ChemSusChem

Research Article
doi.org/10.1002/cssc.202301464

Chemistry Europe
European Chemical Societies Publishing

www.chemsuschem.org

Rapid Lignin Thermal Property Prediction through Attenuated Total Reflectance-Infrared Spectroscopy and Chemometrics

Luke A. Riddell,[a] Jean-Pierre B. Lindner,[b] Peter de Peinder,[c, d] Florian Meirer,*[d] and Pieter C. A. Bruijnincx*[a]

To expedite the valorisation of lignin as a sustainable component in materials applications, rapid and generally available analytical methods are essential to overcome the bottleneck of lignin characterisation. Where features of a lignin's chemical structure have previously been found to be predicted by Partial Least Squares (PLS) regression models built on Infrared (IR) data, we now show for the first time that this approach can be extended to prediction of the glass transition temperature (T_g), a key physicochemical property. This methodology is shown to be convenient and more robust for prediction of T_g than prediction through empirically derived relationships (e.g., Flory-Fox). The chemometric analysis provided root mean squared errors of prediction (RMSEP) as low as 10.0 °C for a botanically, and a process-diverse set of lignins, and 6.2 °C for kraft-only samples. The PLS models could separately predict both the T_g as well as the degree of allylation ($\%_{allyl}$) for allylated lignin fractions, which were all derived from a single lignin source. The models performed exceptionally well, delivering RMSEP of 6.1 °C, and 5.4%, respectively, despite the conflicting influences of increasing molecular weight and $\%_{allyl}$ on T_g. Finally, the method provided accurate determinations of $\%_{allyl}$ with RMSEP of 5.2%.

Introduction

In the necessary resource and materials transitions, in which fossil feedstock is replaced with a renewable one, lignin is a prime candidate to serve as a source of renewable, biobased carbon. It is a polyphenolic component of the plant cell wall, constituting 15–40% of the biomass, and it stands as the most abundant form of renewable aromatic carbon on the planet.[1] However, the complexity and variability of the lignin structure, in particular after isolation from the lignocellulosic biomass, is an impediment for commercial utilisation. The botanical origin of the lignin strongly influences the abundance of differing functionalities which adorn the phenolic rings and impact the nature of the linkages present between these aromatic units.[2] Furthermore, the specific delignification method inevitably alters and further complicates the structure of the lignin, as a result of the typically non-specific and harsh reaction conditions employed in commercial pulp mills, where cellulose is targeted as the main component of value from biomass.[3] Referred to as technical lignins, these isolated lignin-derived products of biomass fractionation processes (e.g. kraft, soda, or organosolv lignins) are more degraded, condensed, and recalcitrant than their native *in planta*, or native-like (e.g., Milled Wood Lignin) counterparts.[4–6] Broadly speaking, this leaves technical lignins with a poorly defined structure and with highly variable, but also often non-desired properties. This hampers direct application, especially since the structural variation between samples leads to variation in material characteristics, even batch-to-batch, preventing it from reaching correct property specification and making quality assurance for industrially produced lignins more tedious. For example, kraft lignins, the most prominent technical lignins, are isolated from rather severe processes, that use high temperatures and strongly alkaline conditions to delignify the cellulose pulp. As a result, a variety of kraft-specific functionalities and linkages are introduced by extensive de- and re-polymerisation, resulting in a highly heterogeneous material.[3,6–8] Consequently, the majority of kraft lignin is simply used as a biofuel, instead of recovered, providing process energy to the biorefinery/pulp mills. Advances in the biorefinery or pulping processes, and in the availability of alternative renewable energy sources, however, provide an opportunity to instead utilise the lignin as a resource in materials applications. As the dominant delignification technology producing 98% of chemical pulps,[3] the valorisation

[a] L. A. Riddell, Prof. Dr. P. C. A. Bruijnincx
Organic Chemistry & Catalysis, Institute for Sustainable and Circular Chemistry, Faculty of Science, Utrecht University, 3584CG Utrecht (The Netherlands)
E-mail: P.C.A.Bruijnincx@uu.nl

[b] Dr. J.-P. B. Lindner
BASF SE, Group Research, Carl-Bosch-Str. 38, 67056 Ludwigshafen am Rhein (Germany)

[c] Dr. P. de Peinder
VibSpec
Haaftenlaan 28, 4006 XL Tiel (The Netherlands)

[d] Dr. P. de Peinder, Dr. F. Meirer
Inorganic Chemistry & Catalysis, Institute for Sustainable and Circular Chemistry, Faculty of Science, Utrecht University, 3584CG Utrecht (The Netherlands)
E-mail: f.meirer@uu.nl

Supporting information for this article is available on the WWW under https://doi.org/10.1002/cssc.202301464

of this carbon reservoir beyond a low value fuel remains a pertinent challenge. To address the problem of lignin heterogeneity, complexity and variability, a range of methods have been utilised. Typically, fractionation and/or modification approaches,[9–13] as well as (partial) depolymerisation approaches[14–16] have been employed to better achieve more narrowly dispersed lignins with reliably tuneable characteristics.

Getting detailed insight into the properties of a technical lignin, and thus an understanding of its potential applications, requires an extensive suite of characterisation methods. For insights into the chemical structure, lignin is typically subjected to an array of spectroscopic methods, including common (^{1}H, ^{31}P, ^{13}C) and multidimensional (^{1}H-^{13}C HSQC, ^{1}H-^{13}C HMBC) NMR techniques,[3,4,7,17–19] as well as vibrational (Fourier Transform Infrared (FTIR), Near-Infrared (NIR), Raman) spectroscopic methods.[20–22] Mass spectroscopy can elucidate lignin fragments and monomer composition ((py)-GC-MS[23,24]), and molecular weight characteristics are determined by GPC/SEC.[25,26] Beyond the structural features, key material properties such as the glass transition temperature (T_g) are obtained from (modulated) differential scanning calorimetry ((M)DSC)[11,27] or DMA measurements,[28] while thermal stability is determined by TGA,[29,30] and viscosity by rheological methods,[31,32] all to assess the performance/application potential of the lignins. The T_g in particular, is an important performance parameter within the material science field, representing a material's temperature of transition between brittle and pliable states. Drawing direct relationships between complex physicochemical parameters, such as T_g, and structural features is non-trivial for highly heterogeneous polymers, such as lignin. Empirical relationships between structure (number average molecular weight, M_n) and performance parameters (T_g), originally developed for more homogeneous polymers (Flory-Fox (FF), Flory-Fox-Ogawa (FFO), and Fox-Loshaek (FL) relationships),[33–35] have, however, been successfully applied to small lignin sample sets, in spite of the heterogeneity of the lignins.[28,36] Such relationships are important as structure-property correlations allow for one to gain a handle on understanding the performance parameters of a lignin based on a given structural descriptor. However, when considering large, diverse datasets, such as in a mixed-feed biorefinery setting, where a range of sources of biomass are processed, it has not yet been established whether these M_n vs T_g relationships would hold for a larger range of lignin samples. Additionally, process variation or post-processing fractionation and/or modification may lead to radically altered lignin samples, leading to different lignins of the same M_n that can display a broad range of T_g values. As these observational correlations are built upon only one structural descriptor, M_n, (n.b. FFO also relies upon M_w) they are anticipated to suffer from the large heterogeneity of lignin. Furthermore, it is unclear as to whether such relationships could be applied as a valid method of predicting complex material properties such as T_g for lignin. Specifically, lignin's T_g is not only tied to molecular weight, but has been suggested to also manifest as a function of the degree of hydrogen bonding,[31,37] methoxy content,[38] and π-π stacking interactions.[39]

Another pertinent limitation is the analytical infrastructure required for measuring lignin's structural and material parameters, as well as establishing relationships between such characteristics. Full sample characterisation clearly constitutes a bottleneck in lignin research, given the need for access to (advanced and thus expensive) equipment as well as the time investment involved. This hampers both lignin research at an academic level as well as industrial application of lignin into, e.g., materials, where quality control and batch checking increase sample load with such a varying lignin feed. In answer to this, application of chemometric, statistical methods can greatly accelerate data acquisition, processing, and even better inform future experimentation. In a recent example by Karlsson *et al.*, a design of experiments approach was applied to better understand the impact of process conditions of their lignin extraction process on the structure and physical properties of the lignin.[40] Another powerful application of chemometrics, is its use in tandem with spectroscopy. Methods such as Principal Component Analysis (PCA) and Partial Least Squares (PLS) can be applied to spectroscopic data of biomass, where the latter method, for example, has been used to predict cellulose oxidation severity,[41] lignin content,[42] and a further range of parameters of lignins. PCA on ATR-FTIR spectra has been used successfully to structurally analyse and group diverse lignin sample sets by botanical origin,[18,43,44] and PLS models have been shown to predict lignin parameters (hydroxyl (OH) content, S/G ratios, emulsion stability, antioxidant parameters)[18,44] from test samples, respectively. Furthermore, ATR-FTIR spectroscopy presents itself as a practical choice for rapid screening of samples and has seen widespread use for qualitative analysis of lignins, for functional group/connectivity identification[18,32] and reaction monitoring.[11,31,45] Lancefield *et al.* investigated PCA based on ATR-FTIR spectra of a broad lignin sample set and uncovered defined clustering of samples by their botanical origin within PC space, similar to an early example of Boeriu *et al.* Samples were also found to be well separated by key structural features, such as β-O-4 content, prompting successful application of PLS modelling as a tool for predictively quantifying linkages, as well as MW characteristics.[43] However, the use of IR-based PLS regression models on the quantification of more complex, and convoluted macromolecular properties of lignin remains unexplored. To further alleviate the burden of lignin analysis, both approaches, MW-based correlation (FFO, and FL), and spectroscopic-chemometric correlation (PLS regression), were assessed as potential predictive methods to add to the growing lignin analytical toolkit. The PLS model proved able to excellently predict lignin T_g for a wide variety of lignin samples, including fractionated and covalently modified ones. With this FTIR-PLS approach, the parameter space that can be predicted by chemometrics was thus expanded to an industrially relevant material parameter.

Results and Discussion

Empirical Relationships as Predictive Tools

Previously, we reported on the prediction of molecular weight (M_n & M_w), linkage abundance (β-O-4, β-O-4', β-5, β-β, SB5, SB1) and hydroxyl group content ([OH]) from the ATR-FTIR spectra of a set of 61 lignins (Table S1 + S2) using PLS. These samples cover a diverse range of botanical origins, as well as delignification methods, and the set also contains some (solvent) fractionated lignins. Here, we used the same sample set and as a part of this study, determined the T_g of each sample by MDSC (Table S1).[43] T_g values were found to span a range of 167.5 °C, from 58.8 to 226.3 °C. These values, along with the samples' reported molecular weight information (Table S1), were then used to construct plots according to the FF, FFO and the FL relationships (Figure 1), which have been shown to work reasonably well for smaller, more focussed lignin sample sets.[28,37,46,47] The aim here, was to first test whether the relatively complex physicochemical parameter, T_g, could still be reasonably predicted using these aforementioned relationships. Interestingly, even for this varied set of lignin samples, each relationship gave a fairly good fit. However, relatively large errors in T_g were observed at a given M_n value (i.e., vertical spread), a consequence of the heterogeneity of the lignin. Surprisingly, not the FFO and the FL models, which were developed to better account for disperse and crosslinked polymers, respectively,[33–35] but the standard FF relation best described this dataset giving the highest R^2 value (0.893), and the lowest root mean squared error (15.6 °C). Consequently, in the following, we focused on the FF relationship only as a potential predictive tool.

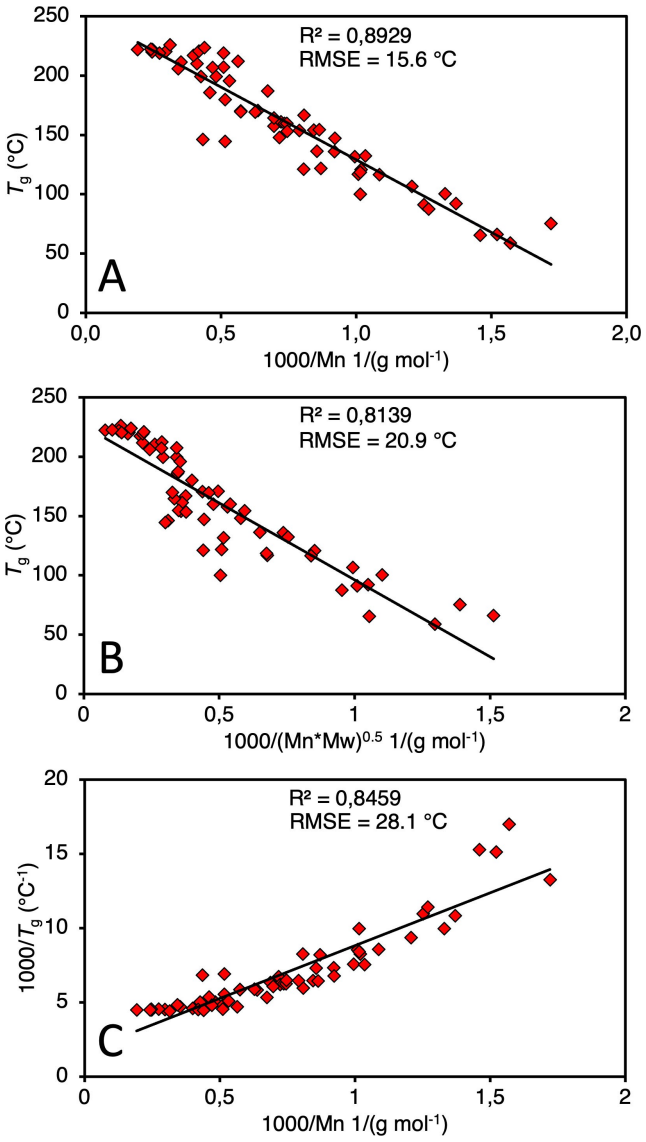

Figure 1. Correlation of T_g with molecular weight characteristics for the set of 61 lignins according to a) FF, b) FFO and c) FL theories, with molecular weight data obtained by Lancefield *et al.*[43] (Table S1 + S2).

The dataset of 61 lignin samples was split into calibration (Cal) and validation (Val) sets such that a ratio of 75:25 (Cal: 46, Val:15) was achieved, and these different divisions of samples were subject to both construction of the FF relationships and the PLS models. As the method of splitting data can be impactful on a (poor) model's performance, be it statistical or otherwise, two different criteria were used to probe influence of the selection of the validation samples on the predictive power of the models. Firstly, we used a Kennard-Stone (KS) algorithm[48] operating on the Mahalanobis distance[49] of the samples' FTIR spectra to select uniformly distributed samples across the set as determined by (FTIR) spectroscopic (dis)similarity, and secondly, by random splits, which cannot ensure even distribution. To further assess the importance of the splitting, random splits of the samples were performed five times in total. From the generated models, the R^2 and root mean squared error of prediction (RMSEP for PLS; $RMSEP_{FF}$ for FF) were calculated and are shown in Table 1. The specific splits of the samples are detailed in Table S7.

The values of $RMSEP_{FF}$ from the random sample selections fluctuate between 15.3—22.9 °C, and with the KS split, it is as low as 7.3 °C. As the error of prediction is dependent on the selection of test samples, this renders FF a poor model for predicting T_g for lignin. Indeed, this is because T_g of a lignin is dependent on a combination of its structural features and not solely its molecular weight, which is the FF's sole predictor variable. Molecular weight of lignin is already a challenging parameter to reliably obtain, with a relatively large error of measurement and large dependence on the measurement setup.[50] To exemplify this, four typical technical lignin samples were subject to duplicate GPC analysis in two separate runs with the same equipment and the M_n was determined twice (Table 2, Figure S2 & S3). The relatively large error in M_n numbers leads to a large discrepancy in the predicted values of T_g. Additionally, due to the FF fit's logarithmic relation, the error is non-stochastic with respect to M_n, i.e., the error in T_g is not constant for different values of M_n with the same measurement error. Specifically, the error in T_g is much larger for smaller M_n values despite having the same measurement uncertainty than larger M_n values. In this respect, a look towards a more suitable method of predicting T_g is indeed warranted. Statistical approaches, such as PLS analysis based on ATR-FTIR spectra, are

Table 1. Details of PLS models predicting T_g values for the 61 lignin samples based on their respective ATR-FTIR spectra. In each model, 1st derivative with SG filtering was applied, followed by MSC and mean centering as standard. Details of the empirical FF relationships used as prediction tools with the same sample splits.

	Captured Variance								FF	
Entry	X	Y	Num LVs	RMSEC	RMSECV	RMSEP	R^2 Cal	R^2 CV	R^2 Pred	$RMSEP_{FF}$	R^2 $Pred_{FF}$
1	92.60	95.48	5	10.20	15.67	14.99	0.955	0.894	0.890	17.06	0.881
2	93.22	94.82	5	11.06	18.03	14.50	0.948	0.865	0.894	15.31	0.894
3	88.55	94.34	4	11.51	16.63	15.59	0.943	0.882	0.877	18.66	0.792
4	86.62	94.85	4	10.74	14.45	19.87	0.949	0.907	0.849	19.59	0.863
5	92.97	95.19	5	10.36	17.82	17.97	0.952	0.862	0.890	22.90	0.820
6	83.6	88.37	3	16.59	19.73	10.00	0.884	0.836	0.939	7.62	0.974

RMSEC, RMSECV and RMSEP are given in °C. All models were split such that Cal: 46 and Val: 15 samples. Data splits for Entries 1–5 were determined randomly, the split for Entry 6 was determined by a KS algorithm on the basis of Mahalanobis distance.

Table 2. Comparison of error in M_n ($g\,mol^{-1}$) determination *via* duplicate GPC measurement, and the subsequent error in T_g as predicted by FF, and the experimentally determined (*via* MDSC) for a selection of 4 lignin samples.

Lignin	M_n1	M_n2	ΔM_n	$_{FF}T_g1$	$_{FF}T_g2$	$\Delta_{FF}Tg$	*Measured* T_g	T_g Difference[a]
SEKL	1061	1392	331	134	161	27	142	5.8
P1000	918	1090	172	117	137	20	136	8.5
IKL	895	1134	239	113	142	29	147	19
LF401	728	880	152	82	111	29	137	40

M_n values are provided in $g\,mol^{-1}$ and T_g values in °C. [a] Calculated by taking the difference between the mean of the FF derived T_g and the experimentally determined T_g

anticipated to better predict T_g, as each spectrum contains a linear combination of the structural features of the sample, thus better informing the model during the calibration. Sample purity is a consideration, as, e.g., presence of residual carbohydrates may negatively impact model performance. Typically, technical lignins have relatively low concentrations of carbohydrate impurities, and for this sample set, we did not find this to have impacted prediction performance for lignin parameters previously.[43] Beyond its intrinsic information advantage, ATR-FTIR spectra are also easy to acquire, demand minimal sample preparation, can be rapidly recorded, and offer reliability of measurement. Given the existing relationship between M_n and T_g, and that M_n has been shown previously to be predicted using PLS,[43] we thus sought to develop a more effective method for prediction of T_g based on ATR-IR/PLS.

Statistical approaches as a means for prediction

The 61 lignin samples (Table S1 + S2) were used 'as-is' (i.e., without any further sample pretreatment) and were each remeasured by ATR-FTIR. New PCA and PLS models were constructed, and spectral data pre-processing parameters were optimised with the objective of minimising the root mean squared error of cross-validation (RMSECV), and RMSECV/RMSEC, where RMSEC is the root mean squared error of calibration, as this avoids construction of a model that is susceptible to overfitting (Table S7). For all models, venetian blinds cross-validation (with 10 splits) was utilised. Recent works from within our group on chemometrics for lignin analysis provided insight into which processing parameters were most suitable for the dataset herein.[43,51] Here, the optimal pre-processing conditions were found when using 1st derivative spectra, then applying a Savitzky-Golay (SG) filter, followed by multiplicative scatter correction (MSC) to mitigate any scaling and baseline effects, and finally, mean centering. Furthermore, it is commonplace to remove the part of the spectrum that primarily contains noise from the diamond crystal (2849-1851 cm^{-1}), along with the lower wavenumbers (749–600 cm^{-1}).[52] It was found here that such steps improved both the RMSECV, and offered an improved RMSECV/RMSEC ratio. These settings were thus used in further modelling (details available in Table S6, Entry 6). Inspection of the first two PCs of the ATR-IR spectra (first two eigenspectra) yielded insights into the important bonds which describe the maximum variance of this sample set (Figure 2). Whilst it is not a true representation of the data used in PCA analysis, visual inspection was aided by plotting the PCs without use of 1st derivative spectra, as interpretation of the spectra derivatives is often more challenging. Here can be seen the peaks at 1266 cm^{-1} and 1100 cm^{-1}, the relative intensities of which have been shown to correlate with the G/S ratio.[43,53] Typical for the S-type (hardwood) is the higher intensity of the 1109 cm^{-1} vibration, being shifted to a slightly lower wavenumber than has been reported elsewhere for transmission and/or DRIFT spectra due to the refractive index effect present in ATR spectra.[54,55] These observed features

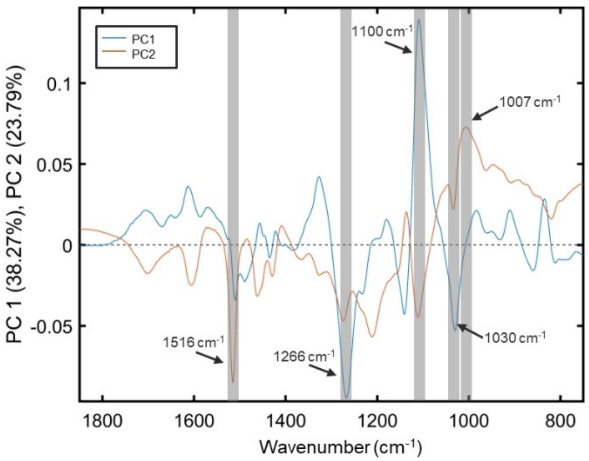

Figure 2. PC1 & PC2 (eigenspectra) loadings for FTIR spectra of the set of 61 lignin samples without application of a 1st derivative or SG smoothing pre-processing, in the window 1850–750 cm^{-1}. Important wavenumbers have been highlighted.

respectively constitute two of the key features in PC1. The vibration at 1266 cm^{-1} is also present in PC2 (with a negative value), along with a broader peak at 1007 cm^{-1} which is most likely related to C–O modes. The typical peak belonging to non-conjugated G-ring vibrations at 1516 cm^{-1} is also present and appears in both PC1 & PC2 with a negative value. These observations provide an indication that whilst the spectral variance is mostly dominated by aromatic-related vibrations, other vibrations relating to ether linkages, methoxy content and other unassigned bands clearly play an important role, given their appearance within these PCs.

The PLS regression modelling was performed with the same sample splits as with the FF modelling (Table S7). The PLS regression provided surprisingly robust predictions of T_g, as denoted by RMSECV/RMSEP ratios relatively close to unity for most models. A low RMSEP/RMSECV ratio is simply an indication that the constructed model better describes the Val set than the Cal set, whereas a high RMSEP/RMSECV value can be an indication of overfitting. However, relatively high RMSECV (leading to a high RMSECV/RMSEP), can be due to the intrinsic nature of cross-validation being a construction of sub-divisions of the Cal set predicting the other samples within the Cal set. Regardless, with the random divisions of the samples (Table 1, Entries 1–5), PLS regression generally offered improvements in RMSEP compared to the FF modelling. More important, however, is the robustness of PLS, as showcased by the models' relative stability in RMSEP, relative to the FF models. Notably, the sample split obtained by the KS algorithm (Table 1, Entry 6; Figure 3) provided the best prediction among the PLS models, with an RMSEP of 10.0 °C. Interestingly, this value is higher than the corresponding $RMSEP_{FF}$ of 7.3 °C, however this is again considered a symptom of the unreliable error of simple empirical models. Similar to the work of Lancefield *et al.*,[43] a more focussed set of kraft-only lignins provided a much-improved prediction power for PLS modelling (Figure 3B). In essence, this served to remove variation from the spectra which is introduced by process-specific chemistry occurring during delignification, allowing the model to 'focus' on the more pertinent spectral features. As most biorefineries operate with a single, fixed feed of softwood or hardwood, such a more concise sample set is more realistic than the more diverse sample set. As anticipated with the contracted sample set, the quality of the modelling was greatly improved. With RMSECV and RMSEP of 10.4 and 6.2 °C, respectively, this represents an exceptionally small relative error (RE) of 3.7 %, based on a T_g as large as 167 °C.

Prediction of T_g for derivatised lignins

Chemical modification of technical (fractionated) lignins are an important and widely implemented means to better prepare it for incorporation into materials applications. Success has been found, for example, by using the (typically phenolic) hydroxyl groups as reactive handles for installing a host of differing functionalities,[11,13] and it is well reported that the capping of

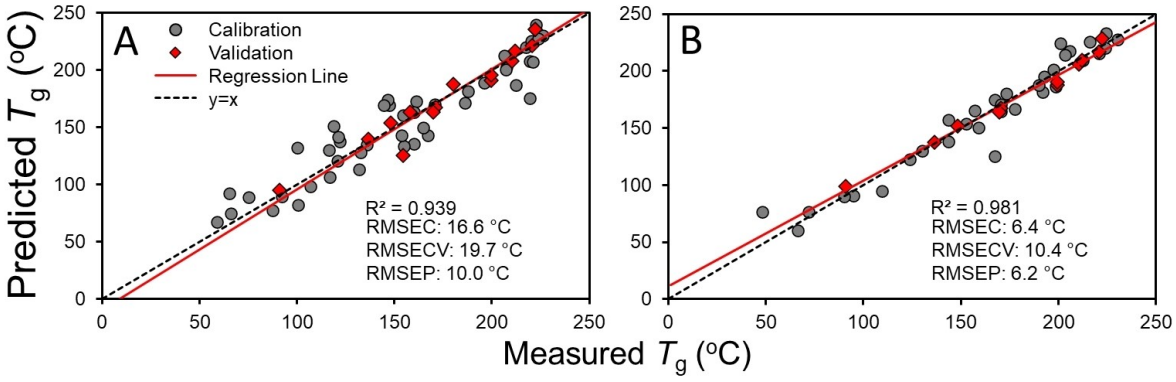

Figure 3. Scatterplots of PLS predicted T_g values vs physically measured T_g values (by MDSC). Model a) was built from the entire sample set of 61 lignins, split 75:25, such that Cal = 46 and Val = 15 samples. Model b) was built from only the kraft samples (44 lignins), split 75:25, such that Cal = 33 and Val = 11 samples.

these hydroxyl functionalities often leads to a reduction in the lignin's T_g.[31,56,57] At the same time, such derivatisation methods, more often than not, subtly increase the molecular weight of the lignin.[11] Such modified lignins would therefore not adhere to the FF (or similar) relationship and require an alternative method for T_g prediction. Previously we showed that fractionation and systematic (partial) modification of a Stora Enso's Lineo Classic softwood kraft lignin (SEKL) yielded a set of lignin samples with a very large (and tuneable) range of T_gs. More specifically, solvent fractionation using a gradient of EtOAc, EtOH, MeOH and acetone gave five specific fractions (including the final insoluble residual fraction).[9] These lignins were well partitioned by molecular weight, with a T_g thermal property space range of 152 °C. Subsequent allylation generated a library of unique samples of lower T_g (relative to the parent fraction), thus expanding upon this property space to a range of 213 °C, from 12 to 225 °C. The relevant thermal data (T_g) and the degree of allylation ($\%_{allyl}$) as determined by MDSC and ^{31}P NMR, respectively, are reproduced in the supporting information (Table S3). This data series therefore provided a good testbed for the use of PLS regression to probe the thermal properties of a modified lignin. Additionally, as $\%_{allyl}$ had been quantified for these samples too, this metric could also be investigated as a potential predictable parameter. For this, the ATR-FTIR spectra of this sample set were measured, and the data pre-processing optimisation was performed as above for the broader sample set (Tables S8 + S9, for $\%_{allyl}$ and T_g, respectively).

As before, prior to quantitative modelling, inspection of the principal components (PCs) of the FTIR dataset was performed, albeit without application of the 1st derivative and SG filter (Figure 4a). For subsequent PCA analysis, however (Figure 4b-d), the 1st derivative and SG filter were applied as the first pre-processing steps (respectively) to yield optimal models, hence the difference in captured variances of PC1&2 in these figures. As can be seen from Figure 4a, one of the major contributions to PC1 is the terminal alkene C–H bending modes at 986 and 922 cm^{-1} (associated with allyl content) with positive loading, but with minimal score in PC2. This therefore indicated that samples with positive PC1 scores will contain increased $\%_{allyl}$ values, whereas PC2 contains minimal information pertaining to $\%_{allyl}$. Indeed, this is confirmed by the PCA plot in Figure 4b wherein the samples are coloured by $\%_{allyl}$. Equally, the aromatic C skeletal ring vibration at 1514 cm^{-1}[54] and the alkyl-aryl ether vibration at 1266 cm^{-1} had a strongly positive score in PC2, providing an indication that a positive score in PC2 is linked to presence of aromatics, and potentially, increasing MW. A

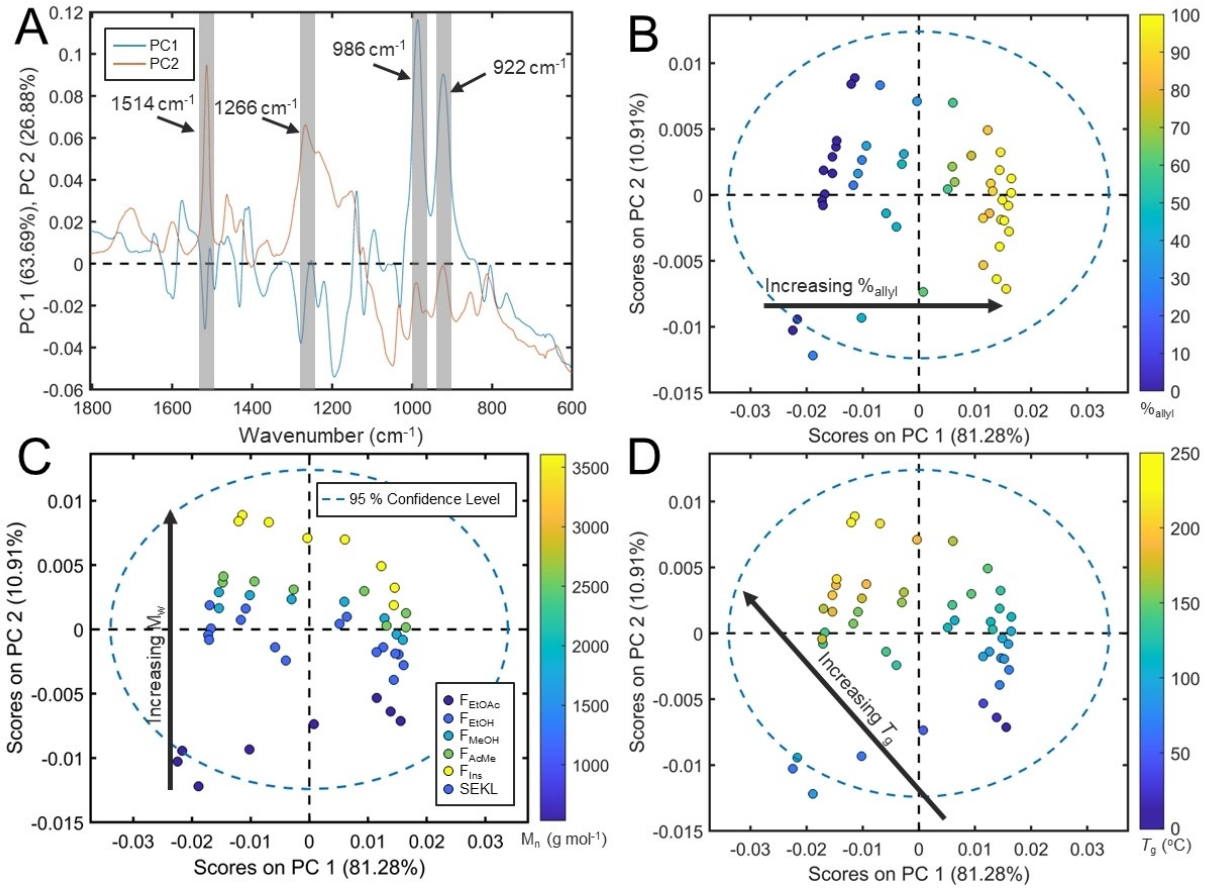

Figure 4. (a) (Top Left) Truncated view of PC1 & PC2 (600-1800-600 cm^{-1}) eigenspectra of the 48 allylated lignin samples with highlighted peaks of interest. Score plots of PC2 vs PC1 of the ATR-FTIR spectra were colour coded according to: (b) (Top Right) $\%_{allyl}$, (c) (Bottom Left) M_n, and (d) (Bottom Right) T_g. All plots were constructed from FTIR data which was pre-processed (see SI). The exception is plot (a) for which neither 1st derivative, nor SG filter were applied. For the loading vector plot where all pre-processing (as outlined above) was applied, see Figure S1. Arrows on b)–d) were added to highlight the emergent trends.

general PLS model was built using all samples in the Cal set and the regression vector (Figure 5) shows that these frequencies are truly utilised by the model. Note that the slight shift in frequencies is from the model's construction on 1st derivative spectra.

Following this, quantitative modelling of both the T_g and $\%_{allyl}$ of the lignins was performed separately (Figure 6). The 48 samples were split into Cal and Val sets of 36 and 12 samples respectively to retain a 75:25 split in the data. To confirm the reliability of the method (beyond the insights from the RMSECV value) a series of PLS models were constructed. Three models were determined with randomised splits in the data, as well as a data split by KS, based on Mahalanobis distance, for both $\%_{allyl}$ and T_g separately (Table S9). Encouragingly for both T_g and $\%_{allyl}$, all models had relatively similar RMSECV, averaging to 5.9% and 10.0 °C for $\%_{ally}$ and T_g, respectively (Table 3). Likewise, the average RMSEP for T_g and $\%_{allyl}$ was found to be 5.2% and 10.2 °C, respectively. The resulting RE values are exceptionally low and are in line with structural parameters which have been predicted previously,[43] reinforcing the robustness of the method for determination of T_g. These results thus provide precedence for the use of IR-PLS methods for the prediction of properties of modified lignins, and more specifically, for the prediction of their T_g.

Table 3. Mean values for RMSECV, RMSEP, R^2 Pred, and RE for the PLS regression models predicting $\%_{allyl}$ and T_g for the set of 48 lignins.

Parameter	$_{avg}$RMSECV	$_{avg}$RMSEP	$_{avg}R^2$ Pred	$_{avg}RE^{[a]}$ (%)
%allyl	5.9	5.2	0.980	5.9
T_g	10.0	10.2	0.969	4.7

[a] RE based on ranges of 98.89% and 213 °C for $\%_{allyl}$ and T_g respectively. Details of the models are provided in Tables S8 + S9.

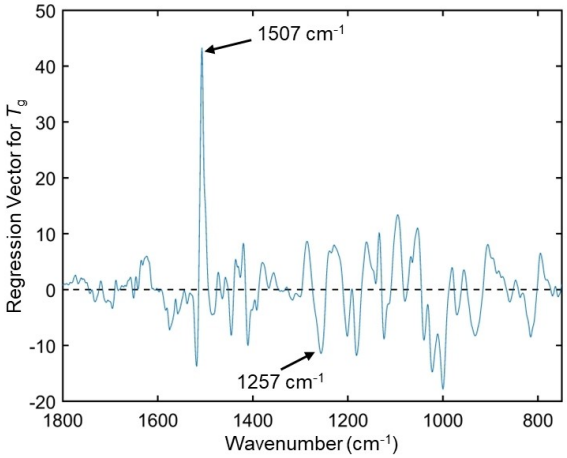

Figure 5. Regression vector for PLS model for the prediction of T_g, constructed with the 48 lignin sample series' ATR-FTIR data. Peaks (derivative) relating to aromatics (1507 cm^{-1}) and alkyl-aryl ethers (1257 cm^{-1}) are highlighted.

Conclusions

Rapid access to key thermal parameters of lignin, and insight into how these properties relate to particular structure descriptors is highly desirable. Here, we have shown that the previously established, empirical Flory-Fox, Flory-Fox-Ogawa and Fox-Loshaek relationships between T_g and molecular weight characteristics also hold reasonably well for large and diverse lignin sample sets. However, whilst M_n is a dominant predictor for lignin T_g, it is not the only factor which determines such a complex parameter, leading to unreliable prediction using such simple observational relationships. Furthermore, the inherent error in molecular weight determinations renders this a less desirable approach to estimating T_g. Instead, an alternative chemometric method, using PLS predictions based on spectroscopic data, acts as a surprisingly powerful, alternative approach, allowing for a variety of occluded structural features to act as predictors, and not just M_n. This allowed for accurate, and robust predictions of the complex physicochemical property, T_g. The PLS models displayed excellent reliability

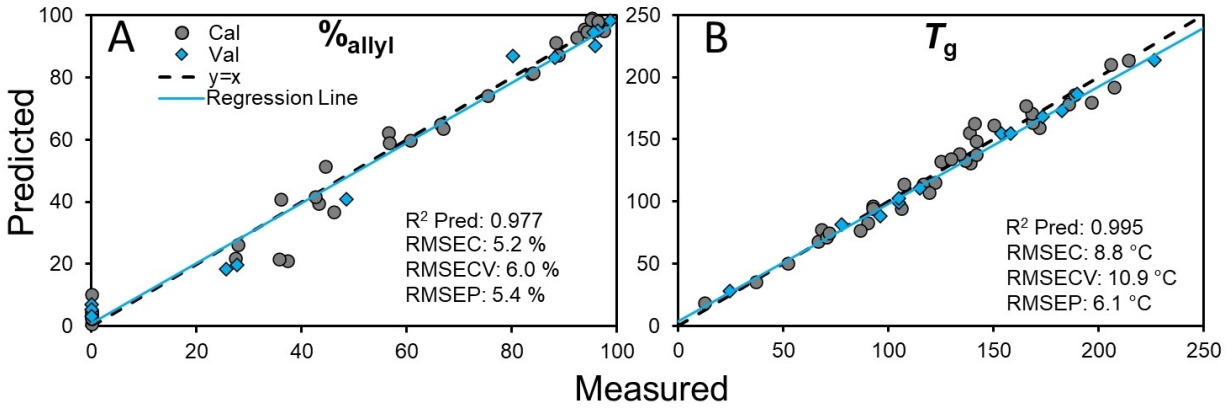

Figure 6. Scatterplots of PLS predicted vs measured values of a) $\%_{allyl}$ as determined by ^{31}P NMR, and b) T_g as determined by MDSC previously.[9] Both models were built from the set of 48 SEKL derivatives, split 75:25, by a KS algorithm on the basis of Mahalanobis distance, such that Cal = 36 and Val = 12 samples.

when adjusting sample splits between Cal and Val sets, showcasing this robustness. Furthermore, after restriction of the sample set to kraft only samples, the relative error of the prediction of T_g is of a similar magnitude to the error for structural features previously predicted by PLS regression. To the best of our knowledge, chemometric analysis with ATR-FTIR has not yet been used for the analysis of functionalised (allylated) lignins. The key spectral variance of this set of lignins was highlighted through PCA, allowing for mapping of the samples in PC space wherein the samples were well separated by their $\%_{allyl}$, molecular weight, and thus, T_g. Furthermore, both T_g and $\%_{allyl}$ were accurately predicted with excellent reliability through PLS regression.

Overall, the use of ATR-FTIR spectroscopy with chemometrics is again shown to be a powerful tool in the 'debottlenecking' of lignin valorisation by expediting its structural analysis. The list of predictable (key) physicochemical properties has again been expanded upon, by inclusion of T_g, a complex material property, to the fold. As the understanding of lignin's potential in material applications continues to grow, so too should the analytical toolbox along with it, to facilitate rapid on-site predictions for industry and academia to assess the nature of their sample, without the typical burden of labour- and equipment-intensive lignin analysis.

Acknowledgements

The authors would like to thank Dr. Christopher S. Lancefield and Floris J. P. A. Enthoven for their assistance with the experimental work and sample acquisition. Furthermore, BASF SE, Netherlands Organisation for Scientific Research (NWO LIFT grant ENPPS.LIFT.019.17) are gratefully acknowledged for supporting this work.

Conflict of Interests

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

[1] A. J. Ragauskas, G. T. Beckham, M. J. Biddy, R. Chandra, F. Chen, M. F. Davis, B. H. Davison, R. A. Dixon, P. Gilna, M. Keller, P. Langan, A. K. Naskar, J. N. Saddler, T. J. Tschaplinski, G. A. Tuskan, C. E. Wyman, *Science (80).* **2014**, *344*, 1246843, DOI 10.1126/science.1246843.
[2] R. Vanholme, B. Demedts, K. Morreel, J. Ralph, W. Boerjan, *Plant Physiol.* **2010**, *153*, 895–905.
[3] C. S. Lancefield, H. J. Wienk, R. Boelens, B. M. Weckhuysen, P. C. A. Bruijnincx, *Chem. Sci.* **2018**, *9*, 6348–6360.
[4] S. Constant, H. L. J. Wienk, A. E. Frissen, P. De Peinder, R. Boelens, D. S. Van Es, R. J. H. Grisel, B. M. Weckhuysen, W. J. J. Huijgen, R. J. A. Gosselink, P. C. A. Bruijnincx, *Green Chem.* **2016**, *18*, 2651–2665.
[5] T. Renders, S. Van Den Bosch, S. F. Koelewijn, W. Schutyser, B. F. Sels, *Energy Environ. Sci.* **2017**, *10*, 1551–1557.
[6] W. Schutyser, T. Renders, S. Van Den Bosch, S. F. Koelewijn, G. T. Beckham, B. F. Sels, *Chem. Soc. Rev.* **2018**, *47*, 852–908.
[7] N. Giummarella, I. V. Pylypchuk, O. Sevastyanova, M. Lawoko, *ACS Sustainable Chem. Eng.* **2020**, *8*, 10983–10994.
[8] C. Crestini, H. Lange, M. Sette, D. S. Argyropoulos, *Green Chem.* **2017**, *19*, 4104–4121.
[9] L. A. Riddell, F. J. P. A. Enthoven, J.-P. B. Lindner, F. Meirer, P. C. A. Bruijnincx, *Green Chem.* **2023**, *25*, 6051–6056.
[10] M. Gigli, C. Crestini, *Green Chem.* **2020**, *22*, 4722–4746.
[11] M. Jawerth, M. Johansson, S. Lundmark, C. Gioia, M. Lawoko, *ACS Sustainable Chem. Eng.* **2017**, *5*, 10918–10925.
[12] A. Tagami, C. Gioia, M. Lauberts, T. Budnyak, R. Moriana, M. E. Lindström, O. Sevastyanova, *Ind. Crops Prod.* **2019**, *129*, 123–134.
[13] C. Gioia, G. Lo Re, M. Lawoko, L. Berglund, *J. Am. Chem. Soc.* **2018**, *140*, 4054–4061.
[14] A. T. Smit, T. Dezaire, L. A. Riddell, P. C. A. Bruijnincx, *ACS Sustainable Chem. Eng.* **2023**, *11*, 6070–6080.
[15] A. Duval, D. Vidal, A. Sarbu, W. René, L. Avérous, *Mater. Today Chem.* **2022**, *24*, 100793, DOI 10.1016/j.mtchem.2022.100793.
[16] L. Cederholm, Y. Xu, A. Tagami, O. Sevastyanova, K. Odelius, M. Hakkarainen, *Ind. Crops Prod.* **2020**, *145*, 112152.
[17] D. S. Argyropoulos, *J. Wood Chem. Technol.* **1994**, *14*, 45–63.
[18] C. G. Boeriu, D. Bravo, R. J. A. Gosselink, J. E. G. Van Dam, in *Ind. Crops Prod.*, **2004**, pp. 205–218.
[19] E. A. Capanema, M. Y. Balakshin, J. F. Kadla, *J. Agric. Food Chem.* **2004**, *52*, 1850–1860.
[20] P. Bock, N. Gierlinger, *J. Raman Spectrosc.* **2019**, *50*, 778–792.
[21] P. Bock, P. Nousiainen, T. Elder, M. Blaukopf, H. Amer, R. Zirbs, A. Potthast, N. Gierlinger, *J. Raman Spectrosc.* **2020**, *51*, 422–431.
[22] U. P. Agarwal, S. A. Ralph, D. Padmakshan, S. Liu, C. E. Foster, *J. Agric. Food Chem.* **2019**, *67*, 4367–4374.
[23] G. Van Erven, R. De Visser, D. W. H. Merkx, W. Strolenberg, P. De Gijsel, H. Gruppen, M. A. Kabel, *Anal. Chem.* **2017**, *89*, 10907–10916.
[24] S. Reale, A. Di Tullio, N. Spreti, F. De Angelis, *Mass Spectrom. Rev.* **2004**, *23*, 87–126.
[25] C. Cui, R. Sun, D. S. Argyropoulos, *ACS Sustainable Chem. Eng.* **2014**, *2*, 959–968.
[26] S. Baumberger, A. Abaecherli, M. Fasching, G. Gellerstedt, R. Gosselink, B. Hortling, J. Li, B. Saake, E. De Jong, *Holzforschung* **2007**, *61*, 459–468.
[27] H. Li, A. G. McDonald, *Ind. Crops Prod.* **2014**, *62*, 67–76.
[28] O. Sevastyanova, M. Helander, S. Chowdhury, H. Lange, H. Wedin, L. Zhang, M. Ek, J. F. Kadla, C. Crestini, M. E. Lindström, *J. Appl. Polym. Sci.* **2014**, *131*, 9505–9515.
[29] M. V. Ramiah, *J. Appl. Polym. Sci.* **1970**, *14*, 1323–1337.
[30] S. Kubo, J. F. Kadla, *J. Wood Chem. Technol.* **2008**, *28*, 106–121.
[31] K. A. Y. Koivu, H. Sadeghifar, P. A. Nousiainen, D. S. Argyropoulos, J. Sipilä, *ACS Sustainable Chem. Eng.* **2016**, *4*, 5238–5247.
[32] B. Shrestha, Y. Le Brech, T. Ghislain, S. Leclerc, V. Carré, F. Aubriet, S. Hoppe, P. Marchal, S. Pontvianne, N. Brosse, A. Dufour, *ACS Sustainable Chem. Eng.* **2017**, *5*, 6940–6949.
[33] T. G. Fox, S. Loshaek, *J. Polym. Sci.* **1955**, *15*, 371–390.
[34] T. G. Fox, P. J. Flory, *J. Appl. Phys.* **1950**, *21*, 581–591.
[35] T. Ogawa, *J. Appl. Polym. Sci.* **1992**, *44*, 1869–1871.
[36] N. Giummarella, P. A. Lindén, D. Areskogh, M. Lawoko, *ACS Sustainable Chem. Eng.* **2020**, *8*, 1112–1120.
[37] R. Ebrahimi Majdar, A. Ghasemian, H. Resalati, A. Saraeian, C. Crestini, H. Lange, *ACS Sustainable Chem. Eng.* **2020**, *8*, 16803–16813.
[38] A. M. Olsson, L. Salmén, *Nord. Pulp Pap. Res. J.* **1997**, *12*, 140–144.
[39] M. E. Jawerth, C. J. Brett, C. Terrier, P. T. Larsson, M. Lawoko, S. V. Roth, S. Lundmark, M. Johansson, *ACS Appl. Polym. Mater.* **2020**, *2*, 668–676.
[40] M. Karlsson, V. L. Vegunta, R. Deshpande, M. Lawoko, *Green Chem.* **2022**, *24*, 1211–1223.
[41] J. Simon, O. Tsetsgee, N. A. Iqbal, J. Sapkota, M. Ristolainen, T. Rosenau, A. Potthast, *Carbohydr. Polym.* **2022**, *278*, 118887.
[42] O. Elle, R. Richter, M. Vohland, A. Weigelt, *Sci. Rep.* **2019**, *9*, 1–11.
[43] C. S. Lancefield, S. Constant, P. de Peinder, P. C. A. Bruijnincx, *ChemSusChem* **2019**, *12*, 1139–1146.
[44] C. G. Boeriu, F. I. Fiţigău, R. J. A. Gosselink, A. E. Frissen, J. Stoutjesdijk, F. Peter, *Ind. Crops Prod.* **2014**, *62*, 481–490.
[45] I. Ribca, M. E. Jawerth, C. J. Brett, M. Lawoko, M. Schwartzkopf, A. Chumakov, S. V. Roth, M. Johansson, *ACS Sustainable Chem. Eng.* **2021**, *9*, 1692–1702.
[46] A. Duval, G. Layrac, A. van Zomeren, A. T. Smit, E. Pollet, L. Avérous, *ChemSusChem* **2021**, *14*, 387–397.

[47] H. Lange, P. Schiffels, M. Sette, O. Sevastyanova, C. Crestini, *ACS Sustainable Chem. Eng.* **2016**, *4*, 5136–5151.
[48] R. W. Kennard, L. A. Stone, *Technomet* **1969**, *11*, 137.
[49] L. C. Lee, C. Y. Liong, A. A. Jemain, *Analyst* **2018**, *143*, 3526–3539.
[50] D. D. Bly, H. J. Stoklosa, J. J. Kirkland, W. W. Yau, *Anal. Chem.* **1975**, *47*, 1810–1813.
[51] K. N. M. Khalili, P. de Peinder, J. Donkers, R. J. A. Gosselink, P. C. A. Bruijnincx, B. M. Weckhuysen, *ChemSusChem* **2021**, *14*, 5517–5524.
[52] C. G. Boeriu, D. Bravo, R. J. A. Gosselink, J. E. G. Van Dam, *Ind. Crops Prod.* **2004**, *20*, 205–218.
[53] K. K. Pandey, *J. Appl. Polym. Sci.* **1999**, *71*, 1969–1975.
[54] H. L. Hergert, *J. Org. Chem.* **1960**, *25*, 405–413.
[55] O. Faix, *Holzforschung* **1991**, *45*, 21–28.
[56] S. Sen, S. Patil, D. S. Argyropoulos, *Green Chem.* **2015**, *17*, 1077–1087.
[57] A. Duval, L. Avérous, *Green Chem.* **2020**, *22*, 1671–1680.

Manuscript received: October 9, 2023
Revised manuscript received: December 26, 2023
Accepted manuscript online: January 9, 2024
Version of record online: January 26, 2024