philos. Trans. R. Soc. London. Ser. A, Math. Phys. Sci.*, 1987, **321**, 523–536.

69 M. Borrega, V. Pihlajaniemi, T. Liitiä, L. Wikström and T. Tamminen, *Biomass Convers. Biorefinery*, 2023, **13**, 7491–7503.

70 M. Guigou, M. N. Cabrera, M. Vique, M. Bariani, J. Guarino, M. D. Ferrari and C. Lareo, *Biomass Convers. Biorefinery*, 2019, **9**, 293–304.

71 J.-M. Lavoie, R. Beauchet, V. Berberi and M. Chornet, in *Biofuel's Engineering Process Technology*, InTech, 2011.

72 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.

73 M. Karlsson, J. Romson, T. Elder, Å. Emmer and M. Lawoko, *Biomacromolecules*, 2023, **24**, 2314–2326.

74 M. Lawoko, R. Berggren, F. Berthold, G. Henriksson and G. Gellerstedt, *Holzforschung*, 2004, **58**, 603–610.

75 M. Lawoko, G. Henriksson and G. Gellerstedt, *Holzforschung*, 2003, **57**, 69–74.

76 M. Lawoko, G. Henriksson and G. Gellerstedt, *Holzforschung*, 2006, **60**, 156–161.

77 X. Wang, R. Wu, Y. Zhang, Y. Fu, M. Qin and C. Xu, *Ind. Crops Prod.*, 2025, **238**, 122363.

78 L. Penín, S. Peleteiro, S. Rivas, V. Santos and J. C. Parajó, *BioResources*, 2019, **14**, 4733–4747.

79 A. Duval, G. Layrac, A. van Zomeren, A. T. Smit, E. Pollet and L. Avérous, *ChemSusChem*, 2021, **14**, 387–397.

80 Y. Sang, G. Li, X. Li, H. Gong, M. Yang, H. Chen and Y. Li, *Chem. Eng. J.*, 2025, **525**, 170324.

81 Z. Zhang, R. Yang, W. Gao and X. Yao, *RSC Adv.*, 2017, **7**, 31009–31017.

82 M. H. Tanis, B. Al-rudainy and O. Wallberg, *Holzforschung*, 2025, 1–12.

83 Å. Henrik-Klemens, F. Caputo, R. Ghaffari, G. Westman, U. Edlund, L. Olsson and A. Larsson, *Holzforschung*, 2024, **78**, 216–230.

84 Z. Ma, Q. Sun, J. Ye, Q. Yao and C. Zhao, *J. Anal. Appl. Pyrolysis*, 2016, **117**, 116–124.

View Article Online
DOI: 10.1039/D6GC02613C

85 M. González Martínez, A. Anca Couce, C. Dupont, D. da Silva Perez, S. Thiéry, X. mi Meyer and C. Gourdon, *Ind. Crops Prod.*, DOI:10.1016/j.indcrop.2021.114350.

86 D. Diment, O. Tkachenko, P. Schlee, N. Kohlhuber, A. Potthast, T. M. Budnyak, D. Rigo and M. Balakshin, *Biomacromolecules*, 2024, **25**, 200–212.

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

88 V. Ponnuchamy, A. Sandak and J. Sandak, *Phys. Chem. Chem. Phys.*, 2020, **22**, 28448–28458.

89 R. M. Dias, L. C. G. Petrin, F. H. B. Sosa, A. M. Da Costa Lopes, J. A. P. Coutinho and M. C. Da Costa, *Ind. Eng. Chem. Res.*, 2020, **59**, 18193–18202.

90 R. M. Dias, A. M. da Costa Lopes, A. J. D. Silvestre, J. A. P. Coutinho and M. C. da Costa, *Ind. Crops Prod.*, 2020, **143**, 111866.

91 Y. Sang, Y. Ma, G. Li, K. Cui, M. Yang, H. Chen and Y. Li, *Chem. Eng. J.*, 2023, **463**, 142256.

92 S. M. Yang, R. I. Muazu, E. Tran, C. B. Talbot, N. Shah, M. S. P. Shaffer and A. Brandt-Talbot, *RSC Sustain.*, 2025, **3**, 3972–3986.

93 Q. Zhang, X. Tang, S. Li, X. Ge, Y. Wei and Z. Xing, *Colloids Surfaces A Physicochem. Eng. Asp.*, 2026, **733**, 139329.

94 S. Laurichesse and L. Avérous, *Prog. Polym. Sci.*, 2014, **39**, 1266–1290.

95 S. Khan, V. Rooni, D. Rauber, N. Sjulander, M. Gallei, C. W. M. Kay, S. Shanmugam and T. Kikas, *Bioresour. Bioprocess.*, 2026, **13**, 54.

96 E. C. Guillen, S. Babaeipour, P. Nousiainen, S. Forssell, N. Forsman, I. Schlapp-Hackl, M. Awais, M. Österberg and L. Dessbesell, *Bioresour. Technol.*, 2026, **444**, 133924.

97 E. Correa-Guillen, K. A. Henn, M. Österberg and L. Dessbesell, *Curr. Opin. Green Sustain. Chem.*, 2025, **54**, 101038.

98 M. Alherech, S. Omolabake, C. M. Holland, G. E. Klinger, E. L. Hegg and S. S. Stahl, *ACS Cent. Sci.*, 2021, **7**, 1831–1837.

99 P. Buono, A. Duval, P. Verge, L. Averous and Y. Habibi, *ACS Sustain. Chem. Eng.*, 2016, **4**, 5212–5222.

100 A. Eraghi Kazzaz and P. Fatehi, *Ind. Crops Prod.*, DOI:10.1016/j.indcrop.2020.112732.

101 A. Eraghi Kazzaz, Z. Hosseinpour Feizi and P. Fatehi, *Green Chem.*, 2019, **21**, 5714–5752.

102 J. Yao, M. Morsali, A. Moreno, M. H. Sipponen and M. Hakkarainen, *Eur. Polym. J.*, 2023, **194**, 112146.

103 K. A. Henn, S. Forssell, A. Pietiläinen, N. Forsman, I. Smal, P. Nousiainen, R. P. Bangalore Ashok, P. Oinas and M. Österberg, *Green Chem.*, DOI:10.1039/d2gc01637k.

View Article Online
DOI: 10.1039/D6GC02613C

104 E. S. Esakkimuthu, D. DeVallance, I. Pylypchuk, A. Moreno and M. H. Sipponen, *Front. Bioeng. Biotechnol.*, 2022, **10**, 1–14.

105 M. Andersson, I. V. Pylypchuk, A. E. Alexakis, L. Y. Liu and M. H. Sipponen, *ACS Appl. Mater. Interfaces*, 2025, **17**, 1931–1941.

106 V. C. Tran, M. Morsali, Z. Khan, R. Crispin, M. H. Sipponen and I. Engquist, *ACS Sustain. Chem. Eng.*, 2025, **13**, 9053–9062.

107 X. Wan, L. Y. Liu, M. A. Karaaslan, Q. Hua, F. Shen, M. Sipponen and S. Renneckar, *Chem. Eng. J.*, 2025, **504**, 158255.

108 B. Bemerw Kassaun and P. Fatehi, *Chem. Eng. J.*, 2024, **493**, 152582.

109 A. Boarino, H. Wang, F. Olgiati, F. Artusio, M. Özkan, S. Bertella, N. Razza, V. Cagno, J. S. Luterbacher, H. A. Klok and F. Stellacci, *ACS Sustain. Chem. Eng.*, DOI:10.1021/acssuschemeng.2c04284.

110 J. Zheng, Y. Wu, C. Guan, D. Wang, Y. Lai, J. Li, F. Yang, S. Li and Z. Zhang, *Carbon Energy*, DOI:10.1002/cey2.538.

111 L. Hu, L. Wang, W. Xu, X. Wang, M. Toivakka, J. Gustafsson and C. Xu, *Cellulose*, 2025, **32**, 5389–5405.

112 S. C. Lee, T. M. T. Tran, J. W. Choi and K. Won, *Int. J. Biol. Macromol.*, 2019, **122**, 549–554.

113 J. E. Q. Quinsaat, E. Feghali, D. J. van de Pas, R. Vendamme and K. M. Torr, *Biomacromolecules*, 2022, **23**, 4562–4573.

114 M. L. Stone, M. S. Webber, W. P. Mounfield, D. C. Bell, E. Christensen, A. R. C. Morais, Y. Li, E. M. Anderson, J. S. Heyne, G. T. Beckham and Y. Román-Leshkov, *Joule*, 2022, **6**, 2324–2337.

115 I. Celada, E. Apellaniz, A. Pérez, P. Jusner, R. Vendamme and M. Comí, *Prog. Org. Coatings*, 2026, **213**, 109931.

116 F. Wang, S. Nithianandam, I. Pylypchuk and M. H. Sipponen, *Sci. Adv.* , DOI:10.1126/sciadv.adr8372.

117 T. Nelis, M. Rolland, C. L. Bourmaud, E. L. M. Vermeirssen, G. Tekleab, H.-A. Klok and J. S. Luterbacher, *Sci. Adv.*, DOI:10.1126/sciadv.adw9912.

118 R. Wu, C. Liu, Y. Zhang, J. Xu, A. Pranovich, J. Hemming, T. Tirri, X. Wang and C. Xu, *Green Chem.*, 2025, **27**, 6764–6775.

119 S. Babaeipour, P. Nousiainen, E. M. Garcia, P. Mohammadi, M. Vuoriluoto, E. Kimiaei, H. Koivula and M. Österberg, *Food Packag. Shelf Life*, DOI:10.1016/j.fpsl.2025.101538.

View Article Online
DOI: 10.1039/D6GC02613C

Data Availability Statement

The data supporting this article have been included as part of the *Supplementary Information*.