Is NMR Combined with Multivariate Regression Applicable for the Molecular Weight Determination of Randomly Cross-Linked Polymers Such as Lignin?

René Burger, Jessica Rumpf, Xuan Tung Do, Yulia B. Monakhova,* Bernd W. K. Diehl, Matthias Rehahn, and Margit Schulze*

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The molecular weight properties of lignins are one of the key elements that need to be analyzed for a successful industrial application of these promising biopolymers. In this study, the use of ^{1}H NMR as well as diffusion-ordered spectroscopy (DOSY NMR), combined with multivariate regression methods, was investigated for the determination of the molecular weight (M_w and M_n) and the polydispersity of organosolv lignins ($n = 53$, *Miscanthus x giganteus*, *Paulownia tomentosa*, and *Silphium perfoliatum*). The suitability of the models was demonstrated by cross validation (CV) as well as by an independent validation set of samples from different biomass origins (beech wood and wheat straw). CV errors of ca. 7−9 and 14−16% were achieved for all parameters with the models from the ^{1}H NMR spectra and the DOSY NMR data, respectively. The prediction errors for the validation samples were in a similar range for the partial least squares model from the ^{1}H NMR data and for a multiple linear regression using the DOSY NMR data. The results indicate the usefulness of NMR measurements combined with multivariate regression methods as a potential alternative to more time-consuming methods such as gel permeation chromatography.
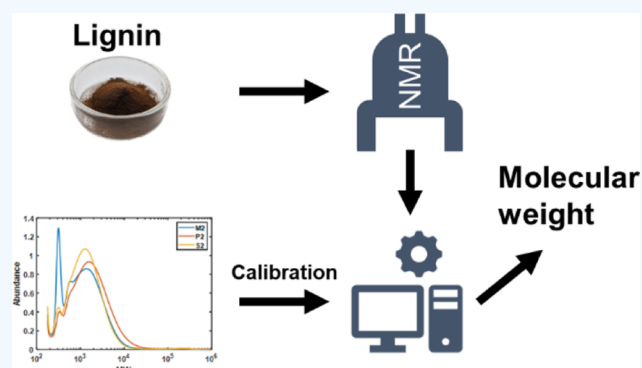
INTRODUCTION

Lignin is a promising renewable raw material to reduce the chemical industry's dependence on fossil fuels. It has a complex heteropolymeric structure, consisting of three different monomers, which are randomly connected via numerous different linkage types (Figure 1). The nature and number of interunit linkages mainly depend on the biomass origin and isolation process.

Thus, spectroscopic techniques with a high resolution such as heteronuclear single-quantum coherence (HSQC) NMR are needed for the analysis of the lignin's chemical structure.[4,5] Although NMR has become an important tool for the structural elucidation, the usual method for the analysis of polymer-specific parameters such as the weight-average molecular weight M_w, the number-average molecular weight M_n, and the resulting polydispersity index PDI = M_w/M_n is gel permeation chromatography (GPC).[2] GPC provides easy access to the whole molecular weight distribution of polymers, not only giving information about an average molecular weight, but also about the dispersity of the sample. The molecular weight of the analyte is determined using the elution time from the GPC column, which is dependent on the hydrodynamic volume of the polymer. A calibration is needed for each type of polymer to obtain the relationship between the hydrodynamic

volume and the molecular weight, making GPC a relative method. Usually, this calibration is simple and reliable, because narrow-distributed, well-characterized standards of the same polymer are readily available (e.g., poly(methyl methacrylate)). Lignin's chemical variability, from sample to sample and also over the molecular weight distribution itself, makes this calibration erroneous.[6] As representative lignin molecular weight standards are still not available, polystyrene or polystyrene sulfonate is used for the calibration. The observed molecular weight is also dependent on the applied experimental conditions such as columns, eluents, and calibrants.[4,7] Thus, GPC can only yield relative results for comparison between different samples.
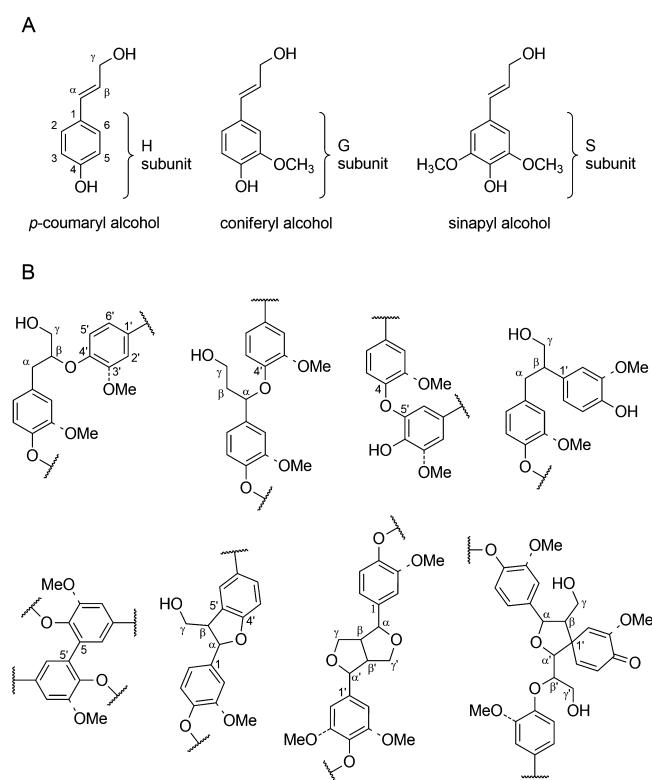
Other methods for the molecular weight determination such as multiangle light scattering (MALLS) or matrix-assisted laser desorption time-of-flight mass spectrometry (MALDI-ToF-MS) are being evaluated to obtain an absolute molecular
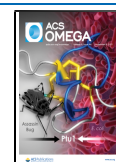
Figure 1. Monolignol structures and linkage types. A: three alcohols are polymerized randomly to form lignin. In the polymer, the aromatic rings are referred to as p-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) subunits. B: first row: β-aryl ether (β-O-4′), α-aryl ether (α-O-4′), biphenyl ether (4-O-5′), and 1,2-diarylpropane (β-1′). Second row: biphenyl (5−5′), phenylcoumaran (β-5′, α-O-4′), resinol (β-β′, α-O-γ′, α′-O-γ), and spirodienone (β-1′, α-O-α′). Dashed lines depict the variability of the number of MeO groups that changes depending on the three different monolignols. For further details of the resulting three-dimensional structure (branching, stereoisomers), see refs 1−3.

weight of lignins. If these were readily available, also GPC measurements could be absolutely calibrated using a second method. Several approaches have been proposed to achieve a universal calibration for GPC measurements of lignins, for example, by combining GPC with MALLS or viscosimetry.[2,8,9] However, as every method has its own advantages and limitations, achieving an absolute molecular weight determination is very complicated and not yet easily applicable to every kind of lignin.[9,10] For example in the case of MALDI-ToF-MS, achieving a constant ionization efficiency of lignin molecules over a broad molecular weight distribution is a challenge, especially for high molecular weights.[5]

Another method currently investigated for polymer analysis is pulsed-field-gradient NMR (PFG NMR), which generates diffusion-ordered spectroscopy (DOSY) NMR spectra. Using DOSY NMR, information about the self-diffusion of the species leading to the NMR signals can be obtained. From this pseudo-2D measurement, the diffusion coefficients (D) can be calculated using a fit on the Stejskal-Tanner equation (eq 1), which describes the signal intensity I as a function of D, the field gradient strength g, the timing parameters Δ and δ, and the gyromagnetic ratio γ.[11]

$$I = I_0 \, e^{-(\gamma g \delta)^2 (\Delta - \frac{\delta}{3}) D} \tag{1}$$

For a NMR signal originating from a single, monodisperse molecular species, the diffusion coefficient can be calculated using a simple least-squares exponential fit of eq 1. When the calculated diffusion values are plotted versus the chemical shifts, a typical 2D DOSY NMR spectrum is obtained. As the resulting diffusion coefficient D is related to the hydrodynamic volume via the Stokes−Einstein equation, a relation between D and the molecular weight of a sample seems consequent. It has been found that a power law similar to the Mark−Houwink−Sakurada (MHS) equation can be used to describe this relation for linear homopolymeric samples (with the scaling parameters K and α)[12]

$$D = KM^{-\alpha} \tag{2}$$

The evaluation of the molecular weight of classical, well-defined polymers such as polystyrene or poly(methyl methacrylate) using DOSY NMR has been shown in multiple publications, also including special fitting techniques to obtain the molecular weight distribution.[13−15]

Also for lignin, approaches have been reported to analyze the molecular weight by DOSY NMR. After solvent fractionation of several bulk lignins, leading to samples with a relatively low polydispersity of ca. 2, the MHS equation could be applied to the relation of M_w and D.[16] Similarly, lignin samples fractionated by GPC have been used to perform a calibration, but here the diffusion coefficients were extrapolated to infinite dissolution (D_0).[17]

This study aims to develop and compare models from DOSY NMR analysis as well as ^{1}H NMR to analyze the molecular weight properties of bulk, non-fractionated lignin samples. Previously, multivariate analysis techniques such as the partial-least-squares (PLS) regression from the DOSY NMR spectra have been applied for the less-complex heparin and low-molecular weight heparin. Heparin is a linear highly sulfated glycosaminoglycan consisting of repeating disaccharides of uronic acid and glucosamine in a 1,4-linkage, and is used as an anticoagulant. As a proof-of-principle, the determination of the M_w of heparin by DOSY NMR could be validated.[18] This approach was extended here to analyze M_w, M_n, and the PDI of lignin samples.

Furthermore, inspired by a recent Fourier transform infrared (FT-IR) study, ^{1}H NMR data with multivariate analysis were also used to obtain information about the molecular weight of lignins.[19] Typically, the molecular weight of lignins is determined via GPC, which was used to calibrate our models. Therefore, an alkaline GPC method using a sulfonated styrene-divinylbenzene copolymer column was chosen to overcome the lignin solubility problem while not needing to derivatize the samples, which could lead to a bias in the molecular weight.[20] All investigated samples in this study were completely soluble in an alkaline medium.

■ RESULTS AND DISCUSSION

In the ^{1}H NMR spectra of the organosolv lignin samples, the typical signals of aromatic protons at ca. δ 6.7 ppm as well as the methoxy signals at δ 3.7 ppm were observed. Also, strong aliphatic signals at δ 1.2 ppm and δ 0.85 ppm were detected (Figure 2). The most prominent lignin signals in the spectrum as well as the tetramethylsilane (TMS) and solvent resonance were picked for DOSY processing.

The aromatic and the methoxy signals showed a similar D value, which is expected for the same molecule (Figure 2). Yet, the aliphatic signals presented significantly higher D values,
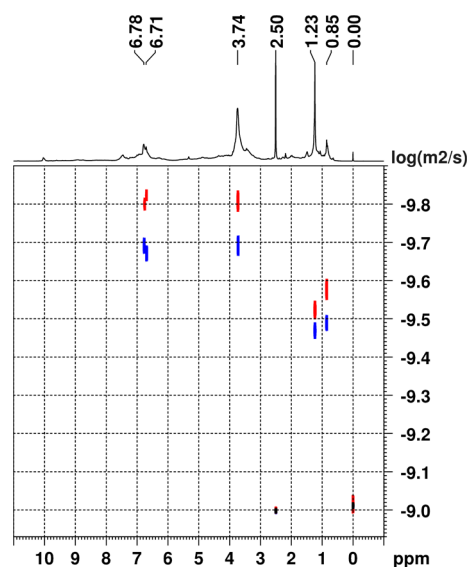
Figure 2. DOSY NMR spectra of a *Miscanthus* (blue) and a *Paulownia* (red) organosolv lignin, with the ^{1}H NMR spectrum of the *Miscanthus* lignin as f2 projection. The *Paulownia* lignin showed a higher molecular weight (M_w = 7123 Da, M_n = 1044 Da) than the *Miscanthus* sample (M_w = 1884 Da, M_n = 672 Da) and thus lower D values.

which correspond to a smaller molecule size. This difference, which was also found in other publications, can be explained by the heteropolymeric properties of the randomly cross-linked lignin (Figure 1).[21] Furthermore, a higher amount of aliphatic protons in smaller lignin molecules would lead to a higher diffusion coefficient, because smaller molecules would be weighted stronger than larger ones. Another explanation could be that the aliphatic signals are somehow comparable to end groups of a linear homopolymer. End group signals show a higher D than the resonances of the repetition units, similar to the different weighting of M_n and M_w.[22]

Power Law Correlation of the DOSY NMR Results with Molecular Weight. In Figure 3, the MHS plot of the measured lignin samples is shown. The aromatic lignin signals were used for the D determination. As reference molecular weight values, the GPC results were used. Although the absolute molecular weights are not accurate, the relative results still contain valuable information. Yet, in contrast to the MHS
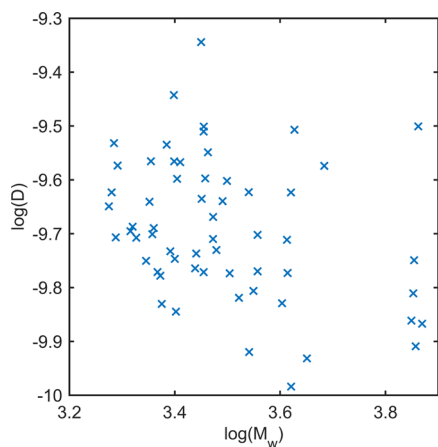
Figure 3. MHS plot of all organosolv lignin samples, using the aromatic signals for D determination.

plots from diffusion measurements of lignins described previously, no significant correlation was achieved here (R^2 = 0.11, log K = −8.7, and α = 0.27).[16,17] Similar results were achieved for the aliphatic and methoxy signals (see Supporting Information, Figure S1).

This effect could be explained by the rather large PDIs of the samples (2.5−7.2). A single diffusion coefficient of one signal might not suffice to describe the large distributional width of non-fractionated lignins. This limitation combined with a chemical heterogeneity over the molecular weight distribution possibly led to high errors.

The determination of D_0, as described by Rönnols et al., yielded acceptable results also for non-fractionated softwood kraft lignin samples.[17] Because a dilution series has to be measured for each sample, and the NMR measuring time increases significantly with stronger dilutions, this laborious method was not tried in this study.

When the needed information cannot be directly obtained by a single variable output, multivariate analysis could help to extract the data. For example, all the peaks evaluated in the PFG NMR experiment, or the whole DOSY NMR spectrum as shown before for heparin, could lead to better results.[18] Next to the PFG NMR experiment, also the data from simple ^{1}H NMR could contain information that is influenced by the molecular weight.

PLS Regression from the ^{1}H NMR Spectra. Looking at classical synthetic polymers, information about the chain length can also be obtained via ^{1}H NMR spectroscopy. If linear homopolymers show a sufficiently intense signal of the end groups, M_n can be calculated using the signal relation of the end group and the repeating unit.[23] This approach cannot be adapted for lignin, because it is heteropolymeric, has no defined end groups, and shows branching due to different types of linkages.[2] However, there is a relation between the line broadening and a higher molecule size, because of the faster T_2 relaxation times.[24] These examples show that the spectrum is influenced by the molecular weight of the sample, which could possibly be extracted by applying a PLS-R.

After the PLS regression using the ^{1}H NMR spectra ([1]H-PLS), the optimal number of included latent variables (LV) into the model was determined by cross validation (CV). The root mean square errors (RMSE) of the calibration set (RMSEC) and validation set (RMSEV) were determined. The optimal number of LVs was determined using the minimal RMSEV (Figure 4).

The results of the regression (Table 1) showed that with a simple and fast ^{1}H NMR measurement, information about the molecular weight of a lignin sample can be obtained in only 2 min. CV errors between 7 and 9% were achieved for all parameters. This is a reasonable accuracy, especially when
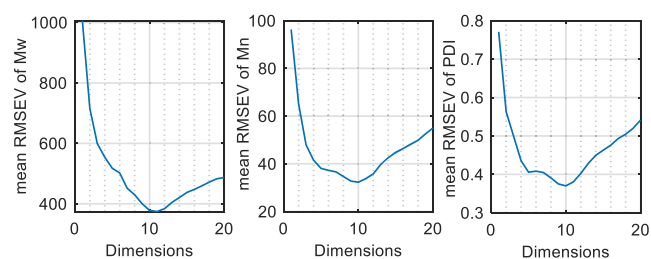
Figure 4. Mean errors of repeated CVs of the PLS regression using ^{1}H NMR spectra. 100 iterations for each CV were performed.

Table 1. PLS-R Results from the ^{1}H NMR Spectra

unit	range	num. LVs	R^2 cal.	R^2 val.	RMSEC	RMSEV	RE[a] (%)
M_w	1884−7405 Da	11	0.99	0.93	187 Da	374 Da	6.9
M_n	664−1052 Da	10	0.97	0.93	21 Da	33 Da	8.7
PDI	2.50−7.23	10	0.98	0.88	0.20	0.37	8.0

[a]Relative error RE = RMSEV/range.

taking the errors of the reference GPC measurement into account, because the results from the PLS regression cannot be better than the method that has been calibrated with. Also, the determination coefficients (R^2) were excellent for both the calibration and validation.

To determine which signals are important for the PLS model, the variable influence on the projection (VIP) scores can be calculated.[25] Figure 5 shows the VIP scores of the optimized models for all three parameters. Variables with a VIP score higher than 1 are usually considered important.
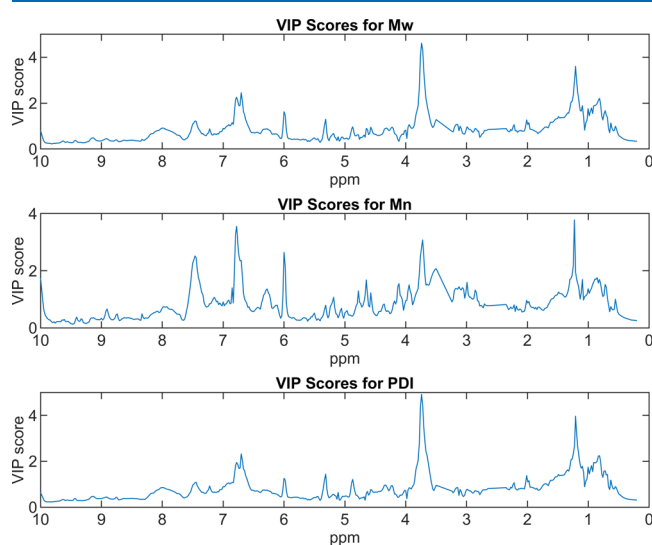

Figure 5. VIP scores for the PLS regression using the ^{1}H NMR spectra.

In all three VIP scores plots, the typical lignin signals are strongly emphasized. This shows that the PLS model is calculated mainly from the NMR signals of lignin itself rather than other artifacts. The most important signals, which were also found in all lignin samples, are the aliphatic peaks at ca δ 0.8 and δ 1.2 ppm, the methoxy signal at δ 3.7 ppm, and the aromatic signals around δ 6.7 and δ 6.8 ppm. These signals were picked and used for the following DOSY analysis.

PLS Regression from the DOSY NMR Spectra. Next to the PLS-R from the one-dimensional (1D) ^{1}H NMR spectra, also a regression using 2D DOSY NMR spectra was performed (DOSY-PLS). The optimal number of LVs was determined using the minimal RMSEV (Figure 6). 4 LVs were optimal for all parameters, which shows that there is a rather direct relation of the DOSY NMR spectra with the molecular weight in comparison to the ^{1}H NMR spectra, where a larger number of LVs were needed. This can be explained by the dependence of D on the hydrodynamic volume of the analyte, which corresponds to the molecular weight (see eq 2).

The results of the modeling showed a reasonable coefficient of determination of more than 0.9, and relative validation errors between 14 and 16% for all three tested parameters
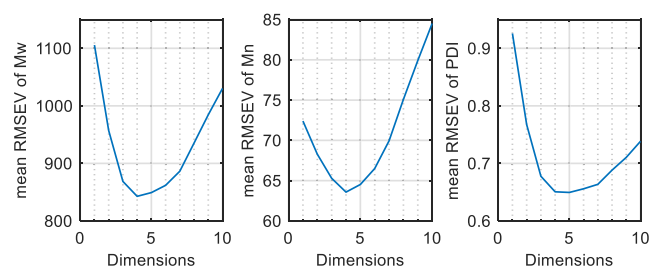

Figure 6. Mean errors of repeated CVs of the PLS regression using the DOSY NMR spectra. 100 iterations of each CV were performed.

(Table 2). For the analyzed samples, also the DOSY NMR evaluation with a monoexponential fit contained information about the dispersity of the samples. This information could be contained in the relations of the five picked lignin signals to each other, similar to the varying diffusion behavior of an end group signal and a signal of a repeating unit.[22]

Regression from the DOSY NMR Fit Results. The DOSY NMR spectrum is generated from the ^{1}H shifts of each picked peak and the resulting fitted D value. The peak width is determined using the ^{1}H peak width and the fit error. Unfolding the 2D spectrum by concatenating all the lines to obtain one large vector results in a big number of variables, which are bucketed afterward. However, the fit results of the PFG NMR measurement can also directly be used for data analysis, without generating a pseudo-2D spectrum in between. Therefore, only the five diffusion values were used as variables for this analysis. Like in the other previous models, 100 iterations of a CV were performed to obtain the optimal number of dimensions. As no dimensional reduction was possible (for RMSEV plots, see Supporting Information, Figure S2), the PLS regression, which is based on extracting LVs and thus decreasing the dimensionality, was not an appropriate tool for this dataset.

Alternatively, a multiple linear regression (MLR) was performed (PFG-Fit-MLR). Similar to the Mark−Houwink approach, it was observed that the model performance increases, when the molecular weights were logarithmized prior to analysis. To be able to compare the errors (RMSE) to the those of other models, the results were first delogarithmized again and then the RMSE was calculated. In Figure 7, the "real" values of the calibration samples are scattered against the values determined using the MLR. A clear correlation is visible in these plots, but still a rather high amount of scattering is visible. Especially the calibration for PDI leads to a large scattering, which can also be seen in the regression coefficients as well as in the RMSEC and RMSEV (Table 3).

The results of the MLR showed similar validation errors to those of the PLS model from the whole DOSY NMR spectrum for M_n and M_w. The PDI determination was slightly better from the DOSY NMR data. In contrast, a significantly lower coefficient of determination from the calibration and also a much higher RMSEC were found in the MLR model. The same trend can be seen when comparing the determination

Table 2. PLS-R Results from the DOSY NMR Spectra

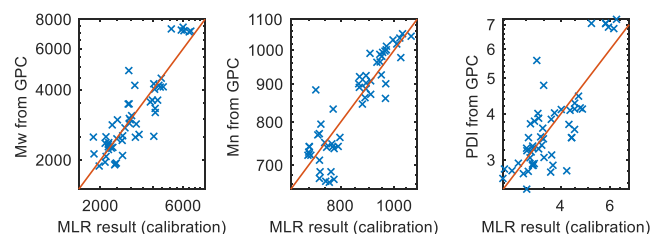| unit	range	num. LVs	R^2 cal.	R^2 val.	RMSEC	RMSEV	RE[a] (%)
M_w	1884−7405 Da	4	0.92	0.69	454 Da	837 Da	15
M_n	664−1052 Da	4	0.93	0.74	32 Da	64 Da	16
PDI	2.50−7.23	4	0.93	0.72	0.34	0.64	14

[a]Relative error RE = RMSEV/range.

Figure 7. Scatterplots of the MLR calibration. The GPC data are plotted against the results of the calibration samples from the MLR on a logarithmic scale due to the model being calculated with the logarithmized data.

Table 3. Multiple Linear Regression Results from the DOSY NMR Fit Results

unit	range	R^2 cal.	R^2 val.	RMSEC	RMSEV	RE[a] (%)
M_w	1884−7405 Da	0.79	0.72	677 Da	804 Da	15%
M_n	664−1052 Da	0.80	0.77	51 Da	62 Da	16%
PDI	2.50−7.23	0.65	0.53	0.71	0.87	18%

[a]Relative error RE = RMSEV/range.

coefficients of both methods. This shows that the model is less stable due to the scattering of the calibration samples. The same measurements processed as the DOSY spectrum resulted in a model with better calibration results. The MLR model does not adapt as much to the calibration samples as the more complex PLS regression, which explains the higher scattering of the calibration set.

The DOSY spectral processing before the PLS regression seems to be superior to a MLR analysis of the diffusion data, with better results of the calibration and comparable CV results.

Variance of Repeated Measurements. There were generally higher errors of the calibration and validation from the DOSY-PLS models than from the [1]H-PLS model. These differences can also be seen in repeated measurements of the same samples. The samples used for this analysis were taken from the calibration set ($n = 3$) and measured in triplicate. The [1]H-PLS model results showed a considerably lower variance than the results from the DOSY-PLS model and the MLR model from the direct PFG NMR analysis (Table 4). However, a considerably lower variance was observed in the MLR model

Table 4. Comparison of the Variation Coefficients from Repeated Measurements Using Different Methods[a]

unit	[1]H-PLS (%)	DOSY-PLS (%)	PFG-Fit-MLR (%)
M_w	0.9	10	5
M_n	0.2	2	1
PDI	1.2	9	4

[a]All values are expressed as the average variation coefficient of the determined parameter, using three repeated measurements of three different samples.

than that in the DOSY-PLS. This shows that the measurement and processing of the same sample has a significant impact on the error of the molecular weight determination using DOSY NMR. The ^{1}H NMR measurement produced, as expected, highly reproducible results.

A threefold determination of the M_w of non-fractionated kraft lignins by Rönnols et al. using a Mark−Houwink approach after D_0 determination led to variation coefficients of ca. 10%.[17] This shows that these errors could possibly be strongly influenced by the instability of the PFG NMR measurement itself, and not only on the method of processing and model calibration. Looking at the determination of D values itself, a mean variance of 4% from repeated measurements was observed. A systematic analysis of the uncertainty of a D value determination of disperse polymers using PFG NMR is needed. This is especially challenging, because strictly speaking, not one D value from a single exponential decay, but a whole distribution is generated.

Model Performance on Different Biomasses. Next to the CV, which depicts the performance of the model on samples most similar to the calibration set, a small independent sample set was used to investigate the performance of the models on lignin samples of different origins. The samples were extracted using the same organosolv process, but the biomass origin was varied. The beech and wheat straw samples showed very similar GPC results, which were within the region of the used calibration samples. Three extractions were performed from wheat straw (W1a-c), and three from beech wood (B1a-c). To be able to apply the samples for the previously created models, they were applied for the PCA analyses of the sample set. All samples were not seen as outliers in the PCA scores plots for all model types and could thus be applied (see Supporting Information, Figure S3−S5).

The results from the [1]H-PLS model were very similar to the reference GPC values for all three parameters. For the [1]H-PLS, the RMSEP of these samples (Table 5) was similar to the

Table 5. External Validation Using Other Biomasses[a]

unit	range	RMSEP of [1]H-PLS	RMSEP of DOSY-PLS	RMSEP of PFG-Fit-MLR
M_w	2759−3156 Da	449 Da	2467 Da	875 Da
M_n	841−1011 Da	49 Da	153 Da	95 Da
PDI	2.83−3.37	0.3	2.0	1.1

[a]The RMSEP values are generated from the difference of the regression model results to the GPC results of the beech wood and wheat straw samples.

RMSEV, which shows that the model could be successfully applied. For the two models generated from the DOSY NMR measurement, it can again clearly be seen that the MLR leads to more stable results than the PLS-R. While for the DOSY-PLS, the RMSEP was much higher than the RMSEV, they were in a similar range for the PFG-Fit-MLR, indicating a successful application of the MLR.

Comparison of Different Methods. When comparing the CV results between different parameters and sample sets, the relative RMSEV values can be used as a benchmark, which are again summarized for all proposed methods in Table 6. For

Table 6. Summarized Results of the Different Applied Models[a]

parameter	range	method	RMSEV	rel. error[b] (%)
M_w	2759–3156 Da	¹H-PLS	374 Da	6.9
		DOSY-PLS	837 Da	15
		PFG-Fit-MLR	804 Da	15
M_n	841–1011 Da	¹H-PLS	33 Da	8.7
		DOSY-PLS	64 Da	16
		PFG-Fit-MLR	62 Da	16
PDI	2.83–3.37	¹H-PLS	0.37	8.0
		DOSY-PLS	0.64	14
		PFG-Fit-MLR	0.87	18

[a]All errors are related to the corresponding GPC measurements. [b]Relative error = Validation error/range.

the ¹H-PLS models, the relative errors were around 8%, both models from the PFG NMR measurements resulted in errors around 15%. Yet, it can be clearly seen that the PFG-Fit-MLR performs better than the DOSY-PLS regarding precision (variance of repeated measurements), and also accuracy of the prediction of new samples (Tables 4 and 5). This concludes that regarding DOSY measurements, the PFG-Fit-MLR model is superior to a PLS regression using the 2D DOSY spectrum. However, the ¹H-PLS model outperforms the DOSY models in all comparisons.

In the literature, only the determination of M_w was described using PFG NMR (Table 7, entries 3 and 4). In contrast to the results found in this study, a Mark–Houwink approach was applied for the calibration, using fractionated samples. The errors of the previously published models have the same order of magnitude (see Table 7). Rönnols et al. also used non-fractionated lignin samples with PDIs from 3 to 5, but focused only on softwood kraft lignins.[17] Montgomery et al. used a solvent fractionation approach resulting in samples with smaller PDIs of around 2, which should significantly simplify the D determination because its distribution is smaller. Also, a parallel analysis of samples from vastly different biomasses and pulping processes was possible. It could be shown that the model is also applicable for bulk samples, which provided

results comparable to the GPC measurements within an error range of ca. ± 1000 Da.[16]

Our models from the PFG data work with samples of high dispersity and additionally provide information about M_n and PDI. A validation with external samples from a different botanical origin showed high errors for the DOSY-PLS, so its applicability to other samples seems to be limited. The ¹H-PLS model as well as the PFG-Fit-MLR were applicable for these samples. When comparing the scattering of the results from repeated measurements, the PFG-Fit-MLR showed a significantly lower variance than the DOSY-PLS. Therefore, the MLR approach seems better suited than the PLS regression using the whole DOSY NMR spectrum. However, the rather high error of the models from PFG NMR measurements imposes the need for optimization of the measurement parameters and processing procedures, to obtain more precise and accurate results. Many factors such as the sample concentration, the used solvent system, and temperature as well as the pulse program and its timing values generally have an impact on the results.

The ¹H-PLS model showed comparable results to the literature data from Lancefield et al., who used the attenuated total reflection (ATR-) FTIR spectra (Table 7, entries 1 and 2). The absolute errors found in our study were significantly lower, but also a smaller set of only three different biomasses was used. Due to the fractionation procedures, which were applied on many samples and led to a higher molecular weight range, the relative errors were similar to those found here.[19]

So far, all publications on PFG NMR to determine the molecular weight of lignin samples have used fit approaches that determine a single D value, and not a distribution.[16,17,21,26,27] The high dispersity of non-fractionated lignins and the chemical heterogeneity over the molecular weight distribution makes finding the optimal fitting approach difficult. A comparative analysis of different fitting methods on lignin samples would be helpful to develop an optimal PFG analysis technique.

■ CONCLUSIONS

Generally speaking, it can be seen that conventional 1D NMR methods such as ¹H NMR combined with multivariate modeling are appropriate to obtain fast and accurate information about the molecular weight of lignins. While it takes about 50 min to acquire a GPC chromatogram, a ¹H NMR spectrum can be generated in 2 min. The proposed

Table 7. Literature Results of Spectroscopic Methods Used for Lignin Molecular Weight Determination[a]

used method	samples	n[b]	parameter	range	validation error	rel. error[c]	refs
ATR-FTIR, PLS-R, CV	kraft and organosolv lignins, from softwood, hardwood, and herbaceous biomass; partially solvent fractionated	54	M_w	29,984 Da	2278 Da	7.6%	19
			M_n	4626 Da	287 Da	6.2%	
			PDI	5.0	1.0	20%	
ATR-FTIR, PLS-R, CV	three softwood kraft lignins, solvent fractionated	28	M_w	29,984 Da	1739 Da	5.8%	19
			M_n	4570 Da	229 Da	5.0%	
			PDI	4.9	0.51	10%	
PFG NMR, MHS equation	dioxasolv and kraft lignins from soft- and hardwoods, solvent fractionated	151	M_w	8800 Da	962 Da (RMSEC)	11% (cal.)	16
PFG NMR, MHS equation	softwood kraft lignins, calibrants from GPC fractionation	5	M_w	8700 Da (Cal)	663 Da	7.6%	17
				3900 Da (val)		17%	

[a]All errors are related to the corresponding GPC measurements. [b]Number of calibration samples. [c]Relative error = Validation error/range.

method of a multiple linear regression from DOSY NMR measurements also leads to accurate molecular weight results. Both approaches work with samples of high dispersity and also provide information about M_w, M_n, and PDI. The proposed methods could be an option for industrial quality control to quickly characterize process streams, where a large number of samples have to be analyzed routinely. Further studies are required to confirm the hypothesis that this technique could be used complementary to the currently most popular GPC.

However, a simple ^{1}H NMR measurement leads to comparable or even better molecular weight results than a DOSY NMR measurement and is faster to measure and easier to process. This is counterintuitive, because PFG NMR provides a rather direct relation between the signal decay due to diffusion, and the molecular weight of the samples. However, the high variance in a PFG measurement of polydisperse samples that was observed in this study, combined with the need to optimize the data processing for disperse samples, seems to lead to errors in the D determination. In order to achieve even better results for the molecular weight determination of lignin using simple spectroscopic methods, multivariate analysis methods applying data fusion, for example, NMR and IR data, are currently under investigation. Also, the use of benchtop NMR spectrometers instead of high-field spectrometers is a possible alternative, which could further simplify the determination of molecular weight characteristics of complex polymers.

■ EXPERIMENTAL SECTION

Lignin Samples. A set of 53 lignin samples, isolated from *Miscanthus x giganteus* (n = 25), *Paulownia tomentosa* (n = 16) and *Silphium perfoliatum* (n = 12), using a catalyst-free ethanol organosolv pulping process was previously prepared and characterized.[28] A representative set of various lignin samples has been collected and analyzed. The differences in the samples were due to different particle size fractions of the biomasses as well as some slight changes in the extraction process like temperature, ethanol concentration, or an added aqueous autohydrolysis step. As validation samples, the same organosolv pulping procedure was applied on a wheat straw sample and a beech wood sample, each in triplicates.

NMR Spectroscopy. NMR measurements were performed on a Bruker AVANCE III 600 MHz spectrometer (Bruker Biospin, Rheinstetten, Germany) with a BBO cryo probe equipped with a Bruker Automatic Sample Changer (B-ACS 120) at 297 K.

^{1}H NMR Spectra were acquired in DMSO-d_6 + 0.03% (v/v) TMS at 70 mg mL^{-1} using a 30° flip angle (zg30 in Bruker language). The relaxation delay (D1) was set to 1 s, time domain: 128k data points, the number of scans: 16, and the acquisition time: 4.54 s.

The PFG NMR experiment was performed using 10 mg of the lignin dissolved in 700 μL DMSO-d_6 + 0.03% (v/v) TMS. The acquisition was performed using the longitudinal eddy current delay with a bipolar gradient pulse pair pulse program (ledbpgp2s) with 16 scans of 32 linear field gradient steps (SW 12 ppm, O1 5 ppm, P30 1750 μs, and D20 300 ms). P30 and D20 were adjusted so that a signal attenuation of more than 95% was achieved.

GPC Measurements. Samples for GPC analysis were prepared in sodium hydroxide solution (0.1 M) with a concentration of 5 mg mL^{-1} and filtered through 0.2 μm polyamide filters prior to analysis. M_w, M_n, and PDI were

determined on a PSS GPC system based on Agilent 1260 Infinity (Mainz, Germany and Waldbronn, Germany, respectively) with a PSS MCX 8 × 50 mm 5 μm precolumn, two PSS MCX 8 × 300 mm 5 μ 1000 Å columns, and one PSS MCX 8 × 300 mm 5 μm 100,000 Å column. A UV detector at 280 nm was used, with sodium hydroxide solution (0.1 M) as an eluent at a flow rate of 1 mL min^{-1} and an injection volume of 50 μL. The calibration was performed with 11 molecular weight standards (benzenesulfonic acid sodium salt and polystyrene sulfonate sodium salts) in the range from 180 Da to 976 kDa using a fifth-degree polynomial function (R = 0.999). The elution times were corrected using ethylene glycol (1 mg mL^{-1}) as the internal standard.

A table of all determined molecular weights (M_w, M_n, and PDI) of all samples can be found in the Supporting Information (Table S1).

Data Analysis. *Processing of the PFG NMR Data.* Bruker Topspin 4.0.7 and Bruker Dynamics Center 2.5.6 were used for processing the spectra. The raw NMR data were Fourier transformed on the f2 axis. A manual phase correction was performed along the ^{1}H direction. The spectra were baseline-corrected with a polynomial of 5th degree. The spectra were further processed by peak picking of the signals at δ 0, δ 0.92, δ 1.30, δ 2.54, δ 3.79, δ 6.76, and δ 6.83 ppm, then a monoexponential fit was applied to calculate D. The DOSY spectra as well as the fit results for each picked signal with the fitting errors were exported.

Chemometric Modeling. Multivariate data analysis was performed using the SAISIR Package within MATLAB R2018b (The Math Works, Natick, MA, USA).[29] The scripts for processing and modeling are available upon request.

The ^{1}H NMR spectra were preprocessed by truncating the spectrum from δ 0.2 to δ 10.0 ppm, also removing DMSO and residual water signals. After bucketing with a width of 0.02 ppm, the spectra were divided by the sum of all their signals to normalize the spectral intensity.

The DOSY NMR spectral region with species diffusing faster than −9.38 log(m^2 sec^{-1}) was removed to eliminate solvent and TMS signals. Along the ^{1}H axis, a bucketing in the signal region from δ 7.5−0.5 ppm was applied. For the M_w and PDI calibration, a bucket width of 0.2 ppm was found as optimal, for M_n a bucket width of 0.02 ppm resulted in minimal CV errors. After bucketing, the spectra were divided by the sum of all their signals to normalize the spectral intensity, and Pareto scaled. As performed in previous work of the group, the 2D spectra were then unfolded slice by slice, to obtain a one-dimensional vector, which was used for the PLS regression.[18] The GPC results were used as reference values.

CVs were performed with 100 repetitions, randomly splitting 25% of the samples into the validation set. The errors and coefficients of determination were calculated as the mean value from all repetitions.

■ ASSOCIATED CONTENT

ⓈI Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.1c03574.

> GPC results of all used samples; MHS plots of all organosolv lignin samples, using the D values of the aromatic, methoxy, and aliphatic signals; RMSEV errors of the PLS model using the PFG NMR fit results; and PCA scores plots of the calibration samples and applied

external beech wood and wheat straw samples, from ^{1}H NMR, DOSY NMR, and PFG NMR Fit data (PDF)

AUTHOR INFORMATION

Corresponding Authors

Yulia B. Monakhova − *Department of Chemistry and Biotechnology, FH Aachen University of Applied Sciences, Jülich 52428, Germany; Institute of Chemistry, Saratov State University, 410012 Saratov, Russia*; Email: monakhova@fh-aachen.de

Margit Schulze − *Department of Natural Sciences, Bonn-Rhein-Sieg University of Applied Sciences, Rheinbach D-53359, Germany;* orcid.org/0000-0002-8975-1753; Email: margit.schulze@h-brs.de

Authors

René Burger − *Department of Natural Sciences, Bonn-Rhein-Sieg University of Applied Sciences, Rheinbach D-53359, Germany;* orcid.org/0000-0002-7801-5734

Jessica Rumpf − *Department of Natural Sciences, Bonn-Rhein-Sieg University of Applied Sciences, Rheinbach D-53359, Germany;* orcid.org/0000-0002-4913-3248

Xuan Tung Do − *Department of Natural Sciences, Bonn-Rhein-Sieg University of Applied Sciences, Rheinbach D-53359, Germany;* orcid.org/0000-0001-9989-2616

Bernd W. K. Diehl − *Spectral Service AG, D-50996 Köln, Germany*

Matthias Rehahn − *Department of Chemistry, Technical University of Darmstadt, Darmstadt D-64287, Germany*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acsomega.1c03574

Author Contributions

This article was written by R.B. with contributions of all authors. All authors have given approval to the final version of the article.

Notes

The authors declare no competing financial interest.

ABBREVIATIONS

ATR-FTIR, attenuated total reflection FTIR; CV, cross validation; DOSY, diffusion-ordered spectroscopy; G, guaiacyl lignin subunit; GPC, gel permeation chromatography; H, *p*-hydroxyphenyl lignin subunit; HSQC, hetero single quantum coherence; LV, latent variables; MALDI-ToF-MS, matrix-assisted laser desorption time-of-flight mass spectrometry; MALLS, multiangle light scattering; MHS, Mark-Houwink-Sakurada; MLR, multivariate linear regression; PDI, poly-dispersity index; PFG NMR, pulsed-field-gradient NMR; PLS-R, partial least squares regression; py-GC/MS, pyrolysis gas chromatography/mass spectrometry; RE, relative error; RMSE, root mean square error; RMSEC, RMSE of calibration; RMSEP, RMSE of prediction; RMSEV, RMSE of validation; S, syringyl lignin subunit; TMS, tetramethylsilane

REFERENCES

(1) Heitner, C.; Dimmel, D.; Schmidt, J. A. *Lignin and Lignans: Advances in Chemistry*; Taylor & Francis, 2010.

(2) Tolbert, A.; Akinosho, H.; Khunsupat, R.; Naskar, A. K.; Ragauskas, A. J. Characterization and analysis of the molecular weight of lignin for biorefining studies. *Biofuels, Bioprod. Bioref.* **2014**, *8*, 836−856.

(3) Lahive, C. W.; Kamer, P. C. J.; Lancefield, C. S.; Deuss, P. J. An Introduction to Model Compounds of Lignin Linking Motifs; Synthesis and Selection Considerations for Reactivity Studies. *ChemSusChem* **2020**, *13*, 4238−4265.

(4) Constant, S.; Wienk, H. L. J.; Frissen, A. E.; Peinder, P. d.; Boelens, R.; van Es, D. S.; Grisel, R. J. H.; Weckhuysen, B. M.; Huijgen, W. J. J.; Gosselink, R. J. A.; Bruijnincx, P. C. A. New insights into the structure and composition of technical lignins: a comparative characterisation study. *Green Chem.* **2016**, *18*, 2651−2665.

(5) Lupoi, J. S.; Singh, S.; Parthasarathi, R.; Simmons, B. A.; Henry, R. J. Recent innovations in analytical methods for the qualitative and quantitative assessment of lignin. *Renewable Sustainable Energy Rev.* **2015**, *49*, 871−906.

(6) Salanti, A.; Orlandi, M.; Zoia, L. Fluorescence Labeling of Technical Lignin for the Study of Phenolic Group Distribution as a Function of the Molecular Weight. *ACS Sustainable Chem. Eng.* **2020**, *8*, 8279−8287.

(7) Lange, H.; Rulli, F.; Crestini, C. Gel Permeation Chromatography in Determining Molecular Weights of Lignins: Critical Aspects Revisited for Improved Utility in the Development of Novel Materials. *ACS Sustainable Chem. Eng.* **2016**, *4*, 5167−5180.

(8) Himmel, M. E.; Tatsumoto, K.; Grohmann, K.; Johnson, D. K.; Chum, H. L. Molecular weight distribution of aspen lignins from conventional gel permeation chromatography, universal calibration and sedimentation equilibrium. *J. Chromatogr. A* **1990**, *498*, 93−104.

(9) Zinovyev, G.; Sulaeva, I.; Podzimek, S.; Rössner, D.; Kilpeläinen, I.; Sumerskii, I.; Rosenau, T.; Potthast, A. Getting Closer to Absolute Molar Masses of Technical Lignins. *ChemSusChem* **2018**, *11*, 3259−3268.

(10) Jacobs, A.; Dahlman, O. Absolute molar mass of lignins by size exclusion chromatography and MALDI-TOF mass spectroscopy. *Nord. Pulp Pap. Res. J.* **2000**, *15*, 120−127.

(11) Sinnaeve, D. The Stejskal-Tanner equation generalized for any gradient shape-an overview of most pulse sequences measuring free diffusion. *Concepts Magn. Reson., Part A* **2012**, *40*, 39−65.

(12) Augé, S.; Schmit, P.-O.; Crutchfield, C. A.; Islam, M. T.; Harris, D. J.; Durand, E.; Clemancey, M.; Quoineaud, A.-A.; Lancelin, J.-M.; Prigent, Y.; Taulelle, F.; Delsuc, M.-A. NMR measure of translational diffusion and fractal dimension. Application to molecular mass measurement. *J. Phys. Chem. B* **2009**, *113*, 1914−1918.

(13) Guo, X.; Laryea, E.; Wilhelm, M.; Luy, B.; Nirschl, H.; Guthausen, G. Diffusion in Polymer Solutions: Molecular Weight Distribution by PFG-NMR and Relation to SEC. *Macromol. Chem. Phys.* **2017**, *218*, 1600440.

(14) Gu, K.; Onorato, J.; Xiao, S. S.; Luscombe, C. K.; Loo, Y.-L. Determination of the Molecular Weight of Conjugated Polymers with Diffusion-Ordered NMR Spectroscopy. *Chem. Mater.* **2018**, *30*, 570−576.

(15) Rosenboom, J.-G.; De Roo, J.; Storti, G.; Morbidelli, M. Diffusion (DOSY) 1H NMR as an Alternative Method for Molecular Weight Determination of Poly(ethylene furanoate) (PEF) Polyesters. *Macromol. Chem. Phys.* **2017**, *218*, 1600436.

(16) Montgomery, J. R. D.; Bazley, P.; Lebl, T.; Westwood, N. J. Using Fractionation and Diffusion Ordered Spectroscopy to Study Lignin Molecular Weight. *ChemistryOpen* **2019**, *8*, 601−605.

(17) Rönnols, J.; Jacobs, A.; Aldaeus, F. Consecutive determination of softwood kraft lignin structure and molar mass from NMR measurements. *Holzforschung* **2017**, *71*, 563−570.

(18) Monakhova, Y. B.; Diehl, B. W. K.; Do, T. X.; Schulze, M.; Witzleben, S. Novel method for the determination of average molecular weight of natural polymers based on 2D DOSY NMR and chemometrics: Example of heparin. *J. Pharm. Biomed. Anal.* **2018**, *149*, 128−132.

(19) Lancefield, C. S.; Constant, S.; de Peinder, P.; Bruijnincx, P. C. A. Linkage Abundance and Molecular Weight Characteristics of Technical Lignins by Attenuated Total Reflection-FTIR Spectroscopy Combined with Multivariate Analysis. *ChemSusChem* **2019**, *12*, 1139−1146.

(20) Andrianova, A. A.; Yeudakimenka, N. A.; Lilak, S. L.; Kozliak, E. I.; Ugrinov, A.; Sibi, M. P.; Kubátová, A. Size exclusion chromatography of lignin: The mechanistic aspects and elimination of undesired secondary interactions. *J. Chromatogr. A* **2018**, *1534*, 101−110. Published Online: Dec. 24, 2017

(21) Cornejo, A.; García-Yoldi, Í.; Alegria-Dallo, I.; Galilea-Gonzalo, R.; Hablich, K.; Sánchez, D.; Otazu, E.; Funcia, I.; Gil, M. J.; Martínez-Merino, V. Systematic Diffusion-Ordered Spectroscopy for the Selective Determination of Molecular Weight in Real Lignins and Fractions Arising from Base-Catalyzed Depolymerization Reaction Mixtures. *ACS Sustainable Chem. Eng.* **2020**, *8*, 8638−8647.

(22) Viéville, J.; Tanty, M.; Delsuc, M.-A. Polydispersity index of polymers revealed by DOSY NMR. *J. Magn. Reson.* **2011**, *212*, 169−173.

(23) Izunobi, J. U.; Higginbotham, C. L. Polymer Molecular Weight Analysis by 1 H NMR Spectroscopy. *J. Chem. Educ.* **2011**, *88*, 1098−1104.

(24) Groves, P. Diffusion ordered spectroscopy (DOSY) as applied to polymers. *Polym. Chem.* **2017**, *8*, 6700−6708.

(25) Wold, S.; Johansson, E.; Cocchi, M. PLS - Partial Least-Squares Projections to Latent Structures. In *3D QSAR in Drug Design*, Kubinyi, H., Ed.; Kluwer/ESCOM, 2000; pp 523−550.

(26) Rönnols, J.; Danieli, E.; Freichels, H.; Aldaeus, F. Lignin analysis with benchtop NMR spectroscopy. *Holzforschung* **2020**, *74*, 226−231.

(27) Montgomery, J. R. D.; Lancefield, C. S.; Miles-Barrett, D. M.; Ackermann, K.; Bode, B. E.; Westwood, N. J.; Lebl, T. Fractionation and DOSY NMR as Analytical Tools: From Model Polymers to a Technical Lignin. *ACS Omega* **2017**, *2*, 8466−8474.

(28) Rumpf, J.; Do, X. T.; Burger, R.; Monakhova, Y. B.; Schulze, M. Extraction of High-Purity Lignins via Catalyst-free Organosolv Pulping from Low-Input Crops. *Biomacromolecules* **2020**, *21*, 1929−1942.

(29) Cordella, C. B. Y.; Bertrand, D. SAISIR: A new general chemometric toolbox. *Trends Anal. Chem.* **2014**, *54*, 75−82.