ORIGINAL RESEARCH

Terahertz time-domain spectroscopy as a novel tool for crystallographic analysis in cellulose

Han Wang · Yoshiki Horikawa · Satoru Tsuchikawa · Tetsuya Inagaki

Abstract Terahertz (THz) radiation is responsive to optical phonons in crystal lattices; therefore, THz time-domain spectroscopy (THz-TDS) has potential utility in crystallographic analysis. In this study, THz-TDS was used to investigate cellulose samples extracted from different sources with varying ratios of I_α and I_β fractions. This is the first study to report that the absorption coefficient in the THz region of cellulose I_α and I_β presents different characteristic absorption peaks around 2 THz. This can be used to distinguish the cellulose I allomorph, which exhibits differences in its crystalline structure. The absorption coefficient at 2.11 and 2.38 THz showed a considerable linear correlation with the I_α fraction of the cellulose samples. The absorption coefficient spectra of 1.88–3.40 THz were further detrended and separated into three Gaussian peaks. The correlations between the parameters of these separated peaks (integrated intensities, peak positions) and the crystalline structural values (d-spacing, crystalline size, crystallinity index) that calculated from the X-ray diffraction (XRD) pattern were investigated.

Keywords THz-TDS · Cellulose · Crystalline allomorph · Crystallinity · Cellulose I_α · Cellulose I_β

Introduction

Cellulose is the most abundant and widely distributed polysaccharide in nature. It is an important component of the structural framework of plant cell walls and is synthesized by cellulose synthetases on cytoplasmic membranes. Cellulose has a linear chain composed of β (1 → 4)-D-glucose units that contain hydrogen bonds arranged in a parallel fashion (Updegraff 1969), where cellulose I is the crystalline cellulose that is naturally produced by plants, algae, and bacteria. Cellulose I has been found to consist of two polymorphs: I_α has a triclinic structure and contains one cellulose chain, and I_β has a monoclinic structure and contains two cellulose chains (Sugiyama et al. 1991). The two polymorphs coexist according to the different proportions of cellulose sources (O'sullvian 1997; Azizi Samir et al. 2005; Nishiyama 2009). For instance, the cellulose found in bacteria and algae such as *Cladophora* spp. and *Valonia ventricosa* is composed mainly of the I_α type, whereas the cellulose of the secondary cell walls of higher plants such as

H. Wang · S. Tsuchikawa · T. Inagaki (✉)
Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya, Aichi 464-8601, Japan
e-mail: inatetsu@agr.nagoya-u.ac.jp

Y. Horikawa
Institute of Agriculture, Tokyo University of Agriculture and Technology, Fuchu, Tokyo 183-8509, Japan

cotton, ramie, and highly crystalline animal cellulose (tunicate cellulose) contains mainly the I_β allomorph (Atalla and VanderHart 1984; Belton et al. 1989; Larsson et al. 1995; Horii et al. 1997). The d-spacings of three lattice planes are approximately 0.61, 0.53, and 0.39 nm, corresponding to lattice planes (100), (010), and (110) in the I_α type and lattice planes (1_10), (110), and (200) in the I_β type (Wada et al. 1997, 2001). Cellulose I_α is metastable and can be converted into I_β allomorph by annealing with saturated steam or incorporating into an aqueous alkaline solution and subjecting to high temperatures (Horii et al. 1987b; Yamamoto and Horii 1993; Horikawa and Sugiyama 2009). The degree of conversion from I_α to I_β can be controlled by adjusting the reaction conditions. To date, there has been no report on the complete conversion from I_α to I_β (Debzi et al. 1991; Yamamoto and Horii 1993). The difference in the structure and stability of I_α and I_β cellulose is noteworthy, as it indicates a difference in the biosynthetic assembly of the cellulose fibrils and the manner of assembly of cellulose molecular chains. Wada et al. have suggested that the I_α/I_β ratio of cellulose is related to the phylogenetic evolution of plants (Wada et al. 1993). It is reasonable to believe that knowledge of the component ratio of the allomorphs in native celluloses is important for determining the biological mechanism of cellulose polymorphisms.

Several techniques are used to determine the I_α/I_β ratio of native cellulose. The most frequently applied technique is cross polarization/magic angle spinning (CP/MAS), which shows the different spectra among algal–bacterial cellulose (I_α-rich cellulose) and cotton-ramie cellulose (I_β-dominant cellulose), as well as carbon 13 nuclear magnetic resonance (^{13}C NMR), where the C1, C4, and C6 bands differ depending on the ratio of cellulose I_α and I_β allomorphs (Atalla and VanderHart 1984; VanderHart and Atalla 1984; Horii et al. 1987a). Aside from NMR and CP/MAS, Wada et al. reported that X-ray diffraction (XRD) can also be used to classify the cellulose extracted from different sources where the cellulose is the I_α-rich or I_β-dominant type (Wada et al. 1993, 2001). Fourier transform infrared (IR) spectroscopy (FTIR) can also be used to classify allomorphs by investigating the characteristic absorption (Imai and Sugiyama 1998) and changes in the crystalline phase after sulfuric acid treatment (Wada and Okano 2001), where the bands near 3240 and 750 cm^{-1} are assigned to the triclinic I_α

cellulose, and bands near 3270 and 710 cm^{-1} can be assigned to the monoclinic I_β cellulose (Persson et al. 1991). The FT-IR spectrum of cellulose in the region 3600—3200 cm^{-1} were corresponding to the OH stretching frequencies in cellulose, on the other hand, the spectrum in the region 700—800 cm^{-1} is not straightforward, Machessault et al reported these vibrations were O–H out-of-plane bending (Marchessault et al. 1960), whereas Blackwell et.al reported the band at 750 cm^{-1} was a CH_2 rocking vibration (Blackwell et al. 1970). Horikawa reported that different spectra of cellulose structural variety from different origins could be observed with near-IR spectroscopy and principal component analysis (Horikawa 2017).

Terahertz (THz) radiation lies in far-IR radiation between microwaves and IR of the electromagnetic spectrum with a frequency range of 0.1–10 THz, corresponding to wavelengths of 3–0.03 mm. The development of a series of new technologies and materials in the 1980s made possible the generation and modulation of stable broadband THz. The earliest commercialized application of this was THz time-domain spectroscopy (THz-TDS). The applications of THz radiation include detecting pharmaceuticals and explosives (Federici et al. 2005), pharmaceutical polymorphs (Taday et al. 2003; Ikeda et al. 2010), determining crystallinity (Vieira and Pasquini 2014), and investigating the physical properties of cellulose-based materials.(Inagaki et al. 2014; Peccianti et al. 2017; Wang et al. 2019) Because THz radiation interacts with optical phonons in crystal lattices and crystalline materials have distinct absorption characteristics due to their varying structure and crystallinity, it is thus expected that THz-TDS could be used to probe the crystal structure, crystalline phases, and crystallinity. For instance, binary mixtures of carbamazepine forms I and III present different absorption peaks at 0.9 and 1.2 THz and can be used to determine the form and quantify the proportion of all components (Strachan et al. 2005). In this sense, THz-TDS can detect differences in the crystal structure of cellulose I.

In this study, THz-TDS was used to investigate cellulose I with different fractions of I_α and I_β extracted from various sources, where the absorption coefficient spectra in the THz region of I_α-rich and I_β-dominant types were each with its own characteristics. The peaks of absorption coefficient spectra in the THz region of cellulose I_α and I_β were 2.38 and 2.11 THz,

and the absorption coefficients were related to the fraction of the two components. Furthermore, the absorption coefficient spectra of 1.88–3.40 THz were separated as three Gaussian peaks, two were fixed at 2.11 and 2.38 THz, respectively; the integrated intensities of these two peaks were both correlated with the fractions of I_α and d-spacing value of the lattice planes of (100), (110) in I_α and (1_10), (200) in I_β types. The other peak was located at around 3 THz, where the peak positions were determined by the curve-fitting process, the integrated intensities were correlated with the crystalline sizes and crystallinity index that calculated from XRD patterns. Moreover, the positions of the peaks were correlated with the fractions of I_α and the d-spacing value of the lattice planes of (010) and (110) in I_α and I_β types. This paper suggests that THz-TDS can be used to distinguish the crystalline structure of cellulose I and used as a new tool for the crystallographic analysis of the cellulose materials.

Experimental

Preparation of the cellulose samples

We utilized the same cell wall samples employed in previously conducted research (Horikawa 2017). Cell wall samples of *Glaucocystis nostochinearum*, *Valonia macrophysa*, *Cladophora* spp., *Boergesenia forbesii*, and *Oocystis apiculata* were boiled in 1% NaOH for 8–10 h and treated in 0.05 N HCl at room temperature. The samples were then bleached in 0.3% NaClO$_2$ with a pH of 4.9 in an acetate buffer at 70 °C for 3 h. The obtained sample was then thoroughly washed with distilled water. At room temperature, *Halocynthia roretzi* was treated with 5% KOH and washed with distilled water. The samples were bleached in 0.3% NaClO$_2$ at pH 4.9 in acetate buffer for 2 h in a 70 °C, and the procedure was repeated several times. After the sample became colorless, it was washed with distilled water. All purified cellulose samples were freeze-dried and made into three tablets for each source with a diameter of 7.3 mm and a thickness of 0.5 to 1.0 mm by hand pressing with a special die set (KYOTO PASTEC Co., Ltd.). The XRD measurement was conducted in 2017 (Horikawa 2017), where the X-ray diffractograms were obtained in reflection mode with Cu-Kα radiation ($\lambda = 0.15418$ nm) by employing an automated multipurpose X-ray diffractometer (SmartLab, Rigaku, Tokyo, Japan) at a voltage of 40 kV and current of 30 mA. After the measurement, all samples were freeze-dried and then stored in a desiccator at all times, since the samples were highly crystalline, biodegradation was even less likely to occur. It can be considered there was no change in the crystal structure.

THz-TDS measurement

THz transmission spectra were acquired using a Tera Prospector-Kit model (NIPPO PRECISION Co., Ltd.). The THz beam was horizontally polarized with a useful spectral range that extended from approximately 0.1 to 4.00 THz, and the spectral resolution in this study was 0.02 THz, determined by the inverse of the temporal scan range (50 ps). The diameter of the THz beam spot on the sample was around 3 mm. Each measurement was recorded by averaging 100 scans to improve the signal-to-noise ratio (one scan takes about 5 s). The reference THz signal without the samples was obtained prior to and after each transmission measurement. For reproducibility, all measurements were conducted three times. Because THz is sensitive to water vapor, the THz optical system was placed in an almost-closed acrylic box that was continuously filled with dry air to ensure the stability of the humidity; all samples were placed in the box for 24 h before measurements to balance the ambient humidity. When the samples were placed in the box, the dry airflow was filled into the box until all the THz measurements were completed. The thickness of samples was measured using a caliper ($\pm$ 0.01 mm), and the mass was measured with an electronic balance ($\pm$ 0.0001 g) before and after all the THz measurements, which was designed to ensure that during the measurements, there was no water absorption.

Results and discussion

Calculation of crystalline structure values with XRD

The XRD patterns were cut out with a scattering angle (2θ) from 10° to 30°, as shown in Fig. 1a, using different colors to distinguish the source of the samples. Where the spectra could be considered
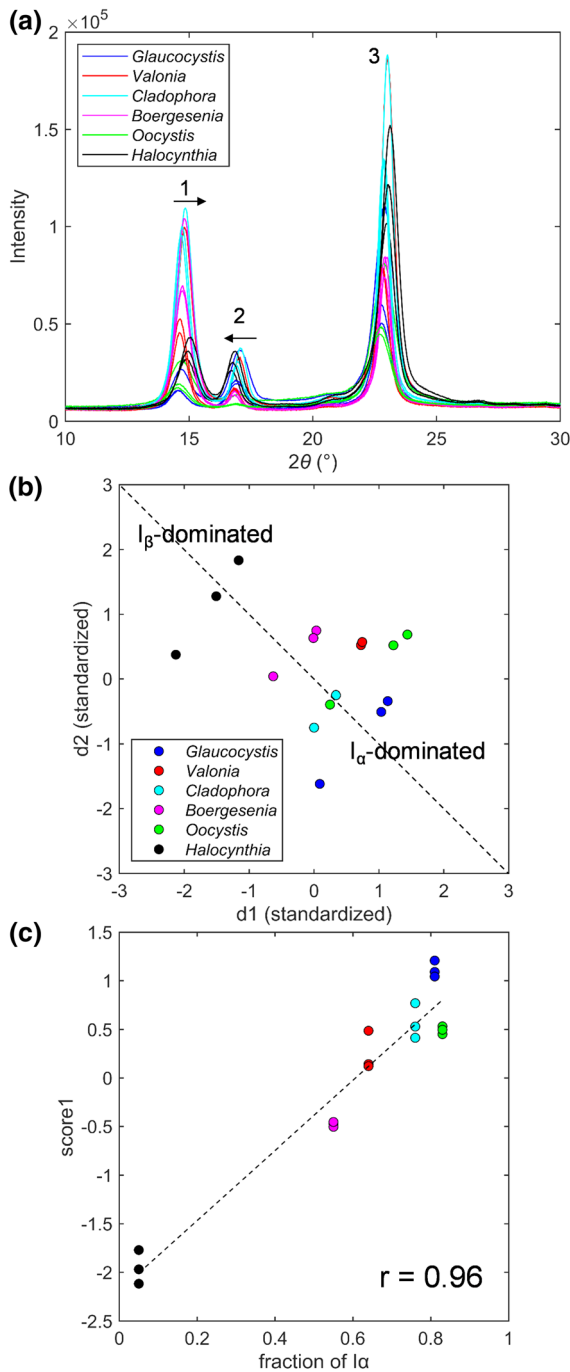
(a)

(b)

(c)

◄**Fig. 1** The original and preprocessing of the XRD patterns of cellulose and classification of the type of cellulose. **a** original XRD patterns of cellulose samples with different I_α fractions, three peaks corresponding to the Miller indices of (100), (010), and (110) in the I_α type and (1_10), (110), and (200) peaks in the I_β pattern, black arrows indicated the directions of peak shifts of *Halocynthia*, **b** scatter plot of standardized d1 and d2, where the top left corner corresponds to the I_β-dominant pattern and the bottom right corner corresponds to the I_α rich pattern, **c** relationship between fraction of I_α and the PC1 score, *r*, correlation coefficient

corresponded to the (1_10), (110), and (200) peaks in the I_β pattern as shown in Fig. 1a that denoted as peak1, peak2 and peak3, and the background intensity curve here was treated as the amorphous scattering for simplification in this study. The positions and FWHM varied for the samples from different sources, which was due to the varying I_α fractions of the cellulose samples. To conduct further analysis, the baseline was first fitted as a first-order polynomial, and the separation of peaks was carried out with Gaussian peaks using the following equation with a curve-fitting process:

$$I_G(2\theta) = I_{\max} \cdot \exp\left\{ -4\ln(2) \cdot \left(\frac{2\theta - 2\theta_{\max}}{\beta} \right)^2 \right\}$$

(1)

where I_{max} is intensity of peaks, $2\theta_{max}$ is the peak position, and β is the FWHM.

The d-spacings, crystalline size (Cs), and crystallinity index (CI) were calculated using the following equations:

$$d = \frac{\lambda}{2\sin\theta}$$

(2)

$$C_s = \frac{K\lambda}{\beta \cdot \cos\theta}$$

(3)

$$CI = \frac{S_{cr}}{S_{cr} + S_{Am}}$$

(4)

where λ is the wavelength of the X-ray radiation (0.1542 nm); θ is the Bragg angle; *K* is the Scherrer constant, which was around 0.9 when assuming Gaussian peaks; and β is the FWHM in radians (Zsigmondy and Scherrer 1912; Park et al. 2010;

composites of a baseline, a background intensity curve, which in most research was attributed to amorphous scattering, but also may due to the overlap of the intensity of adjacent crystalline peaks (French 2014), and the three main crystalline peaks for I_α had Miller indices of (100), (010), and (110), which

Bishnoi et al. 2017). The detail of the processing of the XRD patterns (baseline correction and deconvolution) are presented in Supplementary Information.

In this study, the crystallinity index was calculated with the integrated intensity, where S_{cr} is the sum of the integrated intensity of each crystal peak and S_{Am} is the integrated intensity of the background intensity curve. All the calculated results are summarized in Table 1, where the calculated results are supported by the result of (Wada and Okano 2001; Wada et al. 2001), and the fraction of I_α used in our study was referenced with data estimated from IR spectra (Belton et al. 1989; Imai and Sugiyama 1998; Imai et al. 1999).

For peak 1, the peak position (2θ) range was from 14.53° to 15.03° and the FWHM was from 0.74° to 1.20°. For peaks 2 and 3, the peak position range was 16.70°–17.07° and 22.72°–23.10°, respectively, and the FWHM range was 0.61°–1.20° and 0.58°–0.98°, respectively. As shown in Fig. 1a, the *Halocynthia* (black line) which only had 0.05 fraction of cellulose I_α showed observable peak shift at peak1 and peak2, the directions of peak shifts were indicated by the black arrow in Fig. 1a.

The range of d-spacing values for peaks 1, 2, and 3 were from 0.589–0.610 nm, 0.519–0.531 nm, and 0.385–0.391 nm, respectively. The slight fluctuation occurred in the d-spacing values of the three

Table 1 Calculated values of the crystalline structure from the XRD pattern and published I_α fractions

Position of the peaks (°)			FWHM (°)			d-spacing (nm)			Crystalline size (nm				
2θ	2θ	2θ	$\beta1$	$\beta2$	$\beta3$	d1	d2	d3	cs1	cs2	cs3	CI	fa
Glaucocystis													
14.71	17.04	22.89	1.05	1.00	0.90	0.602	0.520	0.389	7.63	8.01	8.97	76.15	0.81[a]
14.58	16.93	22.78	0.90	0.85	0.86	0.607	0.524	0.390	8.95	9.41	9.45	79.90	
14.57	16.91	22.77	0.93	0.88	0.86	0.608	0.524	0.391	8.61	9.11	9.38	76.81	
Valonia													
14.83	17.02	23.02	0.78	0.79	0.64	0.598	0.521	0.386	10.22	10.15	12.58	84.99	0.64[a]
14.63	16.83	22.82	0.77	0.61	0.63	0.606	0.527	0.390	10.48	13.07	12.87	87.34	
14.62	16.82	22.82	0.74	0.56	0.61	0.606	0.527	0.390	10.86	14.39	13.24	87.27	
Cladophora													
14.85	17.07	23.01	0.77	0.66	0.60	0.597	0.519	0.386	10.42	12.11	13.55	85.59	0.76[b]
14.73	16.95	22.90	0.77	0.66	0.59	0.602	0.523	0.388	10.40	12.20	13.77	86.23	
14.68	16.90	22.85	0.77	0.65	0.58	0.604	0.525	0.389	10.48	12.30	14.06	86.25	
Boergesenia													
14.82	16.87	22.98	0.82	0.70	0.62	0.598	0.525	0.387	9.82	11.48	13.00	88.53	0.55[b]
14.72	16.80	22.90	0.81	0.62	0.63	0.602	0.528	0.388	9.90	12.94	12.94	89.15	
14.73	16.82	22.91	0.82	0.65	0.63	0.602	0.527	0.388	9.79	12.32	12.95	86.53	
Oocystis													
14.69	16.92	22.87	1.20	1.20	0.98	0.603	0.524	0.389	6.68	6.70	8.31	72.02	0.83[a]
14.53	16.81	22.72	1.09	0.93	0.98	0.610	0.527	0.391	7.33	8.60	8.25	75.26	
14.56	16.83	22.75	1.10	1.20	0.94	0.609	0.527	0.391	7.30	6.70	8.59	72.22	
Halocynthia													
14.89	16.70	22.98	0.99	0.79	0.79	0.595	0.531	0.387	8.06	10.23	10.25	79.85	0.05[c]
14.94	16.75	23.04	0.99	0.79	0.78	0.593	0.529	0.386	8.08	10.23	10.46	78.92	
15.03	16.84	23.10	1.02	0.83	0.79	0.589	0.527	0.385	7.88	9.72	10.22	78.36	

fa Cellulose I_α fractions

[a]Imai et al. 1999

[b]Imai and Sugiyama 1998

[c]Mean value of fractions reported by Belton et al. 1989; Larsson et al. 1995

measurements of the same species that can be considered due to the differences among the samples, the fluctuation range was about 0.2%, which is acceptable. The d-spacing value for the I_β-dominant type (i.e., *Halocynthia*) and I_α-rich type (i.e., *Oocystis*) showed a clear difference, where the d1 was lower in value in the I_β-dominant type, and d2 was higher in value than the I_α-rich type. On the other hand, the d3 value showed no obvious changes related to the I_α fractions, and the crystalline size and crystallinity index did not show significant changes related to the I_α fractions. As shown in Fig. 1b, using the plot of the standardized d1 and d2, I_α fractions of the samples could be distinguished, and the results were consistent with previous research (Wada et al. 2001). In order to find the correlation between d1, d2 values and the fraction of I_α, principal component analysis (PCA) was carried out to compres the standardized d1 and d2, the results were shown in Fig. 1c, where the fraction of I_α showed a high correlation with the PC1 score, the correlation coefficient of r was 0.96, which indicated the fraction of I_α can be estimated using XRD patterns.

Absorption coefficient of cellulose with different crystalline structures at the THz region

The measured THz time-domain signal were Fourier transformed, and the absorption coefficient α were then calculated using the following equations:

$$n = -\frac{\varphi c}{2\pi v L} + 1 \qquad (5)$$

$$\alpha = -\frac{2}{L}\ln\left[R\frac{(n+1)^2}{4n}\right] \qquad (6)$$

where φ is the phase difference between the reference and measured samples ($\varphi_{reference}$ -φ_{sample}), R is the ratio of amplitude in the frequency domain of the measured samples to reference, c is the speed of light 3×10^8 m/s, v is the frequency, and L is the thickness of the samples (Reid and Fedosejevs 2006).

Since the samples in this study were with a relatively high cellulose content (about over 85%) and considering that the non-cellulose component is removed and confirmed in the IR spectrum, the absorption coefficients were not normalized with the relative amount of cellulose. The originally calculated absorption coefficient spectra are shown in Fig. 2a,
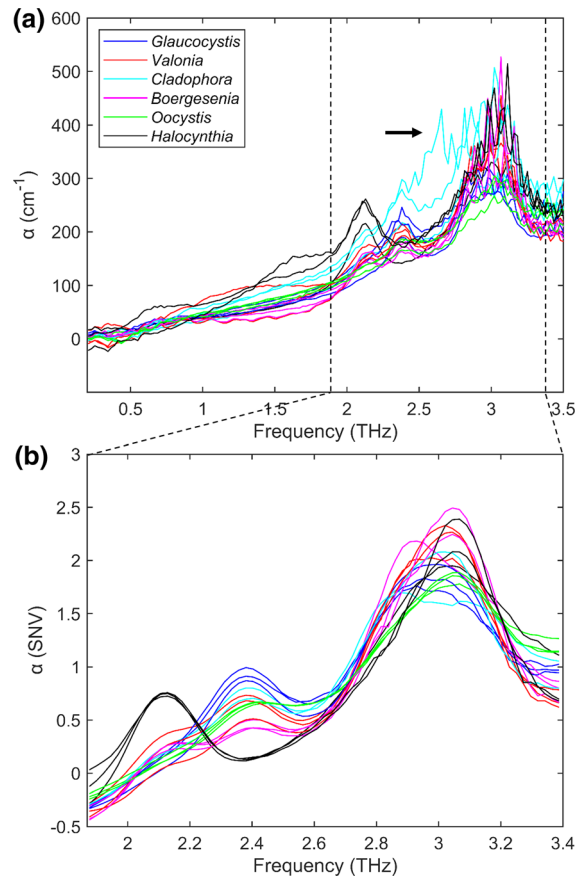

Fig. 2 THz absorption coefficient spectra of cellulose samples. **a** Original THz absorption coefficient spectra, the abnormal spectrum of *Cladophora* is indicated by the black arrows and was eliminated in subsequent processing, **b** SNV and smoothed absorption coefficient spectra of the frequency range of 1.88–3.40 THz

where the peaks around 2.1 and 2.4 THz in the absorption spectra are noteworthy, which may correlate to the cellulose crystalline structure. All the samples showed absorption peaks around 3 THz, where the intensity of the absorption peaks did not show a clear correlation with the I_α fractions. On the contrary, the peak positions showed some fluctuations with the different I_α fractions of the samples, which will be discussed in detail in the next section. Absorption around 3 THz has been reported to be correlated with the cellulose crystallinity index (Vieira and Pasquini 2014). However, one spectrum of *Cladophora* was abnormal (indicated by the arrow in Fig. 2). The abnormal spectrum showed obviously different absorption characteristics with two other *Cladophora*

samples which were considered normal in this study; therefore, the abnormal spectrum was eliminated from the analysis in this study.

All the spectra were treated with a standard normal variate algorithm (SNV) to correct baseline fluctuations of the absorption spectra and the bands from 1.88 to 3.40 THz were cut out and smoothed by applying a Savitzky–Golay filter with a second-order polynomial and fifteen smoothing points. As shown in Fig. 2b, the absorption spectra showed absorption peaks at 2.11 and 2.38 THz due to the different samples, the I_β-dominant type showed absorption peaks at 2.11 THz, whereas the I_α-rich type showed absorption peaks at 2.38 THz. It can be concluded that the absorptions at 2.11 and 2.38 THz were due to cellulose I_β and I_α types, respectively. As mentioned in the introduction, The FT-IR spectra of cellulose showed a similar appearance, where the absorption features near 3240 and 750 cm^{-1} were assigned to I_α, whereas the absorbances near 3270 and 710 cm^{-1} can be assigned to I_β, respectively (Persson et al. 1991).

The absorption coefficients at 2.11 and 2.38 THz showed a considerable linear correlation with the I_α fractions as shown in Figs. 3a and b with correlation coefficient values, r, of -0.94, 0.90, respectively. The values of *Halocynthia* samples which had the least fractions of I_α were separated from the other samples, which may cause the distortion of the correlation coefficient. Therefore, it is necessary to report the correlation coefficient without the *Halocynthia* samples. Where the r values were -0.50 and 0.70 respectively after removing the *Halocynthia* samples of Figs. 3a and b. This indicates that absorption coefficients at 2.11 and 2.38 THz reflected the difference in the proportion of components in cellulose I_α and suggested that the absorption spectra could correspond to the crystalline structure to some extent.

Correlations between THz absorption spectra and the crystalline structural values of cellulose

To further investigate the correlations between the THz absorption spectra and the crystalline structure and crystallinity index of cellulose, the absorption coefficient spectra of the frequency range of 1.88–3.40 THz were first detrended by subtracting the baseline, and the separation of peaks as that in XRD patterns was carried out with three Gaussian peaks using the Eq. (1), where the peak position $2\theta_{max}$
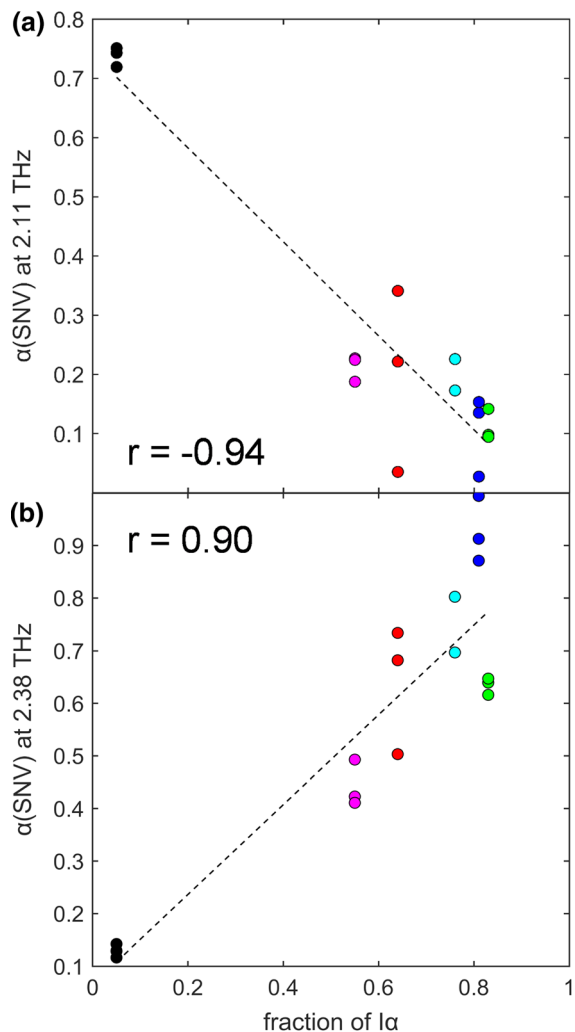

Fig. 3 Relationship between fraction of I_α and absorption coefficients. **a** 2.11 THz, **b** 2.38 THz, r, correlation coefficient

were changed into f_{max} that indicate the peak positions in THz region.

Taking *Valonia* with I_α fraction of 0.64, which showed absorption peaks at both 2.11 and 2.38 THz as an example, the curve-fitting process was shown in Fig. 4a. Peak1 and peak2 were fitted by only using I_{max} and FWHM β, where peak positions were fixed at 2.11 and 2.38 THz that corresponded to the absorption of cellulose I_β and I_α patterns as mentioned above, peak3 were fitted with all three parameters, and the fitted spectra were shown in Fig. 4b, where the peak shift of peak3, which seemed to be correlated with the fraction of I_α can be observed.
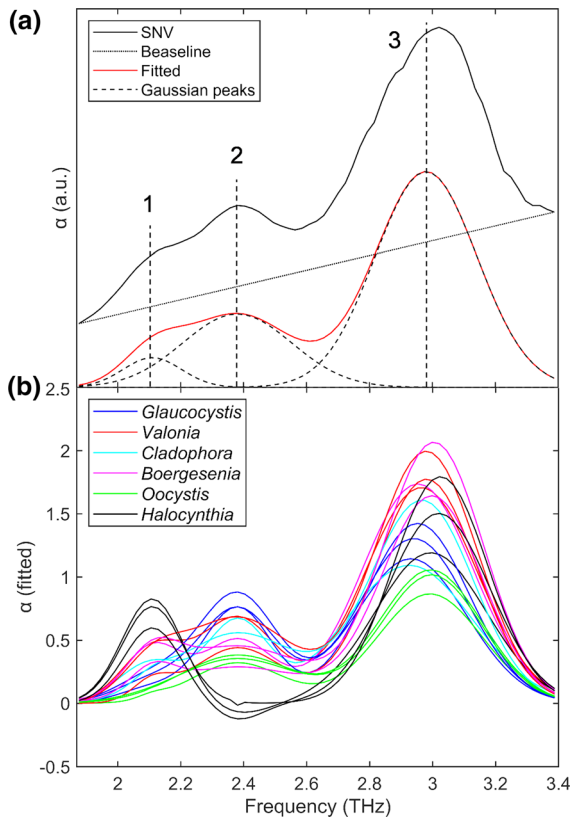
Fig. 4 The processing of THz absorption coefficients spectra, **a** the curve-fitting process, black line: the SNV spectra of *Valonia*. Black dot line: baseline. Black dashed line: separated three Gaussian peaks. Redline: three separated peaks fitted absorption coefficients spectra. **b** the fitted absorption coefficient spectra of cellulose samples with different I_α fractions

The correlations between the integrated intensity of the three peaks, the peak positions of peak3, and the d-spacing, crystalline size and crystallinity index were investigated, and where the correlation coefficient value was more significant than 0.6 were summarized in Table.2.

As shown in Table.2, the integrated intensities of peak1 and peak2 in THz region were correlated with the d-spacing values (d1 and d3), that corresponds to the lattice planes of (100), (110) in the cellulose I_α type and (1_10), (200) in the cellulose I_β type, and these integrated intensities also correlated with the fraction of I_α, where the integrated intensity of peak1 showed higher correlation coefficients of d-spacing values and the fraction of I_α than that of peak2. The correlations between the integrated intensity of peak3 in THz region with the crystalline size and crystallinity index

were different from the peak1 and peak2, where the integrated intensity of peak3 correlated with all three crystalline sizes (Cs1, Cs2 and Cs3) including Cs2 correspond to lattice planes of (010) in the cellulose I_α type and (110) in the cellulose I_β type, and the integrated intensity also showed a good correlation with the crystallinity index of cellulose. It is worth noting that the peak positions of peak3 were correlated with the fraction of I_α and the d2 values which correspond to d-spacing of lattice planes of (010) in the cellulose I_α type and (110) in the cellulose I_β type.

The correlations may indicate that the absorption peaks of 2.11 and 2.38 THz reflected the intermolecular vibrations along the direction between the lattice planes of (100), (110) in the cellulose I_α type and (1_10), (200) in the cellulose I_β type. On the other hand, the absorption peaks near 3 THz might reflect the size of the crystal region, since the integrated intensity of the peak3 was not only correlated with the crystalline size but also correlated with the crystallinity index. It is worth mentioning that the absorption peak near 3 THz showed peak shift, probably due to the absorption peaks here are composed of two peaks that are very close to each other, and the intensity change of the two peaks caused such a slight peak shift. These two peaks may correspond to the cellulose I_α and I_β types, just like the two absorption peaks shown in the 2.11 and 2.38 THz. Under such assumption, the peak position near 3 THz showed a correlation with the fraction of I_α might be reasonable, and the correlation of the d2 values also can be explained by this assumption, if the integrated intensities of the two peaks here are both correlated with the d2 values like the peaks at 2.11 and 2.38 THz correlated with the d1 and d3 values, it may indicate the peaks here near 3 THz reflected the crystal-lattice and intermolecular vibrations along the direction along the lattice planes of (010) and (110) in the cellulose I_α and I_β types (Takahashi 2014).

The partial least squares (PLS) is a common method for analyzing the spectra information. Here the PLS modeling of the THz absorption spectra after SNV and smoothing from 1.88–3.40 THz, the calibration models were constructed using PLS regression with full cross validation. R^2 of predictions, the best coefficients of determination, were 0.98 and 0.68 for the fractions of cellulose I_α and cellulose crystallinity index. For d1-d3, the R^2 were 0.44, 0.28 and 0.22, and for Cs1-Cs3, the R^2 were 0.54, 0.42 and 0.59,

Table 2 Correlations between the THz parameters and the crystallite values calculated from XRD patterns and published I_α fractions

THz parameters	Crystallite values	Correlation coefficient
Integrated intensities of peak1	d1	− 0.86
	d3	− 0.81
	fa	− 0.88
Integrated intensities of peak2	d1	0.71
	d3	0.62
	fa	0.81
Integrated intensities of peak3	Cs1	0.78
	Cs2	0.79
	Cs3	0.76
	CrI	0.86
Peak positions of peak3	d2	0.79
	fa	− 0.61

Cs Crystallite size; *fa* Cellulose I_α fractions; *CrI* Crystallinity index

respectively. Since the purpose of this study did not focus on the prediction, therefore, the detailed discussion about the prediction of crystallite values by using THz-TDS still needs further study.

In this study, we were unable to effectively separate the peaks at 3THz which may due to the relative low S/N ratio. Umesh et al. reported that use the intensity of 93 cm^{-1} Raman band which corresponds to 2.79 THz to calculate the crystallinity of cellulose I_β sample (Agarwal et al. 2018), and the vibrational mode was explained by DFT calculations (Makarem et al. 2019). Thus, without further research such as the quantum mechanical calculation based on density-functional theory (DFT), the assignments of the absorption peaks in the THz region are still not certain.

Conclusions

In this study, various sources of cellulose samples were investigated by THz-TDS and XRD, along with the standardized d-spacing values of the lattice planes of (100), (010) in I_α and (1_10), (110) in I_β, which can be used to determine whether cellulose is I_α or I_β dominant. The d-spacing values were compressed by PCA, the score of PC1 correlated well with the fraction of I_α, which indicated that after PCA processing the XRD pattern can determine the fractions of I_α directly. The THz absorption coefficient spectra were pretreated by SNV to correct baseline fluctuations and smoothed by applying a Savitzky–Golay filter with fifteen points. The absorption coefficients spectra showed two peaks at 2.11 and 2.38 THz, which corresponded to the absorptions due to cellulose I_β and I_α types, respectively; and the intensity at 2.11 and 2.38 THz showed a considerable linear correlation with fractions of I_α. Furthermore, the absorption coefficient spectra in THz region of 1.88–3.40 THz were detrended and fitted as three Gaussian peaks, where two peaks were fixed at 2.11 and 2.38 THz which correspond to the cellulose I_β and I_α types, another was fitted near 3 THz, the integrated intensities of the peaks at 2.11 and 2.38 THz were both showed the correlations with fractions of I_α and the d-spacing values of the lattice planes of (100), (110) in I_α and (1_10), (200) in I_β types. The integrated intensity of the peaks near 3 THz correlated with the crystallinity index and the crystallite sizes of all three lattice planes. In addition, the peak positions of the peaks near 3 THz were correlated with the fractions of I_α and the d-spacing values of the lattice planes of (010) and (110) in I_α and I_β types, respectively. However, the assignments of the absorption peaks in the THz region of cellulose are still not certain. This study demonstrated the potential of THz-TDS as a new technique for investigating I_α fractions and crystalline structures (d-spacing, crystalline size, and crystallinity index) of various sources of cellulose samples. The measurements were rapid and were carried out at room temperature; more importantly, the THz-TDS measurements did not require any pretreatment of the samples, which makes THz-TDS a powerful tool for crystallographic analysis.

References

Agarwal UP, Ralph SA, Reiner RS, Baez C (2018) New cellulose crystallinity estimation method that differentiates between organized and crystalline phases. Carbohydr Polym 190:262–270

Atalla RH, VanderHart DL (1984) Native cellulose: a composite of two distinct crystalline forms. Science 223(4633):283–285

Azizi Samir MAS, Alloin F, Dufresne A (2005) Review of recent research into cellulosic whiskers, their properties and their application in nanocomposite field. Biomacromol 6:612–626

Belton PS, Tanner SF, Cartier N, Chanzy H (1989) High-resolution solid-state ^{13}C nuclear magnetic resonance spectroscopy of tunicin, an animal cellulose. Macromolecules 22:1615–1617

Bishnoi A, Kumar S, Joshi N (2017) Wide-Angle X-ray Diffraction (WXRD). Elsevier 313:337. https://doi.org/10.1016/B978-0-323-46141-2.00009-2

Blackwell J, Vasko PD, Koenig JL (1970) Infrared and raman spectra of the cellulose from the cellwall of valonia ventricosa. J Appl Phys 41:4375–4379

Debzi EM, Chanzy H, Sugiyama J et al (1991) The $I_\alpha \to I_\beta$ Transformation of Highly Crystalline Cellulose by Annealing in Various Mediums. Macromolecules 24:6816–6822

Federici JF, Schulkin B, Huang F et al (2005) THz imaging and sensing for security applications - Explosives, weapons and drugs. Semicond Sci Technol 2(7):S266–S230

French AD (2014) Idealized powder diffraction patterns for cellulose polymorphs. Cellulose 21:885–896

Horii F, Hirai A, Kitamaru R (1987a) CP/MAS^{13}C NMR spectra of the crystalline components of native celluloses. Macromolecules 20:2117–2120

Horii F, Yamamoto H, Kitamaru R et al (1987b) Transformation of Native Cellulose Crystals Induced by Saturated Steam at High Temperatures. Macromolecules 20:2946–2949

Horii F, Yamamoto H, Hirai A (1997) Microstructural Analysis of Microfibrils of Bacterial Cellulose. Macromol Symp 120:197–205

Horikawa Y (2017) Assessment of cellulose structural variety from different origins using near infrared spectroscopy. Cellulose 24:5313–5325

Horikawa Y, Sugiyama J (2009) Localization of crystalline allomorphs in cellulose microfibril. Biomacromol 10:2235–2239

Ikeda Y, Ishihara Y, Moriwaki T et al (2010) A novel analytical method for pharmaceutical polymorphs by terahertz spectroscopy and the optimization of crystal form at the discovery stage. Chem Pharm Bull 58:76–81

Imai T, Sugiyama J (1998) Nanodomains of I_α and I_β cellulose in algal microfibrils. Macromolecules 31:6275–6279

Imai T, Sugiyama J, Itoh T, Horii F (1999) Almost pure I_α cellulose in the cell wall of Glaucocystis. J Struct Biol 127:248–257

Inagaki T, Ahmed B, Hartley ID et al (2014) Simultaneous prediction of density and moisture content of wood by terahertz time domain spectroscopy. J Infrared Millim Terahertz Waves 35:949–961

Larsson PT, Westermark U, Iversen T (1995) Determination of the cellulose $I\alpha$ allomorph content in a tunicate cellulose by CP/MAS ^{13}C-NMR spectroscopy. Carbohydr Res 278:339–343

Makarem M, Lee CM, Kafle K et al (2019) Probing cellulose structures with vibrational spectroscopy. Cellulose 26:35–79

Marchessault RH, Pearson FG, Liang CY (1960) Infrared spectra of crystalline polysaccharides. Biochem Biophys Acta 45:499–507

Nishiyama Y (2009) Structure and properties of the cellulose microfibril. J Wood Sci 55:241–249

O'sullvian AC (1997) Cellulose: the structure slowly unravels. Cellulose 4:173–207

Park S, Baker JO, Himmel ME et al (2010) Cellulose crystallinity index: measurement techniques and their impact on interpreting cellulase performance. Biotechnol Biofuels 3:1–10

Peccianti M, Fastampa R, Mosca Conte A et al (2017) Terahertz absorption by cellulose: application to ancient paper artifacts. Phys Rev Appl 7:1–7

Persson J, Chanzy H, Sugiyama J (1991) Combined infrared and electron diffraction study of the polymorphism of native celluloses. Macromolecules 24:2461–2466

Reid M, Fedosejevs R (2006) Terahertz birefringence and attenuation properties of wood and paper. Appl Opt 45:2766–2772

Strachan CJ, Taday PF, Newnham DA et al (2005) Using terahertz pulsed spectroscopy to quantify pharmaceutical polymorphism and crystallinity. J Pharm Sci 94:837–846

Sugiyama J, Vuong R, Chanzy H (1991) Electron diffraction study on the two crystalline phases occurring in native

cellulose from an algal cell wall. Macromolecules 24:4168–4175

Taday PF, Bradley IV, Arnone DD, Pepper M (2003) Using terahertz pulse spectroscopy to study the crystalline structure of a drug: a case study of the polymorphs of ranitidine hydrochloride. J Pharm Sci 92:831–838

Takahashi M (2014) Terahertz vibrations and hydrogen-bonded networks in crystals. Crystals 4:74–103

Updegraff DM (1969) Semimicro determination of cellulose inbiological materials. Anal Biochem 32:420–424

VanderHart DL, Atalla RH (1984) Studies of microstructure in native celluloses using solid-state ^{13}C NMR. Macromolecules 17:1465–1472

Vieira FS, Pasquini C (2014) Determination of cellulose crystallinity by terahertz-time domain spectroscopy. Anal Chem 86:3780–3786

Wada M, Okano T (2001) Localization of I_α and I_β phases in algal cellulose revealed by acid treatments. Cellulose 8:183–188

Wada M, Sugiyama J, Okano T (1993) Native celluloses on the basis of two crystalline phase ($I\alpha/I\beta$) system. J Appl Polym Sci 49:1491–1496

Wada M, Okano T, Sugiyama J (1997) Synchrotron-radiated X-ray and neutron diffraction study of native cellulose. Cellulose 4:221–232

Wada M, Okano T, Sugiyama J (2001) Allomorphs of native crystalline cellulose I evaluated by two equatorial d-spacings. J Wood Sci 47:124–128

Wang H, Inagaki T, Hartley ID et al (2019) Determination of dielectric function of water in thz region in wood cell wall result in an accurate prediction of moisture content. J Infrared Millim Terahertz Waves 40:673–687

Yamamoto H, Horii F (1993) CP/MAS ^{13}C NMR analysis of the crystal transformation induced for valonia cellulose by annealing at high temperatures. Macromolecules 26:1313–1317

Zsigmondy R, Scherrer P (1912) Bestimmung der inneren Struktur und der Größe von Kolloidteilchen mittels Röntgenstrahlen. Kolloidchem Ein Lehrb 277:387–409