



**HAL**  
open science

## Comparative analysis of molecular weight determination techniques for cellulose oligomers

Wei Li, Gérard Mortha, Issei Otsuka, Yu Ogawa, Yoshiharu Nishiyama

### ► To cite this version:

Wei Li, Gérard Mortha, Issei Otsuka, Yu Ogawa, Yoshiharu Nishiyama. Comparative analysis of molecular weight determination techniques for cellulose oligomers. *Cellulose*, 2023, 30 (13), pp.8245-8258. 10.1007/s10570-023-05447-7. hal-04258080v2

**HAL Id: hal-04258080**

**<https://hal.science/hal-04258080v2>**

Submitted on 25 Oct 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Comparative analysis of molecular weight determination techniques for cellulose oligomers

Wei Li · Gérard Mortha · Issei Otsuka ·  
Yu Ogawa · Yoshiharu Nishiyama

Received: date / Accepted: date

**Abstract** Molecular weight information is essential for comprehending the chemical and physical properties of cellulose. However, traditional methods used to analyze high-molecular-weight cellulose are often unsuitable for cellulose oligomers. In this study, we emphasize the influence of molecular weight distribution on the determination of molecular weight using four characterization methods: liquid and solid-state nuclear magnetic resonance (NMR) spectroscopy, matrix-assisted laser desorption/ionization mass spectrometry (MALDI MS), and size exclusion chromatography (SEC). These techniques were compared using two cellulose oligomer fractions with different molecular weight distributions. Liquid-state NMR was a reliable method for determining the number-average molecular weight but did not provide information on the molecular weight distribution. MALDI MS was more sensitive to the low molecular weight range, while SEC was a preferred technique for cellulose oligomers with relatively higher molecular weight. Among the solution-based characterization techniques, carbanilation was the preferred derivatization method over nitration for its higher scattering power and the increase in molar mass. MALDI MS revealed that cellulose molecules exhibited different degrees of substitution through the same carbanilation reaction, which may explain errors in molecular weight determination by SEC. Our study highlights the importance of considering molecular weight distribution when characterizing cellulose oligomers and demonstrates the strengths and limitations of different techniques for this purpose.

**Keywords** Cellulose · Oligomers · Molecular weight

---

W. Li  
School of Materials Science and Engineering, Beijing Institute of Technology, 100081, Beijing,  
P.R. China  
Univ. Grenoble Alpes, CNRS, CERMAV, 38000 Grenoble, France  
E-mail: weili2021@bit.edu.cn

Y. Ogawa · Y. Nishiyama · I. Otsuka  
Univ. Grenoble Alpes, CNRS, CERMAV, 38000 Grenoble, France

G. Mortha  
CNRS, Grenoble INP, Institute of Engineering Univ. Grenoble Alpes, LGP2, 38000, Grenoble,  
France

## 1 Introduction

2 Cellulose oligomers are low-molecular-weight cellulose comprising repeat units  
3 of  $\beta$ -D-glucopyranose (Zweckmair et al., 2016), typically with a degree of  
4 polymerization (DP) ranging from 3 to 30. These oligomers are excellent model  
5 substrates for investigating the intricate physicochemical properties of cellulose,  
6 including chiral transfer across various scales in cellulose assemblies (Fittolani  
7 et al., 2022) and the conformation studies using theoretical calculations (Queyroy  
8 et al., 2004; Shen et al., 2009). Moreover, cellulose oligomers can serve as the  
9 building blocks for novel functional cellulose nano-objects with well-defined  
10 structures, synthesized via bottom-up self-assembly strategies. (Hiraishi et al.,  
11 2009; Kobayashi et al., 2000; Helbert and Sugiyama, 1998). Although  
12 mono-disperse oligomers can be produced by chemical synthesis, most of the  
13 available cellulose oligomers are generated via depolymerization of  
14 high-molecular-weight cellulose, resulting in a distribution of molecular weights.  
15 The molecular weight information is essential for gaining a better understanding of  
16 their structure-property relationship and identifying optimal applications.  
17 However, current analytical methods for measuring cellulose molecular weight are  
18 primarily optimized for high-molecular-weight cellulose, and oligomers have  
19 received limited attention in cellulose analytics to date. (Oberlerchner et al., 2015).

20  
21 The molecular weight distribution of high-molecular-weight cellulose is  
22 typically quantified using well-established solution-based methods such as  
23 viscometry, light scattering (LS), and size-exclusion chromatography (SEC)  
24 (Oberlerchner et al., 2015). However, these methods require the dissolution of  
25 cellulose, which can be accomplished through direct dissolution in cellulose  
26 solvents or derivatization to make it soluble in common solvents. Among the  
27 available derivatization methods, carbanilation offers distinct advantages over  
28 other methods such as nitration and acetolysis, as it avoids potential hydrolysis of  
29 cellulose in acidic conditions. Additionally, cellulose carbonylate boasts a large  
30 refractive index increment ( $dn/dc$ ) and excellent stability in common SEC eluents,  
31 making it a valuable tool for DP determination of cellulose (Dupont and Mortha,  
32 2004). Unfortunately, the use of phenyl isocyanate, a critical reagent in  
33 carbanilation, is prohibited in certain countries such as China due to anti-narcotics  
34 policies.

35  
36 The determination of the DP for high-molecular-weight cellulose often cannot  
37 be applied to cellulose oligomers. The commonly used viscometry method suffers  
38 from the low viscosity of the oligomer solutions (Evans and Wallis, 1989;  
39 Oberlerchner et al., 2015), while the Mark-Houwink equation contains two  
40 parameters that depend on the solute and solvent properties. To date, no  
41 established Mark-Houwink parameters exist for cellulose oligomers. Static light  
42 scattering (SLS) is a molecular weight determination method based on detecting  
43 light scattered by solute molecules. Multi-angle light scattering (MALS) analysis,  
44 a variation of SLS, determines the absolute molecular weight of the sample  
45 without calibration. For DP determination of cellulose oligomers using the light  
46 scattering method, the primary challenge is the small scattering intensity resulting  
47 from their small molecular weight. Additionally, it is challenging to obtain a set of  
48 well-characterized cellulose oligomers with a narrow DP distribution for column

1 calibration. To address this issue, Zweckmair et al (2016) employed  
2 high-performance liquid chromatography to isolate monodisperse peracetylated  
3 cellulose oligomers. The researchers also demonstrated the utility of a facile  
4 high-performance thin-layer chromatography (HPTLC) technique for efficiently  
5 screening the purity and dispersity of peracetylated cellulose oligomers with a DP  
6 of up to 20. These isolated oligomers can be used to calibrate the oligomeric region  
7 in SEC.

8  
9 Although the molecular weight determination of cellulose oligomers with  
10 specific molecular weight has been previously investigated, little attention has been  
11 given to how the molecular weight distribution of oligomers affects the molecular  
12 weight determination by different characterization techniques. In this study, we  
13 compared four molecular weight determination methods for cellulose oligomers,  
14 namely, liquid and solid-state nuclear magnetic resonance (NMR) spectroscopy,  
15 SEC, and matrix-assisted laser desorption/ionization mass spectrometry (MALDI-  
16 MS). To assess the effects of molecular weight distribution, we used two oligomer  
17 fractions with different average DPs, prepared according to the method developed  
18 by Isogai and Usuda in 1991 with small modifications. Isogai and Usuda reported  
19 two nearly monodisperse cellulose oligomer fractions, namely fraction A and fraction  
20 B, with DP around 15 and 7, respectively (Isogai and Usuda, 1991). In their seminal  
21 work, the DP distributions of both fractions were characterized using intrinsic  
22 viscosity, NMR, and SEC methods. In 2021, Jiang et al. re-examined the preparation  
23 protocol of these two fractions (Jiang et al., 2021). While the DP information  
24 of fraction B was confirmed using techniques such as MALDI-TOF MS, NMR,  
25 small-angle X-ray scattering (SAXS), and small-angle neutron scattering (SANS).  
26 However, the characterization of fraction A was excluded due to its poor solubility  
27 in common solvents. The fractions A and B represent cellulose oligomers exhibiting  
28 a relatively high molar mass and a broad molar mass distribution, and a lower  
29 molar mass with a narrower distribution, respectively. The differences in physical  
30 properties, such as molecular size and solubility in common solvents, between  
31 fraction A and B suggest that these two types of cellulose oligomers may exhibit  
32 different behaviors during molecular weight analysis. Understanding the precise  
33 molecular weight and distribution of cellulose oligomers is a task that presents  
34 numerous intricacies and challenges. This research seeks to shed light on these  
35 complexities rather than providing a definitive solution.

## 36 **Materials and methods**

### 37 **Materials**

38 Microcrystalline cellulose (MCC) Vivapur®101 was derived from purified wood  
39 pulp and purchased from JRS Pharma, Germany. Chemical reagents including  
40 dimethylformamide (DMF,  $\geq 99.5\%$ ), dimethyl sulfoxide (DMSO,  $\geq 99.0\%$ ),  
41 tetrahydrofuran (THF,  $\geq 99.8\%$ ) and pyridine ( $\geq 99.5\%$ ) were bought from  
42 Fisher Chemicals, UK. Fuming nitric acid ( $\geq 90.0\%$ ) was obtained from Acros  
43 Organics, France, while sodium carbonate ( $Na_2CO_3$ , 99.8%) and phosphorus  
44 pentoxide ( $\geq 98.0\%$ ) were purchased from Honeywell Fluka, France. Additionally,  
45 isopropanol ( $\geq 99.8\%$ ) and methanol ( $\geq 99.8\%$ ) were obtained from Biosolve

1 Chimie, Netherlands. Deuterated DMSO ( $\geq 99.9\%$ ) was sourced from Eurisotop,  
2 France, while phosphoric acid ( $H_3PO_4$ , 85%) and phenylisocyanate ( $\geq 98.0\%$ ) were  
3 obtained from Sigma-Aldrich, U.S.A.

#### 4 Preparation of cellulose oligomers

5 A mixture of 3 wt% MCC and 83 wt%  $H_3PO_4$  solution was subjected to  
6 hydrolysis at 23 °C for 8 weeks. The resulting solution was then mixed with an  
7 equal weight of water and allowed to stand for 24 hours, and the water-insoluble  
8 precipitate was collected as fraction A. To obtain fraction B, a 3 wt% MCC/83  
9 wt%  $H_3PO_4$  solution was stored at 23 °C for 6 weeks and then mixed with an  
10 equal weight of water at 23 °C. After 24 hours of decantation, the supernatant  
11 was mixed with 3 times its volume of isopropanol, resulting in the formation of an  
12 isopropanol-insoluble fraction B. Both fractions A and B were subsequently washed  
13 with their respective solvents using centrifugation until the pH of the supernatant  
14 reached neutral.

#### 15 Derivatization of cellulose

16 *Nitration.* Nitrating acid mixture was prepared by mixing 12.12 g of phosphorus  
17 pentoxide with 30.00 g fuming nitric acid in an ice bath. Dry cellulose specimens  
18 (10 mg) were then added to the prepared nitrating acid mixture (400 mg). The  
19 nitration proceeded for 20 minutes at room temperature (approximately 23 °C).  
20 Subsequently, the reaction mixture was neutralized with sodium carbonate in the  
21 presence of ice, and then washed successively with distilled water and ethanol using  
22 vacuum suction filtration.

23 *Carbanilation.* Approximately 20 mg (0.1 mmol) of cellulose samples was vacuum-  
24 dried overnight at 105 °C. Subsequently, the cellulose was mixed with 200  $\mu$ L dry  
25 pyridine. The resulting cellulose suspension was then combined with 100  $\mu$ L of  
26 phenylisocyanate (0.9 mmol, which corresponds to 3.0 molar equivalents to the  
27 -OH groups of cellulose). The reactions were kept at 90 °C for 48 hours and were  
28 stopped by the addition of 100  $\mu$ L methanol.

#### 29 Characterization

30 *Liquid and solid-state NMR.* The  $^1H$  NMR spectra were recorded on a Bruker  
31 Avance III 400 spectrometer at 298 K. The relaxation delay was set as 10 s, and  
32 the number of scans was 16 times. The fraction B was dissolved in deuterated  
33 DMSO at a concentration of approximately 20 mg/mL to facilitate the  $^1H$  NMR  
34 analysis. The solid-state  $^{13}C$  cross-polarization magic-angle spinning (CP/MAS)  
35 NMR spectra were acquired with a Bruker Avance III 400 spectrometer operating  
36 at 100 MHz for  $^{13}C$ . The MAS rate was set to 12 kHz, the sweep width to 29761  
37 Hz, the recycle delay to 2 s, and the cross-polarization contact to 2 ms. The  $^{13}C$   
38 chemical shifts were calibrated with the glycine carboxyl group at 176.03 ppm.  
39 Prior to conducting the solid-state  $^{13}C$  CP/MAS NMR experiments, the cellulose

specimens were rehydrated. For fraction A, the cellulose was immersed in water and then centrifuged to remove excess water. However, as fraction B is partially soluble in water, this method might lead to a partial loss of low-molecular-weight oligomers. Therefore, fraction B was rehydrated in a desiccator at 97% relative humidity for one week.

*Mass spectrometry.* The carbanilated fractions A and B in DMF, and fraction B in DMSO, each at a concentration of approximately  $0.02 \text{ mol} \cdot \text{L}^{-1}$ , were prepared for MALDI-TOF measurements. The measurements were performed on a Bruker Daltonics Autoflex Speed apparatus using 2,5-dihydroxybenzoic acid (DHB) as a matrix. High-resolution mass spectrometry was carried out using a Thermo Scientific LTQ Orbitrap XL (quadrupole hybrid with orthogonal acceleration time-of-flight) mass spectrometer. The integration of peak areas belonging to the same DP was performed using an in-house program.

*Size exclusion chromatography.* The SEC measurements were performed at  $40^\circ \text{C}$  using a Viskotek TDAMax system, consisting of a VE 2001 GPC Solvent/Sample module, a UV detector model 2501 with deuterium lamp with a UV wavelength of 280 nm, as well as TDA 302 Triple array detectors comprising an RI detector, a viscometer detector, and light scattering detectors. The light scattering detectors featured low-angle light scattering (LALS) at  $7^\circ$  and right-angle light scattering (RALS) at  $90^\circ$ , with a laser light wavelength of 670 nm. The SEC instrument was equipped with two series of columns. The first series of columns consisted of a PLGel precolumn and two PLGel MIXED-B LS columns, commonly used for cellulose derivatives in a wide range of molar masses from several thousands to several millions. The columns were run in 0.01M LiCl/DMF at a flow rate of  $1 \text{ mL} \cdot \text{min}^{-1}$ . The second series of columns included a Shodex GF-1G 7B pre-column, a Shodex GF-7 M HQ column (linear,  $7.6 \text{ mm} \times 300 \text{ mm}$ ; pore size, 20 nm; bead size,  $9 \mu\text{m}$ ; exclusion limit,  $4 \times 10^7$ ) and a Shodex GF-310 HQ column (linear,  $7.6 \text{ mm} \times 300 \text{ mm}$ ; pore size, 20 nm; bead size,  $5 \mu\text{m}$ ; exclusion limit,  $4 \times 10^4$ ). These columns were run in THF at a flow rate of  $0.4 \text{ mL} \cdot \text{min}^{-1}$ . A polystyrene standard sample (Mp 1460,  $\text{Đ}$  1.09) was used for calibration by a universal calibration method, where Mp referred to the molecular weight value at which the distribution has the highest frequency or intensity. Subsequently, another polystyrene standard sample (Mp 2900,  $\text{Đ}$  1.07) was subjected to calibration testing, yielding the following results: Mp = 2948, Mw = 2896, and  $\text{Đ}$  = 1.02. The  $\text{dn}/\text{dc}$  values were determined according to the reference (Ono et al., 2016). For the determination of  $\text{dn}/\text{dc}$ , cellulose carbanilates were thoroughly washed with methanol prior to the measurement. Specifically, the  $\text{dn}/\text{dc}$  values of carbanilated cellulose were  $0.131 \text{ cm}^3/\text{g}$  in 0.01M LiCl/DMF and  $0.165 \text{ cm}^3/\text{g}$  in THF for measuring the molecular weight. After carbanilation, pyridine and excess methanol in the mixtures were removed by a vacuum condenser. The precipitates were then dissolved in THF or 0.01M LiCl/DMF at a concentration of about 3 mg/mL for cellulose tricarbonylates and injected for the SEC test. For nitrated samples, cellulose trinitrates were dissolved in 0.01M LiCl/DMF at a concentration of about 3 mg/mL and injected for SEC testing. The first column set was used for the measurements shown in Fig. 4, while the second column set generated the other GPC data in Figs. 5-7 and Table 2. The OmniSEC software, an integrated program

1 that operated in conjunction with the SEC instrument, was used to perform a  
2 linear fit of log M relative to retention time.

### 3 Results and discussion

#### 4 Liquid-state $^1H$ NMR spectroscopy

5 First, we estimated the average DP of cellulose oligomers using liquid-state  $^1H$   
6 NMR spectroscopy. We investigated fraction B as it was readily soluble in common  
7 NMR solvents such as DMSO. Fig. 1 shows a  $^1H$  NMR spectrum of fraction B in  
8 deuterated DMSO. Peak assignments are according to the previous studies (Flugge  
9 et al., 1999; Bernet and Vasella, 2000; Sugiyama et al., 2000; Jiang et al., 2021). The  
10 peaks at 4.89 ppm and 4.39 ppm are attributed to the  $\alpha$ - and  $\beta$ -anomeric protons  
11 at the reducing ends, respectively. Those at 6.31 ppm and 6.65 ppm are assigned  
12 to the hydroxy protons of  $\alpha$ - and  $\beta$ -anomers of the reducing ends. The peaks at  
13 4.31 ppm and 4.33 ppm are the H1 and H1' protons of C1 of the non-reducing  
14 end and repeating residues (inset of Fig. 1). Given the quantitative measurement  
15 of  $^1H$  resonance intensity, the number-averaged DP of cellulose oligomers can be  
16 calculated as follows,

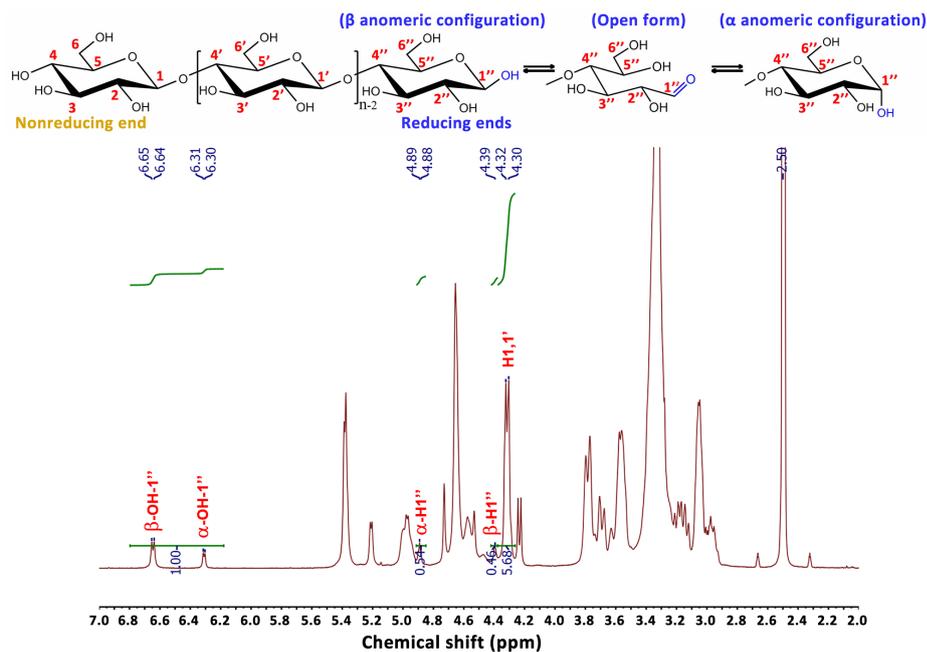
$$17 \quad DP_n = \frac{I_\alpha + I_\beta + I_i}{I_\alpha + I_\beta}$$

18 where  $I_\alpha$  and  $I_\beta$  are the integrated intensities of peaks corresponding to  
19 OH-1" or H1" of reducing ends, while  $I_i$  was the integrated intensity of peak  
20 assigned to H1 and H1'. A number-averaged DP is calculated as 6.7, identical for  
21 two estimations based on O-H and C-H signals.  
22

#### 23 Quantitative DP analysis by solid-state $^{13}C$ CP/MAS NMR

24 As aforementioned, fraction A is insoluble in common mono-component  
25 solvents, making it more difficult to characterize with liquid-state NMR  
26 spectroscopy. Solid-state NMR spectroscopy is thus a more straightforward  
27 method for Fraction A. However, the peak intensities in  $^{13}C$  CP/MAS NMR  
28 measurements are not quantitative due to the difference in the spin relaxation  
29 behavior and the cross polarization efficiency of each carbon. To overcome this  
30 problem, one can restore the quantitative nature of peak intensities of  $^{13}C$   
31 CP/MAS spectra by taking these kinetics effects into account as described in the  
32 section 2 of Supporting information. Here, we made such a correction to estimate  
33 the DPs of both fractions. We used hydrated oligomers conditioned in a humidity  
34 chamber of 97 %RH for the solid-state NMR analysis. The hydrated sample gives  
35 narrower and more resolved peaks without substantial change in chemical shifts  
36 than the freeze-dried one, as shown in Fig. S1, which is beneficial for the peak  
37 deconvolution analysis.

38 The C1 and C1' region of the rehydrated fraction A is deconvoluted into five  
39 peaks, as shown in Fig. 2. The two major peaks at 107.2 ppm and 105.0 ppm with  
40 the same integrated peak area are assigned to the C1 of the center and corner chains

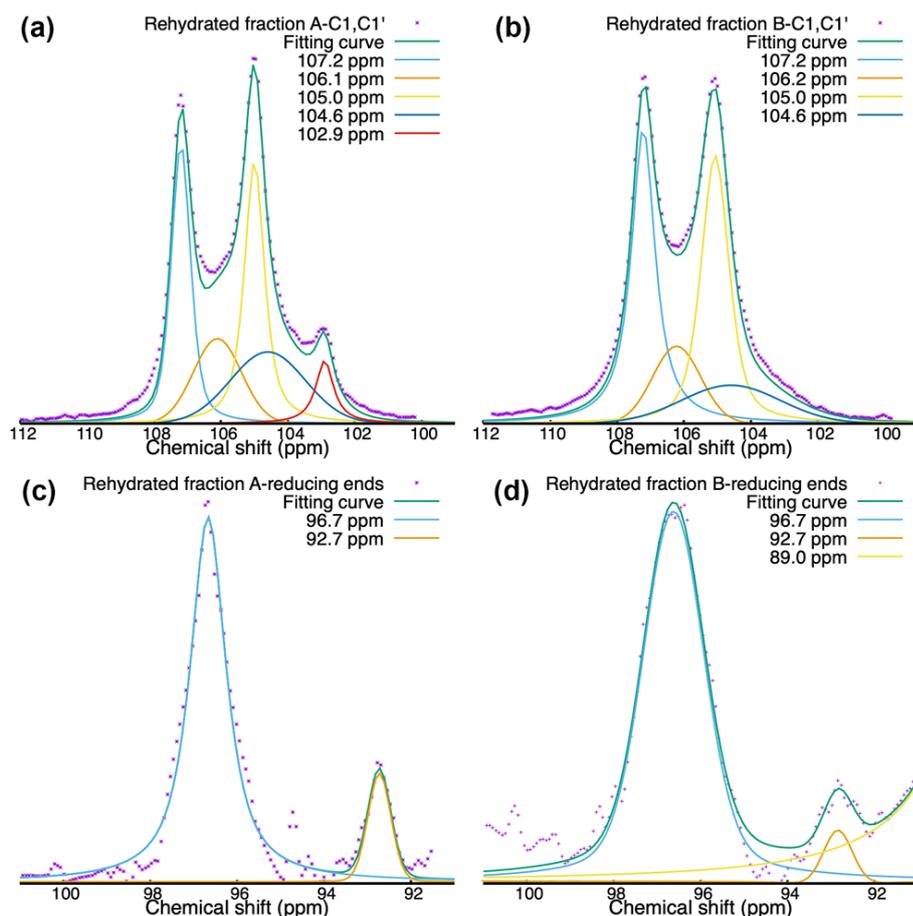


**Fig. 1**  $^1\text{H}$ -NMR spectrum of fraction B in deuterated DMSO. Characteristic chemical shifts corresponding to specific bonds are annotated on the chemical structure of the cellulose molecule.

**Table 1**  $^{13}\text{C}$  chemical shift assignments and the results of fitting the  $^{13}\text{C}$  CP dynamics for rehydrated fraction A and B.

Samples	Assignment	Shift (ppm)	$M_0$	$T_{1H}$ (ms)	$T_{CH}$ (ms)	$M_s$	$M_s/M_0$
Fraction A	C1&C1'	107.2	88012	27.2	0.1	81781	0.9
		106.1	61558	21.9	0.1	56187	0.9
		105.0	88068	27.0	0.1	81788	0.9
		104.6	72368	29.1	0.1	67563	0.9
	Reducing ends	102.9	22369	7.9	0.4	17266	0.8
		96.7	26583	11.5	0.1	22352	0.8
Fraction B	C1&C1'	92.7	9963	2.8	0.1	4905	0.5
		107.2	111665	13.7	0.2	96515	0.9
		106.2	53766	17.6	0.1	47986	0.9
		105.0	112017	13.6	0.2	96718	0.9
	Reducing ends	104.6	89153	2.5	0.1	39943	0.4
		96.7	38182	4.3	0.1	24045	0.6
		92.7	10084	3.2	0.1	5372	0.5

1 of the cellulose II crystal (Idström et al., 2016; Zuckerstätter et al., 2013; Kita et al.,  
 2 2020). The peaks at 106.1 ppm and 102.9 ppm are attributed to the crystalline  
 3 surface signals as they have much faster  $^1\text{H}$  T1 relaxation. The amorphous signals  
 4 correspond to the broad peak centered at 104.6 ppm. For the rehydrated fraction B,  
 5 the resonance region of C1 and C1' is deconvoluted into four peaks two crystalline  
 6 cellulose II peaks at 107.2 ppm and 105.0 ppm, one broad amorphous signal at  
 7 104.6 ppm, and only one crystalline surface signal at 106.2 ppm. The reducing end  
 8 C1 regions have only two peaks at 96.7 ppm and 92.7 ppm, corresponding to the  
 9  $\beta$ - and  $\alpha$ -conformers, respectively. While the two fractions share general spectral



**Fig. 2** (a, b) C1 & C1' region and (c, d) reducing end C1'' region of solid state  $^{13}\text{C}$  CP/MAS NMR spectra together with their spectral deconvolutions of rehydrated fraction A and B.

1 features, there are several differences between two spectra in both the C1-C1' and  
 2 reducing end regions. The two reducing end peaks of rehydrated fraction B are  
 3 broader than that of rehydrated fraction A. The surface peak at 102.9 ppm  
 4 is only visible in the fraction A but not in the fraction B. These differences may  
 5 arise from the different rehydration methods: the fraction A was immersed in bulk  
 6 water, while the Fraction B was conditioned in the saturated water vapor. Thus,  
 7 the fraction A would contain more water than the fraction B after rehydration.

8 Fig. S2-4 presents the result of peak fitting. In the CP experiment, three  
 9 constants of time are involved:  $T_{1H}$ , the proton spin-lattice relaxation time;  $T_{CH}$ ,  
 10 magnetization transfer time constant from  $^1\text{H}$  spin reservoir to  $^{13}\text{C}$  spin reservoir;  
 11 and  $T_{1C}$ ,  $^{13}\text{C}$  spin-lattice relaxation time. As expected from Fig. S2, three constants  
 12 of time are in a following order:  $T_{1C} \gg T_{1H} \gg T_{CH}$ . Since the  $^{13}\text{C}$  T1 (tens to  
 13 a hundred of seconds) is much longer than other two constants (in a millisecond  
 14 order), the rate of variation in  $^{13}\text{C}$  signal intensity ( $M_s$ ), as a function of a CP

1 contact time  $t$  is simplified as (Hill et al., 1994):

$$2 \quad M_S(t) = M_0 \times \left( e^{-\frac{t}{T_{1H}}} - e^{-\frac{t}{T_{CH}}} \right)$$

3 The calculated  $T_{1H}$ ,  $T_{CH}$ , and  $M_0$  are shown in Table 1. The  $M_s/M_0$  describes  
 4 the intensity decay at a given contact time with respect to the extrapolated  
 5 maximum signal intensity  $M_0$  at the CP contact time at 0 s. Then the number-  
 6 averaged DP is calculated as follows,

$$7 \quad DP_n = \frac{\sum M_0 (C1 \& C1') + \sum M_0 (\text{Reducing ends})}{\sum M_0 (\text{Reducing ends})}$$

8 Based on the spectra shown in Fig. 2 and the constants in Table 1, the  $DP_n$   
 9 values of fraction A and fraction B are estimated as 10.1 and 8.6, respectively. The  
 10  $DP_n$  value of fraction B obtained from solid-state NMR is likely to be  
 11 overestimated when compared to its liquid-state NMR counterpart. This  
 12 overestimation can be attributed to errors that arise from peak fitting during data  
 13 analysis.  
 14

## 15 Mass spectrometry

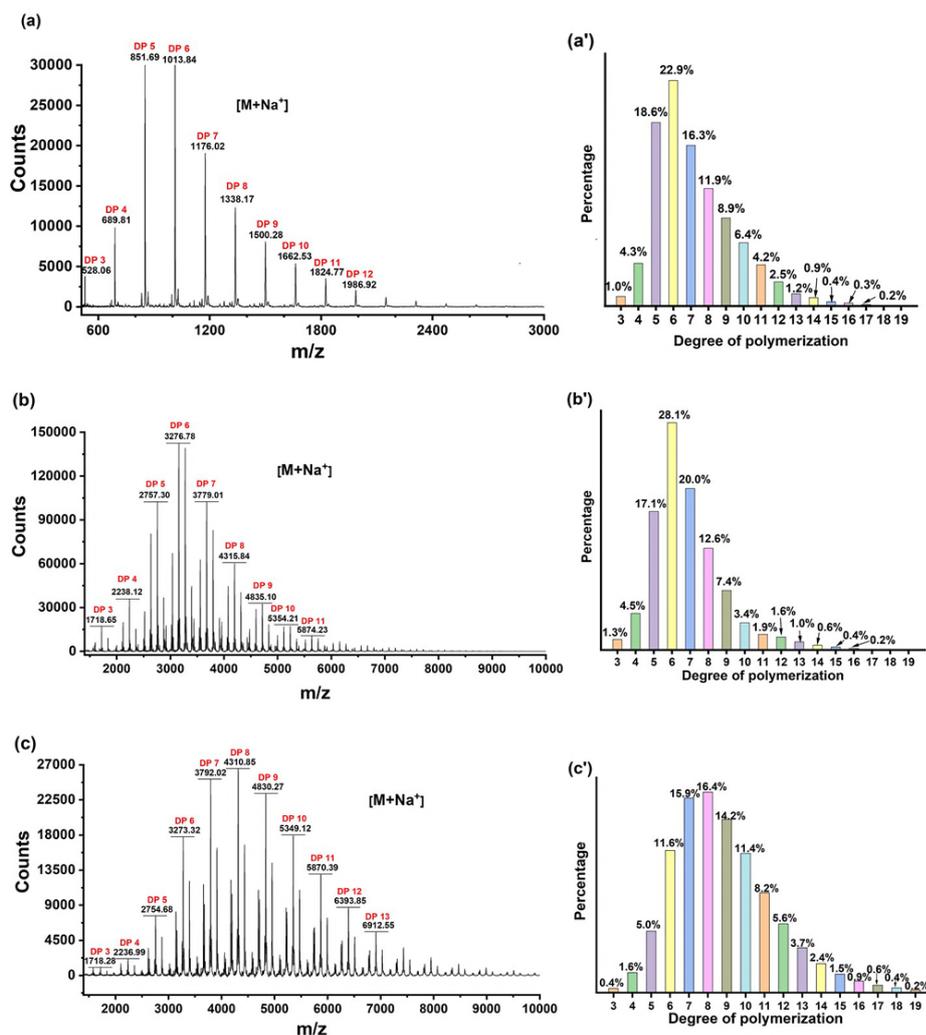
16 The MALDI-TOF MS spectroscopy was applied to the fractions A and B.  
 17 The fraction B could be directly subjected to the experiment after dissolving in  
 18 DMSO. The fraction A was insoluble in DMSO and thus was carbanilated before  
 19 the MS measurement. The fraction B was also carbanilated for comparison. Fig.  
 20 3a-c shows MALDI-TOF MS spectra of the unmodified fraction B, carbanilated  
 21 fraction B and carbanilated fraction A as a function of mass-to-charge ( $m/z$ ) ratio  
 22 in the positive-ion mode, respectively. In the spectrum of the unmodified fraction  
 23 B (Fig. 3a), the major peaks are separated by  $m/z$  of 162, corresponding to one  
 24 anhydroglucose repeating unit. The  $m/z$  values of major peaks are the sum of  
 25 the mass of one cellulose molecule and one sodium ion, expressed in the following  
 26 equation,

$$27 \quad M(n) = 162 \times n + 18 + 23$$

28 where  $n$  is the DP of the cellulose molecule. On the left side of the main peaks,  
 29 the small peaks with a decrease in mass of 18 from the main peaks are ascribed  
 30 to the dehydro-cellulose that may be produced from acid hydrolysis. On the right  
 31 side of the main peaks, the small peaks with a distance of 16 from the main peaks  
 32 are attributed to a cellulose molecule with a potassium ion. For spectra of the  
 33 carbanilated fraction B and fraction A, the main peaks are separated by  $m/z$  of 519,  
 34 corresponding to a tricarbanilted anhydroglucose repeating unit. Again, the  $m/z$   
 35 values of the main peaks are the sum of the mass of one tricarbanilated cellulose  
 36 molecule and one sodium ion, expressed as follows,

$$37 \quad M(n) = 519 \times n + 119 + 17 + 23$$

38 From this equation, it implies that there is one underivatized hydroxy group  
 39 in the cellulose molecule. The adjacent peaks with an increment of 119 from the



**Fig. 3** MALDI-TOF MS spectra (left) as a function of mass-to-charge ( $m/z$ ) ratio in the positive-ion mode and the percentage histogram (right) of different DP in (a, a') the fraction B, (b, b') the carbanilated fraction B and (c, c') the carbanilated fraction A.

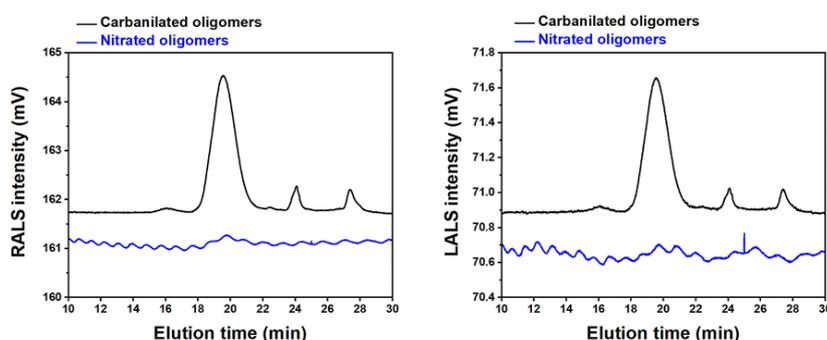
1 main peak are assigned to the fully carbanilated cellulose, and the adjacent peaks  
 2 on the left from the main peaks with a  $m/z$  smaller by 119 and  $119 \times 2$  should  
 3 correspond to cellulose molecules with two and three unmodified hydroxy groups,  
 4 respectively. These four peaks separated by a multiple of  $m/z$  of 119 all belong to  
 5 a cellulose molecule with the same DP value.

6 Based on the peak assignments, the sum of integrated peak areas that belong  
 7 to molecules with the same DP is calculated. Histograms of DP fractions are  
 8 shown in Fig. 3a'-c'. The hexamer (DP = 6) is predominant in the fraction B and  
 9 the carbanilated fraction B. The DP distribution ranges over the DPs between 3  
 10 and 17 for fraction B and 3-16 for carbanilated fraction B. The number-averaged

1 DP, weight-averaged DP, and  $\bar{D}$  value of the fraction B are calculated based on  
 2 the histogram, yielding  $DP_n = 7.2$ ,  $DP_w = 8.0$ , and  $\bar{D} = 1.10$ , respectively. The  
 3  $DP_n$  and  $DP_w$  of the carbanilated fraction B are 0.3 and 0.5 smaller than those of  
 4 the fraction B, respectively. This difference is likely due to different ionization  
 5 efficiency between cellulose oligomers and carbanilated cellulose oligomers. The  
 6 larger carbanilated cellulose molecule may have a lower ionization tendency than  
 7 the non-derivatized cellulose. The carbanilated fraction A gives a DP range of 3-19  
 8 with a predominant DP fraction at  $DP = 8.0$ . The  $DP_n$  and  $DP_w$  of carbanilated  
 9 fraction A are 8.8 and 9.7, respectively. The calculated  $\bar{D}$  values, 1.09, based on  
 10 the MS result are almost the same for the carbanilated fraction A and  
 11 carbanilated fraction B.

12

13 Size exclusion chromatography



**Fig. 4** Light scattering signals of carbanilated and nitrated cellulose oligomers that were hydrolyzed at room temperature for 5 weeks, and the eluent was 0.01M LiCl/DMF.

14 We used two derivatization methods, nitration and carbanilation, before  
 15 subjecting the oligomers to SEC measurements. Cellulose nitrates tended to form  
 16 aggregates in THF (Fig. S5), so 0.01M LiCl/DMF was used as the eluent. The  
 17 carbanilated and nitrated cellulose oligomers were compared using the light  
 18 scattering signals of SEC, as shown in Fig. 4. The RALS and LALS signals of the  
 19 carbanilated cellulose oligomers are detectable with a large signal-to-noise (S/N)  
 20 ratio. The main peak around 20 min is attributed to the carbanilated cellulose  
 21 oligomeric molecules, and a couple of peaks at larger elution times are assigned to  
 22 the side products, phenylisocyanate dimer (24.1 min) and methyl carbanilates  
 23 (27.4 min). In contrast, the RALS and LALS signals of the nitrated cellulose  
 24 oligomers are weak and almost hidden in the background. The  $dn/dc$  values of  
 25 cellulose nitrates and cellulose carbanilates in 0.01M LiCl/DMF are  $0.055 \text{ cm}^3/g$   
 26 and  $0.131 \text{ cm}^3/g$ , respectively. A higher  $dn/dc$  value resulted in higher light  
 27 scattering intensities. Thus, for cellulose oligomers, carbanilation is a better  
 28 derivatization method than nitration for its better performance in light scattering

1 detection.

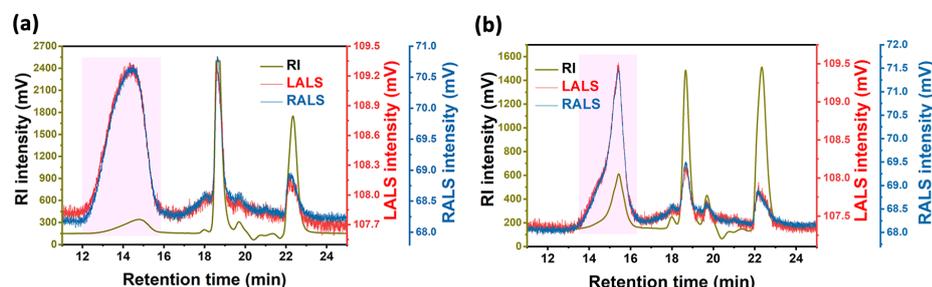


Fig. 5 Chromatograms of (a) carbanilated fraction A and (b) carbanilated fraction B with the eluent of THF.

2

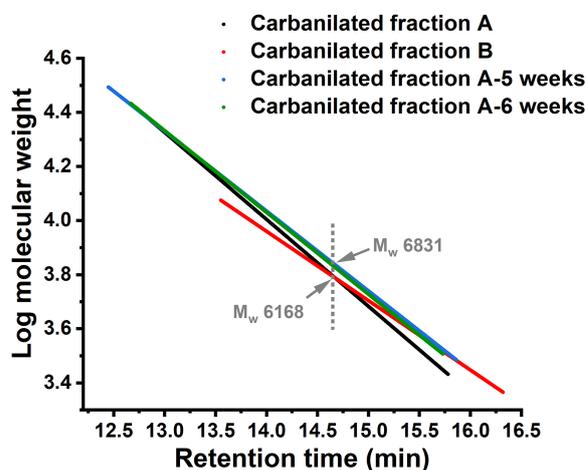


Fig. 6 Overlaid chromatograms of the carbanilated fraction A, fraction B and cellulose oligomers that hydrolyzed at room temperature for 5 weeks and 6 weeks.

3 The refractive index (RI) and light scattering (LALS and RALS) signals, are  
 4 shown in Fig. 5. To enhance the separation of small-sized molecules, the columns  
 5 of the SEC instruments were substituted. Additionally, the eluent was changed to  
 6 THF, owing to the larger  $dn/dc$  value of the tricarbanilated cellulose in THF,  
 7 which is  $0.165 \text{ cm}^3/g$ . The pink regions in Fig. 5 correspond to the carbanilated  
 8 cellulose oligomers that are analyzed, whereas the other peaks outside of these  
 9 regions are assigned to side products of trimers (18.0 min), dimers (18.7 min),  
 10 1,3-dipenylurea (19.7 min), and methyl carbanilates (22.3 min), respectively.  
 11 Regarding carbanilated fraction A, the LALS intensity steadily increased before  
 12 the retention time of 13.8 min, followed by a sharp decrease after 14.7 min. The

1 apex of the RI signal is at around 14.7 min. The resulting values for  $DP_n$ ,  $DP_w$ ,  
2 and  $\bar{D}$ , are 11.5, 14.0, and 1.22, respectively. In contrast, the LALS signal of the  
3 carbanilated fraction B is initially noisy before 14.8 min, attributed to insufficient  
4 scattered light, as the concentration of eluted cellulose molecules is low.  
5 Subsequently, the LALS signal increases more rapidly and becomes less noisy until  
6 the retention time of 15.4 min. This behavior is due to the high concentration of  
7 eluted molecules, as indicated by the RI signal. The RI and LALS signals of  
8 carbanilated fraction B are much sharper than those of carbanilated fraction A,  
9 corresponding to the narrower DP distribution of fraction B, with a  $\bar{D}$  value of  
10 1.07.

11  
12 Fig. 6 summarizes the correlation between retention time and estimated  
13 molecular weight for carbanilated fractions A (black) and B (red). The Log M was  
14 fitted with a single linear function, as otherwise, the molecular weight increased  
15 with retention time when both the molar mass and concentration of the eluted  
16 molecules are small (Fig. S7). This observation indicates a limited accuracy of the  
17 SEC measurements in such a condition (Fig. S7). The black and red curves  
18 intersected at approximately 14.7 min retention time, indicating inconsistent  
19 molecular weight estimations for the two fractions. This inaccuracy in the  
20 molecular weight estimation likely arose from the low scattered light intensity of  
21 the oligomers due to their small molecular size, especially for eluted molecules  
22 with low concentrations. The overlapping part of the black and red curves is  
23 primarily in the region with the highest concentration of oligomers. The uneven  
24 and incomplete carbanilation of the oligomers, as evident in the MALDI-MS  
25 results, may also contribute to the measurement inaccuracy.

26  
27 To verify the relationship between retention time and molecular weight of  
28 tricarbanilates, we examined two new samples: carbanilated fraction A that had  
29 been hydrolyzed for 5 (blue) and 6 (green) weeks at 23 °C. These two new samples  
30 were obtained using the same protocol as the fraction A, except that the  
31 hydrolysis time was shorter. The blue and green curves almost completely overlap  
32 throughout the retention time, indicating a close correlation between retention  
33 time and the log molecular weight of tricarbanilated cellulose oligomers. The black  
34 curve also overlaps with the blue and green curves in the range of 12.8-13.5 min,  
35 suggesting that the estimation of molecular weight for the high molecular-weight  
36 fractions is reproducible and reliable. Similarly, the red curve overlaps with the  
37 blue and green curves in the range of 15.1-15.8 min.

38  
39 The  $DP_w$  values of the carbanilated fractions A-5 weeks, A-6 weeks, A, and B  
40 are calculated to be 18.2, 16.6, 14.0, and 8.6, respectively. These results suggest  
41 that carbanilated cellulose oligomers with relatively smaller molecular weights are  
42 more likely to have errors in the calculation of molecular weight using the LS  
43 method. This is mainly because their small molecular size results in limited  
44 scattered light intensity and inaccurate estimation of molecular weight. When the  
45 proportion of such cellulose molecules is small, as in the cases of the blue and  
46 green curves, this effect is negligible. However, as the smaller molecular weight  
47 fractions increase, the calculated molecular weight by the LS method becomes  
48 more inaccurate. At the retention time of 14.7 min, where the black and red  
49 curves intersect, the molecular weight of the carbanilated fractions A and B is

1 approximately  $6.2 \times 10^3 \text{ g} \cdot \text{mol}^{-1}$ , while that of fractions A-5 weeks and A-6  
 2 weeks is about  $6.8 \times 10^3 \text{ g} \cdot \text{mol}^{-1}$ . The difference of about 600 in molecular  
 3 weight corresponds to approximately 1 tricarbanilated anhydroglucose unit.

4  
 5 The molar mass distribution of the carbanilated fraction A and B from SEC  
 6 and MALDI-TOF MS are compared in Fig. 7. The results of mass spectrometry  
 7 give a histogram of number count for each fraction evenly spaced in molecular  
 8 mass. The intensity is the integration of number count  $N$  and molecular weight  
 9  $M_w$  for each fraction. To plot the  $dWF/d\log M_w$  as a function of  $\text{Log} M_w$  from  
 10 MALDI-TOF MS results, the unit conversion is described as follows,

$$\frac{d \log M_w}{d M_w} = \frac{1}{M_w}$$

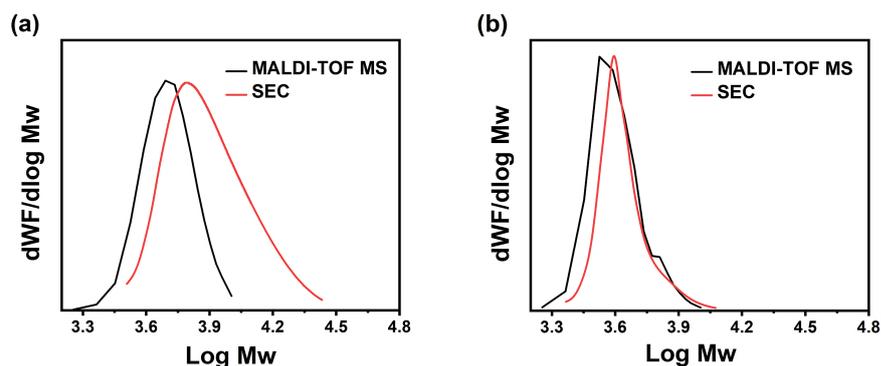
$$dWF = d(N \times M_w)$$

$$\frac{dWF}{d \log M_w} = \frac{M_w \times d(N \times M_w)}{d M_w} = M_w \times \text{Intensity}$$

11  
 12  
 13  
 14  
 15  
 16 Fig. 7 presents the  $dWF/d\log M_w$  of MALDI-TOF MS results, which are  
 17 obtained by multiplying the MS intensity of each fraction by its molecular weight.  
 18 The SEC molar mass distribution shows that the peak top positions of the  
 19 carbanilated fraction A and B are 6250 and 3818  $\text{g} \cdot \text{mol}^{-1}$ , respectively,  
 20 corresponding to  $DP_p$  values of 10.5 and 7.4. The peak shape of the carbanilated  
 21 fraction A is asymmetric, with a long tail in the high-molecular-weight portion.  
 22 The SEC-based distribution of the carbanilated fraction A reveals that  
 23 approximately 84% of the molecules have a molecular weight greater than 5000  
 24  $\text{g} \cdot \text{mol}^{-1}$ . In contrast, this proportion is only 60% for the carbanilated fraction A,  
 25 as shown in the results from MALDI-TOF MS. This discrepancy suggests that the  
 26 mass spectrometry histogram may be missing 20-30% of high-molecular-weight  
 27 molecules, likely due to differences in ionization efficiency between low- and  
 28 high-mass oligomers. Thus, the calculated number-averaged and weight-averaged  
 29 DP values of the carbanilated fraction A from MALDI-TOF MS results may be  
 30 biased toward the lower molecular weight fractions. The SEC molar mass  
 31 distribution of the carbanilated fraction B is narrower and more symmetric than  
 32 that of the carbanilated fraction A. Notably, the SEC-based distribution curve and  
 33 the MALDI-TOF MS curve are superimposed when the molecular weight is larger  
 34 than 3818  $\text{g} \cdot \text{mol}^{-1}$ . However, for the low-molecular-weight molecules on the right  
 35 side of peak top positions, the MALDI-TOF MS results show higher proportions  
 36 than the SEC results for both carbanilated fraction A and B. This finding  
 37 indicates that low-molecular-weight molecules are more likely to be ionized than  
 38 high-molecular-weight ones. Additionally, the SEC method may produce incorrect  
 39 estimations of molecular weight when the molecular size is small, whereas the  
 40 MALDI-TOF MS method is more sensitive to low-molecular-weight molecules.

#### 41 Comparison of estimated DPs from different characterization methods

42 Table 2 summarizes the calculated DP values for fractions A and B obtained  
 43 from various characterization techniques. Notably, significant differences exist in  
 44 the estimated DP values across the different methods employed. Given the greater



**Fig. 7** Overlaid molar mass distribution graphs of (a) carbanilated fraction A and (b) carbanilated fraction B from MALDI-TOF MS and SEC tests.

1 accuracy of SEC analysis for higher molecular fractions, we propose that the  $DP_n$ ,  
 2  $DP_w$ , and  $\bar{D}$  values of fraction A are likely to be 11.5, 14.0, and 1.22, respectively,  
 3 based on SEC analysis. However, the  $DP_n$  value obtained from solid-state  $^{13}\text{C}$   
 4 CP/MAS NMR for fraction A is slightly smaller than that from SEC. For fraction  
 5 B, the average DP value obtained from different techniques is in the range of 6.7  
 6 to 8.6.

7

**Table 2** Summary of molecular weight information of fraction A and B by different characterization techniques.

Techniques		Fraction A	Fraction B
Liquid-state $^1\text{H}$ NMR	$DP_n$	-	6.7
Solid-state $^{13}\text{C}$ CP/MAS NMR	$DP_n$	10.1	8.6
Mass spectrometry	$DP_n$	8.8	7.2
	$DP_w$	9.7	8.0
	$\bar{D}$	1.09	1.10
Size exclusion chromatography	$DP_n$	11.5	8.0
	$DP_w$	14.0	8.6
	$DP_p$	10.5	7.4
	$\bar{D}$	1.22	1.07

8 As presented in Table 2, SEC is reliable in characterizing oligomers with  
 9 relatively higher molecular weight. Conversely, this statement does not hold true  
 10 for low molecular weight oligomers, as the estimation of molecular weight based on  
 11 light scattering is considered unreliable for small oligomers. In contrast, the DP  
 12 information obtained from both MALDI-TOF MS and NMR spectroscopy is  
 13 deemed more reliable for the lower molecular-weight fractions. It should be noted  
 14 that the MALDI-TOF MS data may exhibit bias towards the lower molecular  
 15 weight oligomers, as the ionization efficiency of oligomers is expected to have a  
 16 negative correlation to their size. Additionally, while NMR spectroscopy gives the  
 17 number-average molecular weight, it lacks access to the DP distributions.

18

## 1 **Conclusions**

2 This study has undertaken a comprehensive exploration to clarify the  
3 complexities and subtleties of molecular determination methods of cellulose  
4 oligomers. Liquid state  $^1H$  NMR, solid-state  $^{13}C$  CP/MAS NMR and  
5 MALDI-TOF MS have proven to be more suitable for characterizing cellulose  
6 oligomers possessing lower DPs and narrower DP distributions, such as the  
7 fraction B with an average DP of ca. 8. For cellulose oligomers with a relatively  
8 high molecular weight and broader distribution, such as the fraction A with an  
9 average DP > 10, SEC emerges as a preferred choice in comparison to other  
10 spectroscopic techniques. This study also shed light on the underlying mechanisms  
11 affecting the molecular weight estimation of the oligomers, including their low  
12 light scattering intensities in SEC measurements, uneven carbanilation, and the  
13 varying ionization efficiency in MALDI MS relative to molar mass. The insights  
14 gained from this study will serve as a stepping stone for developing more accurate  
15 and robust techniques for determining the molecular weight of cellulose oligomers.

## 17 **Availability of data and materials**

18 All data generated or analyzed during this study are included in this published  
19 article and its supplementary information files.

20 **Acknowledgements** W.L. thanks the China Scholarship Council due to the financial support.  
21 We acknowledge Mr. Killian Barry and Mr. Jessie Muamba for the size exclusion chromatography  
22 measurements, Ms. Laurine Buon and Mr. Eric Bayma for their assistance in measuring  $dn/dc$   
23 values, Ms. Isabelle Jeacomine for NMR test, and Ms. Laure Fort for the help in mass  
24 spectrometry.

## 25 **Funding**

26 This work was supported by Institut Carnot PolyNat (Investissements d'Avenir  
27 Grant No. ANR-11-CARN-030-01) and Beijing Institute of Technology Startup  
28 Research Fund For Young Scholars (XSQD-202108010).

## 29 **Authors' contributions**

30 W.L. conducted the experiments and wrote the main manuscript, Y.O. and  
31 Y.N. supervised this project. G.M. and I.O. provided valuable help in SEC analysis.  
32 All authors reviewed the manuscript.

## 33 **Competing interests**

34 The authors declare that they have no conflict of interest.

## References

- Bernet B, Vasella A (2000) Intra-and intermolecular H-bonds of alcohols in DMSO,  $^1\text{H}$ -NMR analysis of inter-residue H-bonds in selected oligosaccharides: Cellobiose, lactose, N, N'-diacetylchitobiose, maltose, sucrose, agarose, and hyaluronates. *Helvetica Chimica Acta* 83(9):2055–2071
- Dupont AL, Mortha G (2004) Comparative evaluation of size-exclusion chromatography and viscometry for the characterisation of cellulose. *Journal of Chromatography A* 1026(1-2):129–141
- Evans R, Wallis AF (1989) Cellulose molecular weights determined by viscometry. *Journal of applied polymer science* 37(8):2331–2340
- Fittolani G, Vargová D, Seeberger PH, Ogawa Y, Delbianco M (2022) Bottom-up approach to understand chirality transfer across scales in cellulose assemblies. *Journal of the American Chemical Society* 144(27):12469–12475
- Fluge LA, Blank JT, Petillo PA (1999) Isolation, modification, and nmr assignments of a series of cellulose oligomers. *Journal of the American Chemical Society* 121(31):7228–7238
- Helbert W, Sugiyama J (1998) High-resolution electron microscopy on cellulose II and  $\alpha$ -chitin single crystals. *Cellulose* 5(2):113–122
- Hill DJ, Le TT, Whittaker AK (1994) A technique for the quantitative measurements of signal intensities in cellulose-based transformer insulators by  $^{13}\text{C}$  CP/MAS NMR. *Cellulose* 1(4):237–247
- Hiraishi M, Igarashi K, Kimura S, Wada M, Kitaoka M, Samejima M (2009) Synthesis of highly ordered cellulose II in vitro using cellodextrin phosphorylase. *Carbohydrate Research* 344(18):2468–2473
- Idström A, Schantz S, Sundberg J, Chmelka BF, Gatenholm P, Nordstierna L (2016)  $^{13}\text{C}$  NMR assignments of regenerated cellulose from solid-state 2D NMR spectroscopy. *Carbohydrate polymers* 151:480–487
- Isogai A, Usuda M (1991) Preparation of low-molecular weight celluloses using phosphoric acid. *Mokuzai Gakkaishi* 37:339–344
- Jiang F, Zhang X, Hwang W, Nishiyama Y, Briber RM, Wang H (2021) Oligocellulose from acid hydrolysis: A revisit. *Applied Surface Science* 537:147783
- Kita Y, Kusumi R, Kimura T, Kitaoka M, Nishiyama Y, Wada M (2020) Surface structural analysis of selectively  $^{13}\text{C}$ -labeled cellulose II by solid-state NMR spectroscopy. *Cellulose* 27(4):1899–1907
- Kobayashi S, Hobson LJ, Sakamoto J, Kimura S, Sugiyama J, Imai T, Itoh T (2000) Formation and structure of artificial cellulose spherulites via enzymatic polymerization. *Biomacromolecules* 1(2):168–173
- Oberlerchner JT, Rosenau T, Potthast A (2015) Overview of methods for the direct molar mass determination of cellulose. *Molecules* 20(6):10313–10341
- Ono Y, Ishida T, Soeta H, Saito T, Isogai A (2016) Reliable dn/dc values of cellulose, chitin, and cellulose triacetate dissolved in LiCl/N, N-dimethylacetamide for molecular mass analysis. *Biomacromolecules* 17(1):192–199
- Queyroy S, Müller-Plathe F, Brown D (2004) Molecular dynamics simulations of cellulose oligomers: conformational analysis. *Macromolecular theory and simulations* 13(5):427–440
- Shen T, Langan P, French AD, Johnson GP, Gnanakaran S (2009) Conformational flexibility of soluble cellulose oligomers: chain length and temperature dependence. *Journal of the American Chemical Society* 131(41):14786–14794

- 
- 1 Sugiyama H, Hisamichi K, Usui T, Sakai K, et al. (2000) A study of the conformation
  - 2 of  $\beta$ -1, 4-linked glucose oligomers, cellobiose to cellohexaose, in solution. *Journal*
  - 3 *of Molecular Structure* 556(1-3):173–177
  - 4 Zuckerstätter G, Terinte N, Sixta H, Schuster KC (2013) Novel insight into
  - 5 cellulose supramolecular structure through  $^{13}\text{C}$  CP-MAS NMR spectroscopy and
  - 6 paramagnetic relaxation enhancement. *Carbohydrate polymers* 93(1):122–128
  - 7 Zweckmair T, Oberlerchner JT, Böhmendorfer S, Bacher M, Sauerland V, Rosenau T,
  - 8 Potthast A (2016) Preparation and analytical characterisation of pure fractions
  - 9 of cellooligosaccharides. *Journal of Chromatography A* 1431:47–54