NMR Elucidation of Non-productive Binding Sites of 13C-Labeled Lignin Models with Carbohydrate Binding Module of Cellobiohydrolase I for Efficient Biomass Conversion

Yuki Tokunaga
Kyoto University
Takashi Nagata
Kyoto University
Keiko Kondo
Kyoto University
Masato Katahira
Kyoto University
Takashi Watanabe (twatanab@rish.kyoto-u.ac.jp)
Kyoto University https://orcid.org/0000-0003-0220-4157

Research

Keywords: Cellulase, Lignin, NMR, Carbohydrate-binding module, Cellobiohydrolase I, Enzyme adsorption, Chemical shift perturbation, HSQC, Lignin oligomer model

Posted Date: June 29th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-36005/v1

Version of Record: A version of this preprint was published on October 7th, 2020. See the published version at https://doi.org/10.1186/s13068-020-01805-w.

Abstract

Background: Highly efficient enzymatic saccharification of pretreated lignocellulose is a primary key step in achieving lignocellulosic biorefinery. Cellobiohydrolase I (Cel7A) secreted by *Trichoderma reesei* is an industrially used cellulase possessing carbohydrate binding module 1 (*Tr*CBM1) as the C-terminal domain. Non-productive binding of *Tr*CBM1 to lignin significantly decreases enzymatic saccharification efficiency and enhance cost of biomass conversion due to required additional enzymes. Understanding of the interaction mechanism between lignin and *Tr*CBM1 is essentially required to realize cost-effective biofuels production, but the binding sites in lignin have not been clearly elucidated.

Results: Three types of ^{13}C-labeled b-*O*-4 lignin oligomer models were synthesized and characterized. The 2D ^{1}H-^{13}C HSQC spectra of the ^{13}C-labeled lignin models exhibited that ^{13}C-labels were correctly incorporated in the (1) aromatic rings and b positions, (2) a positions, and (3) methoxy groups, respectively. The *Tr*CBM1 binding sites in lignin were analyzed by observing NMR chemical shift perturbations (CSPs) using the synthetic ^{13}C-labeled b-*O*-4 lignin oligomer models. Obvious CSPs were observed in signals from the aromatic regions in oligomers bound to *Tr*CBM1, whereas perturbations in the signals from aliphatic regions and methoxy groups were insignificant. This indicated that hydrophobic interactions and p–p stacking were dominating factors in non-productive binding. The synthetic lignin models have two configurations whose terminal units were differently aligned and donated $C^{(I)}$ and $C^{(II)}$. The $C^{(I)}$ ring showed remarkable perturbation compared with $C^{(II)}$, which indicated that binding of *Tr*CBM1 is evidently affected by configuration of lignin models. Long-chain lignins (DP 4.16–4.70) clearly bound to *Tr*CBM1. Interactions with short-chain lignins (DP 2.64–3.12) were insignificant, indicating that a DP greater than 4 was necessary for *Tr*CBM1 binding.

Conclusion: The CSP analysis using ^{13}C-labeled b-*O*-4 lignin oligomer models enabled us to identify *Tr*CBM1 binding sites in lignin at the atomic level. This specific interaction analysis will lead to new molecular design of cellulase having controlled affinity to cellulose and lignin for cost-effective biorefinery process.

Background

The enzymatic saccharification of lignocellulose is a key process for the sustainable manufacture of green chemicals and biofuels [1]. *Trichoderma reesei* is a filamentous fungus that is widely used for the production of commercially available cellulolytic enzyme cocktails. Cellobiohydrolase I (Cel7A) accounts for up to 60% of the cellulases secreted by *T. reesei*, and carbohydrate-binding module 1 (*Tr*CBM1) (Fig. 1) is connected to the C-terminus of the Cel7A catalytic domain by a highly glycosylated linker [2]. *Tr*CBM1 enhances enzymatic activity by binding to cellulose [3]. *Tr*CBM1 also has a strong affinity for lignin, so lignin significantly inhibits the enzymatic saccharification of pretreated lignocellulose [4]. Non-productive binding of *Tr*CBM1 to lignin should thus be suppressed to achieve efficient enzymatic saccharification. However, the mechanism of interaction is not entirely understood at the molecular level.

Binding of *Tr*CBM1 to lignin is affected by various saccharification conditions, such as temperature [6], pH [7], and the substrate concentration [8]. The chemical properties of lignin have a critical impact on its cellulase binding affinity. Many pretreatments increase the number of phenolic OH groups in lignin and its degree of condensation, which enhance binding between cellulase and lignin [9, 10]. Electrostatic repulsion due to large numbers of aliphatic OH groups and negatively charged functionalities, such as carboxyl and sulfone groups, interferes with cellulase binding [9, 11]. Softwood lignin is mainly synthesized via the radical coupling of guaiacyl monomers, whereas hardwood lignin is synthesized from both guaiacyl (G) and syringyl (S) monomers. The ratio of syringyl to guaiacyl units, denoted by S/G, is one of the most important factors governing the physicochemical properties of lignified plant cell walls. The S/G ratio affects non-productive binding, but it is not the decisive factor for adsorption. Guo et al. found that a low S/G ratio corresponded to a high adsorption capacity [12]. Yang et al. reported that organosolv lignin isolated from softwood lodgepole pine had a higher adsorption affinity for commercial cellulase than lignin from hardwood poplar [13]. However, genetic engineering studies have suggested the opposite, and comparisons of enzymatic hydrolysis in alfalfa cultivars and *Eucalyptus* mutants with high and low S/G ratios have yielded inconsistent results [14, 15]. Although the relationship between the chemical structure of lignin and its binding affinity for cellulolytic enzymes and CBMs has been researched extensively, evidence for the identification of binding atoms in whole lignin molecules is entirely lacking. This is because no effective method for identifying the interactive sites in lignin, a heterogenous polymer, is available.

NMR chemical shift perturbation (CSP) analysis is a powerful method for the identification of substrate binding sites in proteins [16]. We previously performed CSP analysis to identify the binding amino acid residues in *Tr*CBM1 with softwood and hardwood milled wood lignin (MWL) using ^{15}N-labeled *Tr*CBM1 [17]. The results suggested that the aromatic rings in lignin participated in interactions with amino acid residues in *Tr*CBM1, because the flat plane surface including Y5, Y31, and Y32 in *Tr*CBM1 was a main interaction site. However, the binding sites in lignin have not been characterized at the molecular level. In this study, we synthesized ^{13}C-labeled β-*O*-4 lignin oligomer model compounds with different labelling positions and performed CSP analysis to identify the *Tr*CBM1 binding sites in the lignin models. Pure *Tr*CBM1 without a catalytic domain and linker was expressed and purified and then added to NMR sample tubes containing the ^{13}C-labeled lignin models to monitor perturbations in signals from the binding sites in the model compounds. NMR analysis of the ^{13}C-labeled lignin model compounds provided the first direct evidence for the identification of binding atoms in the linear lignin chains. The results were in good agreement with our previous binding site analysis of the protein counterpart, *Tr*CBM1 with MWL [17]. Herein, we provide direct evidence to reveal the interaction sites in the β-*O*-4 lignin substructure that bind to *Tr*CBM1.

Results

Expression and purification of TrCBM1

Escherichia coli BL21 (DE3) was used to express a histidine (His) tag-*Tr*CBM1-green fluorescent protein (GFP) fusion protein. *Tr*CBM1 was obtained after the addition of enterokinase and thrombin to cleave the His tag and GFP, respectively. We previously performed sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) and 2D ^{1}H-^{15}N HSQC NMR to analyze the molecular mass and conformation of ^{15}N-labeled *Tr*CBM1 [17]. In this work, we performed SDS-PAGE and MALDI-TOF-MS to characterize unlabeled *Tr*CBM1. Pure *Tr*CBM1 was a single protein with a molecular mass of 5195.8 Da. The MALDI-TOF-MS spectrum of *Tr*CBM1 and a full-length SDS-PAGE gel are shown in Figure S1 of the Additional file.

Synthesis and characterization of the lignin models

In recent work, we synthesized the unlabeled β-*O*-4 lignin oligomer model compound **4** (Fig. 2) and fully characterized it by performing 1D ^{1}H NMR, 1D ^{13}C NMR, 2D ^{1}H-^{13}C HSQC, 2D ^{1}H-^{13}C HMBC, 2D ^{1}H-^{13}C long-range heteronuclear single quantum multiple bond correlation (LR-HSQMBC), 2D ^{1}H-^{1}H TOCSY, 2D ^{1}H-^{1}H EXSY and 2D ^{1}H-^{1}H ROESY [18]. We applied the protocol to synthesize three ^{13}C-labeled β-*O*-4 lignin oligomer model compounds. ^{13}C-labels were placed in the aromatic rings and β positions of compound $\mathbf{4}^{(Ar\beta)}$, the α positions of compound $\mathbf{4}^{(\alpha)}$, and methoxy groups in compound $\mathbf{4}^{(m)}$ using ^{13}C-labeled vanillins and *t*-butyl-2-bromoacetate. The ^{13}C-vanillins $\mathbf{1}^{(Ar)}$, $\mathbf{1}^{(\alpha)}$, and $\mathbf{1}^{(m)}$ were individually refluxed in acetone with unlabeled or ^{13}C-labeled *t*-butyl-2-bromoacetate to afford ^{13}C-*t*-butoxycarbonylmethyl vanillins $\mathbf{2}^{(Ar\beta)}$, $\mathbf{2}^{(\alpha)}$, and $\mathbf{2}^{(m)}$, respectively. The ^{13}C-*t*-butoxycarbonylmethyl vanillins were dissolved in anhydrous tetrahydrofuran (THF) and polymerized using lithium diisopropylamide (LDA). The ester groups in polymerized ^{13}C-oligomers $\mathbf{3}^{(Ar\beta)}$, $\mathbf{3}^{(\alpha)}$, and $\mathbf{3}^{(m)}$ were reduced with $NaBH_4$ to obtain ^{13}C-labeled β-*O*-4 lignin oligomer model compounds $\mathbf{4}^{(Ar\beta)}$, $\mathbf{4}^{(\alpha)}$, and $\mathbf{4}^{(m)}$, respectively.

The 2D ^{1}H-^{13}C HSQC spectra of the ^{13}C-labeled lignin models contained signals that corresponded to the ^{1}H-^{13}C correlations (Fig. 3), which was evidence that ^{13}C was incorporated at the designated positions. No HSQC signals attributable to side products were observed in the NMR spectra, which indicated that high-purity lignin models were prepared. HSQC signals at δ_C/δ_H 85.9/4.52–4.64, 72.6/4.06–4.08, and 113.2–123.6/6.81–7.06 ppm in the HSQC spectra of lignin model $\mathbf{4}^{(Ar\beta)}$ with ^{13}C-labeling in the β positions and aromatic rings (Fig. 3a) were assigned to the β positions, Terminal C1, and aromatic regions, respectively. The HSQC spectra of lignin model $\mathbf{4}^{(\alpha)}$ (Fig. 3b) contained three signals at δ_C/δ_H 74.7–75.0/4.92–4.98, 74.8–75.0/4.72–4.76, and 66.2/4.43–4.49 ppm, which were assigned to α positions in the A, B, and C rings, respectively. A single signal was observed at 58.1–58.3/3.57–3.85 ppm in the HSQC spectra of lignin model $\mathbf{4}^{(m)}$ with ^{13}C-labeled methoxy groups (Fig. 3c).

The long- and short-chain lignin models were separated via silica gel chromatography and subjected to NMR and binding analysis. Size exclusion chromatography (SEC) revealed that the ^{13}C-labeled lignin models had narrow molecular weight distributions (Figure S2, Additional file). The degree of polymerization (DP) was calculated for each model compound from its weight-average molecular weight

(*M*w). The DPs of the long-chain lignin models ranged from 4.16 to 4.70, whereas those of the short-chain models ranged from 2.64 to 3.12 (Table 1). Although differences between the *M*ws of the long- and short-chain models were small, their NMR spectra (Fig. 3d) and SEC profiles (Figure S2, Additional file) were clearly distinct. Therefore, the long- and short-chain lignin models could be used to evaluate the effects of *M*w on molecular interactions with *Tr*CBM1.

Table 1

Molecular weight parameters of ^{13}C-labeled and unlabeled β-*O*-4 lignin oligomer model compounds.

Lignin models	Chain length	M_n	M_w	M_w/M_n	DP[a]
Compound **4**$^{(Arβ)}$	Long	843	923	1.095	4.70
(^{13}C-labeled aromatic rings and β positions)	Short	480	520	1.083	2.64
Compound **4**$^{(α)}$	Long	750	918	1.224	4.67
(^{13}C-labeled α positions)	Short	531	614	1.156	3.12
Compound **4**$^{(m)}$	Long	629	818	1.301	4.16
(^{13}C-labeled methoxy groups)	Short	482	587	1.218	2.98
Compound **4**	Long	882	964	1.092	4.91
(unlabeled)	Short	552	560	1.014	2.85

a DP: Degree of polymerization calculated from the *M*w and theoretical molecular mass of the lignin model.

Analysis of interactions between lignin models and TrCBM1

Interactions between the lignin models and *Tr*CBM1 were evaluated via CSP analysis using 2D ^{1}H-^{13}C HSQC NMR. To analyze the binding positions in lignin model compounds **4**$^{(Arβ)}$, **4**$^{(α)}$, and **4**$^{(m)}$ with a high degree of sensitivity, the compounds were separated into high and low molecular mass fractions. Assignment of the ^{13}C-lignin model HSQC signals was accomplished for all of the carbon atoms and non-exchangeable protons. Almost all of the signals from the A, B, and C rings were assigned separately, because the tops of the peaks were distinguishable [18]. We previously found that the lignin models have two diastereomers. Although these A and B rings have similar alignments, the C ring are differently aligned and designated as C$^{(I)}$ and C$^{(II)}$ [18]. Full assignment of the peaks in the lignin model spectra enabled us to analyze the interactions of carbon atoms and protons at the atomic level.

The 2D ^{1}H-^{13}C HSQC spectra of the lignin models (50 μM) in the presence and absence of *Tr*CBM1 were superimposed (Fig. 4). Several HSQC signals acquired in the presence of 200 μM *Tr*CBM1 showed obvious perturbation, particularly those of the aromatic regions in the long-chain lignin spectra. Most of these signals exhibited larger perturbations when the concentration of *Tr*CBM1 was increased to 350 μM

(Fig. 4b). In contrast, there was no significant perturbation of NMR signals from the aliphatic regions and methoxy groups in the presence of *Tr*CBM1 (Fig. 4c). The results clearly showed that the aromatic regions in the long-chain lignin models were the primary sites of interaction with *Tr*CBM1. However, no distinct perturbations were observed in signals from the methoxy groups or any of the aromatic and aliphatic regions when the short-chain lignin models were used for CSP analysis (Figures S3, Additional file). This indicated that the length of the lignin chains was an important factor in *Tr*CBM1 binding.

The change in the chemical shift (Δδ) of each signal was calculated by subtracting the chemical shift value in the spectrum of the lignin model from that of the model recorded in the presence of *Tr*CBM1. The Δδ values obtained in the ^{1}H and ^{13}C-axis are plotted in Fig. 5. Δδ values were large only for the signals from the aromatic regions in the long-chain lignin models, whereas the Δδ values of signals from the aliphatic regions and methoxy groups in both the long- and short-chain lignin models were small. The Δδ values of the A 2, C 2$^{(I)}$, B 5, C 5$^{(I)}$, C 5$^{(II)}$, A 6, C 6$^{(I)}$, and C 6$^{(II)}$ signals from the aromatic regions of the long-chain lignin models were all larger than 0.006 ppm on the ^{1}H-axis, whereas the Δδ values of the A 2 and C 5$^{(I)}$ signals on the ^{13}C-axis were greater than 0.05 ppm (Fig. 5a). For the short-chain lignin models, only the Δδ values of the B 5 signals on the ^{1}H-axis exceeded 0.006 ppm (Fig. 5b). Line broadening was observed in the B 5 signals of the long-chain lignin models in the presence of *Tr*CBM1 (350 μM). The observed line broadening could be attributed to both the on and off rates of complex formation and the multiple binding states of lignin due to non-specific binding [17].

Adsorption experiments with the lignin models

The unlabeled long- and short-chain lignin models of compound **4** were used to evaluate *Tr*CBM1 binding affinity according to the Langmuir adsorption model. The *M*ws and molecular weight distributions of the unlabeled lignin models (Table 1) were nearly identical to these of the ^{13}C-labeled lignin models (Figure S2, Additional file). The *Tr*CBM1–His tag fusion protein was used instead of *Tr*CBM1 without a His tag to conduct the adsorption experiments, because the soluble lignin models could not be separated from the *Tr*CBM1 protein via centrifugation. After incubating the sample solutions containing the lignin models and *Tr*CBM1–His tag at 50 °C for 1 h, cOmplete His Tag Purification Resin (Roche, Basel, Switzerland) was added to the solutions. The unbound lignin models in the supernatant could then be separated from the bound lignin models, which were adsorbed to the precipitated *Tr*CBM1–His tag bound to the His tag resin. We confirmed that all of the *Tr*CBM1–His tag would adsorb to the His tag resin by performing control experiments. We did not observe unwanted binding between the His tag resin and lignin models (Figures S6 and S7, Additional file). The concentrations of the unbound lignin models in the supernatants were determined to calculate the adsorption parameters summarized in Table 2. Although the long- and short-chain lignin models had similar Γ_{max} values, the K_L of the long-chain lignin model was eight times higher than the K_L of the short-chain lignin model. This result was consistent with the results of the NMR interaction analysis. The percentages of the lignin models that bound to *Tr*CBM1 were calculated using the K_L values. In the CSP analysis, 84.6% and 91.4% of the long-chain lignin models were bound to *Tr*CBM1 at *Tr*CBM1 concentrations of 200 and 350 μM, respectively. At *Tr*CBM1 concentrations of 200

and 350 μM, 33.0% and 46.6% of the short-chain lignin models, respectively, were bound to *Tr*CBM1. The chain length was thus an essential factor in binding between *Tr*CBM1 and the lignin chains, which contained β-*O*-4 linkages exclusively. We also found that strong adsorption required a DP above 4.

Table 2
*Tr*CBM1–His tag adsorption parameters of the unlabeled lignin models according to Langmuir adsorption model.

Lignin model	Langmuir affinity constant, K_L (mL/mg)	Amount of adsorption at saturation, Γ_{max} (μg/mg)
Long-chain	36.26	8.77
Short-chain	4.47	9.15

Discussion

Understanding the mechanism of interaction between cellulase and lignin is essential for the efficient enzymatic saccharification of lignocelluloses. *T. reesei* is the most important industrial cellulase-producing filamentous fungus. Cel7A is the most abundant secreted cellulolytic enzyme and contains a catalytic domain and *Tr*CBM1. Non-productive binding between *Tr*CBM1 and lignin decreases the efficiency of enzymatic saccharification of pretreated plant biomass, but the interactive sites in lignin have not been previously elucidated.

The flat plane surface of *Tr*CBM1 binds to both lignin and cellulose, and contains the three tyrosine residues (Y5, Y31, and Y32) and neighboring amino acid residues in the underside of *Tr*CBM1 (Fig. 1) [7, 19]. T17, V18, and T24 residues are present in the cleft, which is located on the opposite side of the flat plane surface. This site also interacts with lignin and cellulosic substrates [17]. Although the *Tr*CBM1 binding sites have been extensively investigated via point mutation analysis and NMR, evidence to identify the lignin atoms that participate in *Tr*CBM1 binding is lacking. Therefore, we conducted CSP analysis to investigate interactions between ^{13}C-labeled β-*O*-4 lignin oligomer compounds and *Tr*CBM1. Interactive sites in the long-chain lignin model were mapped in Fig. 6, which are based on Δδ values observed in CSP analysis. Interactions between the aromatic rings and *Tr*CBM1 were obvious, whereas the aliphatic regions and methoxy groups were not major binding sites. This suggests that *Tr*CBM1 adsorption on lignin occurs via hydrophobic interactions and π–π stacking of the aromatic rings in lignin and the three tyrosine residues on *Tr*CBM1. Rahikainen et al. discussed the importance of *Tr*CBM1 hydrophobicity. They reported that a Y32A *Tr*CBM1 mutant had a lower association constant than wild-type *Tr*CBM1, whereas a Y32W mutant increased the lignin and cellulose binding affinities of *Tr*CBM1 [7]. The hydrophobicity of lignin results in a significant amount of non-productive binding with cellulase [13]. Therefore, hydrophobic interaction can reasonably be interpreted as a dominant driving force behind non-productive *Tr*CBM1 binding by lignin.

Interestingly, there is a variety of Δδ in the aromatic rings of lignin model shown in Fig. 6. The interaction patterns of terminal units on the A and $C^{(I)}$ rings exhibited similar interaction patterns although the A 5 position was not evaluated due to its extremely weak HSQC signal intensity. Internal units, B ring, showed remarkable interactions with *Tr*CBM1 only at the B 5 position, suggesting that the interaction patterns of the terminal and internal units of the lignin model differed. We previously reported two configuration of lignin models whose terminal units were differently aligned and designated them $C^{(I)}$ and $C^{(II)}$ [18]. We demonstrated that the *Tr*CBM1 binding behavior depended significantly on the configurations of the lignin model. The molecular alignment of $C^{(I)}$ apparently resulted in preferential binding to *Tr*CBM1. We are thus the first to reveal that interactive sites in the lignin chains are significantly influenced by their molecular configurations.

Hydrogen bonding and electrostatic interactions are not negligible in non-productive *Tr*CBM1 binding. Phenolic OH groups in lignin promote hydrogen bonding to cellulolytic enzymes [10, 20]. Hydrophilic amino acid residues in *Tr*CBM1, including Q7 and T17 in addition to main chain of H4 and I11 were previously shown to participate in binding with lignin via hydrogen bonds and electrostatic interactions [17]. In our CSP analysis, aromatic rings in the lignin models were the primary sites of interaction with *Tr*CBM1. This was attributed mainly to hydrophobic interactions. However, the results could also be interpreted in terms of hydrogen–π interactions between the aromatic rings in lignin and OH groups on amino acid residues [21, 22]. OH groups in the three tyrosine residues on the flat plane surface of *Tr*CBM1 and the OH group in the T17 residue in the cleft were thought to donate hydrogens for hydrogen–π interactions. Although CSP analysis indicated that interactions at the aliphatic positions were insignificant compared with those at the aromatic positions, signals from the α positions had slightly higher Δδ values than signals from the other aliphatic positions in the lignin models (Fig. 5). This result could be attributed to hydrogen bonding between OH groups near the α positions of the lignin model and *Tr*CBM1.

It is uncertain whether there is a correlation between the molecular weight of lignin and non-productive binding [13, 23]. This is because a decrease in the molecular weight of lignin changes its hydrophobicity and OH content, which are the predominant factors affecting non-productive binding. Changes in these structural features of lignin have disturbed us to estimate the effect of the molecular weight of lignin on cellulase binding. Our CSP results clearly indicated that the long-chain lignin model had a stronger affinity for *Tr*CBM1 than the short-chain lignin model (Table 2). Mattinen et al. found that cellohexaose adsorbed to *Tr*CBM1, whereas short cellooligosaccharides, such as cellobiose and cellotriose, did not bind to *Tr*CBM1 [24]. Based on these results and our observations, stacking on the flat plane surface of *Tr*CBM1 and interactions with the cleft required lignin and cellooligosaccharides with long chains. The low affinity of the short-chain lignin model could also be attributed to its conformation. The short-chain lignin model had a folded conformation in 90% water [18], and the folded short lignin chain could not cover the full length of the three tyrosine residues on the flat plane surface of *Tr*CBM1. Therefore, we concluded that β-*O*-4 lignin chains with DPs above 4 were essential for strong adsorption of *Tr*CBM1.

Conclusions

Highly efficient enzymatic saccharification is hindered by the non-productive binding of lignin to cellulolytic enzymes. Until now, the binding mechanism was not fully understood. We synthesized ^{13}C-labeled β-*O*-4 lignin oligomer model compounds to identify the *Tr*CBM1 binding sites in lignin via NMR CSP analysis. Signals from the aromatic regions in the lignin models exhibited obvious perturbations, whereas signals from the aliphatic regions and methoxy groups were not significantly perturbed. These results suggested that hydrophobic interactions and π–π stacking were the principal forces driving interactions between aromatic rings in the lignin models and tyrosine residues in *Tr*CBM1. *Tr*CBM1 bound differently with terminal and internal units in the lignin models. In addition, the binding patterns associated with the $C^{(I)}$ and $C^{(II)}$ terminal alignments differed. This indicated that binding of the lignin models to *Tr*CBM1 was strongly affected by the molecular configuration. Perturbation of signals from the long-chain lignin models (DP 4.16–4.70) was obvious due to their strong binding affinity relative to that of the short-chain lignin models (DP 2.64–3.12). This indicated that a chain length greater than DP4 was necessary for strong interactions between lignin and *Tr*CBM1. This is the first study to characterize the interactive sites in a lignin model compound at the atomic level using purified *Tr*CBM1. A detailed understanding of non-productive binding will lead to the establishment of a fundamental theory for the structural alteration of lignin and enzymes that are not susceptible to unfavorable binding.

Methods

Materials

E. coli BL21 (DE3) was obtained from Merck (Darmstadt, Germany). Vanillins with ^{13}C-labeled carbonyl group and ^{13}C-labeled aromatic ring were obtained from Cambridge Isotope Laboratories (Tewksbury, MA, USA). Vanillin with ^{13}C-labeled methoxy group and [2-^{13}C]-*tert*-butyl bromoacetate were purchased from Taiyo Nippon Sanso (Tokyo, Japan). Other laboratory chemicals were obtained from Wako Pure Chemical Industries (Osaka, Japan), nacalai tesque (Kyoto, Japan), and Tokyo Chemical Industry (Tokyo, Japan).

Preparation of TrCBM1

The CBM1 of *T. reesei* Cel7A (accession number: CAH10320) was expressed and purified as described previously [17]. *E. coli* BL21 (DE3) was first subjected to heat shock transformation with plasmids for the expression of His-tagged *Tr*CBM1 fused with GFP, hereafter referred to as His tag–*Tr*CBM1–GFP fusion protein. LB medium was inoculated with the transformant, followed by incubation at 37 °C and 200 rpm until the OD_{600} reached 1.2. Protein expression was induced using 1 mM isopropyl β-1-thiogalactopyranoside, and the bacteria were incubated at 37 °C and 200 rpm for 5 h. Following centrifugation and sonication, His tag–*Tr*CBM1–GFP was isolated from the supernatant and purified via Ni affinity chromatography followed by anion exchange chromatography. The His tag and GFP region were removed using enterokinase (New England Bio Labs, MA, USA) and thrombin (GE Healthcare, IL, USA), respectively. The obtained *Tr*CBM1 was concentrated in 50 mM acetic acid-d_4 buffer prepared with

D_2O using a Vivaspin turbo ultrafiltration device (Sartorius, Göttingen, Germany). We skipped the His tag cleavage process to prepare His tag– *Tr*CBM1 for adsorption analysis. The His tag– *Tr*CBM1 was purified via Ni affinity and anion exchange chromatography.

Synthesis of β- O -4 lignin oligomer model compounds

β-*O*-4 lignin oligomer models were synthesized as described in our previous work using a modified protocol originally developed by Katahira et al [18, 25]. Vanillin (**1**) was dissolved in acetone, and the solution was refluxed for 1.5 h in the presence of KI, K_2CO_3, and *t*-butyl-2-bromoacetate to obtain *t*-butoxycarbonylmethyl vanillin (**2**). The monomer (**2**) was dissolved in anhydrous THF, and 1.5 M LDA was added dropwise to the solution under a nitrogen atmosphere over 30 min while maintaining the temperature at − 30 °C. The solution was stirred for 30 min at − 30 °C and for an additional 1.5 h at 0 °C–5 °C to obtain the polymerized oligomer (**3**). Oligomer **3** was dissolved in *t*-butanol containing $NaBH_4$. Methanol was added to the mixture stepwise over 1 h with constant stirring at 65 °C–70 °C, and the mixture was stirred for 6 h to obtain β-*O*-4 lignin oligomer model compound **4**. The synthesized lignin models were separated into long- and short-chain fractions via silica gel chromatography. Sequential elution was performed using ethyl acetate, 5:1 (v/v) ethyl acetate/methanol, and 2:1 (v/v) ethyl acetate/methanol. Three different ^{13}C-labeled lignin model compounds were synthesized for CSP analysis. Vanillin with ^{13}C-labeled aromatic rings (**1**$^{(Ar)}$) and [2-^{13}C]-*tert*-butyl bromoacetate were used as starting materials to synthesize a lignin model with ^{13}C-labeled aromatic rings and β positions (**4**$^{(Arβ)}$). Similarly, vanillins with ^{13}C-labeled carbonyl and methoxy group were used to synthesize a lignin model with ^{13}C-labeled α positions (**4**$^{(α)}$) and a model with ^{13}C-labeled methoxy groups (**4**$^{(m)}$). The lignin models were characterized using 2D ^{1}H-^{13}C HSQC NMR and SEC. SEC was performed using three TSKgel SuperMultipore HZ-M columns (Tosho, Tokyo, Japan) on a Shimadzu instrument equipped with an LC-20AD pump and an SPD-M20A diode array detector (Shimadzu, Kyoto, Japan). THF was used for elution at a flow rate of 0.35 mL/min at 40 °C.

NMR spectroscopy and CSP analysis

All NMR spectra were recorded at 298 K on a Bruker Avance III HD 600 spectrometer equipped with a cryogenic probe and a Z-gradient (Bruker BioSpin, MA, USA). The instrument was controlled using Bruker Topspin NMR software version 3.5. For CSP analysis, 50 µM of the ^{13}C-labeled lignin models with ^{13}C-labeled aromatic rings and β positions (**4**$^{(Arβ)}$), α positions (**4**$^{(α)}$), and methoxy groups (**4**$^{(m)}$) were individually dissolved in 50 mM acetic acid-d_4 buffer, which was prepared with D_2O (pD 5.0) and 10% (v/v) DMSO-d_6. Each NMR sample had a volume of 250 µL and contained 20 µM 2,2-dimethyl-2-silapentane-5-sulfonic acid as an internal standard. To identify the *Tr*CBM1 binding sites in the ^{13}C-labeled lignin models, changes in chemical shift ($\Delta\delta_{C,H}$, ppm) were calculated by comparing chemical shift values in the edited ^{1}H-^{13}C HSQC spectra of the ^{13}C-labeled lignin models in the presence and absence of 200 and 350 µM *Tr*CBM1. CSP analysis of both the long- and short-chain lignin models was

performed to evaluate the effect of molecular weight on *Tr*CBM1 binding to the lignin models. Assignment of the lignin model signals was based on our previous report [18].

Adsorption experiment of lignin models with TrCBM1

The Langmuir adsorption isotherm model was applied to evaluate the *Tr*CBM1 binding affinities of the lignin models. Each unlabeled long- and short-chain lignin model was dissolved in 50 mM acetic acid buffer (pH 5.0) containing 3,200 μg/mL of *Tr*CBM1 with a His tag (His tag– *Tr*CBM1). The lignin models were prepared in 1.5 mL micro-tubes at concentrations of 10, 20, 30, 50, 80, 100, 150, and 200 μM. The total volume of each solution was 50 μL. The sample solutions were incubated at 50 °C for 60 min with constant shaking at 1,000 rpm using a Comfort thermomixer (Eppendorf, Hamburg, Germany). To each solution, 100 μL of a 50% (v/v) suspension of cOmplete His Tag Purification Resin (Roche, Basel, Switzerland) in the same buffer was added. The mixtures were resuspended by vortexing for 10 s and centrifuged at 500 g for 10 s. The free lignin model concentrations in the supernatants were determined by measuring absorbance at 280 nm using a BioSpec-nano spectrophotometer (Shimadzu, Kyoto, Japan). The quantity of each lignin model bound to *Tr*CBM1 was calculated by subtracting the amount of the free lignin model compound from the initially loaded amount. Experiments were conducted in duplicate, and the average values were used to calculate the Langmuir affinity constants using Eq. (1).

$$\Gamma_C = \Gamma_{max}[K_L C/(1 + K_L C)], \quad (1)$$

where Γ_C is the amount of the bound lignin model, Γ_{max} is the amount of the lignin model bound to His tag– *Tr*CBM1 at saturation, K_L is the Langmuir affinity constant, and C is the concentration of the free lignin model in the supernatant. A solution containing only His tag– *Tr*CBM1 without a lignin model was used as a blank.

Abbreviations

CSP
Chemical shift perturbation; DMSO:Dimethyl sulfoxide; DP:Degree of polymerization; EXSY:Exchange spectroscopy; G:Guaiacyl; GFP:Green fluorescent protein; His tag:Histidine tag; HMBC:Heteronuclear multiple bond correlation; HSQC:Heteronuclear single-quantum correlation; LDA:Lithium diisopropylamide; LR-HSQMBC:Long-range heteronuclear single quantum multiple bond correlation; MALDI-TOF-MS:Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry; *M*w:Weight-average molecular weight; MWL:Milled wood lignin; NMR:Nuclear Magnetic Resonance; ROESY:Rotating-frame overhauser effect spectroscopy; S:Syringyl; SDS-PAGE:Sodium dodecyl sulfate-polyacrylamide gel electrophoresis; SEC:Size exclusion chromatography; THF:Tetrahydrofuran; TOCSY:Total correlation spectroscopy.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analyzed during this study are included in this manuscript and its Additional files.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

Y.T., T.N., M.K., T.W. designed the experiments. Y.T. performed the experiments. Y.T., K.K., T.N. analyzed and interpreted the NMR data. Y.T., T.W., T.N., M.K. wrote the paper.

All authors read and approved the final manuscript.

Acknowledgements

We thank Dr. Isamu Katsuyama for supporting chemical synthesis of ^{13}C-labeled lignin oligomer model compounds.

Funding

This work was supported by JSPS KAKENHI Grant Number JP18J20331, a collaboration program of RISH (M2-2) and joint usage/research programs of IAE (ZE30A-36, ZE31A-37).

References

1. Brunecky R, Subramanian V, Yarbrough JM, Donohoe BS, Vinzant TB, Vanderwall TA, Knott BC, Chaudhari YB, Bomble YJ, Himmel ME, Decker SR. Synthetic fungal multifunctional cellulases for enhanced biomass conversion. Green Chem. 2020;22:478–89.
2. Seiboth B, Ivanova C, Seidl-Seiboth V. *Trichoderma reesei*: A Fungal Enzyme Producer for Cellulosic Biofuels. In: Bernardes DS MA, editors. Biofuel Production-Recent Developments and Prospects. Rijeka: Intech; 2011. pp. 309–40.
3. Boraston AB, Bolam DN, Gilbert HJ, Davies GJ. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem J. 2004;382:769–81.
4. Palonen H, Tjerneld F, Zacchi G, Tenkanen M. Adsorption of *Trichoderma reesei* CBH I and EG II and their catalytic domains on steam pretreated softwood and isolated lignin. J Biotechnol. 2004;107:65–72.

5. Kraulis J, Clore GM, Nilges M, Jones TA, Pettersson G, Knowles J, Gronenborn AM. Determination of the three-dimensional solution structure of the C-terminal domain of cellobiohydrolase I from *Trichoderma reesei*. A study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing. Biochemistry. 1989;28:7241–57.
6. Rahikainen JL, Moilanen U, Nurmi-Rantala S, Lappas A, Koivula A, Viikari L, Kruus K. Effect of temperature on lignin-derived inhibition studied with three structurally different cellobiohydrolases. Bioresour Technol. 2013;146:118–25.
7. Rahikainen JL, Evans JD, Mikander S, Kalliola A, Puranen T, Tamminen T, Marjamaa K, Kruus K. Cellulase-lignin interactions-the role of carbohydrate-binding module and pH in non-productive binding. Enzyme Microb Technol. 2013;53:315–21.
8. Pakarinen A, Haven MO, Djajadi DT, Varnai A, Puranen T, Viikari L. Cellulases without carbohydrate-binding modules in high consistency ethanol production process. Biotechnol Biofuels. 2014;7:27.
9. Yu Z, Gwak KS, Treasure T, Jameel H, Chang HM, Park S. Effect of lignin chemistry on the enzymatic hydrolysis of woody biomass. ChemSusChem. 2014;7:1942–50.
10. Huang C, He J, Min D, Lai C, Yong Q. Understanding the Nonproductive Enzyme Adsorption and Physicochemical Properties of Residual Lignins in Moso Bamboo Pretreated with Sulfuric Acid and Kraft Pulping. Appl Biochem Biotechnol. 2016;180:1508–23.
11. Nakagame S, Chandra RP, Kadla JF, Saddler JN. Enhancing the enzymatic hydrolysis of lignocellulosic biomass by increasing the carboxylic acid content of the associated lignin. Biotechnol Bioeng. 2011;108:538–48.
12. Guo F, Shi W, Sun W, Li X, Wang F, Zhao J, Qu Y. Differences in the adsorption of enzymes onto lignins from diverse types of lignocellulosic biomass and the underlying mechanism. Biotechnol Biofuels. 2014;7:38.
13. Yang Q, Pan X. Correlation between lignin physicochemical properties and inhibition to enzymatic hydrolysis of cellulose. Biotechnol Bioeng. 2016;113:1213–24.
14. Dien BS, Miller DJ, Hector RE, Dixon RA, Chen F, McCaslin M, Reisen P, Sarath G, Cotta MA. Enhancing alfalfa conversion efficiencies for sugar recovery and ethanol production by altering lignin composition. Bioresour Technol. 2011;102:6479–86.
15. Papa G, Varanasi P, Sun L, Cheng G, Stavila V, Holmes B, Simmons BA, Adani F, Singh S. Exploring the effect of different plant lignin content and composition on ionic liquid pretreatment efficiency and enzymatic saccharification of *Eucalyptus globulus* L. mutants. Bioresour Technol. 2012;117:352–9.
16. Zuiderweg ERP. Mapping protein-protein interactions in solution by NMR Spectroscopy. Biochemistry. 2002;41:1–7.
17. Tokunaga Y, Nagata T, Suetomi T, Oshiro S, Kondo K, Katahira M, Watanabe T. NMR Analysis on Molecular Interaction of Lignin with Amino Acid Residues of Carbohydrate-Binding Module from *Trichoderma reesei* Cel7A. Sci Rep. 2019;9:1977.

18. Tokunaga Y, Nagata T, Suetomi T, Oshiro S, Kondo K, Katahira M, Watanabe. T. Holzforschung, submitted.
19. Linder M, Mattinen ML, Kontteli M, Lindeberg G, Stahlberg J, Drakenberg T, Reinikainen T, Pettersson G, Annila A. Identification of functionally important amino acids in the cellulose-binding domain of *Trichoderma reesei* cellobiohydrolase I. Protein Sci. 1995;4:1056–64.
20. Sun S, Huang Y, Sun R, Tu M. The strong association of condensed phenolic moieties in isolated lignins with their inhibition of enzymatic hydrolysis. Green Chem. 2016;18:4276–86.
21. Suzuki S, Green PG, Bumgarner RE, Dasgupta S, Goddard WA 3rd, Blake GA. Benzene forms hydrogen bonds with water. Science. 1992;257:942–5.
22. Sipponen MH, Rahikainen J, Leskinen T, Pihlajaniemi V, Mattinen ML, Lange H, Crestini C, Osterberg M. Structural changes of lignin in biorefinery pretreatments and consequences to enzyme-lignin interactions. Nord Pulp Pap Res J. 2017;32:550–71.
23. Pareek N, Gillgren T, Jonsson LJ. Adsorption of proteins involved in hydrolysis of lignocellulose on lignins and hemicelluloses. Bioresour Technol. 2013;148:70–7.
24. Mattinen ML, Linder M, Teleman A, Annila A. Interaction between cellohexaose and cellulose binding domains from *Trichoderma reesei* cellulases. FEBS Lett. 1997;407:291–96.
25. Katahira R, Kamitakahara H, Takano T, Nakatsubo F. Synthesis of β-*O*-4 type oligomeric lignin model compound by the nucleophilic addition of carbanion to the aldehyde group. J Wood Sci. 2006;52:255–60.

Figures

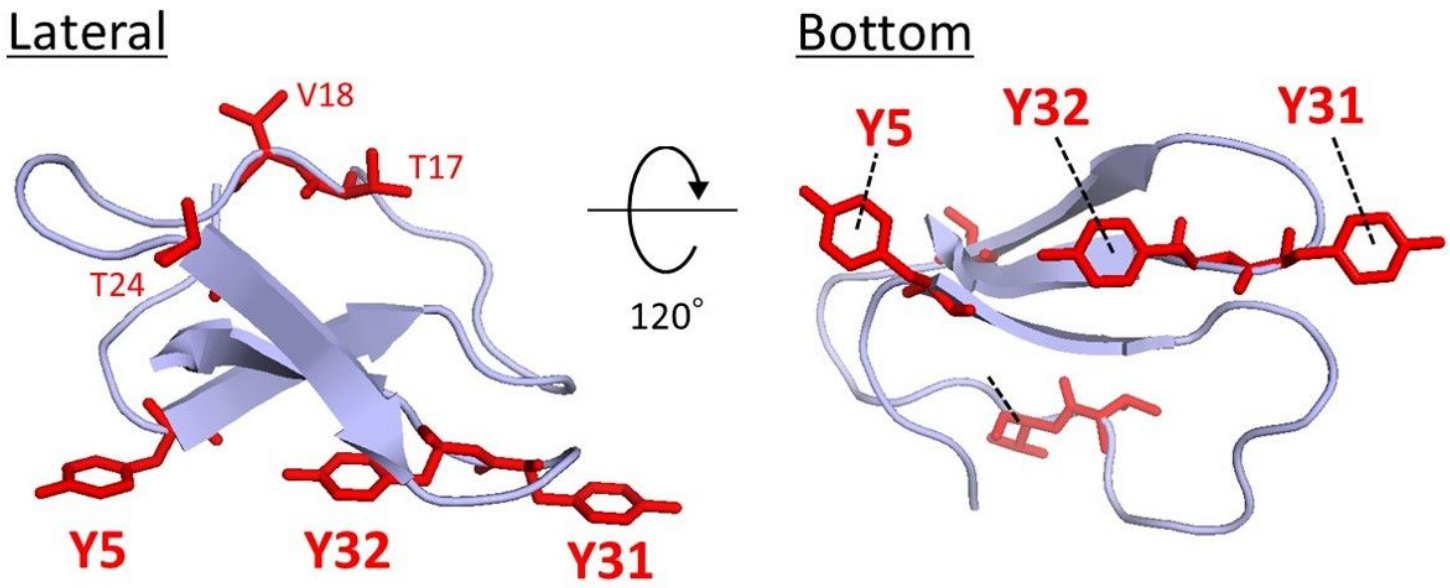

Figure 1

Ribbon models of TrCBM1 viewed from the side and bottom. The structure of TrCBM1 has been determined through NMR analysis (PDB ID 2CBH) [5]. Red stick models of tyrosine residues (Y5, Y31, and Y32) and T17, V18, and T24 residues in the cleft are also displayed.

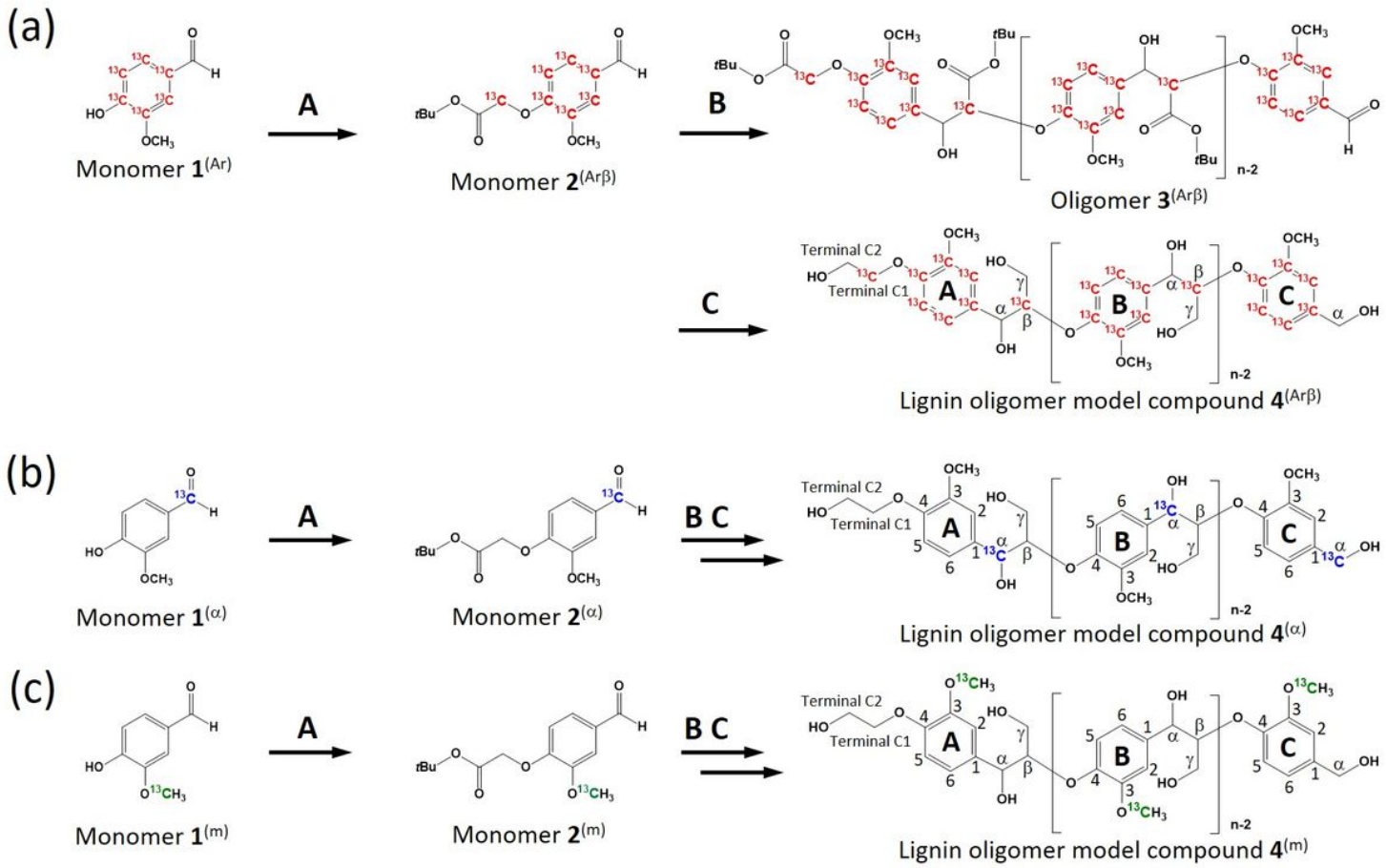

Figure 2

Synthesis of 13C-labeled β-O-4 lignin oligomer model compounds. 13C-labels in (a) aromatic rings and the β positions; (b) α positions; and (c) methoxy groups. A: Vanillin (1) is refluxed with BrCH2COOtBu, KI, and K2CO3 in acetone for 1.5 h. B: 2 is dissolved in anhydrous THF, followed by LDA addition with stirring at −30°C (1 h). The mixture is then stirred at 0°C–5°C for 1.5 h. C: 3 is dissolved in tBuOH with NaBH4, followed by the addition of MeOH and stirring at 65°C–70°C for 7 h.

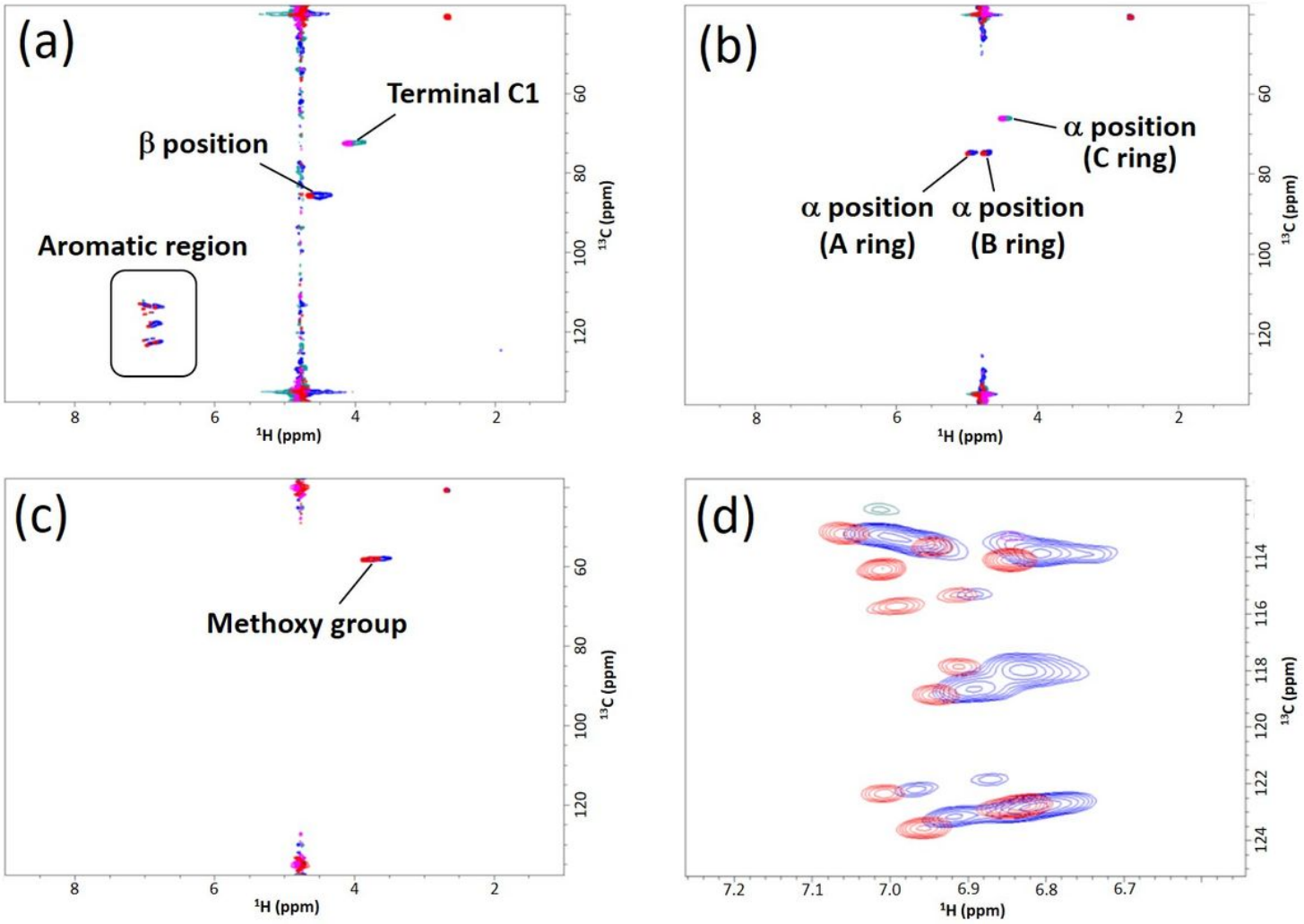

Figure 3

2D 1H-13C HSQC spectra of 13C-labeled β-O-4 lignin oligomer model compounds. Superimposed 2D 1H-13C HSQC spectra of the long- and short-chain models in 50 mM acetic acid-d4 buffer prepared using D2O (pD 5.0) and 10% (v/v) DMSO-d6 are shown in blue and red, respectively. 2D 1H-13C HSQC spectra of lignin oligomer models with (a) 13C-labeled aromatic rings and α positions (4(Arα)); (b) 13C-labeled α positions (4(α)); and (c) 13C-labeled methoxy groups (4(m)). (d) Magnified view of the aromatic region in (a).

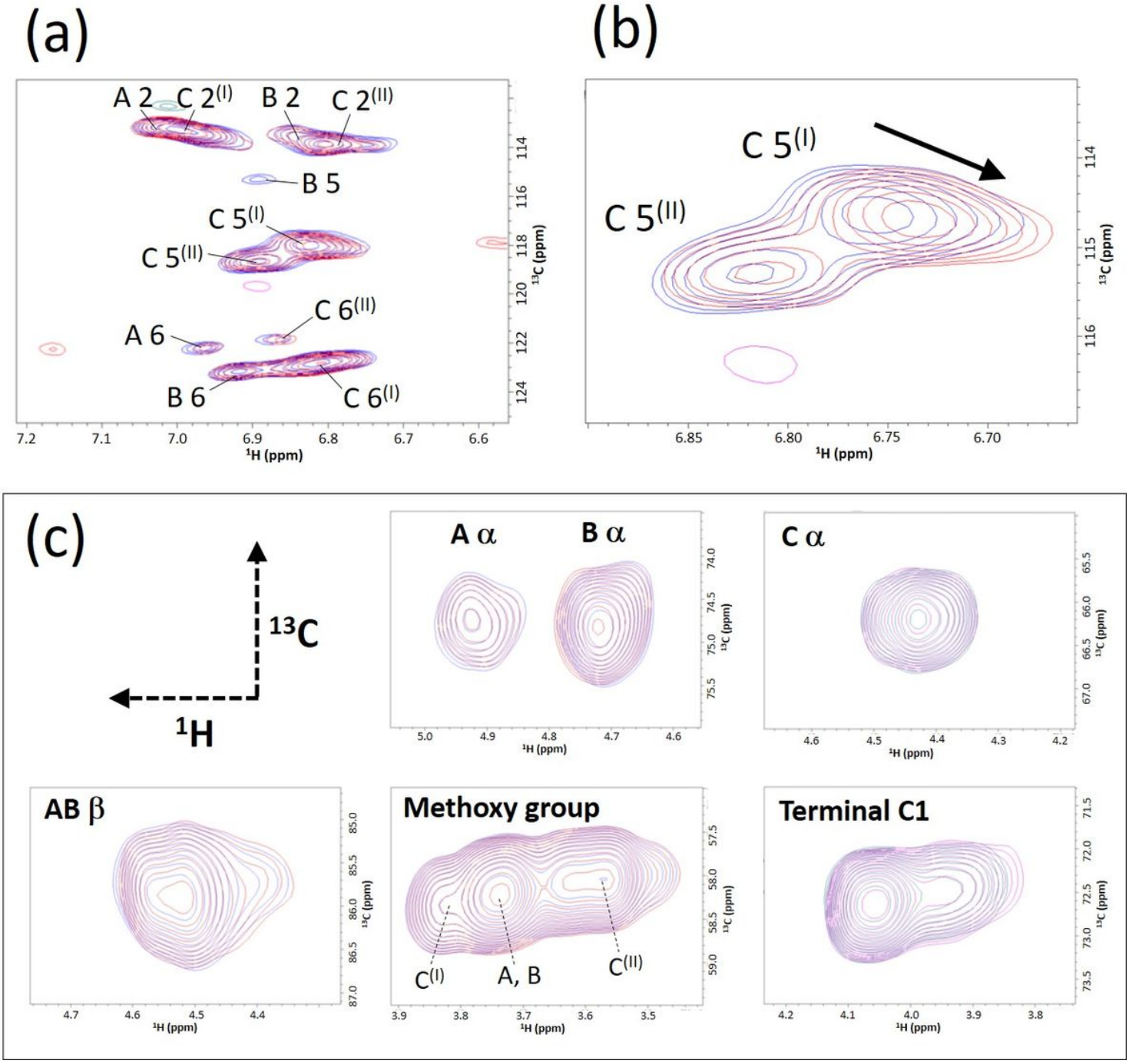

Figure 4

2D 1H-13C HSQC spectra of the long-chain 13C-labeled lignin oligomer models in the CSP experiment. (a) Superimposed 2D 1H-13C HSQC spectra of the aromatic region in compound 4(Ar) (50 M) in the presence (red) and absence (blue) of 350 M TrCBM1. (b) A magnified image of the HSQC signals from the C 5 positions. (c) Superimposed 2D 1H-13C HSQC spectra of aliphatic regions and methoxy group in 4(Ar), 4(), and 4(m) (50 M) in the presence (red) and absence (blue) of 350 M TrCBM1.

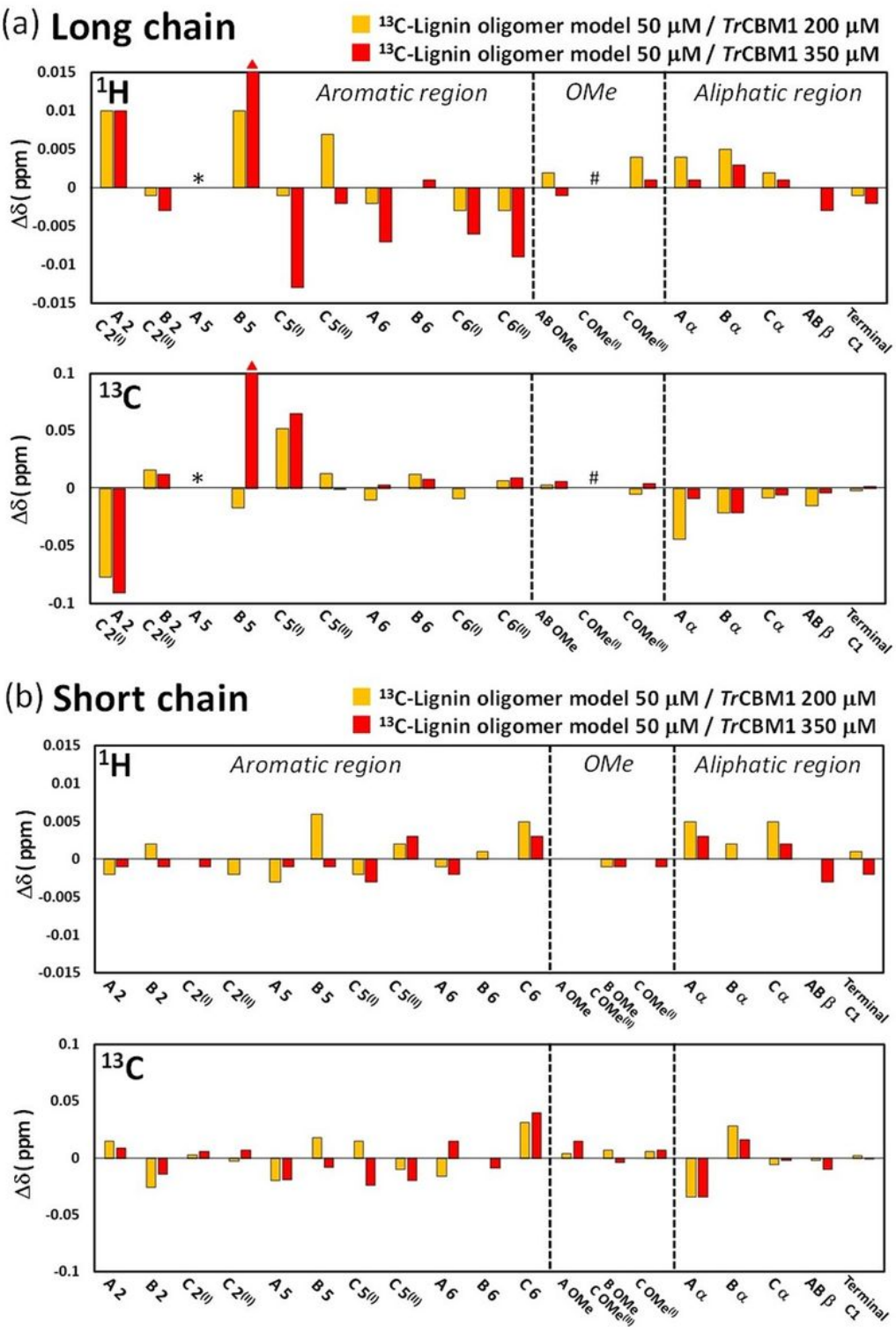

Figure 5

Changes in the chemical shift (□□) observed in the CSP analysis. The CSP analysis of the 13C-labeled lignin oligomer models (50 □M) was performed in the presence of 200 and 350 □M of TrCBM1 in 50 mM acetic acid-d4 buffer prepared using D2O (pD 5.0) and 10% (v/v) DMSO-d6. The □□ values of signals from the (a) long- and (b) short-chain lignin oligomer models in the CSP experiment. The upper panels in (a) and (b) show □□ on the 1H-axis. The lower panels show □□ on the 13C-axis. Line broadening of the NMR

signals from B 5 in the long-chain lignin models was observed in the presence of 350 M TrCBM1. Positions that generated extremely low-intensity signals in the absence of TrCBM1 are indicated by asterisks (*). Positions that generated overlapping signals are indicated by # symbols.

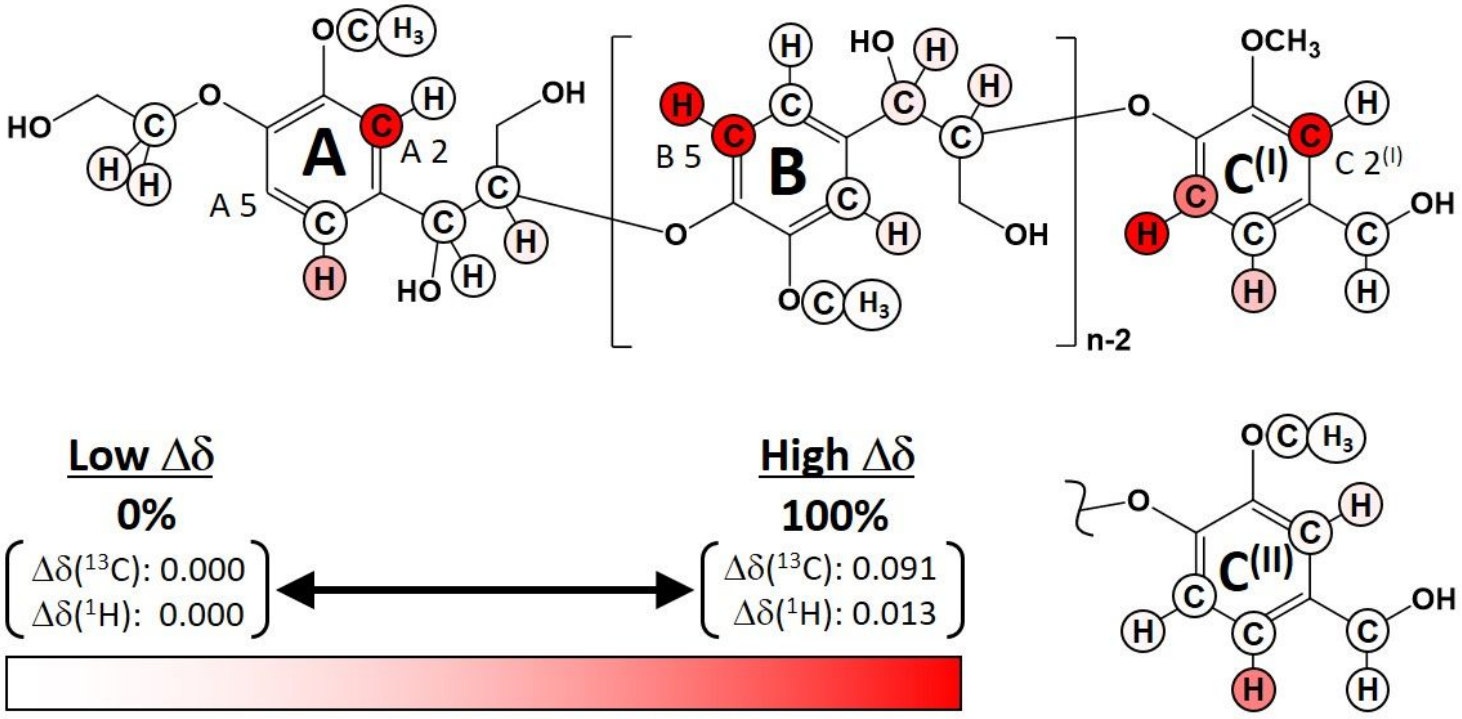

Figure 6

Mapping of lignin model binding sites with TrCBM1. The s observed in CSP experiment using the long-chain lignin model with 350 M of TrCBM1 are shown by red gradation which corresponds to the relative intensities of s. Differently aligned terminal units of two different configurations, C(I) and C(II), are exhibited. In the A 2 and C 2(I) positions, the value is the same because their 2D HSQC signals were not distinguishable due to overlapping.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.docx