

# 10 An In-Depth Understanding of Biomass Recalcitrance Using Natural Poplar Variants as the Feedstock

Xianzhi Meng,<sup>[a]</sup> Yunqiao Pu,<sup>[b, c]</sup> Chang Geun Yoo,<sup>[b, c]</sup> Mi Li,<sup>[b, c]</sup> Garima Bali,<sup>[d]</sup> Doh-Yeon Park,<sup>[d]</sup> Erica Gjersing,<sup>[b, e]</sup> Mark F. Davis,<sup>[b, e]</sup> Wellington Muchero,<sup>[b, c]</sup> Gerald A. Tuskan,<sup>[b, c]</sup> Timothy J. Tschaplinski,<sup>[b, c]</sup> and Arthur J. Ragauskas<sup>\*[a, b, c, f]</sup>

In an effort to better understand the biomass recalcitrance, six natural poplar variants were selected as feedstocks based on previous sugar release analysis. Compositional analysis and physicochemical characterizations of these poplars were performed and the correlations between these physicochemical properties and enzymatic hydrolysis yield were investigated. Gel permeation chromatography (GPC) and <sup>13</sup>C solid state NMR were used to determine the degree of polymerization (DP) and crystallinity index (CrI) of cellulose, and the results along with the sugar release study indicated that cellulose DP likely played a more important role in enzymatic hydrolysis. Simons' stain revealed that the accessible surface area of substrate significantly varied among these variants from 17.3 to

33.2 mg g<sup>-1</sup><sub>biomass</sub> as reflected by dye adsorption, and cellulose accessibility was shown as one of the major factors governing substrates digestibility. HSQC and <sup>31</sup>P NMR analysis detailed the structural features of poplar lignin variants. Overall, cellulose relevant factors appeared to have a stronger correlation with glucose release, if any, than lignin structural features. Lignin structural features, such as a phenolic hydroxyl group and the ratio of syringyl and guaiacyl (S/G), were found to have a more convincing impact on xylose release. Low lignin content, low cellulose DP, and high cellulose accessibility generally favor enzymatic hydrolysis; however, recalcitrance cannot be simply judged on any single substrate factor.

[a] Dr. X. Meng, Prof. A. J. Ragauskas  
Department of Chemical & Biomolecular Engineering  
University of Tennessee Knoxville  
Knoxville, TN 37996 (USA)  
E-mail: aragausk@utk.edu

[b] Dr. Y. Pu, Dr. C. G. Yoo, Dr. M. Li, Dr. E. Gjersing, Dr. M. F. Davis,  
Dr. W. Muchero, Prof. G. A. Tuskan, Prof. T. J. Tschaplinski,  
Prof. A. J. Ragauskas  
BioEnergy Science Center  
Oak Ridge National Laboratory  
Oak Ridge, TN 37831 (USA)

[c] Dr. Y. Pu, Dr. C. G. Yoo, Dr. M. Li, Dr. W. Muchero, Prof. G. A. Tuskan,  
Prof. T. J. Tschaplinski, Prof. A. J. Ragauskas  
Biosciences Division  
Oak Ridge National Laboratory  
Oak Ridge, TN 37831 (USA)

[d] Dr. G. Bali, Dr. D.-Y. Park  
Renewable Bioproducts Institute  
School of Chemistry and Biochemistry  
Georgia Institute of Technology  
Atlanta, GA 30332 (USA)

[e] Dr. E. Gjersing, Dr. M. F. Davis  
National Renewable Energy Laboratory  
Golden, CO 80401 (USA)

[f] Prof. A. J. Ragauskas  
Center for Renewable Carbon  
Department of Forestry, Wildlife, and Fisheries  
University Tennessee Institute of Agriculture  
Knoxville, TN 37996 (USA)

Supporting Information and the ORCID identification number(s) for the author(s) of this article can be found under <http://dx.doi.org/10.1002/cssc.201601303>.

This publication is part of a Special Issue celebrating "10 Years of ChemSusChem". A link to the issue's Table of Contents will appear here once it is complete.

## Introduction

Sustainable energy production has become a global research effort with increasing concerns about diminishing fossil fuels and global warming. Recent advances in genetics, biochemistry, and chemical engineering have led to great progress toward the concept of converting biomass to biofuels.<sup>[1]</sup> This bioconversion process usually involves five steps: feedstock size reduction, pretreatment, enzymatic hydrolysis, fermentation, and product purification/distillation. Owing to the innate biomass recalcitrance, pretreatment has become a necessary process to open the lignin carbohydrate matrix, increasing the efficiency of enzymatic hydrolysis, which is the process of converting the polysaccharides into their 5- and 6-carbon chain sugars.<sup>[2]</sup> Currently, high costs associated with biomass pretreatment and saccharification are the major barrier that significantly hinders the industrial commercialization process. Therefore, cost-effectively overcoming biomass recalcitrance has become one of the most pressing issues in plant-based green technologies.

Historically, natural factors believed to contribute to biomass recalcitrance include particle size, biomass porosity, cellulose degree of polymerization (DP) and crystallinity, cellulose accessibility, and lignin/hemicellulose content, distribution, and structures.<sup>[3]</sup> For example, the strong inter- and intra-chain hydrogen-bonding system in cellulose makes crystalline part highly resistant to enzyme attack. It has been shown that a completely amorphous cellulose sample is hydrolyzed much faster than a crystalline cellulose sample.<sup>[4]</sup> The fact that intimate contact between cellulose and enzymes is the prerequisite step for enzymatic hydrolysis to occur certainly makes cel-

lulose accessibility one of the most important factors. It was reported that drying of lignocellulosic substrates could significantly decrease porosity/surface area, known as fiber hornification, and subsequently reduce the following cellulose saccharification. The result from this drying-induced fiber hornification generated substrates that were chemically identical but varied significantly in cellulose accessibility, suggesting that cellulose accessibility was the dominant and probably the only factor to reduce enzymatic susceptibility of cellulose in this particular scenario.<sup>[5]</sup>

Despite years of efforts that have been focused on correlating substrate characteristics to recalcitrance, much of the literature actually reported conflicting trends. For example, lower cellulose DP was reported to improve enzymatic hydrolysis owing to increasing cellulose reactivity and numbers of cellulose chain reducing ends.<sup>[6]</sup> However, Sinistyn et al. showed that reduction in DP of cotton linters by  $\gamma$ -irradiation actually had a negligible impact on the hydrolysis rate.<sup>[7]</sup> High lignin content is considered as another important factor limiting the rate of enzymatic hydrolysis of biomass, because lignin can physically block the access of enzymes to cellulose as well as adsorb cellulase irreversibly thus decreasing the availability of enzymes.<sup>[8]</sup> Conversely, it was also reported that there was no obvious correlation between lignin content and enzymatic hydrolysis of a large natural population of poplar.<sup>[9]</sup> Another study also suggested that there was no evidence that substantially reduced lignin contents increased saccharification potential of transgenic field-grown poplar.<sup>[10]</sup> Similarly, conflicting conclusions have been reported on the significance of hemicellulose removal in improving sugar releases.<sup>[11]</sup> The lack of consistency could be mainly attributed to the interactive effects between all the structural factors. To highlight the importance of a particular factor in biomass recalcitrance, it would be ideal to just alter that specific cell-wall substrate characteristic of interest, while keeping other factors unchanged. Drying of lignocellulosic substrate to change the cellulose accessibility would be a perfect example because it would only decrease the cellulose accessibility resulting from the irreversible internal pore collapse known as hornification.<sup>[5]</sup> Unfortunately, it is near impossible to alter any other structural features, such as cellulose DP or crystallinity index (CrI), without changing additional ones, such as particle size or specific surface area, given the structural heterogeneity and complexity of the spatial cell wall constituents.<sup>[12]</sup> This interactive effect becomes more prominent when a chemical pretreatment is applied to alter a factor and determine the effect of this particular factor on enzymatic hydrolysis. For example, dilute acid pretreatment (DAP) is known to decrease cellulose DP.<sup>[2]</sup> On the other hand, it has been reported that DAP also increases cellulose accessibility.<sup>[13]</sup> Therefore, analyzing the relative importance of cellulose DP reduction on enzymatic hydrolysis in the absence of cellulose accessibility consideration could easily lead to inadequate conclusions, as later studies confirmed that accessibility increase is the major reason causing high sugar yield for the dilute acid pretreated biomass.<sup>[14]</sup> From this point of view, selecting transgenic plants or sampling natural genetic variants within a particular species as the feedstock without further

pretreatment is probably a better strategy to understand the fundamental mechanisms of biomass recalcitrance.

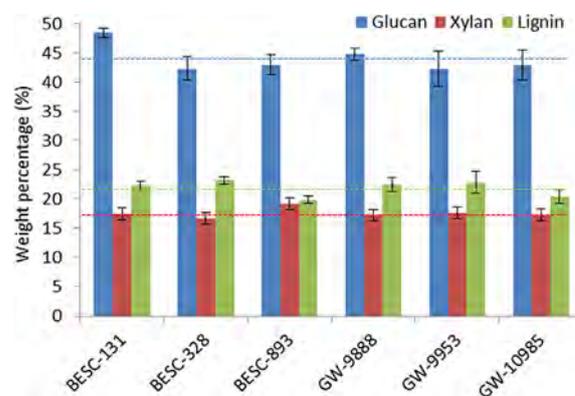
Natural variants could be served as an important resource for studying of gene function and recalcitrance properties in species such as poplar. In a recent study, Muchero et al. demonstrated the power of complementary genetic mapping approaches (i.e., quantitative trait locus) to identify genomic regions associated with cell-wall phenotypes linked to recalcitrance of *Populus*.<sup>[15]</sup> Another study by Studer et al. showed high phenotypic variation among the accessions in recalcitrance measured by lignin content and sugar release using wild *Populus trichocarpa* genotypes collected from northwest Washington to central Oregon.<sup>[9]</sup> The correlations between sugar release and lignin content or the ratio of syringyl and guaiacyl (S/G) was also reported in their study; however, several samples exhibited significantly higher sugar release despite the fact that they showed average values in the analyzed cell wall traits, suggesting that factors beyond lignin content and S/G ratio could influence recalcitrance.

In this study, six 4-year-old natural poplar variants grown and harvested at uniform conditions from Clatskanie, Oregon as previously described by Muchero et al. were selected as the biofuel feedstock.<sup>[15]</sup> Sugar release of these samples were tested using a combined high-throughput pretreatment and enzymatic hydrolysis process as described elsewhere.<sup>[16]</sup> A comprehensive compositional and structural characterization including cellulose DP, CrI, cellulose accessibility, and lignin structural features were performed to provide an in-depth understanding of the roles of these substrate-related factors in biomass recalcitrance.

## Results

### Compositional analysis

Figure 1 shows the glucan, xylan, and Klason lignin contents for each of the poplar variants. Carbohydrate analysis reveals that xylan and glucan are the dominant constituents, while arabinan, galactan, and mannan are found to be negligible. The glucan and xylan contents of these variants range from 42.2% (GW-9953) to 48.4% (BESC-131) and 16.7% (BESC-328)



**Figure 1.** Composition of natural poplar variants. The horizontal lines represent the mean value of glucan, xylan, and lignin content.

to 19.2% (BESC-893), respectively. BESC-893 has the lowest lignin content (19.8%), while all other samples have significantly higher lignin content, with BESC-328 (23.2%) the highest. The chemical composition obtained here is in agreement with other data reported for natural poplar variants in the literature.<sup>[17]</sup>

### Enzymatic hydrolysis yield

The poplar natural variants were subjected to a combined high-throughput pretreatment and enzymatic hydrolysis. Total glucose + xylose release ranged from 0.54 to 0.65  $\text{g g}_{\text{biomass}}^{-1}$  (Table 1). It was interesting to note that poplar with higher glu-

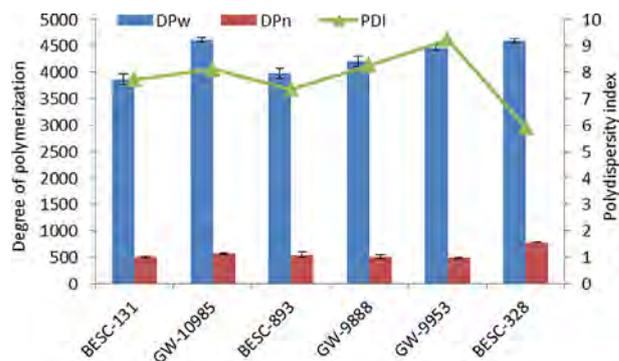
Sample	Amount [ $\text{g g}_{\text{biomass}}^{-1}$ ]		
	Glucose	Xylose	Glucose + Xylose
BESC-131	0.51 (0.008)	0.15 (0.009)	0.65 (0.017)
GW-10985	0.40 (0.037)	0.18 (0.005)	0.58 (0.041)
BESC-893	0.38 (0.003)	0.21 (0.004)	0.59 (0.007)
GW-9888	0.36 (0.006)	0.21 (0.014)	0.57 (0.010)
GW-9953	0.35 (0.021)	0.20 (0.003)	0.55 (0.023)
BESC-328	0.35 (0.017)	0.19 (0.010)	0.54 (0.026)

[a] Data represents the mean value of three replicates under same conditions. Standard deviation is shown in the parenthesis.

cose release did not necessary have higher xylose release. In fact, BESC-131 had the highest glucose release but lowest xylose release, which were 0.51 and 0.15  $\text{g g}_{\text{biomass}}^{-1}$ , respectively. GW-10985 and BESC-893 had very similar total glucose + xylose release (0.59 vs. 0.58  $\text{g g}_{\text{biomass}}^{-1}$ ), but there existed significant differences in each individual sugar release between these two samples according to a student t-test at a confidence level of 95%. In addition, no significant differences in sugar release were found for sample GW-9888, GW-9953, and BESC-328.

### Characterization of cellulose

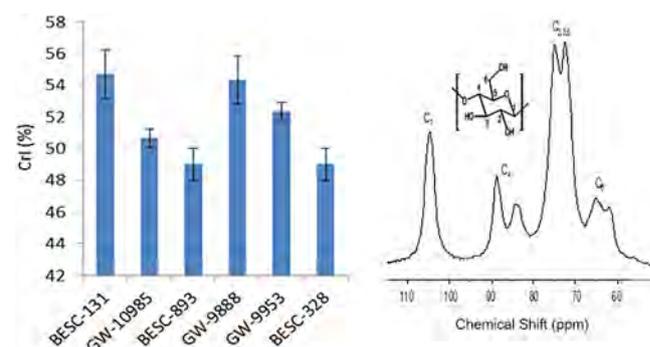
As a prerequisite step, individual biomass components, such as cellulose, hemicellulose, and lignin, need to be isolated and purified prior to analytical characterization. Cellulose isolated from these variants was characterized by gel permeation chromatography (GPC) and solid-state NMR to measure the cellulose DP and CrI, respectively. The weight-average degree of polymerization ( $\text{DP}_w$ ) and number-average degree of polymerization ( $\text{DP}_n$ ) of these cellulose samples are shown in Figure 2. Cellulose  $\text{DP}_w$  ranges from 3862 (BESC-131) to 4609 (GW-10985), whereas the  $\text{DP}_n$  ranges from 482 (GW-9953) to 780 (BESC-328). Two sub-groups could be identified in terms of weight-average molecular size. BESC-328, GW-9953, and GW-10985 with an average  $\text{DP}_w$  of 4554 have obviously higher  $\text{DP}_w$  compared to BESC-131, BESC-893, and GW-9888, which have an average  $\text{DP}_w$  of 4014. Statistically, no significant differences in cellulose  $\text{DP}_w$  were found between BESC-328 and GW-10985.



**Figure 2.** The weight- and number-average DP and PDI of cellulose isolated from natural poplar variants.

BESC-328 has the highest  $\text{DP}_n$ , which is about 50% higher than the average  $\text{DP}_n$  of the rest five samples. Polydispersity index (PDI), a measure of the width of molecular weight distribution in cellulose, ranged from 5.9 (BESC-328) to 9.2 (GW-9953).

Solid-state NMR has been widely used to characterize the ultrastructure of cellulose. Here in this study, solid-state  $^{13}\text{C}$  cross polarization magic angle spinning (CP/MAS) NMR was conducted to determine the cellulose CrI. Figure 3 provides the CrI re-

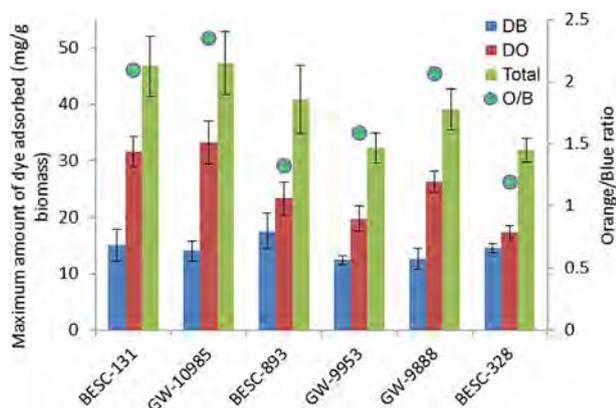


**Figure 3.** CrI and a representative CP/MAS  $^{13}\text{C}$  NMR spectrum of isolated cellulose from natural poplar variants.

sults along with a representative  $^{13}\text{C}$  CP/MAS NMR spectrum of an isolated cellulose sample. As shown in the figure, each of the carbon atoms in the monomeric unit of cellulose are denoted  $\text{C}_1$  through  $\text{C}_6$  and labelled accordingly on the corresponding solid-state NMR spectrum. The  $\text{C}_4$  region ranging from 80 to 92 ppm contained a crystalline domain, which appeared as a fairly sharper signal around 89 ppm, and an amorphous domain that produced broader, slightly shielded resonances.<sup>[18]</sup> The ratio of the area of the crystalline region to the area of the total  $\text{C}_4$  peak area was calculated and designated as the CrI of cellulose. It was found that cellulose isolated from BESC-131 and GW-9888 showed the highest CrI at 55%, which is slightly higher than GW-9953. The rest three samples showed no apparent significant changes in cellulose CrI.

Simons' staining (SS) has been shown to be a promising semi-quantitative technique for the estimation of accessible surface area of lignocellulosic substrate and biomass porosi-

ty.<sup>[19]</sup> Unlike conventional techniques, such as Brunauer–Emmett–Teller (BET) nitrogen adsorption and solute exclusion, SS doesn't require prior drying of the substrates, which would cause irreversible pore collapse, and it could measure both interior and exterior surface area relatively fast by applying two direct dyes: Direct Orange 15 (DO) and Blue 1 (DB).<sup>[8b]</sup> DB dye has a molecular diameter of  $\approx 1$  nm, whereas DO dye is a polymer with a molecular diameter of  $\approx 5$ – $36$  nm, which is similar to a nominal size of  $5.1$  nm representative of the diameter of typical cellulose.<sup>[20]</sup> Owing to the much higher binding affinity of DO dye to the hydroxyl group on cellulosic surface, DB dye molecules will only populate the smaller pores of the fiber, whereas the DO dye can enter the larger pores of the substrate surface. Therefore, the amount of DO dye adsorbed and the ratio of DO and DB (O/B) adsorbed by the substrates can be used to indicate the accessible surface area of cellulose to cellulase and the relative amount of large pores to small pores, respectively. In this study, a modified SS assay based on a previously published procedure was applied on these natural poplar variants to assess cellulose accessibility. As shown in Figure 4, BESC-131, GW-9888, and GW-10985 have relatively



**Figure 4.** SS results for cellulose accessible surface area represented by the amount of adsorbed dye ( $\text{mg g}^{-1}_{\text{biomass}}$ ) and the relative biomass porosity represented by the ratio of the adsorbed DO to DB dye (O/B).

higher DO dye adsorption and O/B ratio indicating that these three samples have larger accessible surface area of cellulose and relatively higher biomass porosity. On the other hand, BESC-328 has the lowest DO dye adsorption ( $17.3 \text{ mg g}^{-1}$ ) and O/B ratio indicating a very limited cellulose accessibility. No significant differences in DO dye adsorption are found between sample BESC-131 and GW-10985 based on a student t-test at a 95% confidence level.

### Characterization of lignin

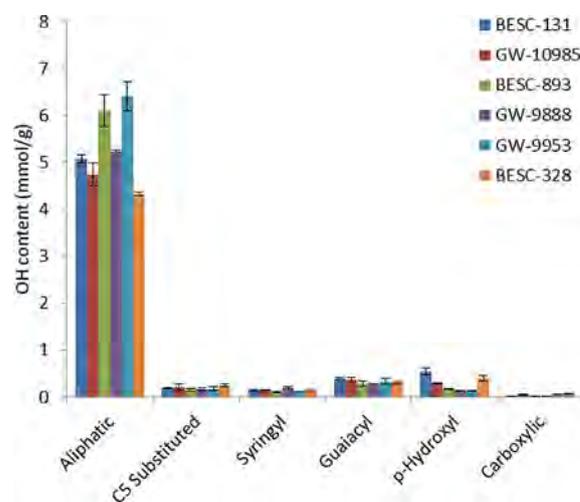
To further characterize the lignin structural variations in these poplar variants, cellulolytic enzyme lignin (CEL) samples were isolated according to the literature.<sup>[21]</sup> Lignin samples were acetylated, dissolved in THF, and subsequently analyzed by GPC to determine the molecular size and molecular weight distribution. As shown in Table 2, lignin number-average molecular

Sample	$M_n \times 10^3$ [ $\text{g mol}^{-1}$ ]	$M_w \times 10^4$ [ $\text{g mol}^{-1}$ ]	PDI
BESC-131	3.08 (0.159)	0.99 (0.037)	3.21 (0.046)
GW-10985	3.72 (0.146)	1.10 (0.020)	2.95 (0.062)
BESC-893	4.32 (0.181)	1.33 (0.017)	3.08 (0.129)
GW-9888	3.27 (0.097)	1.09 (0.040)	3.33 (0.087)
GW-9953	4.63 (0.387)	1.52 (0.029)	3.27 (0.223)
BESC-328	4.39 (0.200)	1.26 (0.039)	2.87 (0.065)

[a] Data represents the mean value of three replicates under same conditions. Standard deviation is shown in the parenthesis.

weight ( $M_n$ ) ranges from  $3.08 \times 10^3$  to  $4.63 \times 10^3 \text{ g mol}^{-1}$  and weight-average molecular weight ( $M_w$ ) ranges from  $0.99 \times 10^4$  to  $1.52 \times 10^4 \text{ g mol}^{-1}$ , which is in consistent with a recent study.<sup>[22]</sup> Lignin isolated from BESC-131 has the lowest  $M_n$  and  $M_w$  while GW-9953 lignin has the highest  $M_n$  and  $M_w$ . The PDI ranged from 2.87 to 3.33, indicating a relatively narrow and wide molecular weight distribution for BESC-328 and GW-9888, respectively. There are no significant differences in lignin  $M_w$  between GW-10985 and GW-9888 according to a student t-test at a 95% confidence level.

<sup>31</sup>P NMR was then used to further determine the amount of various types of hydroxyl groups in lignin, such as aliphatic, guaiacyl, syringyl, *p*-hydroxyphenyl, and carboxylic hydroxyl.<sup>[23]</sup> It involved a phosphorylation of hydroxyl groups in lignin substrate using a <sup>31</sup>P reagent, such as 2-chloro-4,4,5,5-tetramethyl-1,3,2-dioxaphospholane (TMDP), followed by quantitative NMR analysis. Figure 5 summarizes the hydroxyl group contents of lignin isolated from these natural poplar variants. The results show that the aliphatic OH is the dominant hydroxyl type in all the lignin samples, representing  $\approx 81$ – $91\%$  of total free hydroxyl groups. Among the free phenolic hydroxyls, guaiacyl and *p*-hydroxyl are observed to be the most prominent type. For example, *p*-hydroxyl OH represents the major type of free phenolic hydroxyls in BESC-131 and BESC-328, accounting for

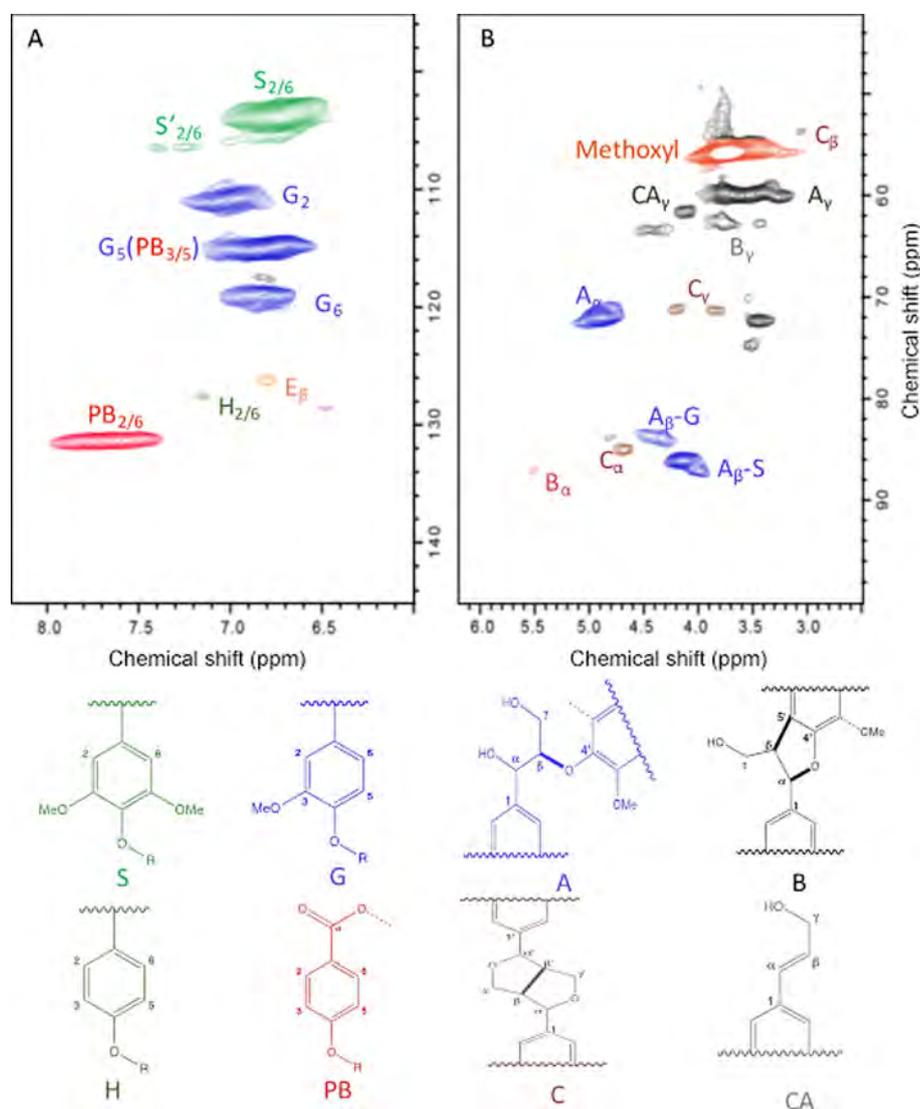


**Figure 5.** Hydroxyl group contents of lignin isolated from natural poplar variants as calculated by <sup>31</sup>P NMR.

approximately 48 and 42%, respectively; guaiacyl OH appears to be the dominant free phenolic hydroxyls in the other samples. Only trace amounts of carboxylic OH is observed with BESC-328 and GW-9953's content slightly higher than the others. BESC-328 has the lowest aliphatic OH around  $4.33 \text{ mmol g}_{\text{biomass}}^{-1}$ , whereas GW-9953 has the highest aliphatic OH content ( $6.42 \text{ mmol g}_{\text{biomass}}^{-1}$ ). In terms of total phenolic OH content, BESC-131 has the highest amount and BESC-893 has the lowest around 1.14 and  $0.61 \text{ mmol g}_{\text{biomass}}^{-1}$ , respectively. The content of G-type phenolic hydroxyl groups is always higher than that of S-type. BESC-131 and BESC-328 contained much more *p*-hydroxyphenyl groups than the rest of samples.

2D HSQC NMR was also used to provide valuable information about the detailed chemical structures of lignin. Lignin, an amorphous, three-dimensional, and cross-linked polyphenolic polymer, is composed primarily of syringyl (S), guaiacyl (G), and *p*-hydroxyphenyl (H) subunits, which are derived from the polymerization of three types of phenylpropane units as mon-

olignols: sinapyl, guaiacyl, and *p*-coumaryl alcohol, respectively. Semi-quantitative analysis of HSQC spectra was used for estimation of monolignol compositions (e.g., the S/G ratio) and relative abundance of inter-unit linkages (e.g.,  $\beta$ -O-4). The HSQC NMR spectra showed that the lignin samples demonstrated similar structural features. Representative HSQC spectra of CEL isolated from BESC-131 are shown in Figure 6, and the relative contents of lignin subunits as well as their interlinkages for all the poplar variants are shown in Table 3. The  $^{13}\text{C}/^1\text{H}$  cross-peaks in aromatic and aliphatic regions of the HSQC spectra were assigned according to literature. Figure 6 clearly shows that the lignin isolated from the natural poplar variant BESC-131 is primarily composed of S and G units as expected in hardwood. The S unit shows major cross peaks for the  $\text{C}_{2,6}/\text{H}_{2,6}$  correlations centered at  $104.8/6.73$  and  $106.5/7.30$  ppm (oxidized units), whereas the G unit shows correlations for  $\text{C}_2/\text{H}_2$ ,  $\text{C}_5/\text{H}_5$ , and  $\text{C}_6/\text{H}_6$  around  $111.0/6.99$ ,  $115.2/6.81$ , and  $119.1/6.83$  ppm, respectively. It also contains



**Figure 6.** A representative aromatic (A) and aliphatic (B) region of the HSQC NMR spectra of lignin isolated from BESC-131. S: syringyl, G: guaiacyl, H: *p*-hydroxyphenyl, PB: *p*-hydroxyphenyl benzoate, A:  $\beta$ -O-4 ether, B:  $\beta$ -5/ $\alpha$ -O-4 phenylcoumaran, C: resinol ( $\beta$ - $\beta$ ), and CA: cinnamyl alcohol.

**Table 3.** Semi-quantitative information for lignin subunits and inter-linkages of natural poplar variants by HSQC NMR technique.<sup>[a]</sup>

Biomass sample	Relative content [%]							
	S	G	Lignin subunits		S/G ratio	Inter-linkages		
			H	PB <sup>[b]</sup>		A	B	C
BESC-131	59.0 (2.83)	40.5 (2.22)	0.48 (0.62)	20.7 (0.73)	1.46 (0.15)	88.7 (3.84)	6.73 (3.63)	4.52 (0.21)
GW-10985	69.8 (1.87)	30.0 (1.79)	0.20 (0.08)	11.2 (0.17)	2.33 (0.20)	86.8 (2.97)	4.99 (1.84)	8.21 (1.12)
BESC-893	72.6 (0.26)	26.5 (0.09)	0.95 (0.35)	4.84 (0.92)	2.74 (0.0004)	88.9 (3.09)	3.44 (3.30)	7.65 (0.21)
GW-9888	76.7 (1.36)	22.7 (0.64)	0.56 (0.72)	2.73 (0.92)	3.37 (0.15)	86.2 (2.49)	3.45 (1.54)	10.4 (0.95)
GW-9953	81.5 (2.65)	17.9 (1.80)	0.73 (0.85)	1.69 (0.44)	4.60 (0.61)	89.2 (1.37)	2.54 (0.91)	8.27 (0.46)
BESC-328	70.3 (0.94)	28.8 (0.86)	0.89 (0.08)	13.7 (0.64)	2.44 (0.11)	89.2 (3.82)	3.81 (2.29)	6.99 (1.53)

[a] Data represents the mean value of three replicates under same conditions. Standard deviation is shown in the parenthesis. S: syringyl, G: guaiacyl, H: *p*-hydroxyphenyl, PB: *p*-hydroxyphenyl benzoate, A:  $\beta$ -O-4 ether, B:  $\beta$ -5/ $\alpha$ -O-4 phenylcoumaran, C: resinol ( $\beta$ - $\beta$ ). [b] PB percentage was expressed as a fraction of total S + G + H.

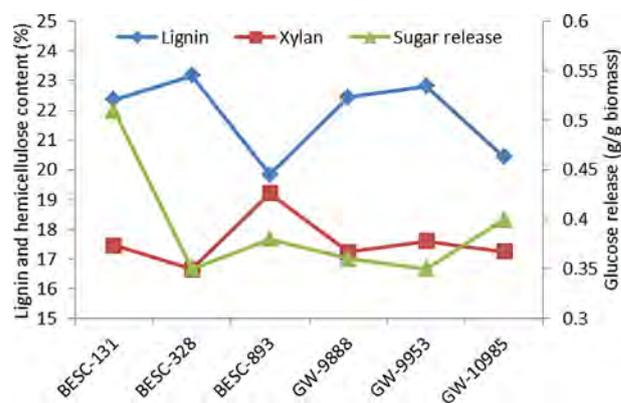
considerable amounts of *p*-hydroxyphenyl benzoate (PB) units, which is observed from  $C_{2,6}/H_{2,6}$  correlation centered around 131.5/7.70 ppm with its  $C_{3,5}/H_{3,5}$  (114.8/6.73 ppm) correlation overlapping with  $G_5$  units. In the aliphatic region, signals associated with methoxyl (55.7/3.76 ppm) and  $\beta$ -O-4 inter-linkages appear to be the most prominent ones in lignin. The C–H correlations in  $\beta$ -O-4 linkages are well recognized for  $\alpha$ ,  $\beta$ , and  $\gamma$  carbons. The presence of phenylcoumaran is also confirmed by C–H correlations for  $\alpha$  and  $\gamma$  carbon positions centered at 87.1/5.51 and 62.6/3.77 ppm, respectively. The resinol subunit is also evidenced by its C/H correlation around 85.1/4.68, 53.5/3.06, and 71.2/3.82 ppm. Relative amounts of lignin subunits, especially the S, G, and PB units, are found significantly different from each other, resulting in a wide range distribution of S/G ratio from 1.46 to 4.60 as shown in Table 3. Among all the samples, much less significant differences are noticed for the amount of lignin inter-linkages including aryl ether bonds ( $\beta$ -O-4) and carbon-carbon linkages ( $\beta$ -5 and  $\beta$ - $\beta$ ). BESC-131 has the lowest content of S units (59.0%) and highest content of G units (40.5%) leading to the smallest S/G ratio, whereas GW-9953 had the highest content of S units (81.5%) and lowest content of G units (17.9%) causing the largest S/G ratio. The amount of PB units expressed as a fraction of total S + G + H units in lignin is found ranging from 1.69% (GW-9953) to 20.7% (BESC-131). BESC-131, GW-10985, and BESC-328 contained much more *p*-hydroxybenzoate substructures than the other three samples as was also revealed by the  $^{31}\text{P}$  NMR analysis.

## Discussion

### Effect of substrate-related factors on biomass recalcitrance

Understanding the fundamentals of biomass recalcitrance by identifying the key structural features that are responsible for low or high sugar release would provide extremely valuable information in overcoming biomass recalcitrance. With the composition, cellulose CrI, cellulose DP, substrate accessibility, and lignin structural-relevant data available for a series of natural poplar variants, an in-depth analysis of the effect of each individual factor on substrate digestibility could be performed.

It has been generally recognized that elementary cellulose fibrils are coated with some non-cellulosic polysaccharides to form microfibrils, and these microfibrils are then cross linked by hemicellulose matrixes to form macrofibrils.<sup>[24]</sup> Some studies have concluded that hemicellulose removal was more important than lignin removal for increasing sugar release, whereas others indicated that lignin removal was much more important.<sup>[9,13]</sup> The relative importance of lignin content versus xylan content on enzymatic hydrolysis yields of natural poplar variants is shown in Figure 7. Xylan doesn't seem to have a promi-

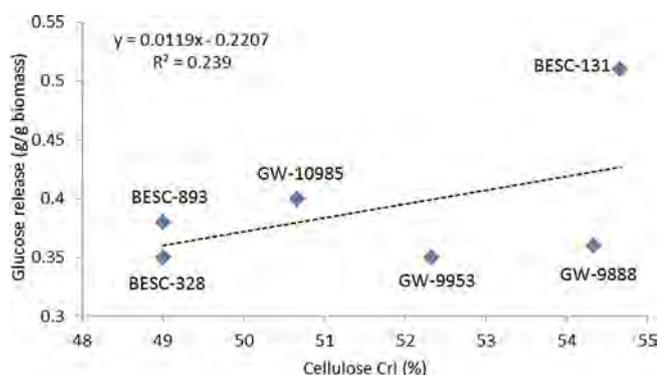


**Figure 7.** Effect of lignin and xylan content on glucose release for a series of natural poplar variants.

nent impact on glucose release, which could be partially because there are not many variations existing between these samples in terms of xylan content. Lignin, on the other hand, has a much clearer pattern. By comparing samples with their neighbors from left (BESC-131) to right (GW-10985) as shown in the horizontal axis of Figure 7, it was found that an increase of the lignin content always reduced the glucose release and a decrease of lignin content always increased the glucose release, indicating lignin likely played a more important role than xylan in poplar recalcitrance. Demartini et al. also showed that xylan removal from switchgrass resulted in nearly 100% glucose yields, whereas chlorite extractions that reduced the lignin content had the most beneficial effect in poplar.<sup>[25]</sup> However, it is also worth mentioning that neither lignin nor hemicellulose

cellulose content could be used to predict glucose release perfectly. For example, BESC-131, which had relatively high lignin and xylan content, released the largest amount of glucose after enzymatic hydrolysis.

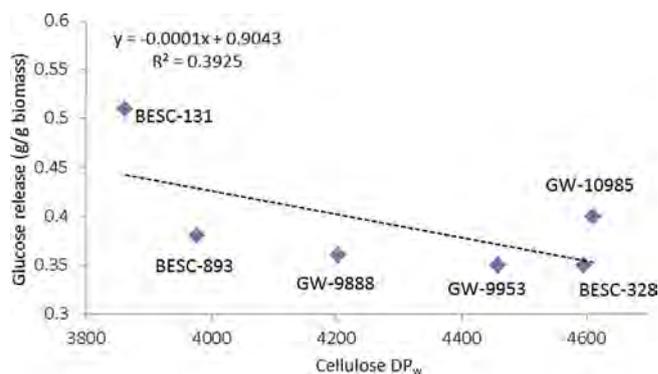
It was generally believed that the hydrolysis rate of crystalline cellulose is much slower than that of amorphous cellulose. However, several studies in literature are not straightforward to provide a clear conclusion that cellulose crystallinity is the key determinant of the hydrolysis rate.<sup>[26]</sup> It was reported that cellulose CrI played a key role in determining the enzymatic hydrolysis rate of Avicel.<sup>[27]</sup> On the other hand, Brienzo et al. suggested that CrI may not have influence in biomass digestibility especially for pretreated materials.<sup>[28]</sup> Figure 8 illustrates



**Figure 8.** Relationship between cellulose CrI and glucose release ( $\text{g g}^{-1}_{\text{biomass}}$ ) for a series of natural poplar variants.

the relationship between cellulose CrI and poplar digestibility in this study. The results suggest that there is no clear correlation between the cellulose CrI and glucose release. For example, the samples that almost have identical cellulose CrI (BESC-131 and GW-9888) could have  $\approx 42\%$  differences in glucose release. These two samples also have very similar lignin content (23.0 vs. 23.2%), suggesting other factors, such as cellulose DP and accessibility, might play some roles here.

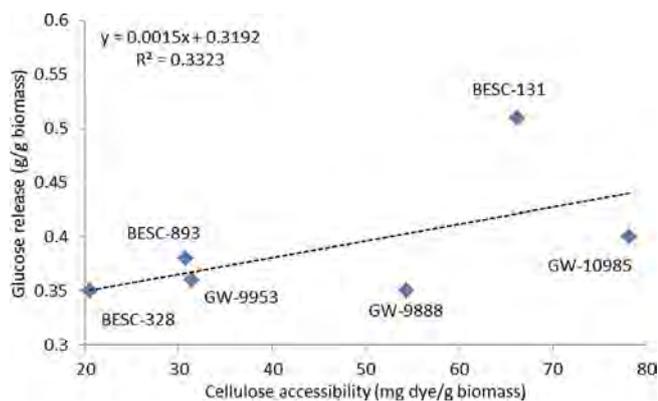
Similar to the cellulose CrI, the exact role of cellulose DP in biomass recalcitrance is still under debate. A shorter cellulose chain contains a weaker hydrogen-bonding system and therefore is believed easier for enzyme to access.<sup>[6]</sup> A lower DP means more reducing ends and consequently a higher exoglucanase activity should be expected during enzymatic hydrolysis. Numerous pretreatments have been shown to reduce biomass recalcitrance partially owing to the significant reduction of cellulose DP.<sup>[2,12]</sup> DP<sub>w</sub> was chosen to represent cellulose DP because of the limited variations in DP<sub>n</sub>, and the glucose release was plotted versus DP<sub>w</sub> as shown in Figure 9. This analysis demonstrates a much stronger inverse relationship between DP<sub>w</sub> and sugar release compared to the cellulose CrI. Clearly, the lowest DP<sub>w</sub> around  $\approx 3862$  resulted in the highest sugar release for BESC-131. This particular study also confirms our previous hypothesis that the DP of cellulose from BESC-131 compared to GW-9888 could be one of the potential reasons causing the  $\approx 42\%$  differences in glucose release. The results demonstrate that the sugar release of these poplar variants de-



**Figure 9.** Relationship between cellulose DP<sub>w</sub> and glucose release ( $\text{g g}^{-1}_{\text{biomass}}$ ) for a series of natural poplar variants.

creases as the cellulose DP<sub>w</sub> increases, with the exception of sample GW-10985, which has the highest cellulose DP<sub>w</sub> but ends up with the second highest sugar release. This could be, in part, a result of the relatively low lignin content and low cellulose crystallinity that GW-10985 exhibits. Again just like the cellulose CrI, samples that have statistically similar cellulose DP<sub>w</sub> (BESC-131 and BESC-893) could have  $\approx 34\%$  differences in glucose release. In this case, neither lignin content nor cellulose CrI could be used to explain the differences because the higher recalcitrant sample BESC-893 actually has the lowest lignin content and lowest cellulose CrI among all the tested samples, suggesting cellulose accessibility or lignin structural features play some roles here.

Although it is quite challenging to assess the effect of any individual factors on enzymatic hydrolysis, because biomass recalcitrance does not come from a single factor and interactive effects naturally exist between these factors, cellulose accessibility has been consistently reported as one of the most important factors. The SS technique is very sensitive to the pore inlet size, and a recent study has proposed that the use of  $A_0(\text{O/B})$  as a correction factor for the pore shape and size distribution in SS technique shows much better accessibility measurement, where  $A_0$  represents the maximum DO dye adsorption and O/B is the ratio of DO and DB adsorbed.<sup>[29]</sup> In this study, the relationship between cellulose accessibility and enzymatic glucose release is analyzed to determine if accessibility is one of the dominant factors affecting saccharification of these natural poplar variants (Figure 10). A general positive relationship between cellulose accessibility and substrate digestibility was obtained. For example, BESC-131 and GW-10985 have relatively higher cellulose accessibility leading to higher glucose releases compared to the rest of samples, whereas BESC-328, BESC-893, and GW-9953 have relatively lower cellulose accessibility and therefore much lower glucose release. However, GW-9888 has very high cellulose accessibility but releases a relatively low amount of glucose, which could be a result of the high lignin content and extremely high cellulose CrI. Just like cellulose CrI and DP, samples that have statistically similar cellulose accessibility (BESC-131 and GW-10985) could have  $\approx 28\%$  differences in sugar release. Cellulose DP could be a likely reason because BESC-131 has the lowest cellulose DP, whereas GW-10985

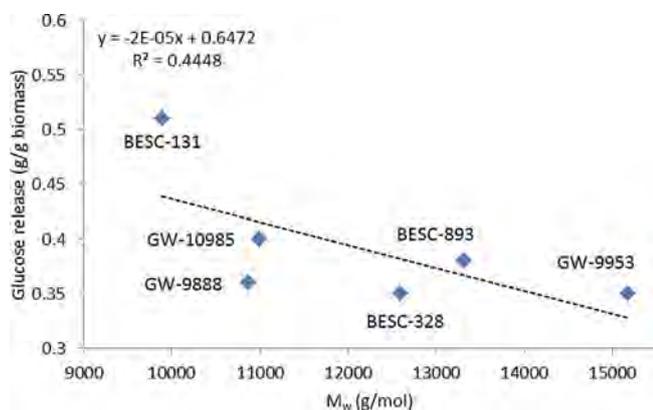


**Figure 10.** Relationship between cellulose accessibility ( $\text{mg}_{\text{dye}} \text{g}_{\text{biomass}}^{-1}$ ) and glucose release ( $\text{g}_{\text{glucose}} \text{g}_{\text{biomass}}^{-1}$ ) for a series of natural poplar variants.

actually has the highest cellulose DP. Figure 10 also confirms the previous hypothesis that cellulose accessibility is the major reason causing the much higher sugar release for BESC-131 compared to BESC-893 despite its extremely low lignin content and low cellulose CrI.

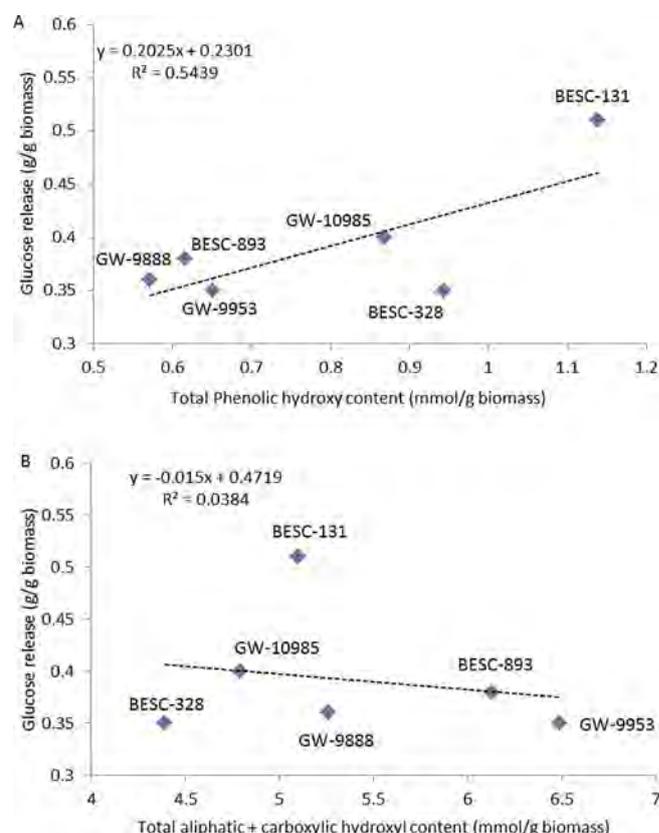
Lignin is considered as the most recalcitrant biopolymers that can be found in the plant secondary cell wall. The molecular weight of lignin has been shown to influence the biomass recalcitrance and lignin valorization.<sup>[30]</sup> Ziebell et al. reported that the molecular weight (MW) of lignin in alfalfa can be decreased by altering the lignin monomer distribution towards increased hydroxyphenyls monomer, and these changes affect the ease in which lignin can be removed by chemical processing.<sup>[31]</sup> The importance of lignin MW and its relationship to recalcitrance was analyzed in Figure 11. Lignin isolated from low recalcitrant variants, such as BESC-131 and GW-10985, tended to have relatively low  $M_w$  whereas high recalcitrant line GW-9953 had the highest  $M_w$  lignin. Berlin et al. reported that the PDI could also be inversely related to the interaction of the lignin with the enzymes.<sup>[32]</sup> However, no clear correlation between lignin PDI and sugar release was obtained in this study.

Besides physically limiting activated cellulose accessible surface, lignin can also unproductively bind to enzymes through



**Figure 11.** Relationship between lignin  $M_w$  ( $\text{g mol}^{-1}$ ) and glucose release ( $\text{g}_{\text{glucose}} \text{g}_{\text{biomass}}^{-1}$ ) for a series of natural poplar variants.

functional groups, such as phenolic hydroxyl groups, and subsequently reduce the sugar release during enzymatic hydrolysis.<sup>[33]</sup> Hydrophobic interaction was identified as one of the major driving forces in the adsorption of enzymes to lignin. Pan et al. reported that the increase of phenolic hydroxyl groups in lignin negatively affected hydrolysis owing to the increase of hydrophobicity, and the hydroxypropylation of these groups can reduce the negative inhibitory effect of lignin.<sup>[34]</sup> Guo et al. and Nakagame et al. reported an increase in aliphatic and carboxylic hydroxyl content within lignin partially alleviated the non-productive binding of cellulases to lignin owing to the decrease of hydrophobicity.<sup>[35]</sup> Content of total phenolic hydroxyl and aliphatic + carboxylic groups was plotted against glucose release to test if there are any correlations (Figure 12).

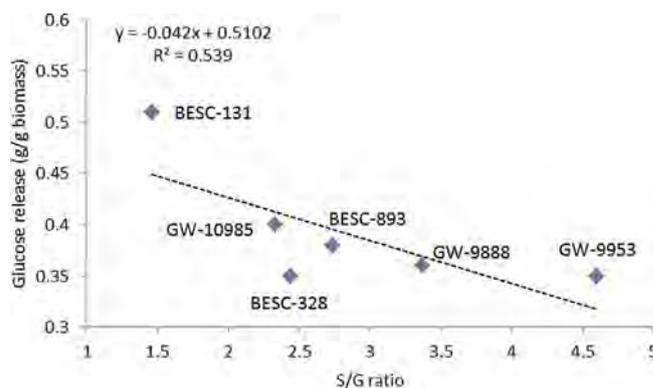


**Figure 12.** Relationship between phenolic hydroxyl content (A) and aliphatic + carboxylic hydroxyl content (B) and glucose release ( $\text{g}_{\text{glucose}} \text{g}_{\text{biomass}}^{-1}$ ) for a series of natural poplar variants.

Interestingly, it appears that the increase of phenolic hydroxyls in lignin is positively linked to the glucose release of these poplar samples. However, no clear correlation is found between aliphatic + carboxylic hydroxyl content and glucose release in this study. One of the reasons causing these inconsistent results compared to literature could be that, because lignin is a heterogeneous polymer, the effect of different factors can be overlapped. Therefore, it becomes extremely difficult to identify which individual lignin structural feature is more important to the non-productive adsorption of enzymes.

The pattern of lignin incorporated with polysaccharides could be another important factor. For example, a more spreadable lignin distribution in plant cell wall could lead to a larger non-productive protein adsorption.<sup>[36]</sup> Cellulose relevant factors might also interfere with these lignin features, if any, in a different way. To eliminate any possible overlapped effect caused by either lignin or cellulose, the majority of literature highlighting the importance of lignin in enzymatic hydrolysis normally used Avicel as a substrate model and simple lignin-like model compounds to simulate complex lignin polymer.<sup>[34,35]</sup> Nonetheless, our results indicate that cellulose-relevant factors probably have a stronger correlation with sugar release, if any, than lignin structural features, at least for the natural poplar variants in this study. For example, despite the fact that lignin isolated from BESC-131 has the highest phenolic hydroxyl content, the lowest cellulose DP with most reducing end and the second highest accessibility of cellulose to cellulase made BESC-131 a low recalcitrant plant.

Lignin S/G ratio is another lignin structural feature that has been reported to affect biomass recalcitrance.<sup>[37]</sup> Studer et al. reported that dilute acid-pretreated natural poplar variants with higher S/G ratios generally had a higher sugar release from enzymatic hydrolysis, and the sugar release was irrespective of S/G ratio for unpretreated substrates.<sup>[9]</sup> It was reported that the G unit of lignin could form a more cross-linked lignin structure than that of the S unit, and subsequently generated a larger physical barrier against the substrate accessibility.<sup>[36]</sup> Interestingly, poplar variants with lower S/G ratio tended to have a higher glucose release in our study (Figure 13). This could be a result of the narrow ranges of lignin content for the samples being studied, because it has been reported that the interaction effect of lignin content and S/G ratio is very significant.<sup>[37]</sup> To eliminate any possible interference caused by cellulose-related factors, the effect of the lignin S/G ratio on xylose release after enzymatic hydrolysis was determined (Figure S11 in the Supporting Information). It is found that xylose release is generally higher for the high S/G ratio lignin sample, which is consistent with other studies.<sup>[9]</sup> Similarly, a strong negative relationship ( $R^2=0.89$ ) between the content of total lignin phenolic hydroxyl group and xylose release is also observed (Figure S12). Wang et al. reported that the xylan conversion efficiency was more sensitively affected by the variation of alkaline pretreatment conditions than glucan conversion efficien-



**Figure 13.** Relationship between lignin S/G ratio and glucose release ( $\text{g g}_{\text{biomass}}^{-1}$ ) for a series of natural poplar variants.

cy.<sup>[38]</sup> Therefore, lignin structural features appear to have a more convincing impact on xylose release than glucose release, which makes sense considering that lignin is the polymer that covalently links to hemicellulose and fills the space between cellulose and pectin matrixes.

All the compositional and structural features for these poplar variants along with its recalcitrant properties are summarized in Table 4. As shown in Table 4, none of these individual potential factors proved to be sufficient to fully explain the differences of sugar release alone. High cellulose accessibility is the major factor causing the high glucose release for low recalcitrant lines, such as BESC-131 and GW-10985. On the other hand, high lignin content, relatively high cellulose/lignin molecular weight, and low cellulose accessibility are suggested to be responsible for the low sugar release of high recalcitrant lines, such as GW-9953 and BESC-328. For samples with high accessibility, the sugar releases are generally higher, and the negative influence of other factors, such as highest cellulose DP (GW-10985) or highest cellulose CrI (BESC-131), are less pronounced. Lignin MW was also found to negatively affect the glucose release to some extent with a  $R^2$  value of 0.368. Surprisingly, samples with high lignin phenolic hydroxyl content tended to have a relatively high sugar release. This is not quite consistent with the majority of literature views, which could be because of the interference of other factors, such as cellulose accessibility, or simply a result of different feedstocks used in different studies. The relative contributions of each factor in

**Table 4.** Effects of compositional and physicochemical structures on enzymatic hydrolysis of natural poplar variants.<sup>[a]</sup>

Sample	Cellulose			Overall	Lignin content		Lignin		Glucose release
	DP <sub>w</sub>	CrI	Accessibility		Phenolic hydroxyl	Aliphatic + Carboxylic hydroxyl	M <sub>w</sub>	S/G	
BESC-131	low <sup>-</sup>	high <sup>+</sup>	high <sup>+</sup>	high <sup>+</sup>	high <sup>+</sup>	low <sup>+</sup>	low <sup>-</sup>	low <sup>-</sup>	high <sup>+</sup>
GW-10985	high <sup>+</sup>	low <sup>-</sup>	high <sup>+</sup>	low <sup>-</sup>	high <sup>-</sup>	low	low <sup>+</sup>	low <sup>+</sup>	high
BESC-893	low <sup>-</sup>	low <sup>-</sup>	low <sup>+</sup>	low <sup>-</sup>	low <sup>-</sup>	high <sup>+</sup>	high	high <sup>-</sup>	high <sup>-</sup>
GW-9888	low <sup>+</sup>	high <sup>+</sup>	high <sup>-</sup>	high <sup>+</sup>	low <sup>-</sup>	high <sup>-</sup>	low <sup>+</sup>	high	low <sup>-</sup>
GW-9953	high	high	low <sup>+</sup>	high <sup>+</sup>	low <sup>-</sup>	high <sup>+</sup>	high <sup>+</sup>	high <sup>+</sup>	low <sup>-</sup>
BESC-328	high <sup>+</sup>	low <sup>-</sup>	low <sup>-</sup>	high <sup>+</sup>	high <sup>-</sup>	low <sup>-</sup>	high <sup>-</sup>	low <sup>+</sup>	low <sup>-</sup>
Linear R <sup>2</sup>	0.393	0.213	0.332	N/A	0.544	0.0384	0.445	0.539	N/A

[a] Key: low<sup>-</sup> < low < low<sup>+</sup> < high<sup>-</sup> < high < high<sup>+</sup>. Samples with same superscript (i.e., - or +) indicates no significant difference among the samples for a statistical significance level of 0.05 as determined by student t-test.

biomass recalcitrance also vary from sample to sample. For instance, cellulose DP plays a much more important role than cellulose CrI and lignin content for the low recalcitrant line BESC-131 and GW-10985, whereas the lignin content appears to be a more important factor causing lower sugar release of BESC-328 compared to GW-10985.

## Conclusions

A comprehensive structural characterization of natural poplar variants along with the enzymatic hydrolysis data were used to provide an in-depth understanding of the fundamentals of biomass recalcitrance over multiple length scales. High cellulose accessibility, low cellulose/lignin molecular weight, and low lignin content favor enzymatic hydrolysis. However, biomass recalcitrance was found to be a multi-variant and multi-scale phenomenon, and it cannot be simply judged on one solely substrate factor. The natural existing interactions between different factors make the mechanisms of biomass recalcitrance very complicated. Nevertheless, the first-order linear correlation analysis allows us to compare the relative contributions of different factors to recalcitrance. Generally, cellulose relevant factors, such as cellulose accessibility and the degree of polymerization, appeared to play more important roles than lignin structural features in glucose release. On the other hand, lignin structural features, such as phenolic hydroxyl group and the ratio of syringyl and guaiacyl units, were found to have a more direct impact on xylose release. The relative contribution of each factor in biomass recalcitrance also varies from sample to sample. Once a factor is no longer limiting, other factors normally become determinant. Moving forward, a multi-variant non-linear statistical analysis of large population sizes of wild type or transgenic plants might be a useful strategy to fully identify, and subsequently overcome, biomass recalcitrance.

## Experimental Section

### Plant materials

Four-year-old *Populus trichocarpa* natural variants were collected from a field site in Clatskanie, Oregon. Stem segments were air-dried to constant weight, debarked, and size reduced through a 40-mesh (0.420 mm) screen (Thomas Scientific, NJ, USA). Details about plant growth conditions and field establishment were described in a previously published manuscript.<sup>[15]</sup>

### Chemical composition analysis

Extractives were subsequently removed by adding  $\approx 5$  g of biomass into an extraction thimble in a Soxhlet extraction apparatus. The extraction flask was filled with dichloromethane and then refluxed at a boiling rate that cycled the biomass for  $\approx 8$  h. Samples were air-dried and stored in a refrigerator. Moisture content was determined by a halogen moisture analyzer. Carbohydrate and acid-insoluble lignin analysis was performed by using the two-stage acid hydrolysis according to the literature.<sup>[39]</sup> In brief, extractive-free samples were treated with 72% sulfuric acid for 1 h at 30 °C and then diluted to 3% using deionized water and subsequently autoclaved at 121 °C for  $\approx 1$  h. The precipitate was filtered

through a G8 glass fiber filter (Fisher Scientific, USA), dried, and weighed to get Klason lignin content. The resulting filtrate was diluted and injected into a high-performance anion exchange chromatograph with pulsed amperometric detection (HPAEC-PAD) using Dionex ICS-3000 (Dionex Corp., USA) with an electrochemical detector, a guard CarboPac PA1 column (2  $\times$  50 mm, Dionex), a CarboPac PA1 column (2  $\times$  250 mm, Dionex), a AS40 automated sampler, and a PC 10 pneumatic controller at room temperature. 0.002 M and 0.004 M NaOH was used as the eluent and post-column rinsing effluent. The total analysis time was 70 min, with a flow rate 0.4 mL min<sup>-1</sup>. Calibration was performed with standard solutions of glucose, xylose, arabinose, mannose, and galactose, and fucose was used as an internal standard.

### Enzymatic hydrolysis

All poplar natural variants were subjected to a high-throughput pretreatment prior to enzymatic hydrolysis based on a 96-well plate format as described elsewhere.<sup>[16]</sup> In brief, biomass (5.00  $\pm$  0.5 mg) was mixed with 250  $\mu$ L of water in triplicate into one of 96 wells in a solid hastelloy microtitre plates, and then sealed with silicone adhesive and Teflon tape. The samples were then pretreated at 180 °C for 17.5 min. Once cooled after pretreatment, enzyme (CTec2, Novozymes), citric buffer, and sodium azide mixture were pipetted into each well without any preceding separation steps. The mixture contained 8% CTec2 with a loading of 70 mg g<sup>-1</sup> biomass in 1 M sodium citrate buffer. The samples are then gently mixed and left to statically incubate at 50 °C for 70 h. After 70 h incubation, an aliquot of the saccharified hydrolysate was diluted and tested using megazymes GOPOD (glucose oxidase/peroxidase) and XDH assays (xylose dehydrogenase). The results are calculated using standard curves created from mixtures of glucose and xylose. For each poplar variant, three analytical replicates were performed and the standard deviation was calculated accordingly.

### Cellulose molecular weight analysis

Poplar ( $\approx 0.6$  g) mixed with peracetic acid ( $\approx 2.10$  g) and deionized water (4.80 mL) was stirred at 25 °C for 24 h to isolate the holocellulose samples. The samples were dried in a vacuum oven at 40 °C overnight. Cellulose was further isolated from holocellulose by extraction with NaOH solution (5.00 mL, 17.5%) at 25 °C for 2 h. The mixture was then diluted to 8.75% NaOH solution by adding of deionized water (5.00 mL) and stirred at 25 °C for additional 2 h. The  $\alpha$ -cellulose was then collected by centrifugation, washed with 50 mL of 1% acetic acid and an excess of deionized water, and air-dried. The  $M_n$  and  $M_w$  of cellulose were measured by Agilent GPC SECurity 1200 system equipped with four Waters Styragel columns (i.e., HR0.5, HR2, HR4, and HR6), refractive index (RI) detector, and UV detector (270 nm) after cellulose tricarbanilation as described elsewhere.<sup>[39]</sup> Briefly, the cellulose derivative was dissolved in THF (1.00 mg mL<sup>-1</sup>), and the solution was filtered through a 0.45  $\mu$ m PTFE filter and placed in an auto-sampler vial. THF was used as the mobile phase (1.00 mL min<sup>-1</sup>) and the injection volume was 30.0  $\mu$ L. Data collection and processing was performed by Polymer Standards Service WinGPC Unity software (Build 6807). The molecular weight was calculated by the software relative to the universal polystyrene calibration curve.

### Cellulose crystallinity measurement

The cellulose sample for solid-state NMR was prepared from holocellulose sample by acid hydrolysis using HCl (2.5 M) as described previously.<sup>[40]</sup> The isolated samples were never dried and stored in a freezer at  $-4^{\circ}\text{C}$  to maintain a moisture content greater than 30%. For NMR analysis, moist cellulose ( $\approx 35\%$ ) was packed into a 4 mm cylindrical ceramic MAS rotor. Solid-state NMR analysis was carried out on a Bruker Avance 400 MHz spectrometer operating at frequencies of 100.55 MHz for  $^{13}\text{C}$  in a Bruker double-resonance MAS probe at spinning speeds of 10 kHz. CP/MAS experiments were carried out with a  $5\ \mu\text{s}$  ( $90^{\circ}$ ) proton pulse, 1.5 ms contact pulse, 4 s recycle delay, and 4–8 k scans.

### Simons' stain

DB (Pontamine Fast Sky Blue 6BX) and DO (Pontamine Fast Orange 6RN) dyes were obtained from Pylam Products Co. Inc. (Garden City, NY). DB was used as received. Although the original staining method developed by Simons utilized both the orange and blue dye as received, later studies suggested that only the high molecular weight fraction of the DO dye was responsible for the increased affinity for cellulose, whereas the low molecular weight part had a very similar affinity for cellulose as the DB dye.<sup>[41]</sup> Therefore, an ultrafiltration of DO dye to remove the low molecular weight part is necessary, and it was done by filtering a 1% solution of DO through a 100 K membrane using an Amicon ultrafiltration apparatus (Amicon Inc., Beverly, MA) under  $\approx 200$  kPa nitrogen gas pressure. To calculate the concentration of the DO dye after ultrafiltration, the solution (1.00 mL) was dried in a  $50^{\circ}\text{C}$  oven for a week and the weight of the solid residue was measured. Fiber samples ( $\approx 100$  mg) were weighed into five centrifuge tubes, and phosphate buffered saline solution (1.00 mL pH 6, 0.3 M  $\text{PO}_4$ , 1.40 M NaCl) was added to each tube. A set of tubes containing a 1:1 mixture of DB and DO dyes at increasing concentrations were prepared by adding same amount of DB and DO dyes in a series of increasing volumes (i.e., 0.25, 0.5, 0.75, 1.00, and 1.50 mL), which could be then used to measure the dye adsorption isotherm. Distilled water was added to each tube to make up the final volume to 10 mL. All the centrifuge tubes were incubated at  $70^{\circ}\text{C}$  for  $\approx 6$  h with shaking at 200 rpm. After that, the absorbance of the supernatant solution was obtained on a Lambda 35 UV/Vis spectrophotometer at 455 and 624 nm, which represent the wavelength of maximum absorbance for DO and DB, respectively. To calculate the concentration of the dye in the supernatant, two Lambert–Beer law equations were solved simultaneously. The amount of each dye adsorbed by the biomass sample was determined using the difference between the concentration of the initial added dye and the concentration of the dye in the supernatant. The maximum amount of either DO dye or DB dye adsorbed to the lignocellulosic substrates was calculated using the Langmuir adsorption equation.

### Cellulolytic enzyme lignin isolation

Cellulolytic enzyme lignin was isolated from poplar variants according to a slightly modified literature.<sup>[21]</sup> Briefly, the extractive-free poplar samples were ball-milled in a porcelain jar with ceramic balls using a rotatory ball milling. The ground powder was then subjected to enzymatic hydrolysis in acetate buffer (pH 4.8,  $50^{\circ}\text{C}$ ) for 48 h. The residue was then isolated and hydrolyzed one more time with freshly added enzyme and buffer. This enzyme-treated lignin-rich residue was then extracted with 96% *p*-dioxane/water

mixture for two times 24 h. The extracts were combined, rotary evaporated to reduce the volume under reduced pressure, followed by freeze drying. The obtained lignin samples were oven dried at  $45^{\circ}\text{C}$  overnight before further analysis.

### Lignin molecular weight analysis

The oven-dried lignin samples ( $\approx 25$  mg) were acetylated using acetic anhydride/pyridine (1:1, v/v, 2.0 mL) at room temperature for 24 h. After 24 h,  $\approx 25$  mL of ethanol was added to the reaction mixture and left for  $\approx 30$  min. The solvent was then removed with a rotary evaporator under reduced pressure. The whole addition and removal of ethanol process was repeated until trace of acetic acid was removed from the sample. The acetylated lignin samples were dried under vacuum at  $45^{\circ}\text{C}$  overnight prior to GPC analysis. The molecular weight distributions of these acetylated lignin samples were analyzed using the GPC SECurity 1200 system. Similar to cellulose MW analysis, THF was used as the mobile phase with a flow rate  $1.0\ \text{mL}\cdot\text{min}^{-1}$  and polystyrene narrow standards were used to prepare the calibration curve.

### HSQC NMR analysis

2D  $^{13}\text{C}$ – $^1\text{H}$  HSQC NMR spectra of isolated lignin samples were acquired in a Bruker Avance 400 MHz spectrometer. A standard Bruker heteronuclear single quantum coherence pulse sequence (hsqcetgp) was used on a 5-mm Broadband Observe (BBO) probe with the following conditions: 13 ppm spectral width in F2 ( $^1\text{H}$ ) dimension with 1024 data points and 210 ppm spectral width in F1 ( $^{13}\text{C}$ ) dimension with 256 data points, a 1.5 s pulse delay, a  $90^{\circ}$  pulse, a  $^1J_{\text{C-H}}$  of 145 Hz, and 32 scans. Lignin samples were prepared as follows: lignin samples ( $\approx 30$  mg) were added to deuterated dimethyl sulfoxide (0.5 mL,  $[\text{D}_6]\text{DMSO}$ ). The residual DMSO solvent peak around 39.5 ppm (C) and 2.50 ppm (H) was used for chemical shift calibration. Relative lignin interunit linkage abundance and monomer compositions were calculated by using volume integration of contours in HSQC spectra semi-quantitatively.<sup>[21]</sup> NMR data and spectra processing was performed using TopSpin 2.1 software (Bruker BioSpin) and Adobe Illustrator CC (Adobe Inc.).

### $^{31}\text{P}$ NMR analysis

$^{31}\text{P}$  NMR experiments were conducted on a Bruker Avance 400 MHz spectrometer. Lignin samples ( $\approx 20$  mg) were dissolved in a solvent mixture of pyridine and deuterated chloroform (1.6/1.0, v/v, 0.50 mL). The lignin solution was then further derivatized with 2-chloro-4,4,5,5-tetramethyl-1,3,2-dioxaphospholane (TMDP). Chromium acetylacetonate and *endo-N*-hydroxy-5-norbornene-2,3-dicarboximide (NHND) were also added into the solution as relaxation agent and internal standard, respectively. The quantitative  $^{31}\text{P}$  spectrum was acquired at a frequency of 161.93 MHz over 32 k data points with an acquisition time of 1.29 s using an inverse gated decoupling pulse sequence (Waltz-16) with a 25 s pulse delay and 128 scans. All the NMR data were processed using the TopSpin 2.1 software (Bruker BioSpin) and MestreNova (Mestre Laboratories) software packages.

## Acknowledgements

This manuscript was authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. This study was supported and performed as part of the BioEnergy Science Center (BESC). The BESC is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

**Keywords:** biomass recalcitrance · cellulose · crystallinity · degree of polymerization · lignin

- [1] A. J. Ragauskas, C. K. Williams, B. H. Davison, G. Britovsek, J. Cairney, C. A. Eckert, W. J. Frederick, Jr., J. P. Hallett, D. J. Leak, C. L. Liotta, J. R. Mielenz, R. Murphy, R. Templer, T. Tschaplinski, *Science* **2006**, *311*, 484–489.
- [2] F. Hu, A. Ragauskas, *BioEnergy Res.* **2012**, *5*, 1043–1066.
- [3] M. E. Himmel, S. Y. Ding, D. K. Johnson, W. S. Adney, M. R. Nimlos, J. W. Brady, T. D. Foust, *Science* **2007**, *315*, 804–807.
- [4] Y.-H. P. Zhang, L. R. Lynd, *Biotechnol. Bioeng.* **2004**, *88*, 797–824.
- [5] X. L. Luo, J. Y. Zhu, R. Gleisner, H. Y. Zhan, *Cellulose* **2011**, *18*, 1055–1062.
- [6] B. B. Hallac, A. J. Ragauskas, *Biofuels Bioprod. Biorefin.* **2011**, *5*, 215–225.
- [7] A. P. Sinititsyn, A. V. Gusakov, E. Y. Vlasenko, *Appl. Biochem. Biotechnol.* **1991**, *30*, 43–59.
- [8] a) C. A. Mooney, S. D. Mansfield, M. G. Touhy, J. N. Saddler, *Bioresour. Technol.* **1998**, *64*, 113–119; b) X. Meng, A. J. Ragauskas, *Curr. Opin. Biotechnol.* **2014**, *27*, 150–158.
- [9] M. H. Studer, J. D. DeMartini, M. F. Davis, R. W. Sykes, B. Davison, M. Keller, G. A. Tuskan, C. E. Wyman, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 6300–6305.
- [10] S. L. Voelker, B. Lachenbruch, F. C. Meinzer, M. Jourdes, C. Ki, A. M. Patten, L. B. Davin, N. G. Lewis, G. A. Tuskan, L. Gunter, S. R. Decker, M. J. Sellig, R. Sykes, M. E. Himmel, P. Kitin, O. Shevchenko, S. H. Strauss, *Plant Physiol.* **2010**, *154*, 874–886.
- [11] a) S.-Y. Leu, J. Y. Zhu, *BioEnergy Res.* **2013**, *6*, 405–415; b) B. Yang, M. C. Gray, C. Liu, T. A. Lloyd, S. L. Stuhler, A. O. Converse, C. E. Wyman, *ACS Symp. Ser.* **2004**, *889*, 100–125.
- [12] Y. Pu, F. Hu, F. Huang, B. H. Davison, A. J. Ragauskas, *Biotechnol. Biofuels* **2013**, *6*, 15.
- [13] X. Meng, T. Wells, Q. Sun, F. Huang, A. Ragauskas, *Green Chem.* **2015**, *17*, 4239–4246.
- [14] M. Foston, A. J. Ragauskas, *Ind. Biotechnol.* **2012**, *8*, 191–208.
- [15] W. Muchero, J. Guo, S. P. DiFazio, J.-G. Chen, P. Ranjan, G. T. Slavov, L. E. Gunter, S. Jawdy, A. C. Bryan, R. Sykes, A. Ziebell, J. Klápště, I. Porth, O. Skyba, F. Unda, Y. A. El-Kassaby, C. J. Douglas, S. D. Mansfield, J. Martin, W. Schackwitz, L. M. Evans, O. Czarnecki, G. A. Tuskan, *BMC Genomics* **2015**, *16*, 24.
- [16] S. R. Decker, R. W. Sykes, G. B. Turner, J. S. Lupoi, C. Doepkpe, M. P. Tucker, L. A. Schuster, K. Mazza, M. E. Himmel, M. F. Davis, E. Gjersing, *J. Visualized Exp.* **2015**, e53163.
- [17] S. Bhagia, W. Muchero, R. Kumar, G. A. Tuskan, C. E. Wyman, *Biotechnol. Biofuels* **2016**, *9*, 1–12.
- [18] M. Foston, A. J. Ragauskas, *Biomass Bioenergy* **2010**, *34*, 1885–1895.
- [19] X. Meng, M. Foston, J. Leisen, J. DeMartini, C. E. Wyman, A. J. Ragauskas, *Bioresour. Technol.* **2013**, *144*, 467–476.
- [20] C. I. Ishizawa, M. F. Davis, D. F. Schell, D. K. Johnson, *J. Agric. Food Chem.* **2007**, *55*, 2575–2581.
- [21] R. Samuel, Y. Pu, N. Jiang, C. Fu, Z.-Y. Wang, A. J. Ragauskas, *Front. Energy Res.* **2014**, *1*, 14.
- [22] Q. Sun, Y. Pu, X. Meng, T. Wells, A. J. Ragauskas, *ACS Sustainable Chem. Eng.* **2015**, *3*, 2203–2210.
- [23] Y. Pu, S. Cao, A. J. Ragauskas, *Energy Environ. Sci.* **2011**, *4*, 3154–3166.
- [24] E. M. Rubin, *Nature* **2008**, *454*, 841–845.
- [25] J. D. DeMartini, S. Pattathil, J. S. Miller, H. Li, M. G. Hahn, C. E. Wyman, *Energy Environ. Sci.* **2013**, *6*, 898–909.
- [26] a) L. R. Lynd, P. J. Weimer, W. H. van Zyl, I. S. Pretorius, *Microbiol. Mol. Biol. Rev.* **2002**, *66*, 506–577; b) S. D. Mansfield, C. Mooney, J. N. Saddler, *Biotechnol. Prog.* **1999**, *15*, 804–816.
- [27] M. Hall, P. Bansal, J. H. Lee, M. J. Reaff, A. S. Bommarius, *FEBS J.* **2010**, *277*, 1571–1582.
- [28] M. Brienzo, S. Ferreira, M. P. Vicentim, W. de Souza, C. Sant'Anna, *BioEnergy Res.* **2014**, *7*, 1454–1465.
- [29] X. Luo, J. Y. Zhu, *Enzyme Microb. Technol.* **2011**, *48*, 92–99.
- [30] A. Tolbert, H. Akinoshio, R. Khunsupat, A. K. Naskar, A. J. Ragauskas, *Biofuels Bioprod. Biorefin.* **2014**, *8*, 836–856.
- [31] A. Ziebell, K. Gracom, R. Katahira, F. Chen, Y. Pu, A. Ragauskas, R. A. Dixon, M. Davis, *J. Biol. Chem.* **2010**, *285*, 38961–38968.
- [32] A. Berlin, M. Balakshin, N. Gilkes, J. Kadla, V. Maximenko, S. Kubo, J. Saddler, *J. Biotechnol.* **2006**, *125*, 198–209.
- [33] G. Cheng, X. Zhang, B. Simmons, S. Singh, *Energy Environ. Sci.* **2015**, *8*, 436–455.
- [34] X. J. Pan, *J. Biobased Mater. Bioenergy.* **2008**, *2*, 25–32.
- [35] a) F. Guo, W. Shi, W. Sun, X. Li, F. Wang, J. Zhao, Y. Qu, *Biotechnol. Biofuels* **2014**, *7*, 1–10; b) S. Nakagame, R. P. Chandra, J. F. Kadla, J. N. Saddler, *Biotechnol. Bioeng.* **2011**, *108*, 538–548.
- [36] Z. Yu, K.-S. Gwak, T. Treasure, H. Jameel, H.-m. Chang, S. Park, *ChemSusChem* **2014**, *7*, 1942–1950.
- [37] B. H. Davison, S. R. Drescher, G. A. Tuskan, M. F. Davis, N. P. Nghiem, *Appl. Biochem. Biotechnol.* **2006**, *130*, 427–435.
- [38] Z. Wang, D. R. Keshwani, A. P. Redding, J. J. Cheng, *Bioresour. Technol.* **2010**, *101*, 3583–3585.
- [39] X. Meng, Q. Sun, M. Kosa, F. Huang, Y. Pu, A. J. Ragauskas, *ACS Sustainable Chem. Eng.* **2016**, *4*, 4563–4572.
- [40] M. B. Foston, C. A. Hubbell, A. J. Ragauskas, *Materials* **2011**, *4*, 1985–2002.
- [41] R. Chandra, S. Ewanick, C. Hsieh, J. N. Saddler, *Biotechnol. Prog.* **2008**, *24*, 1178–1185.

Manuscript received: September 19, 2016

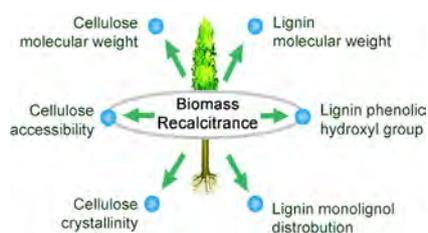
Revised: October 31, 2016

Accepted Article published: November 23, 2016

Final Article published: ■ ■ ■ 0000

## FULL PAPERS

**Back to fundamentals:** Biomass recalcitrance is the major barrier that significantly hindered the industrial commercialization process of converting biomass to bio-ethanol. Cellulose and lignin structural features of various natural poplar variants are characterized, compared, and correlated with their sugar release, providing insight into most fundamental mechanisms of biomass recalcitrance.



*X. Meng, Y. Pu, C. G. Yoo, M. Li, G. Bali, D.-Y. Park, E. Gjersing, M. F. Davis, W. Muchero, G. A. Tuskan, T. J. Tschaplinski, A. J. Ragauskas\**



**An In-Depth Understanding of Biomass Recalcitrance Using Natural Poplar Variants as the Feedstock**

