

1 **Probing Cellulose Structures with Vibrational Spectroscopy**

2 Mohamadamin Makarem,¹⁺ Christopher M. Lee,¹⁺ Kabindra Kafle,¹ Shixin Huang,¹ Inseok
3 Chae,¹ Hui Yang,² James D. Kubicki,³ and Seong H. Kim^{1*}

4 1. Department of Chemical Engineering and Materials Research Institute, Pennsylvania
5 State University, University Park, PA 16802, USA.

6 2. Department of Biology, Pennsylvania State University, University Park, PA 16802, USA.

7 3. Department of Geological Sciences, University of Texas at El Paso, El Paso, Texas
8 79968, USA

9 + Equal contributions

10 * Corresponding author: shkim@enr.psu.edu

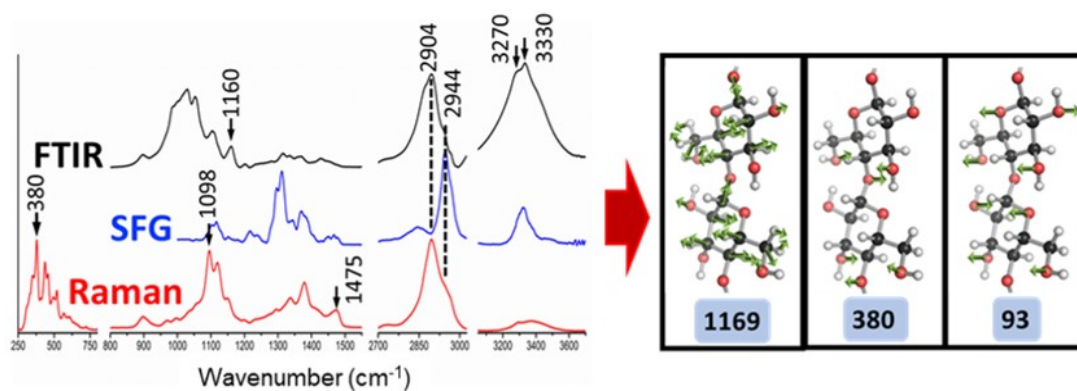
11 Submitted to the 25th Anniversary Special Issue for *Cellulose*

12 **Abstract:** This paper reviews principles, data interpretations, and applications of vibrational
13 spectroscopic methods used for analysis of cellulose in the isolated state and in plant cell walls or
14 lignocellulose biomass. The paper begins with reviewing the crystalline structures of crystalline
15 cellulose polymorphs and the principles of three different vibrational spectroscopy methods –
16 infrared (IR), Raman, and sum frequency generation (SFG) – complemented with density
17 functional theory (DFT) calculations. Then, it discusses the vibrational modes of crystalline
18 celluloses, how the chain orientation in crystalline domain is analyzed in each method, and how
19 the concentration and spatial distribution of crystalline cellulose domains interspersed in
20 amorphous matrices are manifested or analyzed differently in these three methods. Lastly, the
21 paper discusses examples of analyzing crystalline cellulose in plant cell walls or lignocellulose
22 biomass with IR, Raman, and SFG including spectroscopic imaging. One review cannot cover all
23 vibrational spectroscopy literatures on cellulose; this review aims at providing tutorial information,
24 using selected literatures and experimental data, needed to interpret nano-, meso-, and micro-scale
25 structures of cellulose in plant cell walls and lignocellulose biomass.

26 **Keywords:** Cellulose structures, infrared spectroscopy, Raman spectroscopy, sum frequency
27 generation (SFG) spectroscopy, plant cell walls, lignocellulose, biomass.

28 **Graphical Abstract**

29

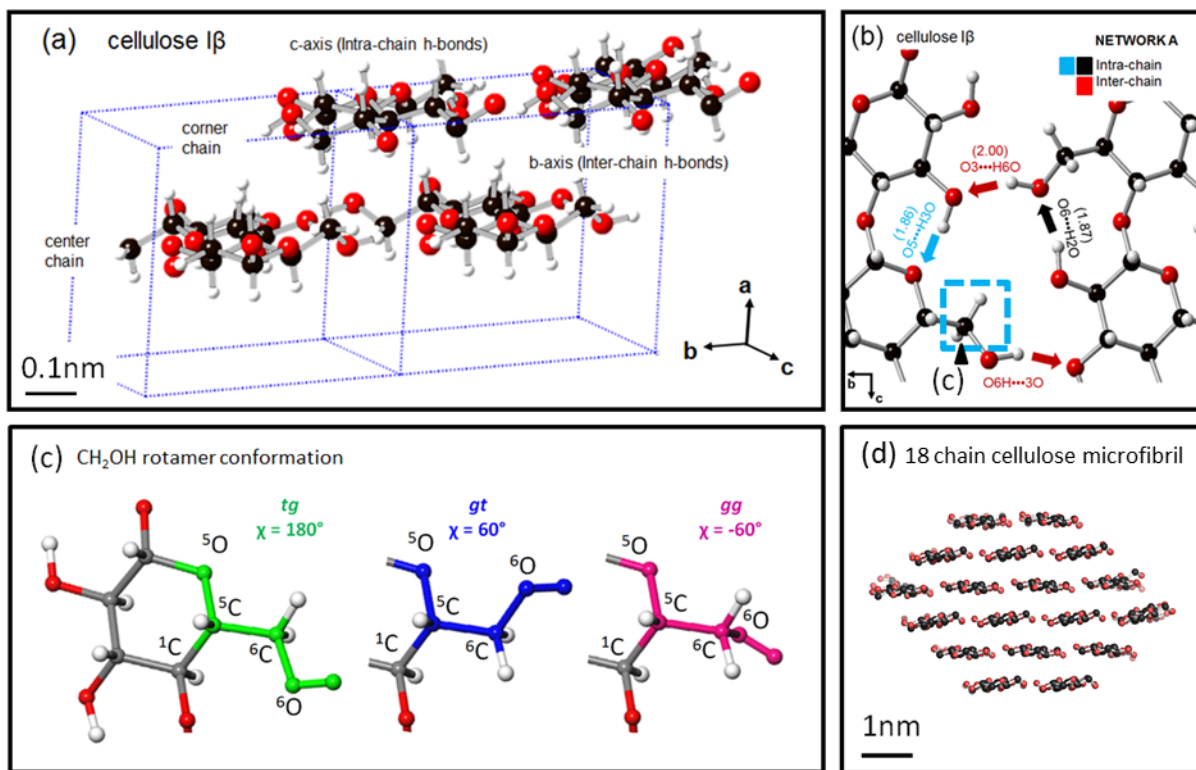


30

31 1. Structures of Crystalline Celluloses

32 1.1 Cellulose chains at the molecular scale

33 Cellulose consists of a linear polymer chain of β -D-glucopyranose monomers connected
34 by 1,4 glycosidic linkages (Figures 1a and 1b). The glucan ring is folded into an arm-chain
35 configuration with all hydroxyl groups (O–H) located on equatorial positions and methine groups
36 (C–H) on the axial positions. The hydroxymethylene group (CH₂OH) at the fifth carbon (C5) from
37 the glycosidic bond can be positioned in three conformations (Figure 1c). In nature, the molecular
38 structure of an individual cellulose chain is intimately related to the polymerization mechanism
39 carried out by subunits of an enzyme complex in the plasma membrane which is usually called
40 cellulose synthase complex (CSC) in the case of plant cells and terminal complex (TC) in the case
41 of algae, bacteria, and tunicates (Nixon et al. 2016; Vandavasi et al. 2016). These enzymes place
42 each successive glucan unit rotated 180° around the chain axis to its nearest neighbor (McNamara
43 et al. 2015). Thus, the repeat unit is composed of the cellobiose moiety (C₁₂H₂₀O₁₀)_n.



44

45 **Figure 1.** (a) Perspective representation of two unit cells of cellulose I β determined from X-ray
46 and neutron diffraction studies. The cellulose chain axis is along the c-direction and the
47 hydrophobic facets of the two alternate sheets are stacked along the a-direction. The C, O, and H
48 atoms are represented as black, red, and light gray, respectively. (b) Hydrogen bonding network
49 with perspective looking down the *bc* plane. The hydrogen bonding network contains three
50 hydrogen bonds 2O-H \cdots 6O (black), 3O-H \cdots 5O (blue), and 6O-H \cdots 3O (red). (c) Three possible
51 conformations, *tg*, *gt* or *gg*, of the OH group in the exocyclic C₆H₂OH group are shown on the
52 right glucose unit. The *t* and *g* characters stand for *trans* and *gauche* conformations, respectively.
53 The first italic character refers to the relative position between O5 and O6 (O5–C5–C6–O6), and
54 the second italic character refers to the relative position between C4 and O6 (C4–C5–C6–O6). χ
55 is the angle between C5–O5 and C6–O6 bonds. (d) Cross section view of a recent 18-chain model
56 for cellulose microfibrils. (Lee et al. 2015d; Lee et al. 2013b; Oehme et al. 2018)

57

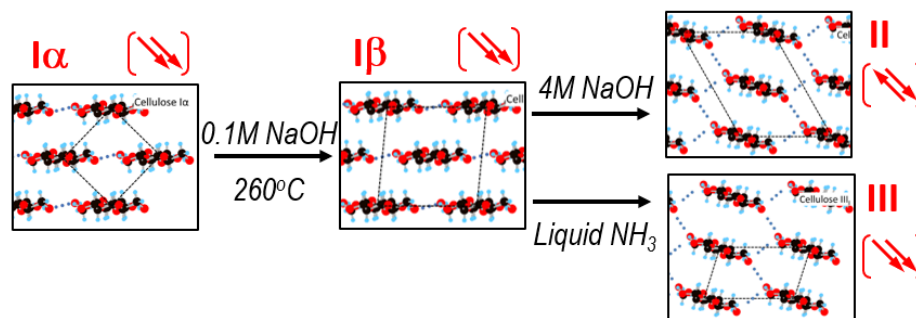
58 ***1.2 Packing of cellulose chains into crystalline orders via hydrogen bonds and van der Waals*** 59 ***interactions at the nanoscale***

60 Multiple cellulose chains are extruded simultaneously from CSCs of the plasma membrane
61 into the extracellular space (Cosgrove 2005; Li et al. 2016; Somerville 2006); because of the
62 uniform structure of linear homopolymers and the close proximity to one another, cellulose chains
63 coming out of CSCs are subsequently packed into crystalline fibrils. Although these fibrils are
64 ~3.5 nm in diameter, (Zhang et al. 2016) they are called microfibrils. The O–H groups form inter-
65 chain hydrogen bonding interactions that stabilize the assembly of cellulose chains side-by-side
66 into sheets (Figure 1b). The C–H groups on the axial positions of the glucan ring are hydrophobic;
67 their van der Waals interactions stabilize the sheet stacking to form a three-dimensional crystal.
68 The number of chains within a single microfibril are approximately 18–36 in plant cell walls; the
69 most recent study elucidated that individual CSCs of plant cells contain 18 cellulose synthesis
70 enzymes (Hill et al. 2014; Nixon et al. 2016; Vandavasi et al. 2016). The most stable structure
71 predicted from computer simulations is shown in Figure 1d (Oehme et al. 2018). In the case of
72 algal cell walls, bacterial pellicles, and tunicate tunics, individual microfibrils contain more chains
73 and up to 15 to 20 nm thick (George and Sabapathi 2015; Moon et al. 2011; Sacui et al. 2014).

74 Cellulose is naturally produced in two crystalline forms – I α and I β . Cellulose produced by
75 algae and bacteria are mostly in the I α allomorph with varying amount of the I β allomorph.

76 Cellulose produced by higher plants and tunicates are the I β allomorph (Atalla and Vanderhart
 77 1984; Lee et al. 2013a). The biosynthesis of cellulose dictates that all chains have their glucan
 78 reducing ends pointing in the same direction, thus cellulose I α and I β crystallites have the parallel
 79 chain polarity within the crystal unit cell (Figure 2). The naturally synthesized cellulose
 80 microfibrils are always mixed with matrix polymers or other components which vary depending
 81 on growth conditions and stages of the cellulose-synthesizing organisms.(Albersheim et al. 2010)
 82 High purity crystalline cellulose can be isolated through acid hydrolysis of cellulose-containing
 83 cell walls or pellicles that relies on differences in kinetics of crystalline and amorphous phase of
 84 cellulose and other polysaccharides (Börjesson and Westman 2015; Habibi et al. 2010).

85



86

87 **Figure 2.** Conversion of cellulose forming different polymorphs. The cross-sectional view shows
 88 the lateral chain packing in the plane perpendicular to the chain axis. The parallel chain polarity is
 89 shown with two red arrows pointing in the same direction, and antiparallel chain polarity is shown
 90 with two red arrows pointing in the opposite directions.

91

92 Most of what is known of the crystal structure of native celluloses has been obtained from
 93 diffraction studies of purified cellulose obtained from algae (I α) or tunicates (I β) (Nishiyama et al.
 94 2002; Nishiyama et al. 2003). These are the largest crystals found in nature (~20 nm in cross
 95 section). The crystallographic models for algal and tunicate celluloses were refined with higher
 96 accuracies because larger crystallites provide a greater number of diffraction peaks (Cullity 1978).
 97 Models available for native cellulose produced in vascular plants are less accurate because
 98 crystallites are an order of magnitude smaller (Fernandes et al. 2011; Thomas et al. 2012). To what

99 extent or accuracy the cellulose crystal structure in land plant cell walls can be represented by the
100 large crystallites of cellulose isolated from aquatic species is debatable (Agarwal et al. 2018).

101 Figure 2 summarizes the polymorphic transition for cellulose reported in the literature.
102 Cellulose I α can be converted to cellulose I β through high-temperature annealing (Sugiyama et al.
103 1990; Wada et al. 2003); but the reverse process has not been reported. Upon treatment with a
104 strong alkali solution (typically, ≥ 4 M NaOH) followed by coagulation, cellulose I α and I β are
105 irreversibly converted to cellulose II (Zugenmaier 2001). When cellulose I (both I α and I β) and II
106 are treated with liquid ammonia, then cellulose III_I and III_{II} are produced, respectively. All
107 allomorphs are chemically identical but differ in the 3-D arrangements of cellulose chains in the
108 unit cell. The most significant differences in crystal structures are the CH₂OH conformation and
109 chain polarity. Figure 1c shows the three possible orientations of the O6 atom of the CH₂OH
110 exocyclic side group; it can be *trans* to O5 and *gauche* to C4 (called *tg*), *gauche* to O5 and *trans*
111 to C4 (called *gt*) or *gauche* to both O5 and C4 (called *gg*). Cellulose I α and I β have the *tg*
112 conformation, whereas cellulose II and III polymorphs have the *gt* conformation (Horii et al. 1983;
113 Wada et al. 2004). Cellulose I α , I β , and III have the parallel directionality for chains in the unit
114 cell, whereas cellulose II has the anti-parallel chain polarity (red arrows shown in Figure 2) (Kim
115 et al. 2006; Langan et al. 2001; Lee et al. 2013a; Nishiyama et al. 2002). The mechanisms for the
116 inversion of chain polarity during conversion from cellulose I to II are debated.

117

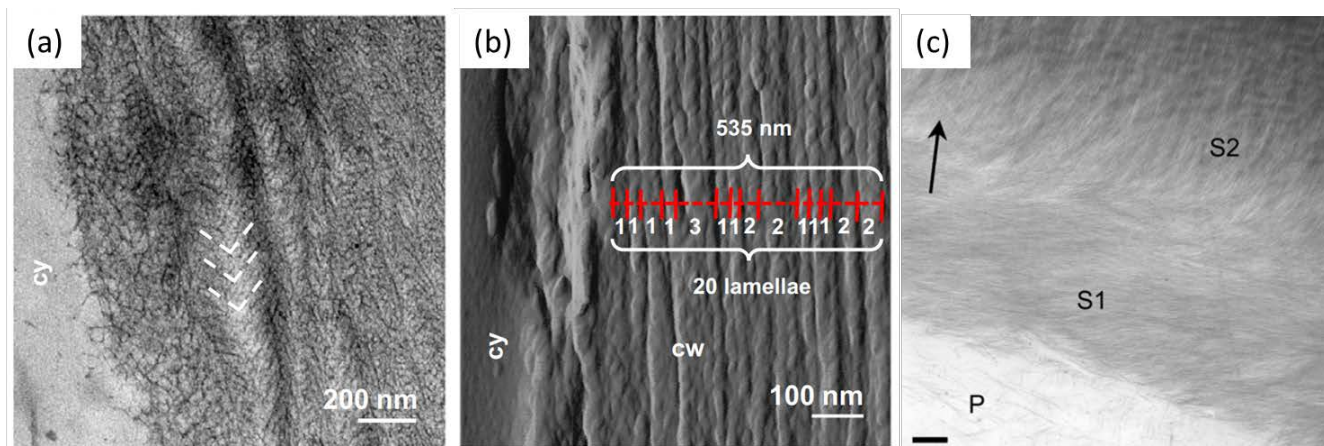
118 ***1.3 Distribution of cellulose crystalline domains or microfibrils in amorphous polymeric*** 119 ***matrices at the mesoscale***

120 Crystalline cellulose is the product of a complex biological machinery. In plants, depending
121 on type and developmental stage, different sets of cellulose synthase (CESA) proteins are activated
122 to produce cellulose microfibrils. In land plants, cellulose is known to be produced by a rosette-
123 shape CSC with six lobes. Each of the lobes has enough space to fit three proteins (Nixon et al.
124 2016; Vandavasi et al. 2016). During cell growth, plants construct thin walls which can support
125 high turgor pressure and at the same time expandable to permit anisotropic cell wall growth; such
126 walls are called primary cell walls (PCWs). Cellulose microfibrils in PCWs are placed or deposited
127 in a multi-lamella structure (Cosgrove 2014). In each lamella, cellulose microfibrils are

128 synthesized with directionality different than the directionality of microfibrils in neighboring
 129 lamella (Figure 3a) (Zhang et al. 2016). Apart from cellulose, other polysaccharide components
 130 such as pectin and hemicellulose, and cell wall proteins can play a major role in plant growth.
 131 Hemicellulose has long been hypothesized to have strong interactions with cellulose microfibrils
 132 surface; however, this hypothesis is contested (Wang and Hong 2015; Wang et al. 2012) . Pectin
 133 molecules forming a gel-like matrix surrounding microfibrils also have close interactions with
 134 cellulose microfibrils (Wang and Hong 2015; Wang et al. 2012).

135 As plant cells stop growing, they start to thicken cell walls to mechanically support the
 136 physical structure of plants; these walls are called secondary cell walls (SCWs). As the SCW
 137 grows, multiple layers named as S1, S2, and S3 are formed in which the dominant angle of
 138 cellulose microfibrils with respect to the main axis of the cell is different. Compared to PCWs, the
 139 packing of microfibrils in SCWs are more aligned along a specific direction, called microfibril
 140 angle (MFA) with respect to the cell main axis (Figure 3b). The MFA varies from species to species
 141 and even among S1, S2, and S2 layers (Barnett and Bonham 2004; Donaldson and Xu 2005).

142



143

144 **Figure 3.** Transmission electron microscopy (TEM) and atomic force microscopy (AFM) of
 145 primary and secondary walls (a) TEM image of onion epidermal cross section, showing angle of
 146 microfibrils in each lamella of primary wall (b) AFM image of onion epidermal cross section,
 147 showing the multi-lamellar structure of primary walls (c) TEM of secondary cell wall in *Radiata*
 148 *pine tracheid*, microfibrils in S2 layer have better organization than S1 (Donaldson and Xu 2005;
 149 Zhang et al. 2016).

150

151 ***1.4 Cellular variances in cellulose distributions at micro-/macro-scale (phenotype)***

152 The type, amount, and mesoscale packing of cell wall polymers can differ. Plants have
153 more than 40 different cell types functioning various roles (Albersheim et al. 2010). The
154 orientation and packing of cellulose microfibrils and their interactions with other cell wall
155 components are intimately related with the functionality of individual cell types (Fujita and
156 Wasteneys 2014). Depending the cell function, it may form PCW only or PCW and SCW layers.
157 The wall formation also depends on the developmental stage of the cell. Cells with load bearing
158 functionality such as interfascicular fibers (IFFs) in angiosperms and xylem cells in gymnosperm
159 plants often develop thicker SCWs (Déjardin et al. 2010). The packing of microfibrils in IFF and
160 xylem secondary walls in Arabidopsis are different. Microfibrils in xylem cells are, in general,
161 more transversely packed compared to interfascicular fiber cells (Terashima et al. 2004; Zhong et
162 al. 2002). This could be related to the functionality of these walls. Epidermis cells that protect
163 plants from surrounding environments develop PCWs only, so that they can keep expanding while
164 plants grow. Although the epidermal wall is known as PCW, its matrix formation is different than
165 other types of PCW such as cortex and pith cells (Fujita et al. 2013). Understanding these
166 dissimilarities in various cell types can lead to better understanding on how matrix polymers are
167 packed and are interacting with each other to form a specific cell wall.

168

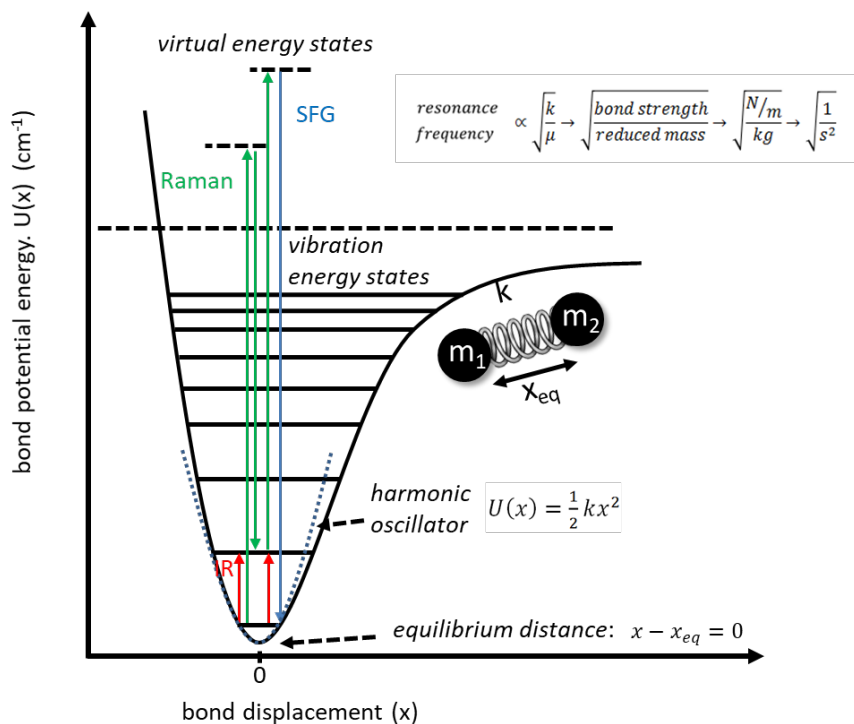
169 **2. Basic Principles of Vibrational Spectroscopy – IR, Raman, and SFG**

170 Infrared (IR) and Raman spectroscopy have been the most widely used vibrational
171 spectroscopic techniques for structural characterization of cell wall polymers. IR and Raman are
172 linear spectroscopy techniques where the induced polarization is linearly related to the electric
173 field strength. IR and Raman have been widely used to study cellulose, pectin, and lignin in PCWs
174 and SCWs. If specific components can be identified or represented with characteristic bands free
175 from interferences from other components, then their intensities can be used to estimate
176 concentrations in the sample using the Beer-Lambert principle (Banwell and McCash 1994).
177 However, if several components have common vibrational features and cannot be distinguished

178 with characteristic peaks, then such identification and quantification of individual components
 179 become difficult (Jolliffe 2011). This has been the case for IR and Raman analysis of cellulose in
 180 plant cell walls and lignocellulose biomass.

181 A new type of vibrational spectroscopy method, sum frequency generation (SFG), has been
 182 employed to cellulose. SFG is a nonlinear optical process in which the induced polarization is
 183 proportional to the square of the electric field strength of light. If the system has random or
 184 inversion symmetry, it cannot generate any SFG signal; only non-centrosymmetric system can be
 185 SFG-active (Lambert et al. 2005; Wang et al. 2005). Thanks to this unique symmetry rule, SFG
 186 can selectively detect crystalline cellulose without spectral interferences from other matrix
 187 components in plant cell walls and lignocellulose biomass (Barnette et al. 2011). In plant cell walls,
 188 crystalline cellulose is the only species that has no inversion symmetry and all other amorphous
 189 matrix polymers cannot produce SFG signals. This section provides brief theoretical background
 190 of these linear (IR and Raman) and non-linear (SFG) vibrational spectroscopy techniques. Figure
 191 4 schematically illustrates photon absorption and emission processes as well as typical
 192 experimental geometries involved in IR, Raman, and SFG spectroscopy.

193



194

195 **Figure 4.** Illustration of electron transitions among vibrational and virtual electronic states
 196 corresponding to IR absorption, Raman scattering, and SFG processes. The potential energy shown
 197 here is for a simple diatomic molecule as a function of distortion of bond (x -axis) from the
 198 equilibrium distance (x_{eq}). The harmonic oscillator approximation is shown in the dotted line. The
 199 vibration energy states and electronic transitions involved in IR absorption (red-up arrow on the
 200 left) and Raman scattering (a pair of green-up/down arrow in the middle) and SFG (a combination
 201 of red-up, green-up and blue-down arrows on the right).

202

203 **2.1 Infrared and Raman spectroscopy – Linear Spectroscopy**

204 The vibration energy states of a di-atomic molecule can be understood using a harmonic
 205 oscillator approximation, where the chemical bond is modeled as a spring connecting two atomic
 206 constituents with masses m_1 and m_2 at an equilibrium bond distance x_{eq} (Figure 4). The spring
 207 with a spring constant (bond stiffness) of k obeys the Hooke's law. In other words, with small
 208 displacements by dx from their equilibrium positions, particles are subject to a restoring force
 209 $F_{res}(dx) = -k \cdot dx$. The harmonic potential energy is the restoring force times the displacement,
 210 $U(dx) = F_{res}dx$. The potential energy deviating from the equilibrium state is $U(x) =$
 211 $\frac{1}{2}k(x - x_{eq})^2$. The potential energy function of a chemical bond is described as a parabola when
 212 undergoing small displacements near the equilibrium position, and the curvature of the potential
 213 energy function is equal to the spring constant $k = d^2U/dx^2$ (dotted line in Figure 4). Solving the
 214 Schrödinger equation for an electron under the harmonic potential gives the quantum-mechanically
 215 allowed vibrational energy states with the energy level:

$$216 \quad E_v = \left(v + \frac{1}{2} \right) \frac{h}{2\pi} \sqrt{\frac{k}{\mu}} \quad (1)$$

217 where h is the Planck constant and v is the vibration quantum number of the allowed state ($v =$
 218 $0,1,2, \dots$), and μ is the reduced mass of the system ($m_1m_2/m_1 + m_2$).

219 In IR spectroscopy, the IR beam interacting with the sample is attenuated due to absorption
 220 of the IR photon (red up arrow in Figure 4) corresponding to the electronic excitation from the
 221 vibrationally ground state ($v = 0$) and to the first vibrationally excited state ($v = 1$). Thus, the IR
 222 absorption peak is observed at a frequency:

223
$$\omega_{IR} = (E_{v=1} - E_{v=0})/h = (1/2\pi)\sqrt{k/\mu} \quad (2)$$

224 The wavelength of ω_{IR} is in the mid-IR region (25 – 2.5 μm ; 400 – 4000 cm^{-1} in the wavenumber
225 unit) for vibrations of organic molecules. This simple di-atomic model shows the relationships
226 between the frequency (energy) of a vibrational mode and the bond strength (through k) and
227 reduced mass of atoms (through μ) involved in that vibrational mode. Note that in the gas-phase,
228 each bond has a specific bond strength and hence k . In condensed systems such as plant cell walls,
229 a given bond strength can be altered via interactions with other atoms, so changes in k indicate
230 changes in bond strength that are induced by plant cell wall interactions. This is especially
231 important for hydrogen bonding interactions of O-H bonds (Libowitzky 1999; Novak 1974; Wang
232 et al. 2014a).

233 The quantum-mechanical selection rule states that the vibrational mode inducing a net
234 dipole moment change can be observed in IR spectroscopy.(Banwell and McCash 1994) Due to
235 this requirement, molecules containing polar functional groups (e.g., O–H and N–H) possessing
236 large permanent dipoles strongly absorb IR light and are readily detected in IR spectroscopy. In
237 an IR spectrum, the attenuation of IR is plotted in the unit of percentage transmittance ($\%T =$
238 $(I_T/I_0) \times 100\%$) or absorbance ($A = -\log(I_T/I_0)$). The intensity in the absorbance unit can be
239 related to the concentration of analyte through the Beer-Lambert Law (Robinson 1996).

240 In near-IR (NIR) spectroscopy, the electronic transition is from the ground state to a higher
241 excited state (typically, $v = 2$ or 3) or involves excitation of two vibrational modes; these are
242 called overtone or combination bands, respectively. NIR spectroscopy will not be covered in this
243 paper; readers with interests in NIR can find other review papers (Siesler et al. 2008).

244 A transient dipole can be induced in the molecule through irradiation of photons in the
245 visible range (ω_{probe}) that promotes the electronic excitation from the ground state ($v = 0$) to a
246 virtual energy state (green up arrow in Figure 4). In that case, the molecule is mostly relaxed back
247 to the ground state. This is the Rayleigh scattering and is elastic in nature; meaning that the energies
248 of the absorbed and emitted photons are the same. In contrast, the Raman scattering is an inelastic
249 process (green down arrow in Figure 4) where the molecule is relaxed from the virtual energy state
250 to the vibrationally-excited state ($v = 1$), emitting a $\omega_{scatter}$ photon at a frequency lower than the

251 probing photon. The energy level difference between the absorbed and emitted photons coincides
252 with the energy level of a vibrational mode: $\omega_{Raman} = \omega_{probe} - \omega_{scatter} = (1/2\pi)\sqrt{k/\mu}$
253 (Figure 4). A fundamental requirement for the Raman scattering is the net change in bond
254 polarizability during the scattering process.(Banwell and McCash 1994) In a Raman spectrum
255 (Figure 4c), the detected intensity is plotted as a function of the Raman shift units
256 ($\omega_{probe} - \omega_{scatter}$). Similar to IR, the Raman intensity can be linearly related to the concentration
257 of analyte through the Beer-lambert Law.

258

259 **2.2 Sum frequency generation (SFG) vibrational spectroscopy – Nonlinear spectroscopy**

260 SFG is a second-order nonlinear optical process that can occur in non-centrosymmetric
261 media (Shen 1984). Because the second-order nonlinear susceptibility of molecule is extremely
262 small compared to the linear susceptibility (such as dielectric constant involved in IR and Raman
263 spectroscopy) (Boyd 1999), SFG requires very high intensity electric fields; normally, pico- or
264 femto-second laser pulses with a high peak power are used. In a typical SFG setup, two laser beams
265 at different frequencies ω_1 and ω_2 are overlapped spatially and temporally at the sample surface.
266 If the non-centrosymmetric selection rules are satisfied at the sample surface or in the sample, a
267 SFG photon can be emitted at the frequency satisfying energy conservation: $\omega_{SFG} = \omega_1 +$
268 ω_2 .(Boyd and Boyd 2003) For vibrational spectroscopic analysis, one of the input photon energy
269 is tuned to the mid IR range ($\omega_1 = 2.5\text{--}10 \mu\text{m}$). The second input photon (ω_2) is typically 532 nm
270 or 1064 nm if a pico-second laser Nd:YAG is used or 800 nm if a femto-second Ti:sapphire laser
271 is used. Mid-IR laser pulses are produced from these laser pulses through an optical parametric
272 generation and amplification (OPG/OPA) process.

273 As shown in the energy level diagram in Figure 4, the absorption of ω_{IR} induces the
274 transition to the vibrationally-excited state with a change in dipole moment (red up arrow in Figure
275 4) and the absorption of ω_{VIS} provides up-conversion to a virtual energy state (green up arrow in
276 Figure 4). The transition back to the ground state (blue down arrow in Figure 4) is accompanied
277 by the emission of photon at ω_{SFG} . Upon scanning ω_{IR} across the frequency of a vibrational mode

278 (ω_q) of a molecule, the ω_{SFG} photon emission is enhanced when $\omega_{IR} = \omega_q$. The ω_{SFG} emission
 279 intensity can be expressed as (Boyd and Boyd 2003):

$$280 \quad \frac{I(\omega_{SFG})}{I(\omega_{VIS})I(\omega_{IR})} \propto |\chi_{eff}^{(2)}|^2 \text{sinc}^2\left(\frac{\Delta k \cdot L}{2}\right) L^2 \quad (3)$$

281 where $I(\omega_{VIS})$ and $I(\omega_{IR})$ are the intensities of the visible and infrared laser pulses and $\chi_{eff}^{(2)}$ is
 282 the second-order nonlinear susceptibility:

$$283 \quad \chi_{eff}^{(2)} \propto N \frac{\sum_{\alpha\beta\gamma} \langle M_{\alpha\beta} A_{\gamma} \rangle}{\varepsilon_0 (\omega_q - \omega_{IR} - i\Gamma_q)} \quad (4)$$

284 where N is the number density of non-centrosymmetrically arranged vibrational groups, ε_0 the
 285 dielectric constant of vacuum, $\langle M_{\alpha\beta} A_{\gamma} \rangle$ the angle-average product of the Raman polarizability
 286 tensor and IR transition dipole moment vector, ω_q the frequency of the q -th vibrational mode, and
 287 Γ_q the damping constant of the mode. The last term in Eq.(3), $\text{sinc}^2\left(\frac{\Delta k \cdot L}{2}\right)$, is the synchronization
 288 factor which is a function of the phase mismatch between SFG, visible, and IR beams ($\Delta k =$
 289 $\vec{k}_{SFG} - \vec{k}_{VIS} - \vec{k}_{IR}$ where \vec{k} is the momentum of each light) and the sample size (L).

290 Eq.(4) shows that when ω_{IR} approaches the ω_q , the denominator of the $\chi_{eff}^{(2)}$ term
 291 becomes small and thus the $\chi_{eff}^{(2)}$ value becomes large. In other word, the SFG process is
 292 resonantly enhanced when $\omega_{IR} = \omega_q$. In addition, this equation shows that the vibrational mode
 293 should be both Raman- and IR- active; otherwise, the numerator of the $\chi_{eff}^{(2)}$ term becomes zero.
 294 If the molecule has an inversion symmetry, then its vibrational modes are either Raman-active
 295 only or IR-active only; such molecules cannot be SFG-active. In addition to the non-
 296 centrosymmetry at the molecular lever, the non-centrosymmetric distribution of molecules or
 297 vibrational functional groups in space over a so-called coherence length is also important. The
 298 coherence length (L_c) characteristic to SFG originates from the momentum conservation, also
 299 called “phase matching”, of the nonlinear optical process mixing three waves (ω_{SFG} , ω_{VIS} , and
 300 ω_{IR}). (Boyd and Boyd 2003) It can be estimated from the synchronization factor: $L_c \approx 2/\Delta k$
 301 (Boyd and Boyd 2003). The SFG coherence length varies depending on the dispersion of refractive
 302 index of the material being probed and the SFG detection geometry (Lee et al. 2014).

303 In summary, each vibrational spectroscopic technique (IR, Raman and SFG) provides
304 complementary structural information due to differences in optical processes being responsible for
305 measured signals. For IR detection, vibration modes must have a net change in its dipole moment
306 during the vibrational excitation. For Raman detection, vibrational modes should have a net change
307 in the polarizability. In SFG, vibration modes must be non-centrosymmetric at the molecular level
308 and arranged without inversion symmetry over the SFG coherence length scale. Advantages and
309 limitations inherent to each technique in structural analysis of crystalline cellulose will be
310 discussed in the sections following the next sub-section.

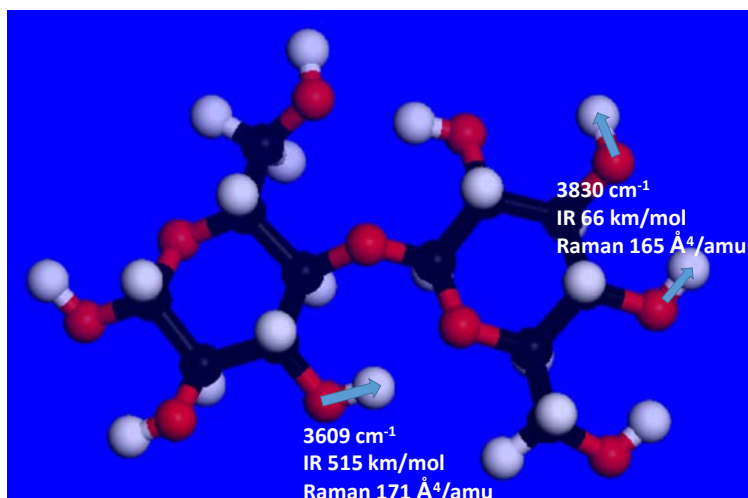
311

312 *2.3 Theoretical calculations of vibrational modes and their frequencies*

313 For theoretical analysis of vibrational modes of cellulose, the most widely used quantum
314 mechanical method is density functional theory (DFT). DFT calculates structural and electronic
315 properties of a many-electron system by using functionals based on the spatially-dependent
316 electron density (Koch and Holthausen 2015). In organic crystals, van der Waals attractions play
317 important roles but are difficult to reproduce with conventional DFT calculations because DFT
318 does not represent the temporary fluctuations in electron density that occur in nearby orbitals. To
319 take van der Waals interactions into account, DFT with dispersion corrections (with acronyms of
320 D2, D3, DF1) based on empirical fitting of training sets from simple gas-phase molecules are
321 added (Dion et al. 2004; Grimme 2006; Grimme et al. 2010).

322 To obtain DFT vibrational frequencies, the molecular structure of the system is first
323 energy-minimized with respect to all atoms, then the second derivatives of the energy with respect
324 to all atomic coordinates (i.e., the Hessian matrix) are calculated (Koch and Holthausen 2015).
325 Finite-difference approximations or analytical solutions (Wilson et al. 1955) are used to calculate
326 to the Hessian matrix from which the characteristic resonance frequencies are derived. Periodic
327 modeling programs typically use finite-difference methods, but analytical solutions are also
328 available to calculate IR (Yamaguchi et al. 1986) and Raman intensities (Frisch et al. 1986) of
329 non-periodic models. The normal modes of vibrations are predicted from the eigenvectors of the
330 matrix and the vibrational frequencies are calculated from the eigenvalues of the matrix (Bučko et
331 al. 2011; Kubicki et al. 2013; Li et al. 2011). An example is shown in Figure 5.

332



333

334 **Figure 5.** Frequency calculation of cellobiose performed in Gaussian 09 (Frisch et al. 2010) with
335 analytically calculated frequencies and IR/Raman intensities. Note that the C2-OH group is H-
336 bonded to the O5 atom which results in a lower vibrational frequency for this O-H stretch due to
337 weakening of the O-H bond. The 3830 cm^{-1} mode involves two O-H groups, neither of which is
338 H-bonded. The positive charges on H and the negative charges O atoms change as the O-H bond
339 stretches. This vibration thus changes the dipole moment and the polarizability which affects the
340 IR and Raman intensities, respectively.

341

342 Constructing the initial structural model of crystalline cellulose for DFT-D2 calculations,
343 the atomic coordinates of the structural model determined from x-ray diffraction (XRD) can be
344 used (Nishiyama et al. 2002; Nishiyama et al. 2003). After the energy minimization, testing the
345 accuracy of the DFT calculations to the available experimental data is necessary. To test the energy
346 minimization step, the atomic positions could be compared to the XRD data. The benchmark XRD
347 experimental values include: bond lengths, bond angles and torsion angles as well as the unit cell
348 lengths and interaxial angles (Bučko et al. 2011; Kubicki et al. 2013; Li et al. 2011). Once the DFT
349 protocol for energy minimization is optimized, other calculations can be performed that include:
350 IR and Raman vibrational frequencies and ^{13}C NMR chemical shifts (Kubicki et al. 2013). The
351 predicted frequencies and NMR chemical shifts can then be compared to observed values.

352 Limitations of DFT calculations include the fact that the theoretical frequencies calculated
353 using the Hessian matrix under the harmonic approximation need to be re-scaled to match with the

354 experimental values to account for the anharmonic nature of observed vibrational frequencies
355 (Johnson III et al. 2010; Laury et al. 2012). The OH stretch frequency is sensitive to changes in
356 hydrogen bonding interactions (Libowitzky 1999; Novak 1974; Wang et al. 2014a). Thus, the
357 accuracy in predicting the OH stretch frequency is limited.

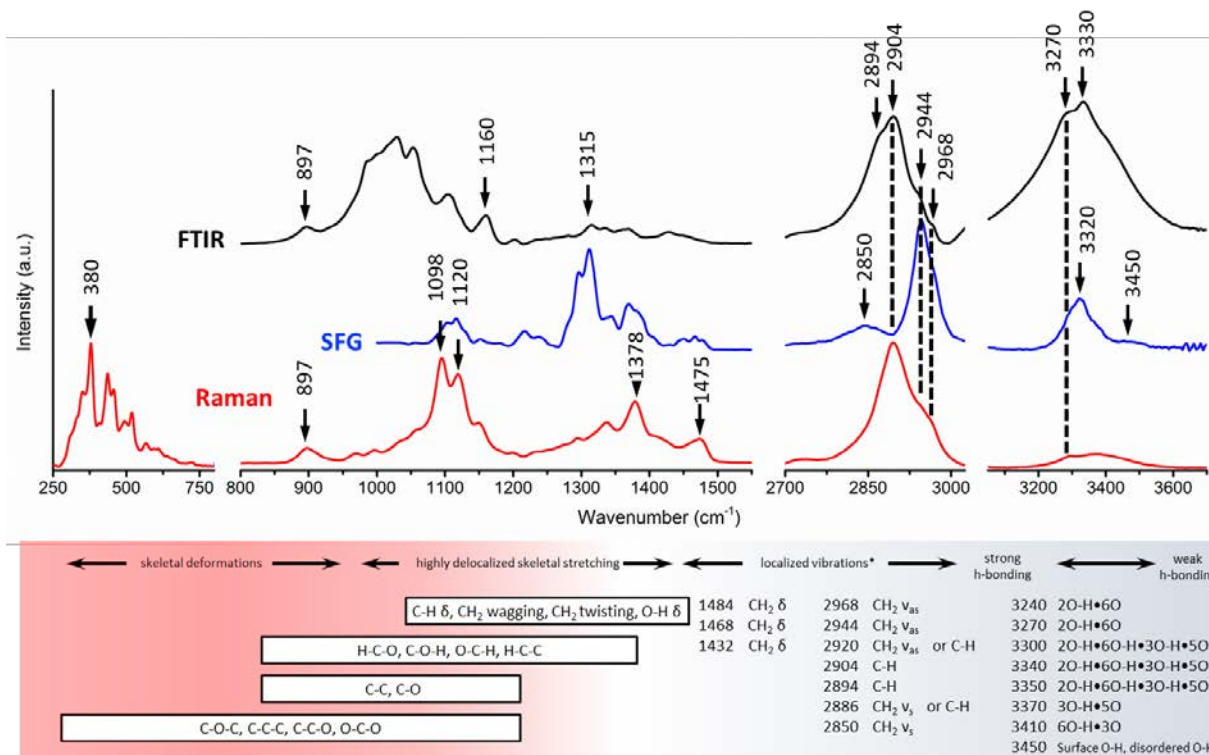
358 The Hessian matrix analysis of the DFT-calculated structure can provide normal modes
359 and their frequencies (ω_q in eq.4); however, it cannot predict the peak intensity (the numerator
360 term in eq.4) and width (Γ_q in eq.4). To predict the peak intensity, one should use a method that
361 can calculate the electronic structure at the excited state in addition to the ground state. This is
362 done with time-dependent DFT (TD-DFT) calculations (Marques and Gross 2004). TD-DFT can
363 calculate the polarizability and dipole derivatives at the equilibrium geometry distorted along the
364 normal mode coordinates (Bishop et al. 1997; Reiher et al. 2003). Then, theoretical peak intensities
365 can be calculated. The peak width is a function of dephasing dynamics of vibrational polarization
366 (Velarde and Wang 2013). In most computational predictions of vibrational spectra using DFT
367 and TD-DFT, the peak width is typically chosen to match experimental spectral features rather
368 than being calculated with first principles models.

369

370 **3. Vibrational Modes of Crystalline Cellulose**

371 Figure 6 compares the IR, Raman, and SFG spectra of cellulose I β . Some of characteristic
372 peak positions are marked in each spectrum. The vibrational modes involving polar groups such
373 as OH stretches in the 3200 – 3600 cm^{-1} and C-O stretches near 1000 cm^{-1} are relatively stronger
374 in IR, but weak in Raman. In both IR and Raman spectra, the CH and CH₂ stretches in the 2800 –
375 3000 cm^{-1} region are prominent. The 2904 cm^{-1} peak is dominant in both IR and Raman spectra of
376 this region, but that peak is absent in SFG and the peak at 2944 cm^{-1} and the shoulder at 2968 cm^{-1}
377 are prominent in the SFG spectrum. Based on the Beer-Lambert law, the most dominant peak in
378 IR and Raman spectra could be assigned to the most abundant species. In the cellobiose monomeric
379 unit, (C₁₂H₂₀O₁₀)_n, of cellulose, there are 10 axial CH groups and 2 exocyclic CH₂ groups; thus,
380 the 2904 cm^{-1} main peak in IR and Raman is likely due to the vibrational modes involving these
381 axial CH groups. In that structure, the equal number of axial CH groups are pointing the opposite

382 sides of the molecular plane. Thus, it could be argued that the symmetric arrangement of axial CH
 383 groups may make the 2904 cm⁻¹ peak inactive in SFG (Barnette et al. 2011). Further details of
 384 peak assignments or interpretation require comparison with computational calculations.



References for assignments are: a(Cael et al. 1974) b(Wiley and Atalla, 1987) c(Lee et al. 2013b) d(Lee et al. 2015d)

385
 386 **Figure 6.** Comparison of IR, Raman, and SFG spectra of pure cellulose I β film. The SFG
 387 polarization combination is *ssp*; details of SFG polarizations are explained in Section 3.3.3.
 388 (Barnette et al. 2011; Lee et al. 2015d; Lee et al. 2013b)

389
 390 **3.1 Peak assignments for native cellulose (I β and I α allomorphs)**

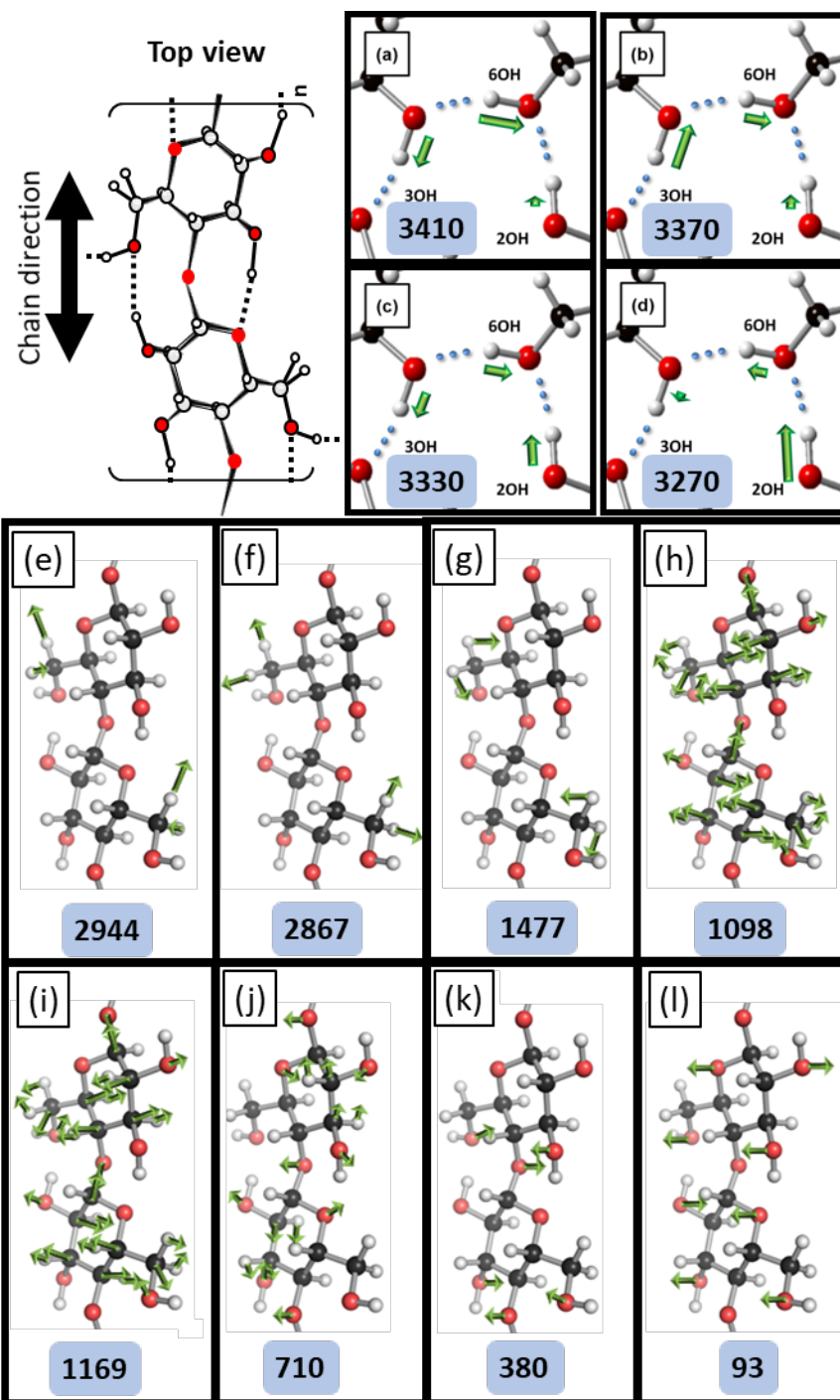
391 The vibrational modes of crystalline phases are complicated because all functional groups
 392 are connected through the inter-molecular hydrogen bond networks as well as the intra-chain
 393 covalent bonds, imposing structural constraints of all atoms to be positioned at specific locations
 394 in the crystalline lattice. Assignments of cellulose vibrational peaks cannot be done through
 395 comparisons with similar peaks in the same frequency range observed for small molecule
 396 analogues. In isolated molecules in the gas or liquid phase, the internal vibrations of molecules are

397 independent from of other molecules. However, in the crystalline domain, no molecular units can
398 vibrate independently from others. The vibrations of functional groups are all coupled through
399 covalent bonds and other constraints (such as hydrogen bonds). In other words, the vibrational
400 modes of cellulose detected in IR, Raman, and SFG should be considered like phonon modes of
401 inorganic crystalline solids (Ruppin and Englman 1970).

402 An attempt has been made to assign each peak in the vibrational spectrum of cellulose to a
403 specific single functional group based on spectral changes after selective OH/OD exchanges or the
404 directionality of OH group in the unit cell with respect to the chain axis (Fan et al. 2012; Hofstetter
405 et al. 2006; Maréchal and Chanzy 2000). However, it is inappropriate to consider, for example, the
406 isolated vibration of the O6-H group of one chain. That is because, when it vibrates, it will affect
407 the O3-H group of the next chain to which the O6-H group is hydrogen-bonded, which may affect
408 the motions of the C atoms in the arm-chair ring because the O3-H group is hydrogen bonded to
409 the O5 atom of the ring (see Figure 1b). A more advanced approach would be computer simulation
410 approaches. The first attempt was made by Blackwell and co-workers (Cael et al. 1975); however,
411 the crystal structure used in that study was not accurate.

412 DFT-D2 calculations with the cellulose I α and I β structures using the best known crystal
413 structures from x-ray and neutron diffractions were carried out recently (Lee et al. 2015d; Lee et
414 al. 2013b). In these studies, all atoms were initially allowed to fully relax from the experimentally-
415 determined coordinates during the energy minimization process and then found that the peak
416 positions for the OH stretch bands were significantly lower than the experimentally observed
417 values (Lee et al. 2013b). This was due to the maximization of the hydrogen bonding interactions
418 for energy minimization to the lowest possible value. When the C and O atoms in the crystal
419 structure were fixed to the coordinates determined by the diffraction study and only H atoms were
420 allowed to relax during the energy minimization, then the calculated OH peak positions were in
421 better agreement with the experimental values (Lee et al. 2015d). Figure 7 displays some of the
422 vibrational modes simulated from these DFT-D2 calculations.

423



424

425 **Figure 7.** Vibrational modes of cellulose I β simulated from DFT-D2 calculations. (a-d) represents
 426 OH stretch modes which is a combination of different OH groups motions. (e-f) CH₂ asymmetric
 427 and symmetric stretch which are coupled with CH stretch on the ring. (g) CH₂ bend mode which
 428 is fairly isolated showing very weak coupling with other vibrations. (h-i) skeletal vibrations,
 429 showing bending and stretching in armchair atoms and glycosidic C-O-C bonds. (j) vibration mode
 430 used in IR spectroscopy to measure cellulose crystallinity. (k-l) vibration modes used to calculate

431 the crystallinity in Raman spectroscopy. Movies of all vibrational modes shown here can be found
432 in the Supporting Information (Kafle et al. 2017; Lee et al. 2015d; Lee et al. 2013b).

433

434 Figures 7a – 7d show four modes in the OH stretch region. No mode can be assigned to a
435 single OH group; all OH stretch modes visualized from DFT-D2 calculations show collective
436 motions of the O2H, O3H, and O6H groups with varying degrees of contribution for each OH
437 group. This vibrational coupling is due to hydrogen bonding interactions. Figures 7e and 7f display
438 the CH stretch modes coupled with the CH₂ asymmetric and symmetric stretches, respectively.
439 Even the stretch vibrations of the CH₂ group of the exocyclic side chain are coupled with those of
440 the CH groups in the six-membered ring. Figure 7g show the CH₂ bend mode. This mode appears
441 to be fairly isolated from or negligibly coupled with the motions of other atoms in the cellulose
442 crystal. Figures 7h and 7i exhibit two modes involving stretching and bending motions of atoms
443 in the armchair rings and the glycosidic C-O-C bonds. In the case of skeletal vibrations in the 1000
444 – 1200 cm⁻¹ region, the vibrational peak cannot be assigned to a single C-C or C-O bond. The peak
445 at ~1160 cm⁻¹ has been attributed to the C-O-C glycosidic bond stretch (Chen et al. 2010; Popescu
446 et al. 2007; Wilson et al. 2000); however, Figure 7h shows that it cannot be assigned the glycosidic
447 bond only; when the C-O-C links are vibrationally excited, then all other atoms in the rings
448 attached by that linkage are affected (Kafle et al. 2017).

449 Figures 7j -7l display vibrational modes used for crystallinity calculation (see Section 4.1).
450 These are the modes existing only in the crystalline phase because they are subject to specific
451 structural constraints of the crystalline lattice to which amorphous phase chains are not subjected.
452 The 710 cm⁻¹ mode shown in Figure 7j includes out-of-plane bending of C-O-C glycosidic bonds
453 and O2H and O3H groups and in-plane distortion of armchair rings (movement of each bond can
454 be seen in movie S7J in the Supporting Information). The 380 cm⁻¹ mode shown in Figure 7k has
455 been used in Raman spectroscopy for measuring crystallinity of cellulose. It involves out-of-plane
456 bending of C-O-C glycosidic bonds and O6H groups coupled with simultaneous expansion and
457 contraction in all rings. The 93 cm⁻¹ mode shown in Figure 7l is also used in Raman spectroscopy
458 for cellulose crystallinity calculation. The twisting of the entire chain along the c-axis of cellulose
459 crystal is obvious from movie S7L in the Supporting Information. The alternative out-of-plane
460 vibration of C5-O6-C1 of the six-membered rings creates a twist of chains in the crystalline lattice.

461 Comparing these DFT-D2 calculation results (Lee et al. 2015d; Lee et al. 2013b) with
462 experimental spectra reported in the literature (Abidi et al. 2008; Blackwell 1977; Blackwell et al.
463 1970; Kacurakova et al. 2000; Kafle et al. 2017; Lee et al. 2013a; Szymanska-Chargot and Zdunek
464 2013; Tsuboi 1957; Wiley and Atalla 1987), tentatively peak assignments for vibrational peaks
465 observed in IR, Raman, and SFG spectra of crystalline cellulose I α and I β are made and
466 summarized in Table 1. It is difficult to accurately describe complicated motions of atoms in script.

467

468 **Table 1.** Positions of vibrational peaks of cellulose I α and I β observed in IR, Raman, and SFG
469 spectroscopy with tentative assignments.

Frequency	Assignment	Reference
93 cm ⁻¹	Twist along the chain direction (Figure 7l) (SI, Movie S7L)	This paper
380 cm ⁻¹	Stretching of glucose ring (Figure 7k) (SI, Movie S7K)	Lee et al. 2013b
710 cm ⁻¹	Twisting along chain direction, Iβ (Figure j) (SI, Movie S7J)	Lee et al. 2013b
740 cm ⁻¹	Twisting along chain direction, Iα	Lee et al. 2013b
896 cm ⁻¹	Glucose ring deformation	Blackwell 1977
970-1050 cm ⁻¹	C-C and C-O vibration	Tsuboi 1957
1090-1120 cm ⁻¹	Skeletal deformation (Figure 7h) (SI, Movie S7H)	Lee et al. 2013b
1160 cm ⁻¹	Skeletal deformation (Figure 7i) (SI, Movie S7I)	Lee et al. 2013b
1315 cm ⁻¹	C-C-H, C-O-H, and CH ₂ bending	Blackwell 1977
1338-1378 cm ⁻¹	C-C-H, C-O-H, and O-C-H	Wiley and Atalla 1987
1470 cm ⁻¹	CH ₂ bending (Figure 7g) (SI, Movie S7G)	Lee et al 2013b
2850-2880 cm ⁻¹	CH ₂ symmetric stretching (Figure 7f) (SI, Movie S7F)	Lee et al 2013b
2894-2904 cm ⁻¹	CH stretching	Lee et al 2013b
2944 cm ⁻¹ , 2968 cm ⁻¹	CH ₂ asymmetric stretching (Figure 7e) (SI, Movie S7E)	Lee et al 2013b
3240 cm ⁻¹	Stretching of intrachain 2O-H...6O, Iα	Lee et al 2013b
3270 cm ⁻¹	Stretching of intrachain 2O-H...6O, Iβ (Figure 7d) (SI, Movie S7D)	Lee et al. 2015d
3300-3330 cm ⁻¹	Stretching of coupled 2O-H...6O-H...3O-H...5O (Figure 7c) (SI, Movie S7C)	Lee et al. 2015d
3370 cm ⁻¹	Stretching of intrachain 3O-H...5O (Figure 7b) (SI, Movie S7B)	Lee et al. 2015d
3410 cm ⁻¹	Stretching of intrachain 6O-H...3O (Figure 7a) (SI, Movie S7A)	Lee et al. 2015d
3450 cm ⁻¹	Weakly hydrogen bonded OH groups on the surface of crystal	Lee et al. 2015d

470

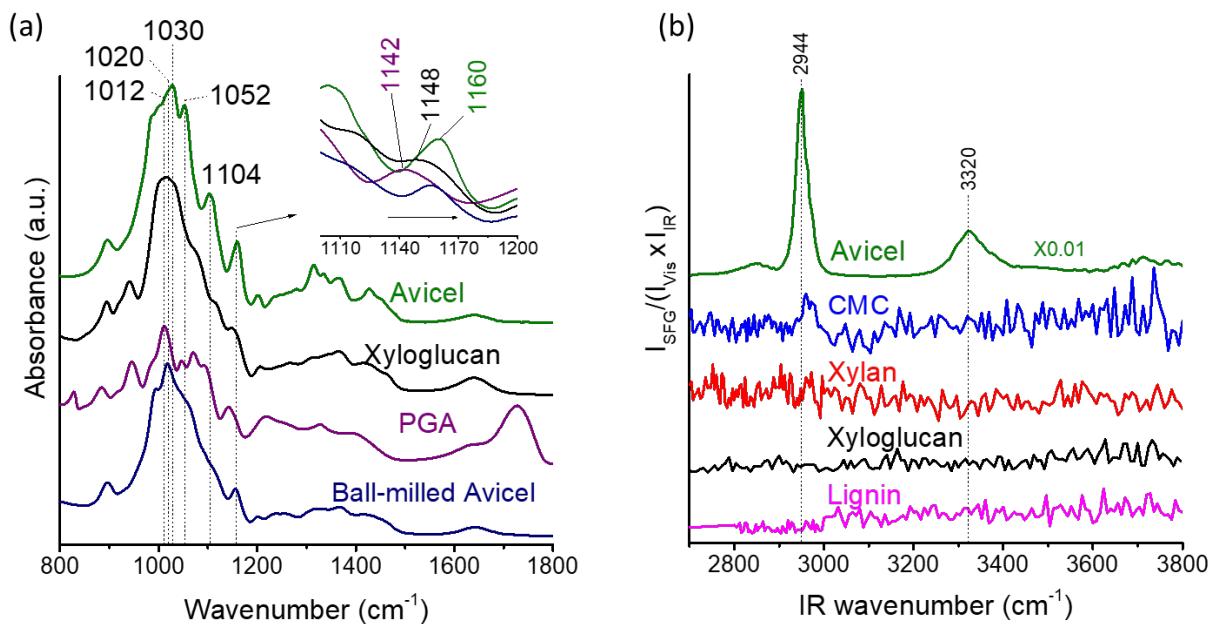
471

472 If IR and Raman spectra of pure components are compared, it is possible to distinguish
473 cellulose from other matrix polymers in plant cells such as pectin, hemicellulose, and lignin.
474 However, this distinction is challenging for IR and Raman spectra of plant cell walls although it
475 has been attempted. The main reason is that the linear spectroscopy detects everything in the beam

476 path and the differences between crystalline cellulose and other matrix polymers are subtle. This
477 is because matrix polymers have the same functional groups as cellulose – C-H, C-C, C-O, and O-
478 H.

479 Figure 8a compares the fingerprint region of the IR spectra of crystalline cellulose (modeled with
480 Avicel), amorphous cellulose (produced by ball milling), xyloglucan, and polygalacturonic acid
481 (PGA, as a model for pectin). Studies (Blackwell et al. 1970; Schwanninger et al. 2004; Wilson et
482 al. 2000) have observed the peak at 1160 cm^{-1} as a characteristic peak for the glycosidic bond
483 vibration in the crystalline cellulose. However, the data shown in the inset of Figure 8a show that
484 all polysaccharides have the peaks at the similar location; thus, resolving them into different
485 species would be extremely difficult. The band in the $1140 - 1160\text{ cm}^{-1}$ region is sharper than the
486 band in the $1000 - 1000\text{ cm}^{-1}$ region but not sharp enough to distinguish the subtle differences
487 among species. This band is not purely due to the C-O-C glycosidic bond vibration only (Figure
488 7i). In contrast, IR and Raman are useful to detect pectin and lignin species, respectively. The
489 carboxyl groups ($-\text{COOH}$, $-\text{COOCH}_3$) of pectin molecules have characteristic bands near 1750 cm^{-1}
490 $^{-1}$ in IR (Sene et al. 1994; Synytsya et al. 2003). The aromatic groups of lignin have characteristic
491 bands at $1600 - 1650\text{ cm}^{-1}$ in Raman (Agarwal and Ralph 1997; Northey et al. 2000).

492



493

494 **Figure 8.** (a) IR spectra of Avicel (crystalline cellulose), xyloglucan, polyglutrunoic acid, and ball-
495 milled Avicel (amorphous cellulose). (b) SFG spectra of Avicel, carboxymethyl cellulose (CMC),
496 xylan, xyloglucan, and lignin. Note that except the Avicel spectrum (the intensity of which is
497 reduced to compare with others), there are no discernable peaks above the noise level for
498 amorphous polymers. The SFG polarization combination is *ssp*. (Barnette et al. 2011; Kafle et al.
499 2017)

500

501 In SFG analysis, the situation is opposite. Figure 8b compares the SFG spectra collected
502 for Avicel, carboxymethyl cellulose (CMC; cellulose gum), xylan, xyloglucan, and lignin. All
503 matrix polymers in plant cell walls are amorphous; thus, they cannot produce any detectable SFG
504 signals. Cellulose derivatives are amorphous, so they cannot generate SFG signals either. Only
505 crystalline cellulose can generate SFG signals effectively due to the non-centrosymmetry in its
506 crystalline structure (see Section 2.2). Thus, SFG is highly efficient for selectively probing
507 cellulose in plant cell walls. Note that SFG cannot detect pectin, hemicellulose, and lignin in plant
508 cell walls; however, if their compositions or structures are modified in plant cell walls through
509 genetic mutations or environmental stresses, then the cellulose deposition pattern in plant cell walls
510 may be altered and such changes can be detected with SFG (see Section 5.3)

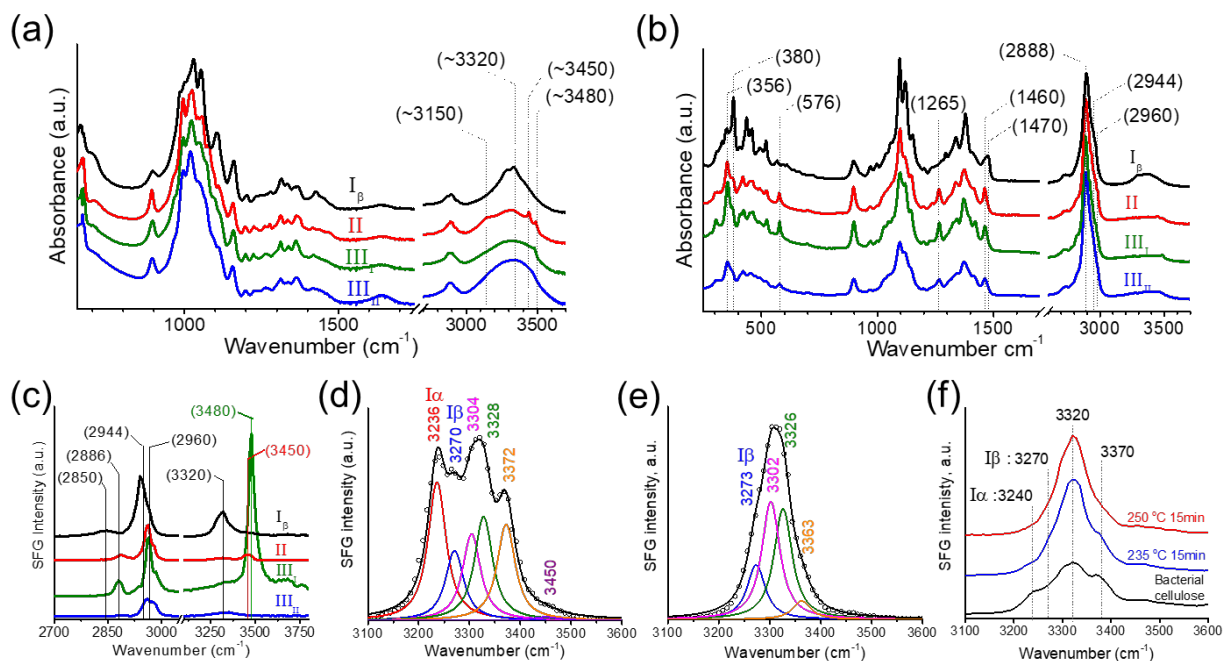
511

512 **3.2 Detection of cellulose polymorphism**

513 Polymorphic structures of crystalline cellulose can be distinguished with IR, Raman, and
514 SFG. Figure 9a compares the IR spectra of cellulose I β , II, III_I and III_{II} (Lee et al. 2013a). The
515 most prominent difference is the OH stretch bands that are sensitive to the hydrogen bond networks
516 in the crystal structures (Figures 1 and 5). Cellulose I β can be distinguished with the peak 3270
517 cm⁻¹ and cellulose I α has a peak at 3240cm⁻¹ (Table 1). The characteristic feature of cellulose II is
518 two peaks at 3450 cm⁻¹ and 3480 cm⁻¹ as well as a weak shoulder at ~3150 cm⁻¹. Cellulose III_I
519 show only one peak at 3480 cm⁻¹ (Lee et al. 2013a). In Figure 9a, the characteristic feature of
520 cellulose III_{II} is not clearly seen due to a lower crystallinity.

521 In Raman spectra of cellulose polymorphs (Figure 9b) (Lee et al. 2013a), the OH stretch
522 intensities are relatively weak compared to the IR spectra. This is because the Raman signal is

523 sensitive to the polarizability change, whereas the IR signal is to the dipole moment change (see
 524 Section 2.2). In the Raman spectrum of cellulose I β , the peak with a large contribution from the
 525 CH₂ bending mode can be seen at 1470 cm⁻¹, while that peak is shifted to ~1460 cm⁻¹ in spectra of
 526 other allomorphs. This can be attributed to the difference in the conformation of the CH₂OH
 527 exocyclic side chain (Figure 1c). Cellulose I β (and cellulose I α as well) has the *tg* conformation,
 528 while other allomorphs have the *gt* conformation (Lee et al. 2013b). Another difference is that
 529 cellulose I β show no peak at 576 cm⁻¹; other crystalline allomorphs have a noticeable peak at 576
 530 cm⁻¹. Also, cellulose I β has a strong peak at 380 cm⁻¹, and other allomorphs have a strong peak at
 531 356 cm⁻¹.



532
 533 **Figure 9.** (a) IR, (b) Raman, (c) SFG spectra of cellulose I β , II, III_I, and III_{II} (Lee et al. 2013a).
 534 (d,e) Comparison of cellulose in *Glaucozystis* and *Halocynthia* (Lee et al. 2014). (f) SFG spectra
 535 of cellulose produced by *Glucanacetobacter* upon heating in 0.1M NaOH. In (c-f), the SFG
 536 polarization combination is *ssp*.

537
 538 In the case of SFG study, the CH/CH₂ stretch and OH stretch regions have been studied
 539 and reported, while the IR and Raman studies include comparison of spectral features in the
 540 fingerprint region (<1500 cm⁻¹). The high degree of vibrational mode coupling in the fingerprint

541 regions (Figure 7) makes the interpretation of the SFG non-linear optical signals complicated.
542 Figure 9c displays the SFG spectra of cellulose I β , II, III_I, and III_{II} (Lee et al. 2013a). The main
543 peak in the CH/CH₂ stretch region is 2944 cm⁻¹ for cellulose I β , while it is ~2964 cm⁻¹ for other
544 allomorphs. Similar to the 1470 vs. 1460 cm⁻¹ band in Raman, this difference could be attributed
545 to the difference in the *tg* and *gt* conformations of the CH₂OH exocyclic side chain in the crystal
546 unit cells (Lee et al. 2013a). In the OH stretch region, cellulose I β and III_I have strong bands, and
547 cellulose II and III_{II} have very weak or negligible bands. The relative intensities of the OH bands
548 can be correlated with the chain polarity in the unit cell. The cellulose chains in the unit cell of I β
549 (and I α as well) and III_I allomorphs are parallel, while those in of II and III_{II} are antiparallel
550 (Langan et al. 2001). Thus, the OH dipoles in the cellulose II and III_{II} are likely to be cancelled
551 making them SFG-inactive (see Section 2.2). The symmetry cancellation effect is also observed
552 for theoretical calculations of SFG spectra of cellulose I β and I α crystals packed in the antiparallel
553 direction (Chen et al. 2017) (see Section 4.2). The symmetry cancellation effect for the antiparallel
554 chain packing in the unit cell is also observed in the comparison of amylose-A/B (parallel) versus
555 amylose-V (antiparallel) (Kong et al. 2014) and β -chitin (parallel) versus α -chitin (antiparallel)
556 (Ogawa et al. 2016).

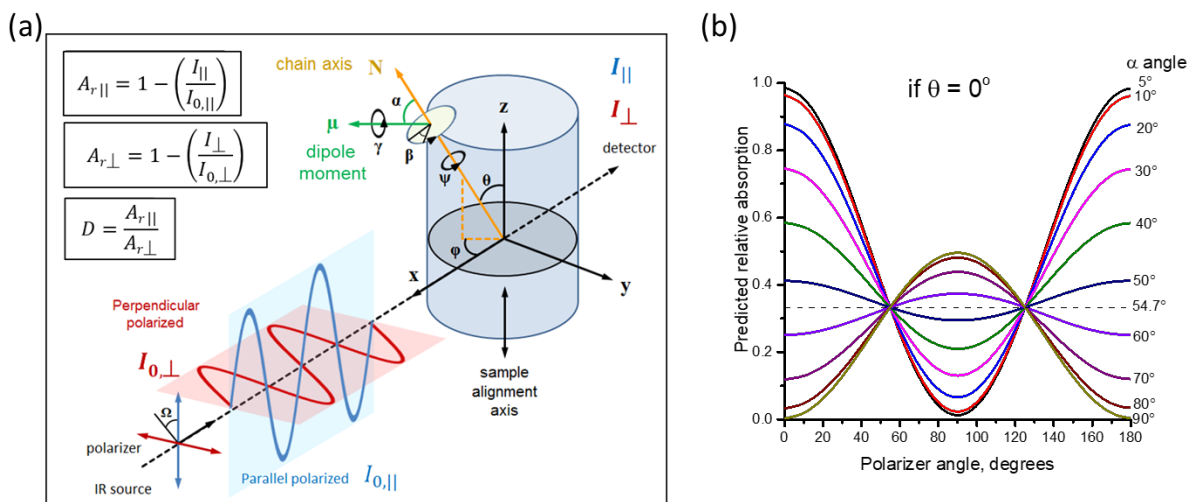
557 Figures 9d and 9e compare the SFG bands in the OH stretch region of cellulose produced
558 by *Glaucozystis* and *Halocynthia* (Lee et al. 2014). The cellulose structures produced by two
559 species are often used as references for cellulose I α (with a minor contribution of I β) and cellulose
560 I β , respectively (Imai et al. 1999; Watanabe et al. 2007). The spectrum in Figure 9d clearly shows
561 the 3240 cm⁻¹ peak (characteristic of I α), while the spectrum in Figure 9e has the 3270 cm⁻¹ peak
562 only (characteristic of I β) and no peak at 3240 cm⁻¹. Figure 9f monitors changes in the OH-stretch
563 SFG spectral feature upon thermal annealing of bacterial cellulose which typically contains >70%
564 of cellulose I α . As cellulose I α is converted to I β (Figure 2), the 3240 cm⁻¹ component disappears.
565 The ~3370 cm⁻¹ component is also disappearing which may indicate that this mode is more
566 prominent in the I α structure than the I β structure (Lee et al. 2014).

567

568 3.3 Polarization dependence of vibrational modes of crystalline cellulose

569 3.3.1 Polarization dependence in IR

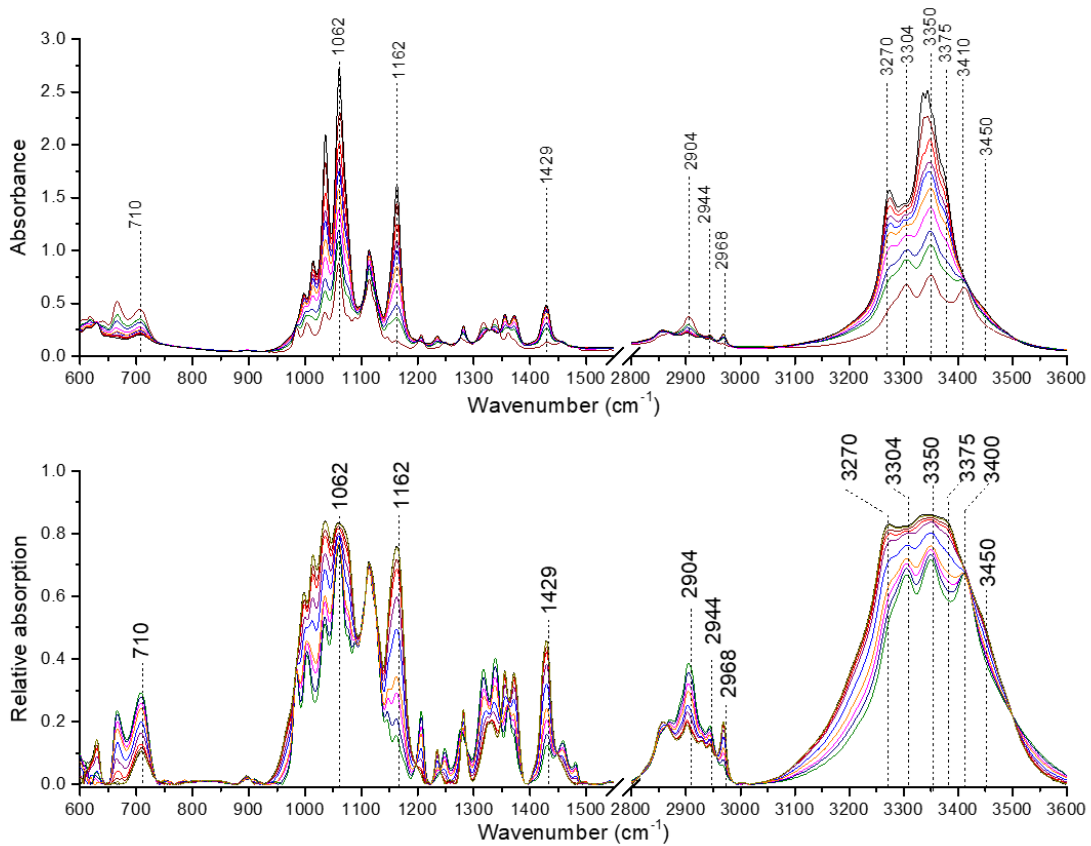
570 In IR spectroscopy, the absorption intensity is greatly enhanced when the direction of the
 571 net dipole moment change occurring during the vibrations coincides with the IR polarization
 572 (Kafle et al. 2017). Figure 10a depicts a situation where the main axis of a sample is aligned along
 573 the z-axis in the laboratory coordinate (x,y,z). Here, the cellulose chain axis (c-direction of
 574 cellulose I β /I α unit cell) can be defined with a set of Euler angles (θ, φ, ψ) with respect to the
 575 laboratory coordinate. The dipole moment of a specific vibrational mode is defined with another
 576 set of Euler angles (α, β, γ) with respect to the chain axis. Figure 10b plots the theoretically-
 577 calculated relative absorption ($A_r = 1 - I/I_0$) as a function of IR polarization angle (Ω) for
 578 various angles (α) between the dipole moment and the z-axis (i.e., when $\theta = 0^\circ$). When $\alpha = 0^\circ$, A_r
 579 = 1 at $\Omega = 0^\circ$ and 0 at $\Omega = 90^\circ$. The ratio of relative absorptions at these two polarization angles is
 580 the dichroic ratio, $D = A_{r\parallel} / A_{r\perp}$, representing the polarization anisotropy. Thus, $D = \infty$ when $\alpha = 0^\circ$.
 581 When $\alpha = 90^\circ$, A_r is 0 at $\Omega = 0^\circ$ and 0.5 at $\Omega = 90^\circ$; thus, $D = 0$. When $\alpha = 54.7^\circ$, then the IR
 582 absorption intensity does not vary with Ω ; thus, $D=1$. This angle is called the magic angle. In this
 583 case, it is not possible to distinguish if the angle between the dipole moment and the chain axis is
 584 actually at 54.7° or completely random.



585
 586 **Figure 10.** (a) Polarized FTIR experimental setup for orientation analysis of uniaxially-aligned
 587 cellulose crystallites where IR propagates along the x-direction and the main axis of sample is
 588 placed along the z-direction of the laboratory coordinate (x,y,z). When the electric field of IR
 589 oscillates along in the x-z plane of the laboratory coordinate ($\Omega = 0$ or 180°), it is parallel to the

590 sample axis ($\mathbf{I}_{0,\parallel}$); when it oscillates along the x-y plane ($\Omega = 90$ or 270°), the polarization is
 591 perpendicular to the sample axis ($\mathbf{I}_{0,\perp}$). The chain axis can be described with three Euler angles
 592 with respect to the laboratory coordinate: θ (tilt angle), φ (azimuth angle), and ψ (twist angle). In
 593 the molecular frame, the Euler angles (α,β,γ) relates the orientation of the transition dipole (μ) of
 594 a specific vibration mode to the cellulose chain axis. The intensity of transmitted IR is denoted as
 595 \mathbf{I}_{\parallel} (parallel polarization) or \mathbf{I}_{\perp} (perpendicular polarization). (b) Simulation of the relative
 596 absorption as a function of polarization angle (Ω) for various tilt angles (α) between the dipole
 597 moment vector and the z-axis when the chain is aligned along the z-direction (i.e., $\theta = 0^\circ$). (Kafle
 598 et al. 2017)

599



600

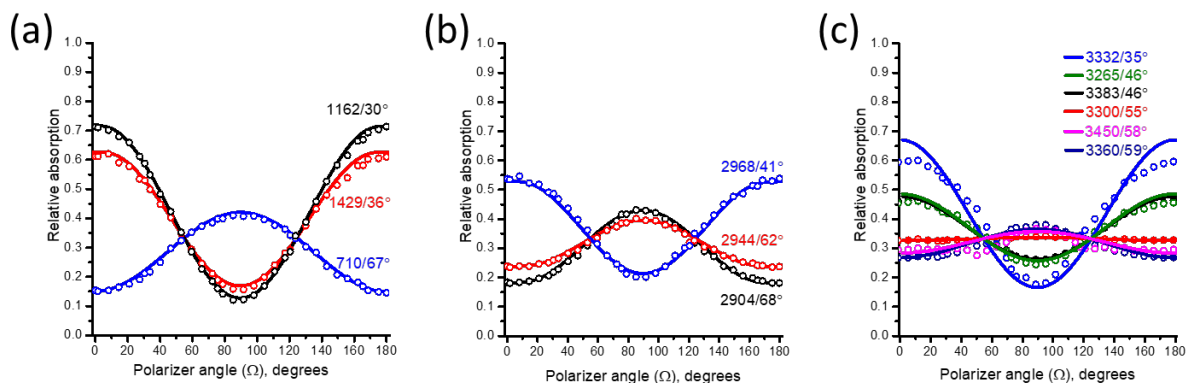
601 **Figure 11.** Transmission polarized infrared spectra, in (a) absorbance scale ($Abs = -\log(I/I_0)$)
 602 and (b) relative absorption scale (A_r , as defined in Figure 10a), of cellulose I β crystallites
 603 uniaxially-aligned along the z-direction as various polarization angles (Ω).

604

605 Figure 11 displays the transmission IR spectra collected at various Ω angles using a thin
 606 film in which cellulose I β crystallites (isolated from *Halocynthia*) are uniaxially-aligned along the
 607 z-direction ($\theta = 0^\circ$). For the orientation analysis, it is important to process the polarization
 608 dependence data in the relative absorption scale (A_r) as defined in Figure 10a because the IR
 609 absorption intensity varies with the cosine of the angle between the dipole moment (α) and the IR
 610 polarization (Ω). The quantitative analysis of the polarization dependence should not be done with
 611 the absorbance data; this is an important difference from the concentration analysis using the Beer-
 612 Lambert law in which absorbance is related to the concentration: $Abs = -\log\left(\frac{I}{I_0}\right) = abc$ where a
 613 = absorptivity, b = beam path length, and c = concentration (Robinson 1996).

614 Figure 12 plots the magnitudes of relative absorption for selected bands of uniaxially-
 615 aligned cellulose I β as a function of IR polarization angle (Ω) shown in Figure 11b. Because many
 616 of these bands are not well-resolved, the analysis requires peak deconvolution. The deconvolution
 617 can be done with Lorentzian functions (Lee et al. 2014). The deconvolution results are shown in
 618 the Supporting Information. Here, it is important to note that no band has the dipole moment
 619 perfectly parallel or perpendicular from the chain axis. Again, this is the consequence of the
 620 collective vibrational motions involving many groups in the crystalline lattice.

621



622

623 **Figure 12.** Plots of the relative absorption (A_r) of various IR bands detected with the uniaxially-
 624 aligned cellulose I β crystallites as a function of the IR polarization angle (Ω). The symbols are
 625 experimental data and the lines are fit results using the theoretical model. The peak position and
 626 tilt angle of each band are marked in each data.

627

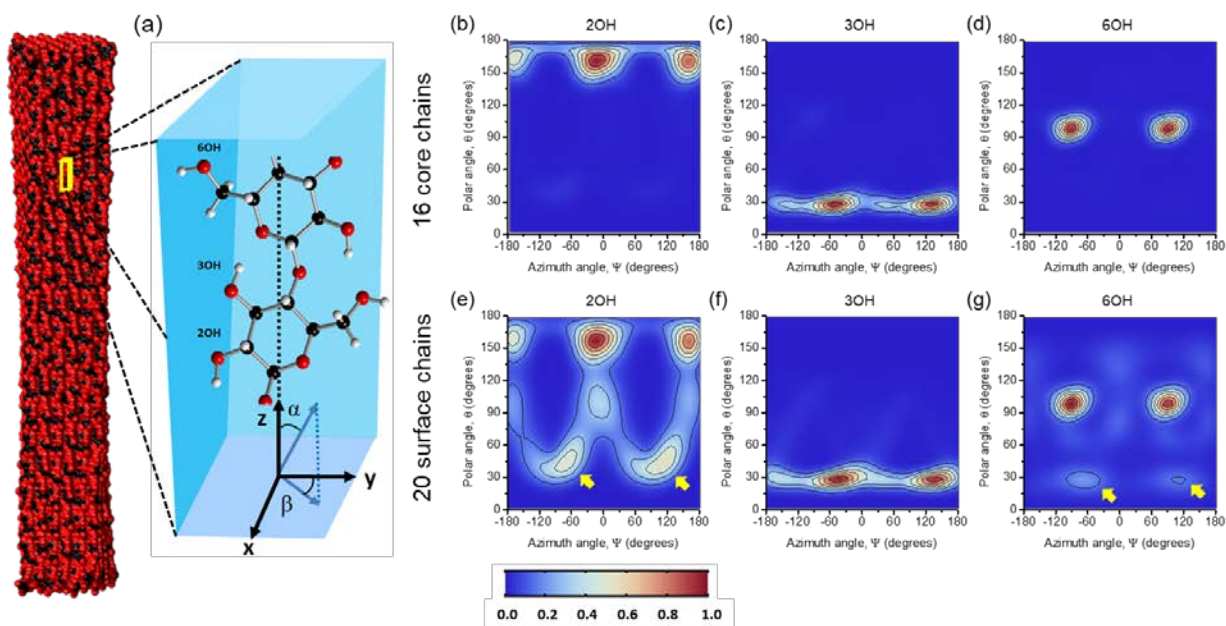
628 In Figure 12a, the dipole moment of the 710 cm^{-1} band is determined to be $\sim 67^\circ$ with respect
629 to the chain axis. This high tilt angle larger than the magic angle from the chain axis is consistent
630 with the normal mode predicted from DFT-D2 calculations shown in Figure 7j; the net dipole
631 moment of chain backbone vibrations in this mode appears to be close to the direction normal to
632 the chain axis. The 1430 cm^{-1} mode involving CH_2 bending motions (Figure 7g) has its dipole
633 moment about 36° tilted from the chain axis. In the past, the 1162 cm^{-1} mode involving the C–O–
634 C glycosidic bond stretch coupled with the backbone vibrations was assumed to be along the chain
635 axis (Suslov and Verbelen 2006; Suslov et al. 2009; Wilson et al. 2000). Although it intuitively
636 made sense, the data shown in Figure 12a indicates that its dipole moment is actually $\sim 30^\circ$ away
637 from the chain axis. This is understandable if one sees the movie associated with the mode shown
638 in Figure 7h (Supporting Information). In this mode, the glycosidic bonds along the chain
639 alternatively expand and contract, so the net dipole change of these vibrations along the chain axis
640 must be cancelled. At the same time, it can be seen that C2, C3, C5 and O5 atoms in the six-
641 membered arm-chair rings are moving as well as the $\text{C}_6\text{H}_2\text{OH}$ side chains. The sum of all dipoles
642 associated with these skeletal vibrations must be off the chain axis by $\sim 30^\circ$.

643 Figure 12b shows the Ω -dependence of three CH/ CH_2 stretch modes. The 2904 cm^{-1} is
644 attributed to the collective motions of the axial C-H groups; thus, its dipole moment would be close
645 to the normal direction. The data in Figure 12b indicates that the dipole moment of the 2904 cm^{-1}
646 mode (Figure 7e) is $\sim 68^\circ$ from the chain axis. The 2944 cm^{-1} mode also has the dipole moment
647 closer to the chain normal direction ($\sim 62^\circ$). In contrast, the dipole moment of the 2968 cm^{-1} mode
648 is closer to the chain direction ($\sim 41^\circ$). Although not shown in Figure 12b, the data in Figure 11b
649 reveal that the intensity of the 2870 cm^{-1} mode (Figure 7f) does not change with Ω , indicating that
650 its dipole moment is at the magic angle (54.7°) from the chain axis. It is unlikely that the 2870 cm^{-1}
651 dipole is completely random in the uniaxially-aligned crystallites.

652 Figure 12c plots the Ω -dependence of six OH stretch modes. After deconvolution of the
653 OH bands, the 3300 cm^{-1} mode (Figure 7c) appears to have the dipole moment close to the magic
654 angle ($\sim 55^\circ$). In Figure 11b, it can also be seen that the 3400 cm^{-1} mode (Figure 7a) has the dipole
655 moment at 54.7° from the chain axis. The 3270 cm^{-1} (Figure 7d), 3330 cm^{-1} , and 3380 cm^{-1} (Figure

656 7b) modes have the dipole moment tilt angle of 40° , 35° , and 46° , respectively, from the chain axis.
657 In contrast, the 3360 cm^{-1} and 3450 cm^{-1} components are tilted about $58\text{-}59^\circ$ from the chain axis.
658 Although the dichroic ratio at these two locations in the raw spectra looks larger than 1, the areas
659 of these two bands obtained from the peak deconvolution gives the dichroic ratio slightly less than
660 1. The intensity reading from the raw spectra can be affected by the overlap of nearby components
661 that have the Lorentzian line shape (for an example of peak deconvolution, see Supporting
662 Information). In other words, the dipole tilt angles of OH bands determined from the Ω -
663 dependence could vary slightly depending on the peak deconvolution.

664 Overall, the dipole moment tilt angles measured with the polarization-dependent IR
665 analysis of the uniaxially-aligned cellulose I β are reasonably consistent with the collective
666 vibrational motions depicted from DFT-D2 calculations (Figure 7) (Lee et al. 2015d). These
667 transition dipole angles determined from the polarized IR analysis of the uniaxially-aligned
668 cellulose I β are different from those expected from the unit cell structure of cellulose I β . Figure
669 13 shows the molecular dynamics (MD) simulation results for a cellulose I β microfibril (Lee et al.
670 2015d). From MD simulations, the geometry of each hydroxyl group can be assessed by plotting
671 the relative abundance of the OH bond polar angle (α) at each cross section as a function of azimuth
672 angle (β) (Figure 13b – 13g). Along the azimuth (β) axis, there are two peaks due to the P2₁
673 symmetry of the unit cell. Along the tilt (α) axis, the 2OH group is pointing downward ($\sim 160^\circ$)
674 and the 3OH group is upward ($\sim 30^\circ$), while the 6OH group is nearly normal to the chain axis
675 ($\sim 100^\circ$). All OH groups in the core have a narrow angle distribution in α (Figures 13b – 13d). The
676 2OH and 6OH groups at the surface of the cellulose microfibril have broader distributions in α
677 (Figures 13e and 13g); in contrast, the surface 3OH group involved in intra-chain hydrogen bond
678 ($3\text{O}-\text{H}\cdots 5\text{O}$) has still a narrow distribution because of the stiffness of the microfibril (Figure 13f).
679 These simulation results imply that if the experimentally-observed OH bands for uniaxially-
680 aligned cellulose I β crystals were to be attributed to individual OH groups in the chain, then the
681 measured α values would be around $20\text{-}30^\circ$ and $\sim 80^\circ$. However, the experimental values shown
682 in Figure 12c are between 35° and 59° . This discrepancy confirms that the OH stretch bands of
683 cellulose cannot be attributed to individual OH groups in the crystal unit cell; they are collective
684 motions of all OH groups in the crystal (as shown from DFT-D2 simulations, see Figures 7).



686

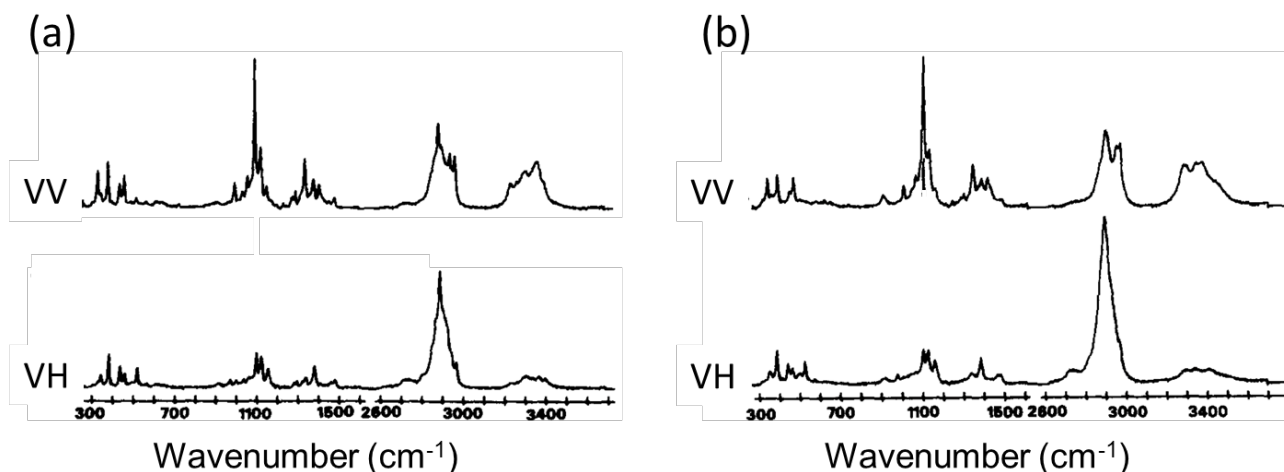
687 **Figure 13.** Hydroxyl groups geometry from molecular dynamics (MD) simulations of a cellulose
 688 microfibril containing 36 chains: (a) Energy-minimized structure of the microfibril (C=black,
 689 O=red), hydrogen atoms are omitted for clarity. The value of α is the angle between the OH bond
 690 axis and the microfibril axis (z-axis in (a); $\theta = 0^\circ$ in Figure 10a). The value of β is the azimuth
 691 angle of the cross-section plane with respect to the b-axis of the bottom unit cell (y-axis in (a)). (b-
 692 g) Angular distributions of the O-H groups averaged along the cellulose microfibril for the 16
 693 interior chains (c-d) and the 20 exterior chains (e-g). The color scale represents the relative
 694 abundance, with red signifying the highest occurrence. (Lee et al. 2015d)

695

696 3.3.2 Polarization dependence in Raman

697 Although the polarization-dependent IR analysis of the uniaxially-aligned cellulose give
 698 the dipole moment tilt angle with respect to the chain axis, the polarization-dependent Raman
 699 analysis provides the orientation of the polarizability (Atalla et al. 1980; Wiley and Atalla 1987).
 700 When the sample is excited with the vertically (V) polarized light, then the emitted Raman signal
 701 can be recorded at the vertical (VV) or horizontal (VH) polarizations. The relative intensities in
 702 the VV and VH spectra reflect the polarized and depolarized components, respectively, of a
 703 vibrational mode. Figure 14 displays the VV and VH Raman spectra of *Valonia* and ramie fibers

704 (Wiley and Atalla 1987). These spectra show that the OH bands as well as the 1096~1098 cm^{-1}
705 skeletal vibration mode are highly polarized (stronger in VV), while the CH stretch peak at ~2900
706 cm^{-1} is highly depolarized (stronger in HV). The CH/CH₂ stretch modes at 2944 cm^{-1} and 2968
707 cm^{-1} appear to be polarized (stronger in VV).



708

709 **Figure 14.** Polarized Raman spectra of (a) *Valonia* fiber and (b) ramie fiber. In these fibers,
710 cellulose crystallites are preferentially aligned along the fiber direction. The excitation beam is
711 vertically polarized (parallel to the fiber direction); the VV and VH means that the emitted signal
712 is vertically and horizontally polarized, respectively (Wiley and Atalla 1987).

713

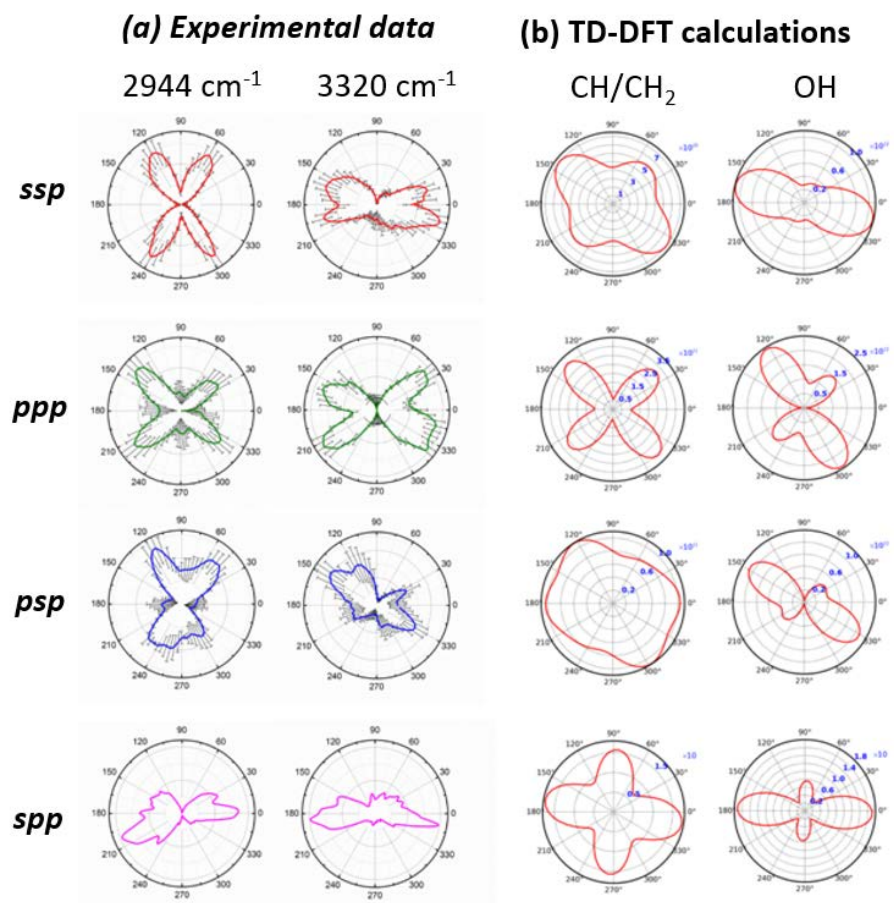
714 3.3.3 Polarization and azimuth angle dependence in SFG

715 The polarization dependence of vibrational modes in SFG spectroscopy is much more
716 complicated than IR and Raman spectroscopy. That is mainly because SFG is a non-linear optical
717 process where the electric fields of three photons are mixed. In SFG, the polarization is defined
718 with respect to the laser incidence plane. The *p*- and *s*-polarizations means the electric vector of
719 the photon within and perpendicular, respectively, to the laser incidence plane. Since the laser
720 beam is irradiated from a certain angle from the surface normal, the *p*-polarized light interacts with
721 the surface normal and parallel tensors in the laser incidence plane, while the *s*-polarized light
722 interacts with the surface parallel tensor perpendicular to the laser incidence plane (Wang et al.
723 2005). In SFG experiments, the polarizations are expressed with a combination of three letters (*p*
724 or *s*) – the first letter is the polarization of SFG signal, the middle letter is the polarization of up-

725 conversion light (532 nm or 800 nm), and the last letter is the polarization of tunable IR beam
726 whose energy is in resonance with the vibrational excitation. For example, the notation *ssp* means
727 that the SFG signal is detected in the *s*-polarization when the polarizations of the up-conversion
728 and tunable IR input beams are in the *s*- and *p*-polarizations, respectively. If the uniaxial alignment
729 axis is parallel to the sample surface ($\theta = 90^\circ$ in Figure 10a), then the azimuth angle (φ) between
730 the alignment axis (chain axis) and the laser incidence plane should be considered in the analysis.
731 Also, if the cross-section of cellulose crystallites is in a rectangular shape, there could be some
732 preferential packing orientation of the facets with respect to the sample surface and this anisotropy
733 should also be considered (Chen et al. 2017).

734 Figure 15 compares the experimentally-observed azimuth angle dependences of the 2944
735 cm^{-1} and 3320 cm^{-1} modes at four different polarization combinations with the TD-DFT calculation
736 results for representative CH/CH₂ and OH stretch modes (Chen et al. 2017). In the experimental
737 data shown in Figure 15a, the polar plots along the azimuth angle show four lobes. Except the *spp*
738 polarization case, the maxima of four lobes are not aligned parallel or perpendicular to the chain
739 axis. This might be because none of CH/CH₂ and OH stretch modes has the dipole moment parallel
740 or perpendicular to the chain axis (see Figure 12). In the *ssp* polarization, the 3320 cm^{-1} intensity
741 is large when $\varphi = 0^\circ$ and 180° (i.e. the chain axis is parallel to the *p*-polarized IR field) and almost
742 zero when $\varphi = 90^\circ$ and 270° (i.e. the chain axis is perpendicular to the *p*-polarized IR field). In
743 contrast, the 2944 cm^{-1} peak has finite intensities at $\varphi = 90^\circ$ and 270° and has very weak intensities
744 at $\varphi = 0^\circ$ and 180° . This could mean that for a sample where cellulose microfibrils are preferentially
745 aligned, whether the predominant direction of microfibrils could be determined qualitatively from
746 the comparison of CH/CH₂ and OH SFG intensities at two azimuth angles – one parallel to the
747 laser incidence plane and the other perpendicular to the laser incidence plane. At different
748 polarization combinations of the SFG experiment, the azimuth angle dependence will be different
749 (Figure 15a). If the chain axes of cellulose crystallites (*c*-axis of the unit cell) are randomly
750 arranged in the sample, then there will be no specific azimuth angle dependence in SFG (Barnette
751 et al. 2011).

752



753

754 **Figure 15.** Comparison of azimuth angle (ϕ) dependence of dominant CH/CH₂ (2944 cm^{-1}) and
 755 OH (3320 cm^{-1}) SFG peaks of the uniaxially-aligned cellulose I β crystals at *ssp*, *ppp*, *psp*, and *spp*
 756 polarizations. The c-axis of cellulose crystal is perpendicular to the azimuth rotation axis. (a)
 757 Experimental data. (b) TD-DFT simulation data for the case where 72% of crystallites has the (1 $\bar{1}$ 0)
 758 plane parallel to the sample surface and 28% have the (110) plane parallel to the sample surface.
 759 (Chen et al. 2017)

760

761 Figures 15b compares the TD-DFT simulation results calculated for the uniaxially aligned
 762 cellulose I β cases. The simulation was carried out for the case where 72% of the crystallites have
 763 the (1 $\bar{1}$ 0) facet parallel to the sample surface and other 28% have the (110) facet parallel to the
 764 sample surface. This ratio was chosen based on the XRD analysis of the uniaxially-aligned
 765 cellulose I β film (Chen et al. 2017). Although the TD-DFT simulation results are not in perfect
 766 agreement with the experimental data, the general trends are comparable with the experimental
 767 data (Figure 15a).

768

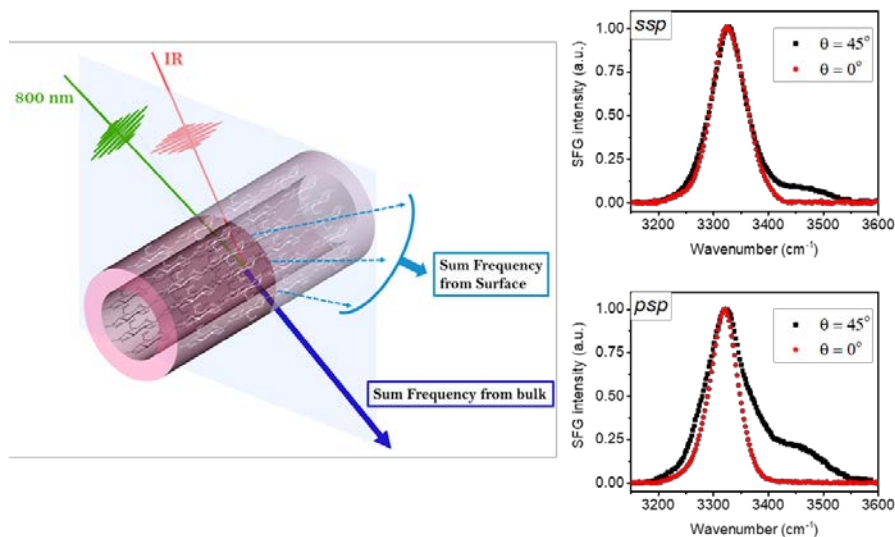
769 *3.5 Distinguishing bulk and surface hydroxyl groups of crystalline cellulose*

770 In studying the roles of cellulose microfibrils in plant cell walls or properties of polymer
771 composites engineered with cellulose nanocrystals, a question arises about interactions between
772 the cellulose and matrix polymers because they play critical roles in mechanical properties of the
773 system. Probing surface or interface species in the presence of the same species in the crystalline
774 bulk or in the amorphous matrix phase is often extremely difficult in IR and Raman spectroscopy.
775 Because the signal intensity in linear spectroscopy is governed by the Beer-Lambert law and the
776 surface or interfacial species in such systems are less than the bulk species, the IR and Raman
777 spectra are usually dominated by the bulk signals.

778 This problem can be addressed in SFG using the phase matching condition in non-linear
779 spectroscopy (Roke et al. 2004; Roke et al. 2003). At angles away from the phase-matched
780 direction of SFG, up-conversion, and tunable IR beams, the phases of these beams across the
781 scattering object are shifted, which ensures that signals from functional groups with opposite
782 orientations on either side of an object will not completely cancel out each other (de Beer et al.
783 2010; de Beer and Roke 2007; de Beer and Roke 2009). The theoretical framework describing the
784 SFG scattering has been developed for arbitrary shapes particles (de Beer et al. 2011). Using this
785 principle, it is possible to distinguish the contributions from the OH groups inside the crystalline
786 cellulose domain and those from the OH groups at the surface of crystalline cellulose.

787 Figure 16a illustrates the SFG scattering experiment carried out to demonstrate this
788 principle (Makarem et al. 2018). At the phase-matched direction, the OH signals inside the
789 crystalline cellulose are dominant. As the SFG detection angle moves away from this phase-
790 matched direction, the OH signals from the surface of crystalline cellulose will increase while the
791 bulk signal contribution diminishes. Figure 16b compares the *ssp* and *psp* SFG spectra collected
792 at the phase matched direction (0°) and a scattering angle of 45° (Makarem et al. 2018). The data
793 shows that at the 45° scattering angle, the SFG intensity at $>3400\text{ cm}^{-1}$ is significantly enhanced
794 compared to the main peak centered at $\sim 3320\text{ cm}^{-1}$. It is well known that the OH stretch band
795 position decreases as the hydrogen bond strength increases (Gorman 1957; Libowitzky 1999;
796 Novak 1974; Wang et al. 2014a). Thus, the data in Figure 16b indicates that the OH bands at

797 $>3400\text{ cm}^{-1}$ in SFG spectra of cellulose can be attributed to the OH groups at the crystalline
798 cellulose surface which have hydrogen bonding interactions weaker than the OH groups in the
799 highly-packed crystalline bulk.



800
801 **Figure 16.** (a) Experimental geometry to measure SFG signals of a cellulose crystal at the phase-
802 matching and scattering angles. (b) SFG spectra of a uniaxially-aligned cellulose I β fiber collected
803 with the *ssp* and *psp* polarization combinations at the phase-matched direction (0°) and a scattering
804 angle of 45° (Makarem et al. 2018).

805
806 When crystalline cellulose is immersed in D_2O at room temperature, the SFG signal at
807 $>3400\text{ cm}^{-1}$ is suppressed due to the OH/OD exchange, while the signal at $<3400\text{ cm}^{-1}$ is not
808 affected (Makarem et al. 2018). That implies that the OH groups responsible for the SFG signals
809 at $>3400\text{ cm}^{-1}$ are readily accessible by D_2O solvent molecules, but those at $<3400\text{ cm}^{-1}$ are not
810 accessible. A similar observation was also reported in IR analysis of crystalline cellulose (Tsuboi
811 1957). This difference in the OH/OD exchange activity also supports that the weakly-hydrogen
812 bonded OH groups exposed at the crystalline cellulose surface are detected at $>3400\text{ cm}^{-1}$. In
813 Figure 12c, the polarization IR analysis shows that the dipole moment of the 3450 cm^{-1} component
814 is about 58° from the chain axis (close to or slightly above the magic angle). This may imply that
815 the surface OH groups are pointing away from the crystal surface, which would provide favorable
816 geometry for hydrogen bonding interactions with surrounding molecules.

817 **4. Concentration and Spatial Distribution of Crystalline Cellulose in** 818 **Amorphous Matrix**

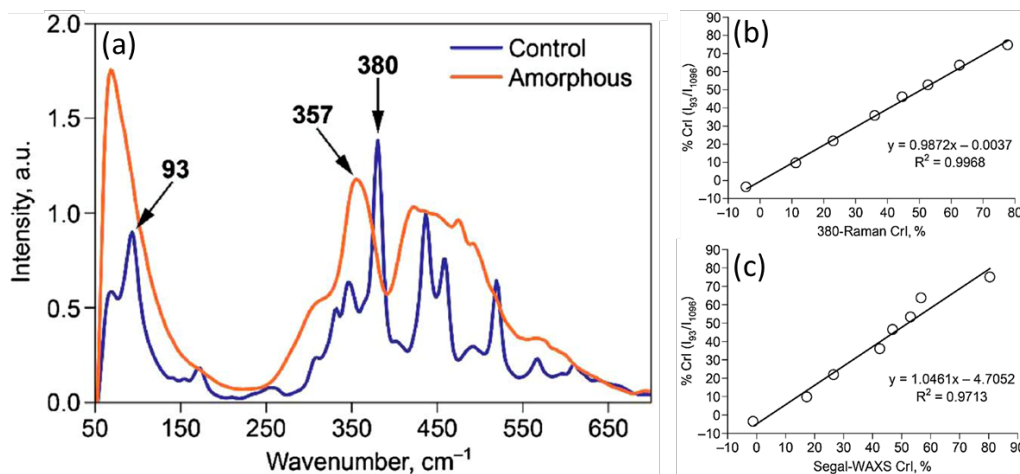
819 *4.1 Concentration Dependence of Vibrational Peak Intensities*

820 One of the key questions studied with vibrational spectroscopy is the quantification of
821 crystalline cellulose in the sample. The same question can be addressed with nuclear magnetic
822 resonance (NMR) and XRD (Lee et al. 2015a; Park et al. 2010; Wickholm et al. 2001). If
823 vibrational peaks specific to the crystalline order of cellulose can be identified without any
824 interference from amorphous cellulose and other matrix polymers, then this task can be done most
825 efficiently using the linear spectroscopy (IR and Raman) because the signal intensity can be
826 linearly related with the concentration of crystalline cellulose in the sample using the Beer-
827 Lambert law. Often peaks overlap with broad backgrounds originating from the amorphous phase
828 (for example, see Figure 8a). In that case, deconvolution of the crystalline peak from the
829 amorphous background is necessary. Then, the peak height or area of the deconvoluted crystalline
830 peak component can be correlated with the concentration of crystalline cellulose in the sample.

831 When reference samples with known crystallinity index (XRD) are used, then a simple linear
832 equation can be obtained from least squares fitting. Here, it should be noted that the XRD
833 crystallinity index (CI) is not an absolute value and varies depending on what method is used (Park
834 et al. 2010). Interferences between the crystalline and amorphous scattering signals in the
835 reciprocal lattice spacing are not a simple additive function. Changes in the crystal size can also
836 affect the CI value calculated from XRD calculation (Lee et al. 2015a). Depending on which
837 method is used to calculate XRD CI values (Park et al. 2010), the correlation with the vibrational
838 peak intensity may vary. Also, the CI represents the fraction of crystalline cellulose in the entire
839 sample, not the ratio between crystalline versus amorphous cellulose (unless the entire sample
840 consists of cellulose only). Thus, if the XRD CI is calculated to be higher than the cellulose content
841 measured with chemical composition analysis methods, then the calculate CI value is obviously
842 overestimated. The peak height method (also known as the Segal method) tends to overestimate
843 the CI value compared to other methods (Park et al. 2010).

844 Figure 17a compares the Raman spectra of crystalline and amorphous cellulose samples
845 (Agarwal et al. 2018). The crystalline phase show two sharp peaks at 93 cm^{-1} and 380 cm^{-1} that
846 can be distinguished from the amorphous background. The 93 cm^{-1} peak is attributed to twisting
847 of crystal along the c-axis of the crystalline cellulose (Figure 7l) and the 380 cm^{-1} peak is assigned
848 to the out-of-plane breathing of glucose rings along the chain axis. (Figure 7k). The signal intensity
849 of those two peaks correlate to each other very well (Figure 17b). By monitoring their intensities
850 of the samples with known CI values, an empirical equation can be obtained to convert their
851 intensities (either peak height or area) to the cellulose crystallinity (Figure 17b,c) (Agarwal et al.
852 2018; Agarwal et al. 2010).

853



854

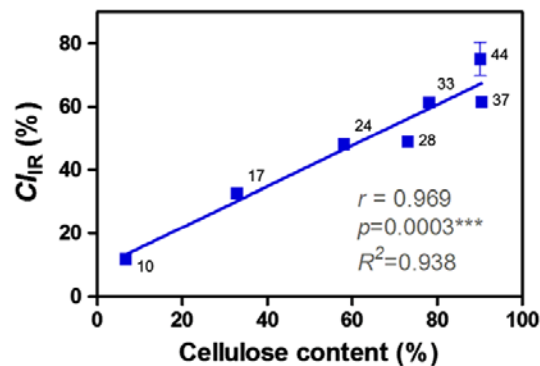
855 **Figure 17.** (a) Raman spectra of crystalline sample (Whatman CC31) and amorphous sample (ball-
856 milled for 120 min). (b) Correlations between the 93 cm^{-1} and 380 cm^{-1} Raman band intensities.
857 (c) Correlation between the 93 cm^{-1} intensity and the XRD crystallinity calculated with the XRD
858 peak height method for a calibration sample set (Agarwal et al. 2018).

859

860 In IR, several peaks have been reported to be used for CI calculations. In typical Fourier
861 transform IR instruments using the KBr beam splitter, the cut-off frequency is 400 cm^{-1} . So, the
862 low wavenumber crystalline peaks seen in Raman are not observed in IR. The wavenumber
863 differences between the crystalline and amorphous peaks tend to become smaller as the vibration
864 energy increases (Oh et al. 2005); this makes the distinction between the crystalline and amorphous

865 phase more difficult. In the case of cellulose I β , the 710 cm⁻¹ band shown in Figure 7j could be
866 used for the CI calculation. For cellulose I α , the corresponding band is at 740 cm⁻¹ (Table 1). After
867 deconvoluting the background signal and normalizing with a peak that is not sensitive to
868 crystallinity, the intensities of these peaks could be used to calculate CI. Figure 18 displays an
869 example of using the 710 cm⁻¹ intensity of cotton fibers calculating the CI and comparing it with
870 amount of crystalline cellulose in the sample determined with the Updegraff method at different
871 developmental stages of cotton (Kim et al. 2018).

872



873

874 **Figure 18.** Correlation between the CI value of cotton fibers calculated with the 710 cm⁻¹ mode in
875 IR and the cellulose content determine from a biochemical assay. (Kim et al. 2018)

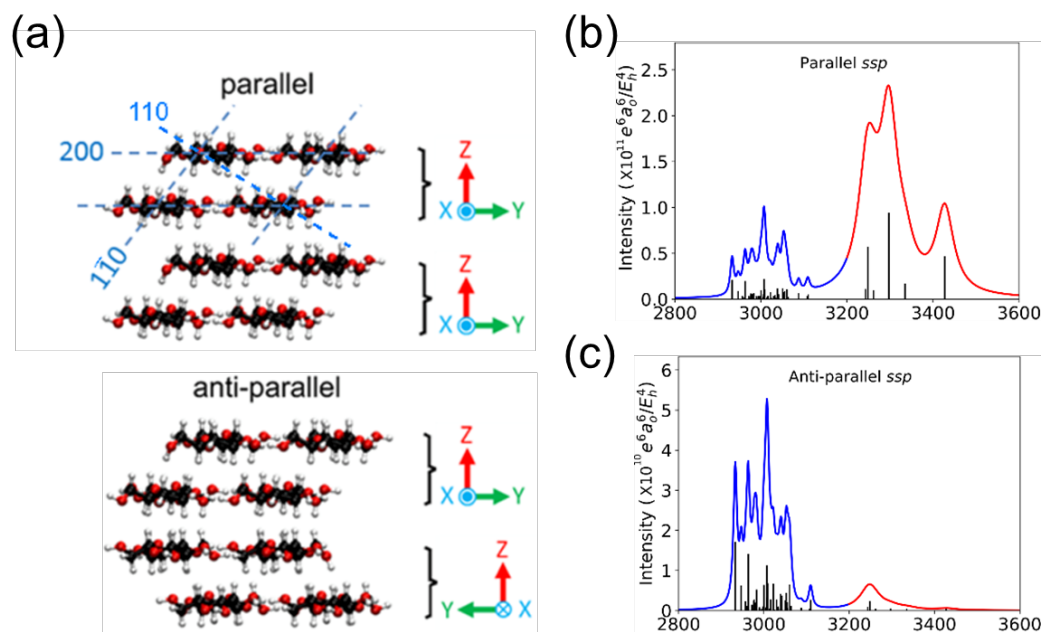
876

877 Although the Raman- and IR-based CI calculation methods are demonstrated with the
878 control samples containing mostly cellulose or purely cellulose only, the background subtraction
879 becomes difficult when the cellulose content in the sample is low. Also, the peak deconvolution
880 becomes challenging when any of amorphous matrix polymers have strong absorption bands near
881 the cellulose peaks used for the CI calculation (for example, see Figure 8a). Compared to these
882 linear spectroscopic methods, SFG does not suffer from such interferences because only crystalline
883 cellulose is detected in SFG (Figure 8b). However, the quantitative conversion of the SFG signal
884 intensity to the crystalline cellulose content in the sample is complicated.

885 As discussed in Section 2.2, the SFG intensity is proportional to the square of $\chi_{eff}^{(2)}$ (eq.
886 3), which is in turn proportional to the density (concentration) of the non-centrosymmetrically

887 arranged vibrational modes (functional groups). In other words, if all SFG-active modes are
 888 arranged in the noncentrosymmetric way, the SFG intensity of such modes will be increase in
 889 proportional to the square of their concentration. However, if they are arranged with the inversion
 890 symmetry (as in the case of antiparallel packing of crystallites in the uniaxial alignment), then the
 891 SFG intensity would be extremely small or negligible.

892



893

894 **Figure 19.** (a) Computational models of cellulose Iβ crystallites packed in the parallel (top) and
 895 anti-parallel (bottom) directionality. The crystallites are viewed along the *c*-axis. In the top panel,
 896 the (200), (110), and (1 $\bar{1}$ 0) surfaces of cellulose Iβ are marked. In the laboratory frame, the laser
 897 incident plane lies in the XZ plane. (b) Simulated *ssp* SFG spectra of uniaxially-aligned cellulose
 898 Iβ crystallites with the parallel (top) and antiparallel (bottom) directionality. Note that the y-axis
 899 scale of the top panel is larger by a factor of 4 than that of the bottom panel (Chen et al. 2017).

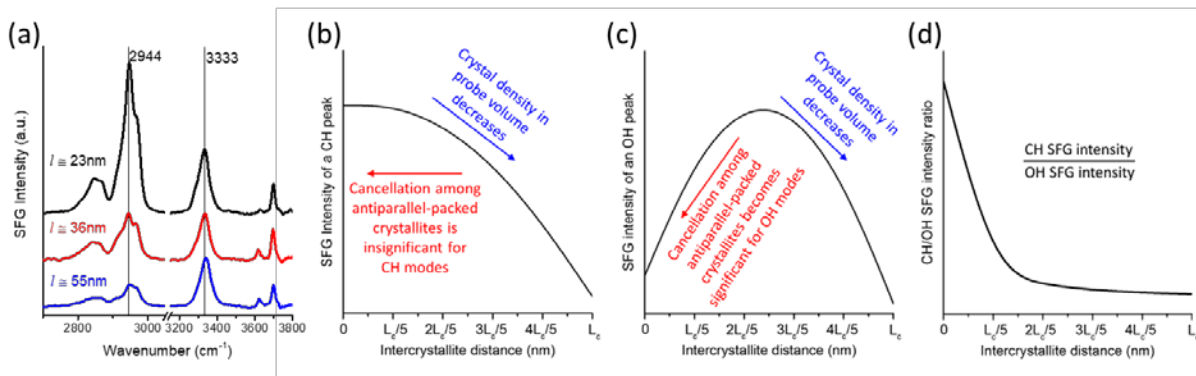
900

901 This principle is explained with TD-DFT calculations (Chen et al. 2017). Figure 19a shows
 902 two crystalline domains aligned and stacked along the *c*-axis of cellulose Iβ unit cell. Because the
 903 OH dipoles are within the *bc* plane of the cellulose Iβ unit cell (Figure 1b), when one domain is
 904 rotated by 180° along the *a* axis, the two domains are antiparallel to each other. TD-DFT

905 calculations show that the OH SFG intensity of the antiparallel-packed domains (Figure 19b,
 906 bottom) is about 50 times smaller than that of the parallel-packed domains (Figure 19b, top). This
 907 is because the OH dipoles are cancelled in the antiparallel-packed domains and enhanced in the
 908 parallel-packed domains. The CH/CH₂ SFG intensity is comparable in these two cases. That is
 909 because the dipoles of coupled vibrational motions of CH and CH₂ groups are pointing toward the
 910 *a*-axis of the unit cell. Thus, the CH/CH₂ SFG intensity could be related to the concentration of
 911 cellulose if cellulose crystallites are uniaxially aligned within the SFG coherence length.

912 The non-centrosymmetric addition and centrosymmetric cancellation effects are not always
 913 equal; because of the synchronization factor in eq.(3), they also depend on the distance between
 914 such domains. Figure 20a displays the SFG spectra of uniaxially-aligned cellulose I β crystallites
 915 dispersed in a fibrin protein matrix (Makarem et al. 2017). Because the cellulose I β crystallites
 916 were randomly dispersed in the matrix initially and then aligned uniaxially through fiber pulling,
 917 their directionalities are statistically random along the fiber direction; in other words, equal
 918 probabilities of pointing the opposite directions. As the cellulose concentration increases in the
 919 matrix, the inter-crystallite distance (*l*) decreases. The average value of *l* can be estimated assuming
 920 perfect and equal separations among the crystallites. The spectra in Figure 20a show that as the
 921 inter-crystallite distance decreases, the CH/CH₂ SFG intensity increases monotonically, whereas
 922 the OH SFG intensity slightly increases and then decreases.

923



924

925 **Figure 20.** (a) SFG spectra of cellulose I β –fibrin composite fibers collected with the *ppp*
 926 polarizations at the 45° scattering direction. The cellulose I β crystallites are fully dispersed and
 927 uniaxially aligned in the fibrin matrix. The distance between cellulose crystallites (*l*) is estimated
 928 from the concentration of cellulose I β in the composite. Dependence of the *ppp* SFG intensity of

929 (b) a CH mode and (c) an OH mode on the intercrystallite distance (l) simulated for cellulose
930 crystals uniaxially aligned without net polarity (antiparallel in average). The CH/OH ratio shown
931 in (d) is obtained by dividing the data in (b) with that in (c) (Makarem et al. 2017).

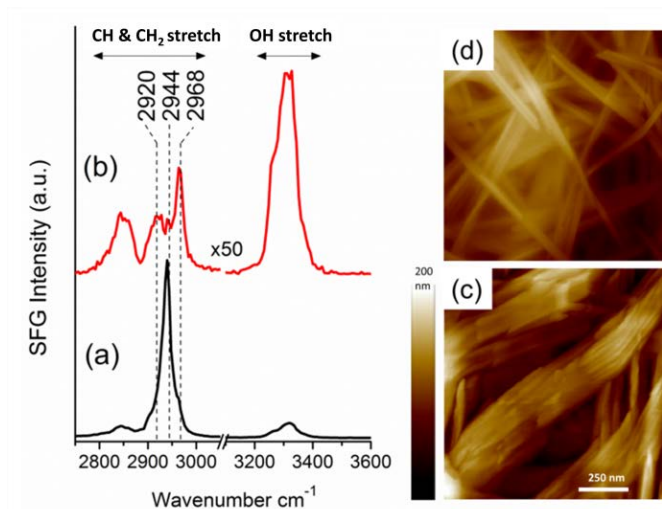
932

933 Using the symmetry cancellation effect shown in Figure 19 and the synchronization factor
934 shown in eq.(3), the dependence of the CH/CH₂ and OH SFG signals on the inter-crystallite
935 separation distance could be simulated (Figures 20b and 20c). When the separation distance among
936 cellulose crystallites is large, the OH SFG intensity appears to be governed by the density effect.
937 But at shorter distances, the symmetry cancellation effect appears to be dominant over the density
938 effect. However, this effect is insignificant for the CH/CH₂ SFG signals. Because of this difference
939 in the symmetry cancellation effect for these CH/CH₂ and OH modes, the 2944 cm⁻¹/3320 cm⁻¹
940 intensity ratio also changes as the concentration of cellulose crystallites varies (Figure 20d). This
941 simulation also shows that the CH/CH₂ SFG intensity could be used for quantification purpose,
942 but not OH SFG intensity, if crystallites are uniaxially aligned with equal probabilities of pointing
943 two opposite directions.

944 When the crystalline cellulose domains are larger than the SFG coherence length and they
945 are mixed with amorphous matrix polymers, then the symmetry cancellation effect among those
946 gigantic crystallites is negligible. In such cases, it is possible to prepare reference samples with
947 known concentrations by mixing micron-size crystallites (such as Avicel or α -cellulose) and other
948 amorphous matrix polymers. When such samples are analyzed with SFG, an empirical calibration
949 curve can be constructed and used for quantification purposes (Barnette et al. 2012; Kafle et al.
950 2015a). Cautions must be taken because the degree of light scattering at the sample surface could
951 vary with the texture size. Also, because of the fourth-power dependence of the scattering process
952 on the wavelength of the light, the scattering loss of the CH/CH₂ and OH SFG signals inside the
953 sample would be different (Lee et al. 2016). In addition, the attenuation of the probing IR and
954 visible beams inside the sample is different for the CH/CH₂ and OH regions. Because of these
955 complexities in vibrational signal measurements as well as x-ray scattering processes, correlations
956 between the CI values calculated from vibrational spectroscopy and XRD analyses could be non-
957 linear (Lee et al. 2015a).

958

959 4.2 Dependence of SFG Intensities on Meso-scale Assembly Patterns



960

961 **Figure 21.** Comparison of SFG spectra (a,b) and AFM images (c,d) of uniaxially-aligned (a,c) and
962 randomly-packed (b,d) cellulose Iβ crystals. Scale bar is 250 nm. Note that the SFG spectrum of
963 the randomly-packed crystals is magnified by a factor of 50 (Lee et al. 2014).

964

965 The non-centrosymmetry and phase synchronization requirements make SFG sensitive to
966 not just the concentration (inter-crystallite distance), but also the assembly pattern. When the
967 assembly pattern among cellulose crystallites varies within the SFG coherence length (L_c), then
968 the symmetry cancellation effect also varies. For example, the OH SFG intensity becomes weak
969 when cellulose crystallites are packed in the antiparallel fashion due to the symmetry cancellation
970 effect (Figure 19b, bottom). However, when they are packed randomly in space, such cancellation
971 effect is relatively weak because the phases of SFG signals from randomly arranged SFG-active
972 domains would be different.[doi:10.1038/nature03027] At the same time, the enhancement effect
973 of the CH/CH₂ SFG signals will be reduced and could alter the peak shape too. So, the overall SFG
974 intensity of the randomly packed cellulose crystallites is much weaker than the antiparallel-packed
975 crystallites and the 2944 cm⁻¹/3320 cm⁻¹ intensity ratio is much smaller (because the OH signal is
976 large), compared to the antiparallel-packed case. This principle can be seen in the experimental
977 data shown in Figure 21 (Lee et al. 2014). Because L_c is on the order of a few hundred nanometers,

978 this makes SFG useful to analyze the spatial packing or assembly pattern of cellulose crystallites
979 or microfibrils over the meso-scale length in the sample (see Section 5.3).

980

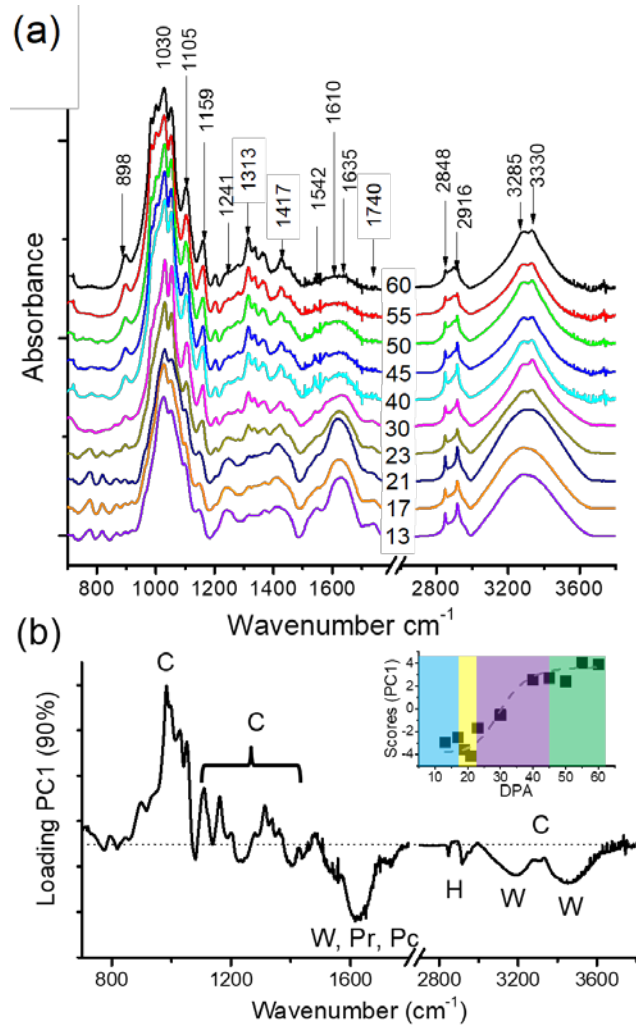
981 **5. Applications of Vibrational Spectroscopy to analysis of cellulose in plant cell** 982 **walls and lignocellulose biomass**

983 *5.1 IR analysis of cellulose in plant cell walls*

984 IR spectroscopy has been applied to study cell wall components, in particular, the
985 composition of matrix polymers and their variations at different stages of plant growth (Abidi et
986 al. 2008; Butler et al. 2015; Cintrón and Hinchliffe 2015; Sene et al. 1994). However, due to the
987 presence of polysaccharides with similar chemical compositions and structures in plant cell walls,
988 assigning an individual peak to a vibration of a specific group of a specific component is
989 challenging. During the cell growth and expansion, the primary wall is made of multilamellar
990 cellulose microfibrils with other matrix polysaccharides and proteins dispersed between fibers
991 (Cosgrove 2005). As cell growth stops, the secondary cell wall starts to form, thickening the wall
992 with cellulose-rich layers, and then eventually lignin will be deposited (Kumar et al. 2015). IR
993 spectroscopy can be a useful tool to screen the variation in polysaccharides and proteins during
994 cell growth and maturation.

995 Figure 22 shows IR spectra of *Gossypium Hirsutum* cotton fibers in different
996 developmental stages (Lee et al. 2015c). Cotton fibers are highly-elongated epidermal cells. The
997 developmental steps of cotton fibers include elongation during which only primary wall is
998 deposited, then start of cell wall thickening in a transition state and eventually formation of a
999 thicker secondary cell wall during maturation. In Figure 22, the 13 and 17 DPA samples contain
1000 primary walls only, the 18 – 22 DPAs represent the transition stage, and the 23-60 DPAs are the
1001 secondary cell wall formation and maturation. The IR spectra of cotton fiber show a large number
1002 of peaks associated with crystalline cellulose (1100 cm^{-1} - 1300 cm^{-1}), pectin (1740 cm^{-1}), proteins
1003 (1520 cm^{-1} - 1620 cm^{-1}), and water (bending of HOH at 1640 cm^{-1} and OH stretch peak between
1004 3150 cm^{-1} - 3600 cm^{-1}).

1005



1006

1007 **Figure 22.** (a) IR spectra of *G. Hirsutum* cotton fibers in different developmental stages from 13
1008 to 60 DPA. (b) loading plot of principle component 1 (PC1) showing the most influential changes
1009 in the spectra as the cotton fiber grows. C: Cellulose, W: Water, Pr: Protein, Pc: Pectin, and H:
1010 Hydrocarbons. The inset plots the scores of PC1 versus DPA colored regions represent different
1011 developmental stages: Blue: Primary wall synthesis, Yellow: Transitional layer synthesis, Purple:
1012 Secondary wall synthesis, Green: Maturation stage. (Lee et al. 2015c)

1013

1014 The main variation in these vibration modes can be analyzed with principle component
1015 analysis (PCA). The principle component 1 (PC1) shown in Figure 22b represents 90% of the
1016 variance of the all spectra. The maturation of cell wall intensifies the peaks between 1100 cm^{-1} and

1017 1300 cm⁻¹ which most probably belong to skeletal deformations of crystalline cellulose. The peak
 1018 at 1160 cm⁻¹ becomes prominent between 21 DPA and 30 DPA, indicating the secondary cell wall
 1019 formation with highly crystalline cellulose. The 1740 cm⁻¹ peak, which is characteristic of pectin,
 1020 weakens as the secondary cell wall starts to form in 21 DPA and becomes negligible at 40 DPA.
 1021 The amide bands of proteins appearing between 1520 cm⁻¹ - 1620 cm⁻¹ also decrease in their
 1022 intensities. These changes are observed in the transition state; they are reflected as negative peaks
 1023 in the PC1 spectrum (Figure 22). The OH region of the PC1 spectrum has two components – one
 1024 is a broad peak at 3000 – 3600 cm⁻¹ with a negative intensity and the other is a sharp peak at 3250
 1025 – 3400 cm⁻¹ with a positive intensity. The former represents the decrease in the water content in
 1026 the cell wall and the latter indicates the increase in the cellulose component. The score plot of PC1
 1027 versus DPA (Figure 22b, inset) clearly shows the development of cell wall growth stage from
 1028 primary, transition, and secondary thickening and maturation (Lee et al. 2015c).

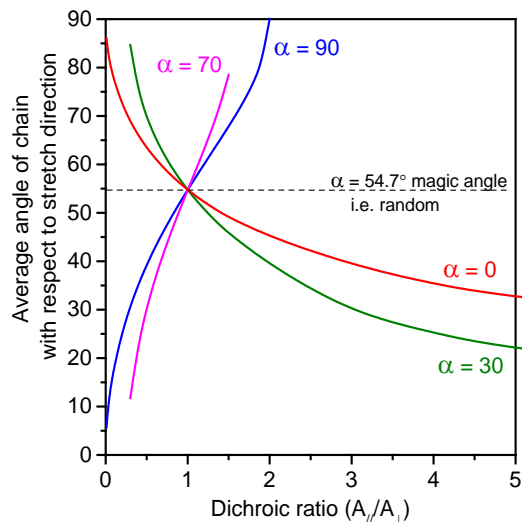
1029 If cellulose microfibrils in the sample have a preferential orientation, the averaged
 1030 orientation direction can be found from the dichroic ratio calculated from polarized IR
 1031 measurements at two angles, $\Omega=0^\circ$ and 90° . Let's assume that cellulose microfibrils are distributed
 1032 around a preferential orientation axis tilted by θ from the reference axis of the sample without
 1033 specific constraints in φ (see Figure 10a). Then, the average tilt angle (θ) described by the second-
 1034 order Legendre polynomial term, $P_2(\cos \theta)$ can be related to the measured dichroic ratio (D)
 1035 (Ward 1985; Wilkes 1971):

$$1036 \quad P_2(\cos \theta) = \frac{3(\cos^2 \theta) - 1}{2} = C \frac{(D-1)}{(D+2)} \quad (5)$$

$$1037 \quad \text{where} \quad C = \frac{2\cot^2 \alpha + 2}{2\cot^2 \alpha - 1} \quad (6)$$

1038 For a vibration mode with a transition moment parallel to the cellulose chain axis, $\alpha = 0^\circ$ and $C =$
 1039 1. For a vibration mode with transition moment perpendicular to the chain axis, $\alpha = 90^\circ$ and $C =$
 1040 -2 .

1041



1042

1043 **Figure 23.** Theoretical calculation of the average tilt angle (θ) of cellulose chain with respect the
 1044 IR polarization axis ($\Omega=0^\circ$) as a function of dichroic ratio (D) for four dipole tilt angles with respect
 1045 to the chain axis ($\alpha = 0^\circ, 30^\circ, 70^\circ$, and 90°). See Figure 10a for the graphical description of angles
 1046 (Kafle et al. 2017).

1047

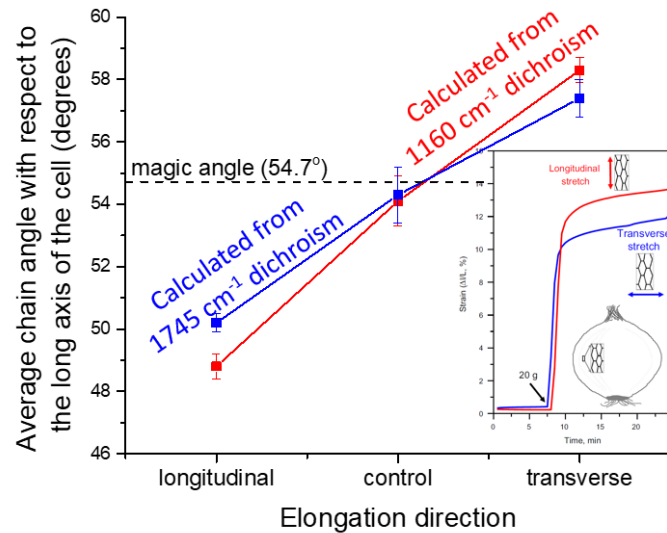
1048 Figure 23 shows how the dichroic ratio (D) is related to the average tilt (θ) of cellulose
 1049 chains with respect to the IR polarization axis ($\Omega=0^\circ$) for four dipole moment angles from the
 1050 chain axis: $\alpha = 0^\circ, 30^\circ, 70^\circ$, and 90° (Kafle et al. 2017). In solving eq.(5) for any measured D , there
 1051 are two unknowns (θ and α); thus, the degree of freedom is +1. This means that it cannot be solved
 1052 for a sample with unknown orientation and unknown molecular structure. This can be seen
 1053 graphically in Figure 23; for any given D , the solution for θ varies depending on α . In the case of
 1054 cellulose I β , the data shown in Figure 12 provide the α values for various IR bands. Some of the
 1055 corresponding vibrational modes can be seen in Figure 7. Thus, the preferential microfibril angle
 1056 (θ) can be determined from the dichroic ratio measurement. If the α value of the vibrational mode
 1057 of interest is known to be less than or larger than the magic angle (54.7°), the only information that
 1058 can be obtained from the dichroic ratio measurement is if the preferential orientation is above or
 1059 below the magic angle – when $\alpha < 54.7^\circ$, $D > 1$ means $\theta < 54.7^\circ$; when $\alpha > 54.7^\circ$, $D > 1$ means $\theta >$
 1060 54.7° (Suslov and Verbelen 2006; Suslov et al. 2009).

1061 Using the relationship shown in Figure 23, the average orientation of cellulose microfibrils
1062 and pectin molecules in an epidermal cell wall of the onion scale fifth from the outside skin (Kafle
1063 et al. 2017), which is often used as a model primary cell wall (Park et al. 2014; Zhang et al. 2013;
1064 Zhang et al. 2017; Zhang et al. 2016). For the $\sim 1160\text{ cm}^{-1}$ band which involves vibrations of the
1065 C-O-C glycosidic bond and the arm-chair backbone of cellulose, the dipole moment is tilted about
1066 30° from the chain axis (Figure 12a). But, it should be kept in mind that all hemicellulose and
1067 pectin molecules show the band in this region (Figure 8a) and cannot be distinguished from the
1068 cellulose band. Although the α value of the $\sim 1160\text{ cm}^{-1}$ band is not known for hemicellulose and
1069 pectin, it could be assumed to be similar to that of cellulose. Then, the chain orientation calculated
1070 from the dichroic ratio measured from the $\sim 1160\text{ cm}^{-1}$ band must be considered as the weighted
1071 average of all polysaccharides in the epidermal cell wall. Pectin molecules have a characteristic
1072 peak at 1745 cm^{-1} (from the C=O stretch). The dipole moment tilt angle from the chain axis is not
1073 known for this band; but it can be assumed that its angle is higher than the magic angle since the
1074 C=O stretch dipole is pointing away from the chain. The α value for the 1745 cm^{-1} could be
1075 assumed to be $\sim 70^\circ$ (if it is assumed to be 90° , then the calculated θ value is slightly different).

1076 Figure 24 shows the chain orientations calculated from the dichroism of the $\sim 1160\text{ cm}^{-1}$
1077 and 1745 cm^{-1} bands for an abaxial epidermal wall of the fifth scale of an onion bulk in the relaxed
1078 (control) and two different mechanically-stretched states. The chain orientation values ($54.1\pm 0.8^\circ$
1079 from the 1160 cm^{-1} dichroism and $54.3\pm 0.9^\circ$ from the 1745 cm^{-1} dichroism) in the relaxed state
1080 are close to the magic angle (54.7°). This could mean two cases – (i) the molecular chain axes are
1081 random in the cell wall or (ii) there exists some degree of preferential orientations with an average
1082 angle of $\sim 54^\circ$ from the long axis of the cell. The onion cell walls often exhibit anisotropic
1083 elongation under mechanical stress (Suslov and Verbelen 2006; Suslov et al. 2009). Also, SFG
1084 analysis of similar onion epidermal walls suggested some degree of preferential orientations
1085 toward the short axis direction of the cell (Kafle et al. 2014c). Combining these results, one might
1086 say that the average angle of molecular chains in the epidermal cell wall is in fact close to $\sim 54^\circ$,
1087 not completely random. The data shown in Figure 24 also show that when the onion epidermal cell
1088 wall is stretched in one direction, the average orientations of all polymers (both cellulose
1089 microfibrils and pectin matrix molecules) in the cell wall are tilted toward the stretch direction by
1090 similar amounts. These results imply that all cell wall polymers (not just cellulose microfibrils, but

1091 also matrix polymers) are passively reoriented or tilted along the stretch direction by external
1092 mechanical force.

1093



1094

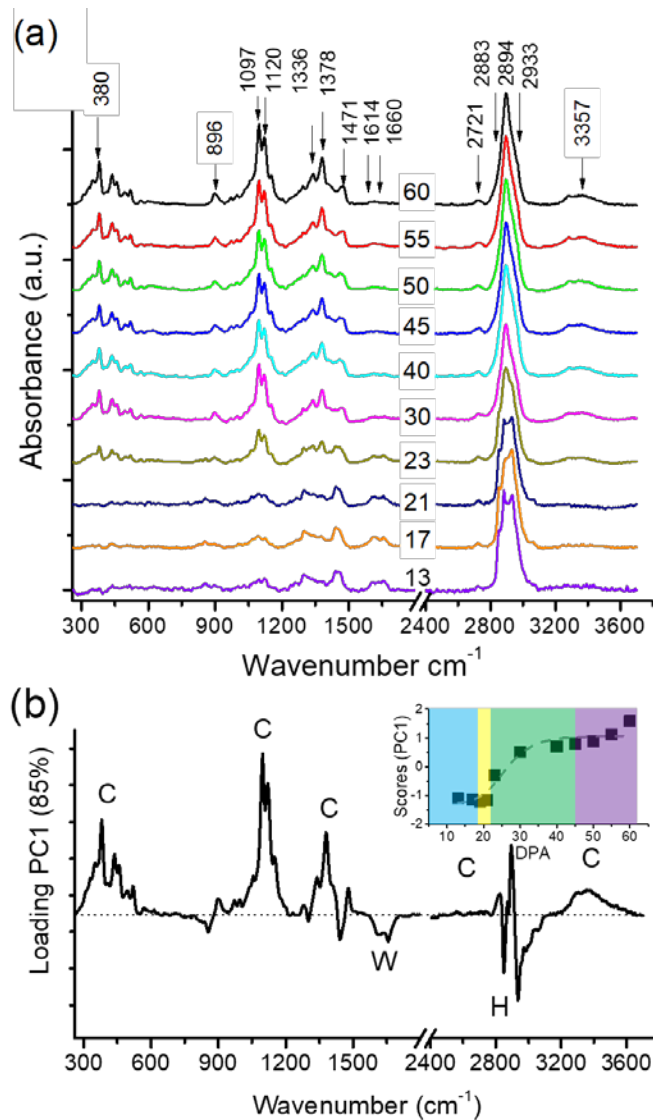
1095 **Figure 24.** Average orientation angle of cellulose microfibrils and pectin molecules in the control
1096 and mechanically-stretched onion epidermal wall calculated from the $\sim 1160\text{ cm}^{-1}$ and 1745 cm^{-1}
1097 dichroism. The data were collected in the transmission mode, thus the molecular chains parallel to
1098 the surface can be measured in this analysis (Kafle et al. 2017).

1099

1100 The linear dichroism analysis of IR absorption (eq.(5)) can also be used to find the
1101 transition dipole angle with respect to the reference axis ($\Omega=0^\circ$). In the celery collenchyma, it is
1102 known that cellulose microfibrils are aligned along the cell elongation direction (Thomas et al.
1103 2012). The polarized IR spectra of cellulose in the celery collenchyma at $\Omega = 0^\circ$ and 90° showed
1104 the spectral features similar to the ones in Figure 11 (Lee et al. 2015d). When the dipole moment
1105 angles $\mu(\alpha)$ of the OH stretch modes are calculated using eq.(5) and the distribution of cellulose
1106 chain orientation from the 2D XRD analysis, the results are in good agreement with the data shown
1107 in Figure 12c (Lee et al. 2015d).

1108

1109 **5.2 Raman analysis of cellulose in plant cell walls**



1110

1111 **Figure 25.** (a) Raman spectra of cotton fiber in different developmental gradients from 13 to 60
 1112 DPA. (b) loading plot of principle component 1 (PC1) representing the major changes in Raman
 1113 spectrum as the cotton fiber grows. C: Cellulose, W: Unsaturated hydrocarbon, and H: Long chain
 1114 alkyl groups. Inset plots the scores of PC1 versus DPA. The inset is score plot from 13-60 DPA.
 1115 The colored region shows different stages from primary wall synthesis to maturation. Blue:
 1116 Primary wall synthesis, Yellow: Transitional layer synthesis, Purple: Secondary wall synthesis,
 1117 Green: Maturation stage. (Lee et al. 2015c).

1118

1119 The developmental change of plant cell wall can also be monitored using Raman
1120 spectroscopy (Atalla and Vanderhart 1988; Lee et al. 2015c; Sene et al. 1994). Raman
1121 spectroscopy is particularly useful for detection of lignin components because the Raman bands
1122 of aromatic rings in lignin do not overlap with other polysaccharide components (see Section 5.4).
1123 Similar to IR, the Raman analysis can monitor changes in composition of amorphous matrix and
1124 crystallinity of cellulose in plant cell wall. Figure 25 compares Raman spectra of *G. Hirsutum*
1125 cotton fibers, the same set shown in Figure 22. During the primary cell wall stage (13-21 DPAs),
1126 there is no peak discernable at 380 cm^{-1} , indicating that the crystalline cellulose content is low. As
1127 the secondary cell wall starts growing after 23 DPA, the 380 cm^{-1} peak becomes prominent. In the
1128 PC1 spectrum obtained from PCA analysis of these Raman spectra (Figure 25b), the peaks
1129 belonging to crystalline cellulose are pointing positive and the peaks belonging to water and
1130 proteins ($1600 - 1700\text{ cm}^{-1}$) and long alkyl chains of cuticle (two sharp peaks at 2850 cm^{-1} and
1131 2920 cm^{-1}) are pointing negative. Note that the OH stretch band of water (broad negative band
1132 from $3000 - 3600\text{ cm}^{-1}$) is not seen in Raman; this is just because Raman is not highly sensitive to
1133 the OH stretch of water. The score plot of PC1 of Raman bands (Figure 25b, inset) as a function of
1134 DPA vary similar to that of IR bands (Figure 22b, inset).

1135

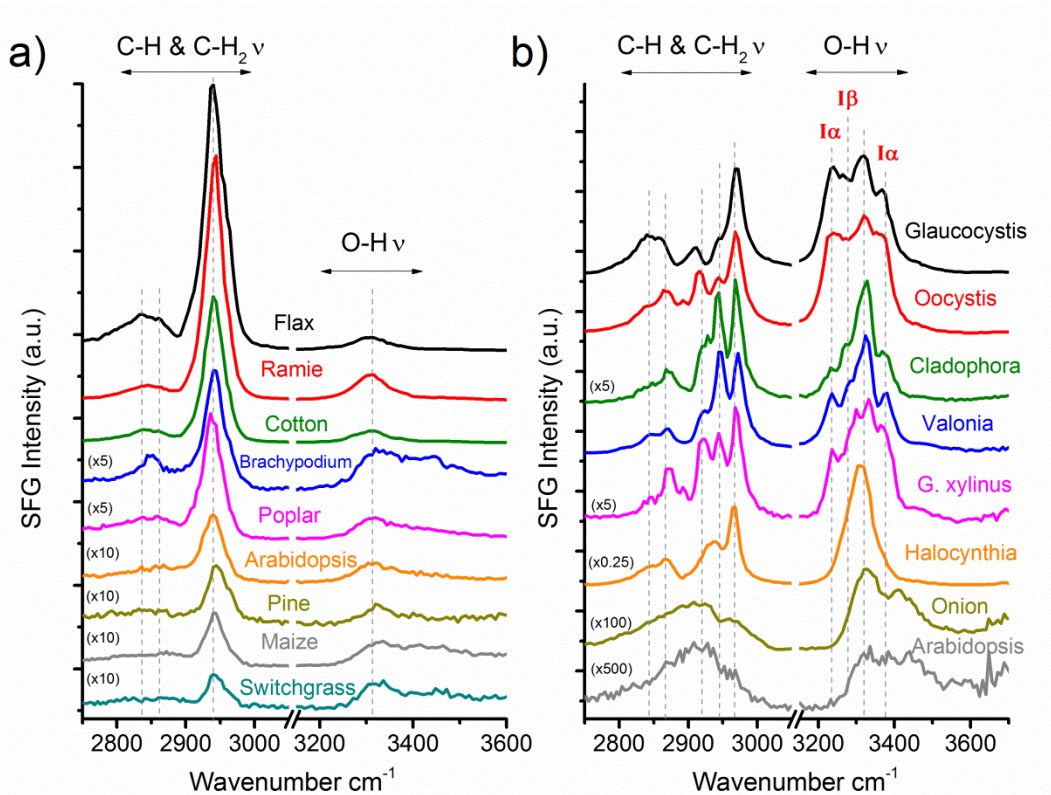
1136 ***5.3 SFG analysis of cellulose in plant cell walls***

1137 Due to the non-centrosymmetry requirement of non-linear optical process (see Section 2.2),
1138 SFG can selectively probe crystalline cellulose in plant cell walls and lignocellulose biomass
1139 without spectral interferences from amorphous matrix polymers such as pectin, hemicellulose, and
1140 lignin. This is the main advantage of SFG over IR and Raman analysis, if the main goal is just to
1141 detect cellulose (Barnette et al. 2011). However, due to the same non-centrosymmetry requirement
1142 as well as the phase synchronization condition for three photons with different wavelengths, the
1143 interpretation of SFG spectral features is not as simple as analysis of IR and Raman spectra which
1144 requires only peak identification, deconvolution, and quantification. There has been a great deal
1145 of understanding on SFG responses of crystalline cellulose interspersed in amorphous matrices
1146 (Kafle et al. 2015b; Lee et al. 2015b; Lee et al. 2015c; Lee et al. 2014; Makarem et al. 2018;
1147 Makarem et al. 2017; Park et al. 2014; Park et al. 2013) With recent advances in this subject, SFG

1148 became a powerful analytical tool for structural analysis of crystalline cellulose in intact plant cell
1149 walls and lignocellulose biomass. SFG is now capable of determining polymorphic structures
1150 governed by chain packing (structure at the nanoscale; see Figures 9c–9f), and spatial
1151 arrangements of cellulose microfibrils in plant cell walls (structure at the mesoscale; reviewed in
1152 this section), and variances among different cells (structure at the microscale; see Section 5.4).

1153 Figure 26 compares SFG spectra of cellulose microfibrils in various types of plants as well
1154 as a bacterial pellicle (*G. xylinus*) and a tunic of *Halocynthia* (Lee et al. 2014). The SFG spectral
1155 features of cellulose microfibrils in air-dried, but otherwise intact, cell walls can be divided into
1156 two groups – (i) a relatively large single CH/CH₂ stretch peak between 2900 cm⁻¹ and 3000 cm⁻¹
1157 and relatively weak OH peak in the 3200 – 3500 cm⁻¹ region (Figure 26a) and (ii) multiple peaks
1158 in the 2900 – 3000 cm⁻¹ region and relatively large peaks in the OH stretch region (Figure 26b).
1159 The SFG spectral pattern shown in Figure 26a is observed for cell walls of fully grown land plants
1160 that are categorized as secondary cell walls – stems, woody tissues, and vegetative fibers. The
1161 pattern shown in Figure 26b is observed for cellulose in algal cell walls, bacterial pellicles, tunics
1162 of tunicates, and walls of young growing cell walls (called primary cell walls) in land plants.

1163 Using the insights obtained from in Figure 19 (Chen et al. 2017) and Figure 21 (Lee et al.
1164 2014), the spectral difference between these two groups can be explained with the meso-scale
1165 packing of cellulose microfibrils. Secondary cell walls of land plants are known to have S1, S2,
1166 and S3 layers with different thicknesses (Figure 3) (Barnett and Bonham 2004). In each layer,
1167 cellulose microfibrils are aligned along a preferential axis. The angle between this axis and the
1168 main cell axis is called microfibril angle (MFA). Each layer can have different MFAs (Barnett and
1169 Bonham 2004). In each layer, cellulose microfibrils are deposited along this axis without specific
1170 preference in polarity. In other words, the number of cellulose microfibrils pointing to one
1171 direction would be equal to that pointing to the other direction. Thus, the net dipole would be zero.
1172 In such cases, the OH SFG intensity becomes small due to the symmetry cancellation (Figure 19)
1173 (Chen et al. 2017). Thus, the OH SFG intensities of secondary cell walls are generally weak and
1174 often significantly weaker than the CH/CH₂ SFG intensity when the cellulose concentration is high
1175 (Figure 26a). In this structure, the 2944 cm⁻¹ peak is the main component in the CH/CH₂ stretch
1176 region (Figure 26a).



1178

1179 **Figure 26.** SFG spectra of cellulose in biological tissues. **a)** Mature secondary cell walls of land
 1180 plants: flax, ramie, cotton, *Brachypodium*, poplar, *Arabidopsis*, pine, maize and switchgrass. **b)**
 1181 Algal cell walls (*Glaucozystis*, *Oocystis*, *Valonia*, and *Cladophora*), cellulose biofilm produced
 1182 from *G. xylinus*, *Halocynthia* mantle, Onion epidermis and *Arabidopsis* aerial tissue. (Lee et al.
 1183 2014)

1184

1185 Primary cell walls typically consist of multiple lamellae; each lamella is a few tens of
 1186 nanometer thick (much thinner than the SFG coherence length) (Anderson et al. 2010; McCann et
 1187 al. 1990; Zhang et al. 2016). In each lamella, cellulose microfibrils are laid along some preferential
 1188 orientation, but that orientation is progressively changing among adjacent lamellae (Anderson et
 1189 al. 2010; Zhang et al. 2016). So, many possible orientations are likely to be found within the SFG
 1190 coherence length, which is similar to the situation of the randomly-packed cellulose crystals in
 1191 Figure 21 (Lee et al. 2014). Similarly, cellulose microfibrils in bacterial pellicles are deposited in

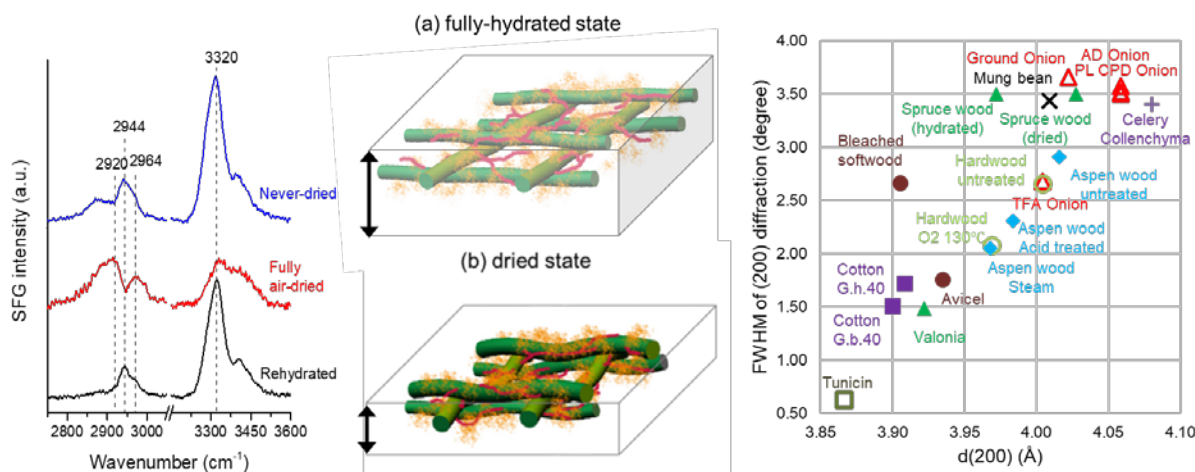
1192 random directions in two-dimensional lamellae (Lee et al. 2015b). In algal cell walls, cellulose
1193 microfibrils are often deposited in the cross-lamellar fashion. In such situations, the symmetry
1194 cancellation effect of OH dipoles within the SFG coherence length is relatively small compared to
1195 the uniaxial and anti-parallel packing of fibrils. Thus, the OH SFG intensity is high – comparable
1196 to or even larger than to the CH/CH₂ SFG peaks (Figure 26b). For these non-uniaxial packing
1197 structures, the multiple SFG peaks are observed in the 2920 – 2970 cm⁻¹ region and sometimes the
1198 2944 cm⁻¹ is a minor component (Figure 26b).

1199 In IR spectroscopy, the attenuation of IR by water can interfere with detection of absorption
1200 peaks of cellulose and other matrix polymers in primary cell walls in the fully hydrated state (Chen
1201 et al. 1997; McCann et al. 1992; Wilson et al. 2000). Similarly, the amorphous scattering of water
1202 makes it difficult to use wide-angle XRD for detection of diffraction peaks of crystalline cellulose
1203 in native PCWs (Newman et al. 2013; Thomas et al. 2012). For that reason, primary cell walls
1204 have often been analyzed in the dehydrated state. However, dehydration alters the gel-like
1205 structures of pectin in PCWs (Ha et al. 1997; Jarvis 1992; Jarvis 1984); thus, dehydration can
1206 change the porosity, thickness and rigidity of the cell walls, which may result in alteration of
1207 cellulose – matrix polymer interactions (Wang et al. 2015) Such dehydration will not change the
1208 molecular structure of cellulose microfibrils, but it can change the crystalline order or packing of
1209 cellulose microfibrils because the (200) width of elementary microfibrils is only ~3.6 nm
1210 (Donaldson and Xu 2005; Zhang et al. 2016) and a large portion of chains are exposed at the
1211 microfibril surface interacting with water molecules and other matrix polymers (see Figure 1d).

1212 Figure 27 shows how the SFG spectral features are changed upon dehydration and
1213 rehydration of an onion epidermis cell wall used as a model primary cell wall (Huang et al. 2018b).
1214 The most significant changes in SFG spectral features upon dehydration are the decrease in the
1215 OH stretch band intensity and the disappearance of the 2944 cm⁻¹ accompanied by the appearance
1216 or dominance of the 2920 and 2964 cm⁻¹ peaks in the CH/CH₂ stretch region. These changes are
1217 associated with the decrease in the cell wall thickness by ~60% upon dehydration due to the
1218 collapse of pectin network from the fully-hydrated hydrogel-like state to the collapsed glassy state
1219 (Huang et al. 2018b). The collapse of pectin networks would not be uniform because of the
1220 presence of load-bearing biomechanical hot spots in the cellulose microfibril network (Park and
1221 Cosgrove 2012); this could result in inhomogeneous strains in cellulose microfibrils (Figure 27c).

1222 During this physical collapse, it is possible that cellulose microfibrils interspersed in the originally
 1223 swollen pectin network are subject to inhomogeneous physical deformation, as illustrated in
 1224 Figures 27b and 27c. Such deformation can decrease the coherency of OH stretch vibration modes
 1225 along the microfibril axis, which will reduce the OH SFG intensity. Inhomogeneous local strains
 1226 can also alter the coherency of CH/CH₂ stretch modes such that symmetry-forbidden bands in the
 1227 relaxed state can become SFG-active in the strained state (Lee et al. 2014; Lee et al. 2013b).

1228



1229

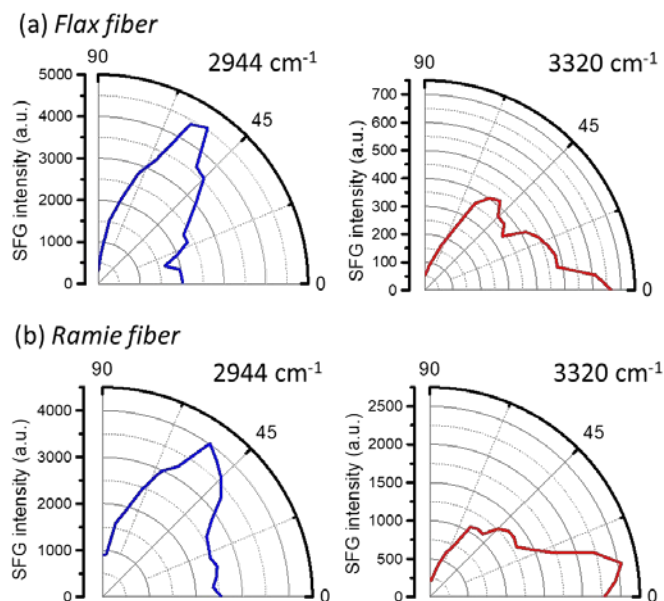
1230 **Figure 27.** (a) Comparison of SFG spectra of cellulose microfibrils in the abaxial epidermal wall
 1231 of the fifth scale of onion bulb in the never-dried, air-dried, and re-hydrated states. (b,c) Schematic
 1232 representation of cellulose microfibrils in (b) fully-hydrated and (b) dehydrated primary cell walls.
 1233 cellulose microfibrils (green rods) are embedded in the pectin network (orange splatters) which
 1234 also contains a hemicellulose (red lines). (d) Relationship between full-width-half-maximum and
 1235 d-spacing of the (200) plane diffraction determined from XRD of cellulose in various dehydrated
 1236 plant cell wall samples (Huang et al. 2018b).

1237

1238 The fact that cellulose microfibrils in the dehydrated primary cell walls have locally
 1239 inhomogeneous strains, which was revealed by SFG, has a significant implication in XRD analysis
 1240 (Huang et al. 2018b). In XRD analysis, the full-width-half-maximum and d-spacing of (200)
 1241 diffraction peak of cellulose I β are substantially increased for the primary cell walls compared to
 1242 cellulose in secondary cell walls (Figure 27d). The larger d-spacing might be due to strain in the
 1243 unit cell. The inhomogeneity in strain also increases the FWHM of the diffraction peak

1244 (Cammarata and Eby 1991; Cammarata 1994; Rivnay et al. 2011). This means that the FWHM of
1245 the (200) diffraction peak is a function of the degree of inhomogeneous strains as well as the crystal
1246 size. Thus, using the Scherrer equation with the FWHM value of the (200) diffraction peak
1247 measured from XRD data of the dehydrated sample will lead to an error in estimation of the crystal
1248 size or diameter of cellulose microfibrils in primary cell walls. It is possible to deconvolute the
1249 inhomogeneous deformation effect in the FWHM of diffraction peaks, but it requires detection of
1250 multiple sharp peaks for regression (Williamson and Hall 1953). Unfortunately, the XRD of
1251 cellulose in primary cell walls gives only two or three broad peaks, which are too few to do
1252 deconvolution.

1253



1254

1255 **Figure 28.** Azimuth angle dependence of the 2944 cm⁻¹ (CH/CH₂) and 3320 cm⁻¹ (OH) SFG peaks
1256 of uniaxially-aligned cellulose microfibrils in (a) flax and (b) ramie fibers. SFG signals were
1257 collected with the *ssp* polarization and the azimuth angle is defined as the angle between the fiber
1258 axis (perpendicular to the azimuth rotation axis) and the laser incidence plane.

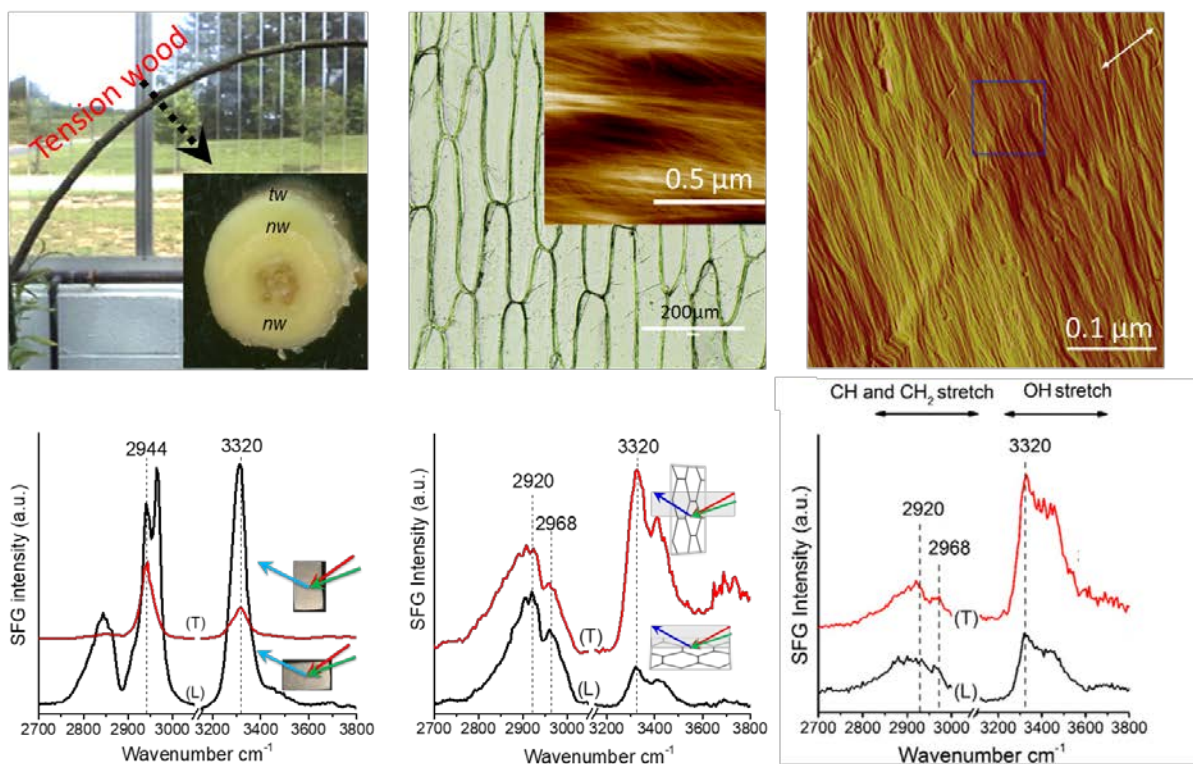
1259

1260 SFG can provide the cellulose microfibril orientation information in plant cell walls. The
1261 basic principle is already explained in Figure 15 (Chen et al. 2017). Figure 28 displays the azimuth

1262 angle (φ) dependence of the *ssp* SFG signals of the 2944 cm^{-1} (CH/CH₂ stretch) and 3320 cm^{-1}
1263 (OH stretch) peaks of cellulose microfibrils in flax and ramie fibers. In these vegetative fibers,
1264 cellulose microfibrils are highly aligned along the fiber axis. The average microfibril angle
1265 calculated using XRD is $\sim 9^\circ$ for flax fibers and $\sim 7.5^\circ$ for ramie fibers (Bourmaud et al. 2013; Goda
1266 et al. 2006). The 2944 cm^{-1} peak has the maximum intensity around $\varphi \approx 56^\circ$ (consistent with Figure
1267 15a). The 3320 cm^{-1} intensity is maximum at $\varphi = 0^\circ$ and extremely small at $\varphi = 90^\circ$ at the *ssp*
1268 polarization. Based on this data, it can be said that cellulose fibers are aligned more along the *ssp*
1269 incidence plane when the 3320 cm^{-1} / 2944 cm^{-1} intensity ratio increases as φ decreases to 0° .

1270 This simple empirical rule has been employed to find if the preferential orientation angle
1271 of cellulose microfibrils averaged over the SFG probe volume is tilted toward to the main axis of
1272 the cell or closer to the perpendicular direction of the cell main axis. The long and short axes of an
1273 anisotropic cell are called longitudinal and transverse directions, respectively. Figure 29a displays
1274 a picture of tension wood formed in a poplar stem grown under gravitropic stress and the *ssp* SFG
1275 spectra collected at the parallel ($\varphi = 0^\circ$) and perpendicular ($\varphi = 90^\circ$) to the stem axis. The OH
1276 signal is significantly stronger at $\varphi = 0^\circ$, indicating that cellulose microfibrils in the tension wood
1277 section are highly aligned along the stem direction (Kafle et al. 2014b). Figure 29b exhibits an
1278 optical image of the abaxial epidermal cell wall of the second scale of an onion bulb and an AFM
1279 image of cellulose microfibrils exposed and accessible with AFM tip from the cytoplasm side; it
1280 also compares the *ssp* SFG spectra of the onion epidermal cell wall collected at the longitudinal
1281 ($\varphi = 0^\circ$) and transverse ($\varphi = 90^\circ$) directions (Park et al. 2014). The AFM image clearly showed
1282 that the newly-deposited cellulose microfibrils exposed at the cytoplasm side are highly oriented
1283 in the transverse direction, but it cannot not provide any information about microfibrils that are
1284 not accessible by the AFM tip. The SFG spectra clearly showed that the OH SFG intensity is
1285 significantly enhanced at the transverse direction, indicating that the orientation averaged over the
1286 entire cell wall thickness is closer to the transverse direction. Similarly, the comparison of *ssp* SFG
1287 spectra of cellulose collected for a hypocotyl cell wall of *Arabidopsis jia-1* mutant also shows a
1288 similar pattern as the onion cell wall (Figure 29c). The *jia-1* mutant tends to deposit cellulose
1289 microfibrils preferentially along the transverse direction, similar to the onion cell wall, than the
1290 longitudinal direction (Lei et al. 2014).

1291



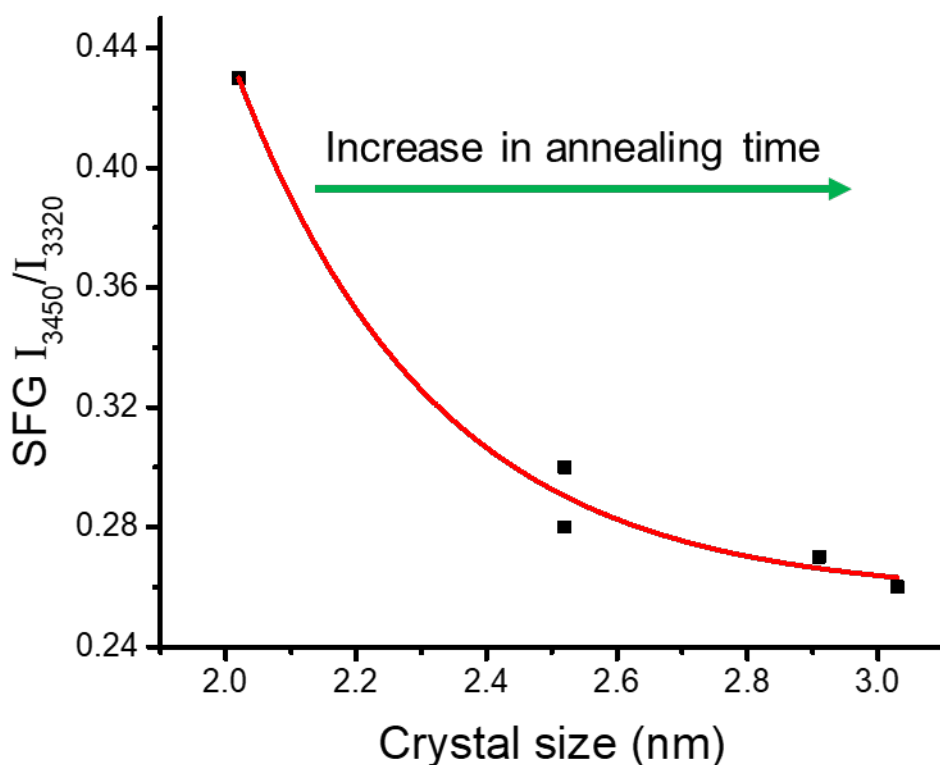
1293

1294 **Figure 29.** Dependence of SFG spectral features on preferential orientation of cellulose
 1295 microfibrils in (a) tension wood, (Kafle et al. 2014b) (b) onion epidermal cell wall (Park et al.
 1296 2014), and (c) hypocotyl of *jia-1* mutant of *Arabidopsis* (Lei et al. 2014).

1297

1298 Another possibility of using SFG is to distinguish the OH groups exposed at the crystalline
 1299 domain surface interacting with surrounding molecules versus the OH groups inside the domain
 1300 keeping the integrity of the crystalline order. From the SFG scattering experiment (Figure 16), it
 1301 is found that the surface OH groups tend to have weaker hydrogen bonding interactions, so their
 1302 OH stretch bands appear in the region higher than 3400 cm^{-1} , while the bulk OH groups are
 1303 involved in strong hydrogen bonds and thus their vibrations are in the lower wavenumber region
 1304 ($\leq 3400\text{ cm}^{-1}$). The OH stretch dipole is also different for the surface versus bulk OH groups (see
 1305 Figures 12c, 13e, and 13g). These findings allow the deconvolution of OH stretch bands into the
 1306 surface and bulk components. Figure 30 plots the ratio between the 3450 cm^{-1} (surface) and 3320

1307 cm^{-1} (bulk) SFG intensities of hardwood samples treated with oxygen delignification at 130°C
1308 (Kafle et al. 2015a). As the delignification time at 130 °C increases, the (200) diffraction peak in
1309 XRD becomes sharper and the surface-to-bulk SFG intensity ratio decreases. Based on the
1310 Scherrer equation, the decrease in XRD peak width means the increase in crystal size; this might
1311 be due to the annealing effect upon heating at 130°C. Thus, this data set supports that changes in
1312 the surface OH and bulk OH groups can be monitored with SFG.



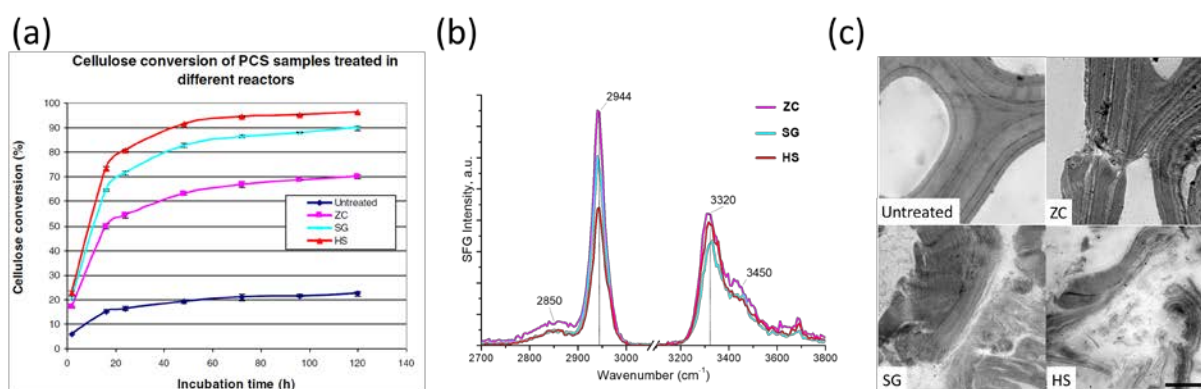
1313
1314 **Figure 30.** Correlation between the surface-to-bulk OH peak intensity ratio (I_{3450}/I_{3320}) in SFG and
1315 the domain size of the (200) crystallographic plane estimated from XRD using Scherrer equation.
1316 for hardwood samples treated with oxygen delignification at 130°C. The data are taken from (Kafle
1317 et al. 2015a).

1318
1319 The amount of crystalline cellulose in plant cell walls or lignocellulose biomass can be
1320 estimated if proper calibration curves can be constructed and used (Barnette et al. 2011; Kafle et

1321 al. 2015a). Such calibration curves can be obtained by measuring SFG signals of reference samples
1322 prepared with known concentrations. In such approaches, the crystalline cellulose components
1323 chosen for preparation of reference samples must have spatial distributions of crystalline cellulose
1324 domains similar to the cellulose microfibril packing patterns of the sample of interest. If this
1325 condition is uncertain or not confirmed independently, then the crystalline cellulose amount
1326 estimated with the calibration curve should be considered only as a qualitative guide or the value
1327 equivalent to the concentration of model compound (Handakumbura et al. 2013; Park et al. 2013).

1328 If the total amount of cellulose content in the sample can be determined through
1329 independent methods, then changes in the SFG spectral features can be interpreted in terms of
1330 spatial distribution of crystalline cellulose domains in the sample. Figure 20 is a good example
1331 where the relative intensity of CH and OH stretch modes in SFG can be correlated with the average
1332 separation distance between crystalline domains that are uniaxially aligned with antiparallel
1333 directionality in average over the SFG coherence length (Makarem et al. 2017). This principle has
1334 been employed to explain the origin for enhanced efficiency of pre-treated lignocellulose biomass
1335 (Wang et al. 2014b) (Figure 31), superior material properties of *G. barbadense* cotton fibers
1336 over *G. hirsutum* fibers (Lee et al. 2015c), inferior properties of *immature (im)* mutants of cotton
1337 compared to the Texas marker (TM) species (Kim et al. 2017), and degradation mechanism of
1338 woody cell walls by Brown Lot fungi (Goodell et al. 2017).

1339



1340

1341 **Figure 31.** (a) Enzymatic digestion, (b) SFG spectra, and (c) TEM images of acid-pretreated corn
1342 stover in zipper-clave (ZP), steam Gun (SG), and horizontal-screw (SC) reactors. In (a) and (c), the
1343 untreated sample data are also shown for comparison. The severity of physical shear during the

1344 pretreatment increases in the order of ZP<SG<HS. The large increase in enzymatic digestion yield
1345 of cellulose with increasing the shear action is attributed to larger separation among cellulose
1346 microfibrils as suggested from a smaller CH/OH ratio in SFG and a more disintegration of cell
1347 walls. (Wang et al. 2014b)

1348

1349 *5.4 Vibrational spectroscopic imaging of cellulose in plant cell walls*

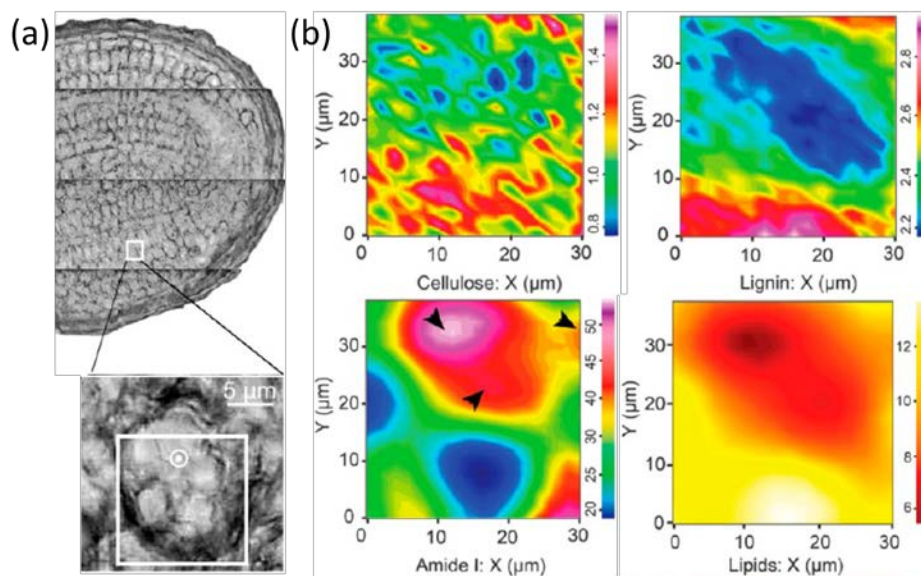
1350 Plants have diverse cell types and the arrangement of cell wall components can be
1351 inhomogeneous inside each cell type (Li et al. 2016; Zhang et al. 2016; Zhong et al. 2002). For
1352 example, in 4-day old Arabidopsis hypocotyls, there are different cell types (Fagard et al. 2000;
1353 Hohm et al. 2014) and each cell type can have its own way of arranging cellulose microfibril.
1354 These variances are directly related to diverse functions of different cell types and such vital
1355 information could be missed if one relies only on analysis techniques that average the signal over
1356 a large volume encompassing many different types of cell walls. Thus, it is critical to employ
1357 analytical techniques with sufficiently high spatial resolutions to conduct spectroscopic
1358 characterizations at the individual cellular level. By combining the vibrational spectroscopy with
1359 optical microscopy, it is possible to obtain vibrational spectrum of each cell walls (Gierlinger
1360 2018). Chemical mapping can be done either by plotting the intensity of a particular peak or
1361 drawing chemometric map of the scanning area.

1362 In the case of IR and Raman spectroscopy, the spatial resolution is determined by the
1363 Abby's diffraction limit (Lasch and Naumann 2006): $\Delta x \geq 0.61 \frac{\lambda}{n \sin \theta}$ where λ is the wavelength
1364 of the probe beam, n is the refractive index and θ is the half-angle of the focusing beam (note that
1365 $n \sin \theta$ is known as numerical aperture, NA). In addition to diffraction limitation, the optical
1366 design, detector, and signal intensity can also influence the spatial resolution. Because of long
1367 wavelength in IR spectroscopy, the spatial resolution varies typically from 2-3 μm in the CH and
1368 OH stretch region to $\sim 10 \mu\text{m}$ in the finger print region (Matthäus et al. 2008). In the case of Raman
1369 spectroscopy, the spatial resolution could go down to 300-500 nm in ideal cases (Matthäus et al.
1370 2008). For SFG microscopy, the spatial resolution is limited by the overlap of the focused visible
1371 and tunable IR beams. A sub-micron spatial resolution has been demonstrated for SFG microscopy

1372 (Li et al. 2017; Raghunathan et al. 2011). But, typical systems deliver a few microns resolution
1373 (Hoang Chi et al. 2011; Kouyama et al. 2014).

1374 Figure 32 shows an example of IR microscopy analysis to understand the distribution of
1375 cellulose, lignin, protein, and lipid in the cotyledon parenchyma cells inside Arabidopsis seed
1376 (Withana-Gamage et al. 2013). The strong intensity of amide-I band (1650 cm^{-1}) near the center
1377 of the cell indicates that the parenchyma cell is rich in proteins. The spatial distribution of the 1740
1378 cm^{-1} intensity, assigned to lipid, correlates negatively with the protein distribution. In contrast, the
1379 peak intensities associated with cellulose (1059 cm^{-1}) and lignin (1513 cm^{-1}) are negligible in the
1380 scanning area. Here, caution must be taken because the peak assignments used to draw the chemical
1381 maps can sometime be incorrect. For example, the 1059 cm^{-1} peak cannot be attributed solely to
1382 cellulose; all polysaccharides in the cell wall show broad peaks in the $1000\text{-}1100\text{ cm}^{-1}$ region (see
1383 Figure 8a).

1384



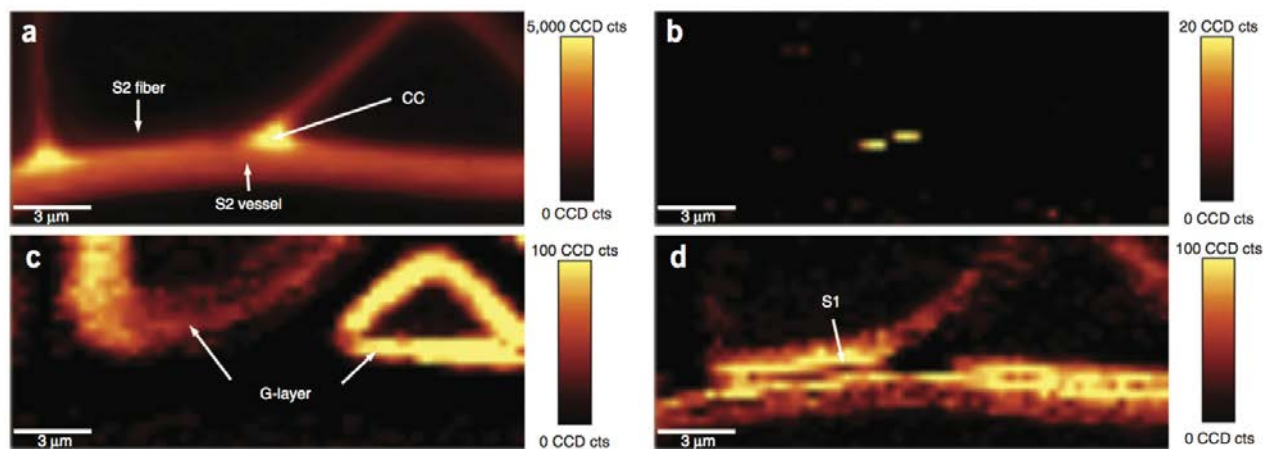
1385

1386 **Figure 32.** Infrared microscopy imaging of cotyledon parenchyma cells inside Arabidopsis seed.
1387 The cross section of an individual cell is subject to chemical mapping. (a) Optical image of the
1388 cross section of Arabidopsis seed. (b) IR intensity map of cellulose (1059 cm^{-1}), lignin (1513 cm^{-1})
1389 1), lipids (1740 cm^{-1}) and amide-1 (1650 cm^{-1}) bands. (Withana-Gamage et al. 2013)

1390

1391 Raman microscopy can be more advantageous comparing to IR microscopy when higher
1392 resolution chemical mapping is needed. If non-resonance background due to strong fluorescence
1393 can be suppressed effectively (Agarwal 2006), Raman microscopy can particularly be valuable to
1394 study lignin because of its high sensitivity and distinct peak position in Raman analysis. Figure 33
1395 is an example of Raman microscopy mapping of cellulose, lignin, and pectin species in the
1396 secondary cell wall of a poplar tension wood (Gierlinger 2018). The peaks between 1530-1715
1397 cm^{-1} , assigned to lignin, is strong in the middle region of the cell wall. The distribution of the 834-
1398 855 cm^{-1} peak, assigned to pectin, is negligible except at the middle lamella region (marked as cc
1399 in Figure 33a). That is probably because the thickness of the primary cell walls containing pectin
1400 is too thin compared to the thickened secondary cell walls. The distributions of the 1374 and 1096
1401 cm^{-1} peaks are different which can both be attributed to cellulose peaks (see Figure 6). The 1374
1402 cm^{-1} intensity is dominant in the S3 (G-layer) region and the 1096 cm^{-1} intensity is strong in the
1403 S1 region. The 1096 cm^{-1} mode is highly polarized, while the 1374 cm^{-1} mode is polarization-
1404 insensitive (see Figure 14). Based on this polarization dependence, it was interpreted that the
1405 orientation of cellulose microfibrils is different in the S3 and S1 regions.

1406



1407
1408 **Figure 33.** Microscopic Raman map of a tension wood section of poplar: (a) 1530-1715 cm^{-1}
1409 assigned to lignin, (b) 835-855 cm^{-1} assigned to pectin, (c) 1374 cm^{-1} assigned to cellulose, and
1410 (d) 1096 cm^{-1} assigned to orientation sensitive cellulose. (Gierlinger 2018)

1411

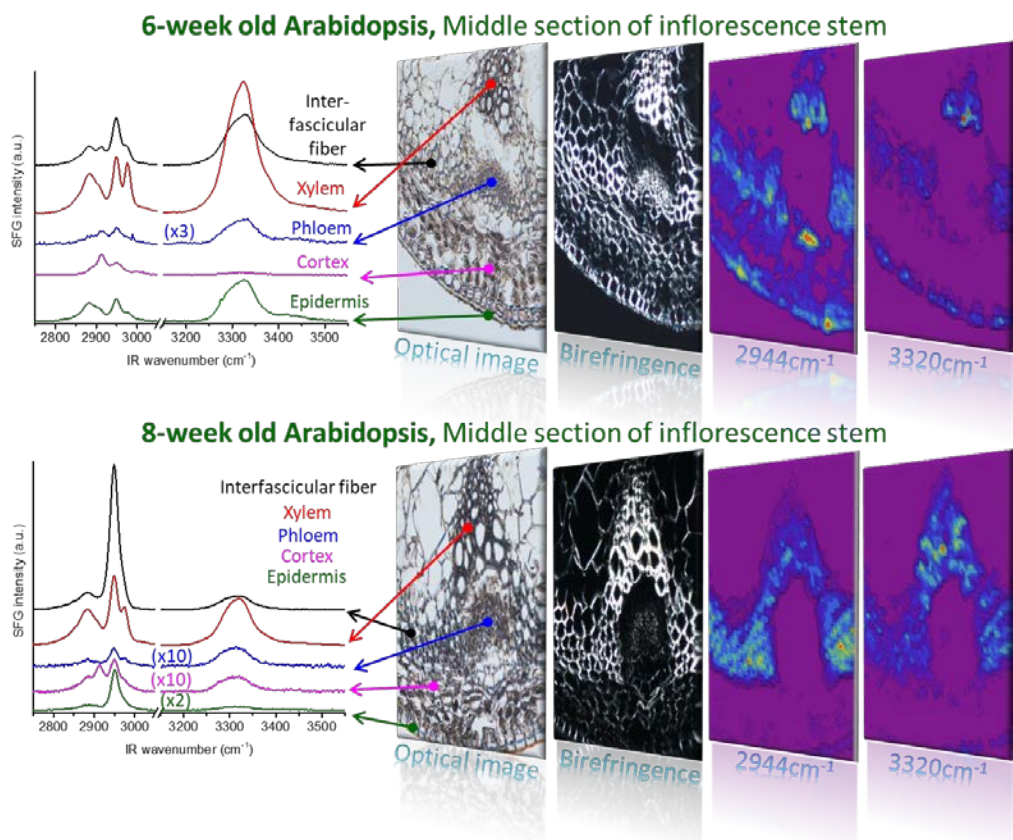
1412 As in the case of IR, it should be noted that the 1096 and 1373 cm^{-1} peaks cannot be
1413 attributed to cellulose only. Although the overall cellulose content is higher in tension wood than
1414 normal wood, it still contains a significant amount of hemicellulose such as xylan (Kafle et al.
1415 2014b). Because the G-layer of tension wood is known to have a very high cellulose content, the
1416 strong intensity at 1373 cm^{-1} in the G-layer region can be safely attributed to cellulose. However,
1417 the 1096 cm^{-1} peak in the S1 layer region cannot be solely due to cellulose. For example, xylan
1418 also gives strong peaks at 1095–1120 cm^{-1} and 1370–1383 cm^{-1} region, just as does cellulose (Shen
1419 et al. 2004; Zeng et al. 2016).

1420 The problem of peak overlaps between cellulose and hemicellulose species in IR and
1421 Raman microscopy analysis is circumvented if SFG microscopy is used (Huang et al. 2018a).
1422 Although SFG microscopy can detect cellulose only (Figure 8b), the SFG spectral features of
1423 cellulose are highly coupled with spatial distributions of cellulose microfibrils in the amorphous
1424 matrix (for example, see Figures 19, 20, 21, 26, 27, and 29). So, changes in the matrix polymer
1425 composition and amount will be reflect in the cellulose SFG spectral features. Also, SFG can
1426 distinguish polymorphic structures of crystalline cellulose (Figures 9c-9f).

1427 SFG microscopy has been applied to analysis of cellulose structures at the tissue-level
1428 resolution for the inflorescence stems of Arabidopsis (Huang et al. 2018a). Figure 34 displays SFG
1429 spectra of cellulose in interfascicular fiber (IFF), xylem, phloem, cortex, and epidermis regions of
1430 the 6-week and 8-week old Arabidopsis stems. Based on the CH/OH intensity ratio, it can be seen
1431 that cellulose microfibrils in the IFF are not highly packed in the uniaxial direction in the stem of
1432 a 6-week old plant and are fully developed into the uniaxial direction in the 8-week old stem. The
1433 meso-scale packing of cellulose microfibrils in the xylem region is quite different from that in the
1434 IFF region. The hyper-spectral maps also show clearly the difference in spatial distributions of the
1435 CH/CH₂ (2944 cm^{-1}) and OH (3320 cm^{-1}) stretch peak intensities. Two primary walls, cortex and
1436 phloem, do not show drastic changes in their SFG intensity meaning the wall thickness and
1437 cellulose content do not change drastically in these walls. The change in the ratio of CH OH stretch
1438 intensities could mean some alterations in the average orientation of cellulose microfibrils in these
1439 walls. The SFG spectra of epidermal cell wall in 6-week and 8-week plants show the similar trend
1440 observed in the IFF region. The strong 2944 cm^{-1} peak and weaker OH region peaks in 8-week
1441 section indicates better organizations of microfibrils in epidermal cell wall. At this fully-grown

1442 stage, the epidermis cell walls may not need to expand anymore and the well-organized cellulose
1443 microfibrils in the cell wall may enhance mechanical strength of the wall for better protection from
1444 environmental stimuli or invasions.

1445



1446

1447 **Figure 34.** SFG spectra of five different cell types in the inflorescence stem of (a) 6-week and (b)
1448 8-week old Arabidopsis. Also shown are optical and birefringence images as well as SFG
1449 hyperspectral images at 2944 cm⁻¹ and 3320 cm⁻¹ of the stem cross-section. (Huang et al. 2018a)

1450

1451 5. Summary and Perspective

1452 The experimental data available from IR, Raman and SFG spectroscopy complemented
1453 with DFT calculations are of great value for understanding the structure of cellulose and surface
1454 interactions (e.g., water adsorption, cellulose bundling, etc.). Although DFT calculations have

1455 limitations in terms of accuracy (i.e., agreement to $\leq 1 \text{ cm}^{-1}$ cannot be expected) and the ability to
1456 predict intensities and bandwidths, they are reliable enough to provide an objective method for
1457 interpreting vibrational spectra. For example, the coupling of O-H stretches in hydrogen-bonded
1458 hydroxyl groups is clear from animations of DFT-calculated modes but would be hard to envision
1459 via experimental study alone. When combined with comparisons against other experimental
1460 observables such as XRD and NMR spectroscopy, DFT can be benchmarked to distinguish among
1461 competing models (Kubicki et al. 2013). Once reliable models are in hand, the same DFT methods
1462 can be used to provide details not apparent from experimental data and to explore potential
1463 chemistries that made be difficult to probe via analytical methods. Advancements in theoretical
1464 methods, modeling software and computational hardware will continue to improve the accuracy
1465 of DFT and allow for greater insight into molecular-level behavior hidden within complex systems.

1466 Once the vibrational modes corresponding to observable peaks are well understood, then
1467 vibrational spectroscopy can be highly useful and powerful for structural analysis and
1468 quantification of cellulose (and other wall matrix components) in plant cell walls and
1469 lignocellulose biomass. The newly introduced SFG spectroscopy can provide mesoscale structural
1470 information of crystalline cellulose in *intact* cell walls that cannot be obtained with other
1471 characterization techniques such as IR, Raman, NMR, and XRD. Small angle x-ray and neutron
1472 scatterings could provide mesoscale spatial dimensions such as particle sizes or separation
1473 distances (Saxe et al. 2014; Thomas et al. 2012; Ye et al. 2018), but they are not molecular specific
1474 and cannot be combined with microscopic imaging. AFM and SEM can provide high-resolution
1475 images of cellulose microfibrils at the surface exposed outside (Zhang et al. 2017; Zhang et al.
1476 2016; Zheng et al. 2017). In contrast, SFG provides mesoscale structural information averaged
1477 over the entire cell wall thickness. Another advantage of SFG is that it can be applied to intact cell
1478 walls in the fully-hydrated state, which is extremely difficult for IR and Raman analysis. It is
1479 important to combine several analytical techniques that are complementary to each other to obtain
1480 more accurate and complete structural information.

1481 With better understanding of advantages and limitations of each vibrational spectroscopy
1482 technique, one can more effectively and accurately study biosynthesis of cellulose in plant cell
1483 walls (Cho et al. 2015; Liu 2013; Norris et al. 2017), bacterial synthesis of cellulose (Lee et al.
1484 2015b), cellulose-matrix interactions during cellulose synthesis in model systems (Park et al.

1485 2014), polymorphic structures of cellulose in algal cell walls (Park et al. 2015). With the
1486 development of more advanced imaging techniques coupled with vibrational spectroscopy (Ma et
1487 al. 2017), it would be possible to investigate the cellulose content and structure at the cellular level.
1488 Such developments will be particularly useful in studies of plant cell walls upon genetic mutations
1489 or under endogenous and exogenous stimuli, because different tissues may respond differently to
1490 such conditions. The vibrational spectroscopy analysis can also advance mechanistic
1491 understanding of biomass pretreatments and enzymatic conversion (Kafle et al. 2015b; Karimi and
1492 Taherzadeh 2016; Wang et al. 2014b) and well as textile processing (Kafle et al. 2014a).

1493

1494 **Supporting Information.**

1495 movies of vibration modes:

1496 Movie S7A.avi corresponding to 3410 cm^{-1} vibration mode

1497 Movie S7B.avi corresponding to 3370 cm^{-1} vibration mode

1498 Movie S7C.avi corresponding to 3330 cm^{-1} vibration mode

1499 Movie S7D.avi corresponding to 3270 cm^{-1} vibration mode

1500 Movie S7E.avi corresponding to 2944 cm^{-1} vibration mode

1501 Movie S7F.avi corresponding to 2867 cm^{-1} vibration mode

1502 Movie S7G.avi corresponding to 1477 cm^{-1} vibration mode

1503 Movie S7H.avi corresponding to 1098 cm^{-1} vibration mode

1504 Movie S7I.avi corresponding to 1169 cm^{-1} vibration mode

1505 Movie S7J.avi corresponding to 710 cm^{-1} vibration mode

1506 Movie S7K.avi corresponding to 380 cm^{-1} vibration mode

1507 Movie S7L.avi corresponding to 93 cm⁻¹ vibration mode

1508

1509 **Acknowledgements.** This work was supported by The Center for Lignocellulose Structure and
1510 Formation, Energy Frontier Research Center funded by the U.S. Department of Energy, Office of
1511 Science, Basic Energy Sciences, under Award Number DE-SC0001090. TD-DFT calculations of
1512 theoretical SFG spectra were from the work supported by the Air Force Office of Scientific
1513 Research (AFOSR) (Grant No. FA9550-16-1-0062).

1514

1515 **References**

- 1516 Abidi N, Hequet E, Cabrales L, Gannaway J, Wilkins T, Wells LW (2008) Evaluating cell wall
1517 structure and composition of developing cotton fibers using Fourier transform infrared
1518 spectroscopy and thermogravimetric analysis *J Appl Polym Sci* 107:476-486
- 1519 Agarwal UP (2006) Raman imaging to investigate ultrastructure and composition of plant cell
1520 walls: distribution of lignin and cellulose in black spruce wood (*Picea mariana*) *Planta*
1521 224:1141
- 1522 Agarwal UP, Ralph SA (1997) FT-Raman spectroscopy of wood: identifying contributions of
1523 lignin and carbohydrate polymers in the spectrum of black spruce (*Picea mariana*) *Appl*
1524 *Spectrosc* 51:1648-1655
- 1525 Agarwal UP, Ralph SA, Reiner RS, Baez C (2018) New cellulose crystallinity estimation method
1526 that differentiates between organized and crystalline phases *Carbohydr Polym* 190:262-
1527 270
- 1528 Agarwal UP, Reiner RS, Ralph SA (2010) Cellulose I crystallinity determination using FT-Raman
1529 spectroscopy: univariate and multivariate methods *Cellulose* 17:721-733
- 1530 Albersheim P, Darvill A, Roberts K, Sederoff R, Staehelin A (2010) *Plant cell walls*. Garland
1531 Science, Taylor & Francise Group, New York
- 1532 Anderson CT, Carroll A, Akhmetova L, Somerville C (2010) Real-time imaging of cellulose
1533 reorientation during cell wall expansion in *Arabidopsis* roots *Plant Physiol* 152:787-796
- 1534 Atalla RH, Vanderhart DL (1984) Native cellulose: a composite of two distinct crystalline forms
1535 *Science* 223:283-285
- 1536 Atalla RH, Vanderhart DL (1988) Studies on the structure of cellulose using Raman spectroscopy
1537 and solid state C NMR. In: *Cellulose and wood-chemistry and technology*. Wiley, New
1538 York, pp 169–188
- 1539 Atalla RH, Whitmore RE, Heimbach CJ (1980) Raman spectral evidence for molecular orientation
1540 in native cellulosic fibers *Macromolecules* 13:1717-1719
- 1541 Banwell CN, McCash EM (1994) *Fundamentals of molecular spectroscopy* vol 851. McGraw-Hill
1542 New York
- 1543 Barnett JR, Bonham VA (2004) Cellulose microfibril angle in the cell wall of wood fibres *Biol*
1544 *Rev Cambridge Philos Soc* 79:461-472

- 1545 Barnette AL et al. (2011) Selective detection of crystalline cellulose in plant cell walls with sum-
1546 frequency-generation (SFG) vibration spectroscopy *Biomacromolecules* 12:2434-2439
- 1547 Barnette AL et al. (2012) Quantification of crystalline cellulose in lignocellulosic biomass using
1548 sum frequency generation (SFG) vibration spectroscopy and comparison with other
1549 analytical methods *Carbohydr Polym* 89:802-809
- 1550 Bishop DM, Kirtman B, Champagne Bt (1997) Differences between the exact sum-over-states and
1551 the canonical approximation for the calculation of static and dynamic hyperpolarizabilities
1552 *J Chem Phys* 107:5780-5787
- 1553 Blackwell J (1977) Infrared and Raman spectroscopy of cellulose. In: *Cellulose Chemistry and*
1554 *Technology*, vol 48. ACS Symposium Series, vol 48. American Chemical Society, pp 206-
1555 218
- 1556 Blackwell J, Vasko P, Koenig J (1970) Infrared and Raman spectra of the cellulose from the cell
1557 wall of *Valonia ventricosa* *J Appl Phys* 41:4375-4379
- 1558 Börjesson M, Westman G (2015) Crystalline nanocellulose—preparation, modification, and
1559 properties. In: *Cellulose—fundamental aspects and current trends*. InTech,
- 1560 Bourmaud A, Morvan C, Bouali A, Placet V, Perre P, Baley C (2013) Relationships between
1561 micro-fibrillar angle, mechanical properties and biochemical composition of flax fibers *Ind*
1562 *Crops Prod* 44:343-351
- 1563 Boyd RW (1999) Order-of-magnitude estimates of the nonlinear optical susceptibility *J Mod Opt*
1564 46:367-378
- 1565 Boyd RW, Boyd RW (2003) *Nonlinear Optics*. Elsevier Science, Burlington, UNITED STATES
- 1566 Bučko Ts, Tunega D, Ángyán JG, Hafner Jr (2011) Ab initio study of structure and interconversion
1567 of native cellulose phases *J Phys Chem A* 115:10097-10105
- 1568 Butler HJ, McAinsh MR, Adams S, Martin FL (2015) Application of vibrational spectroscopy
1569 techniques to non-destructively monitor plant health and development *Anal Methods*
1570 7:4059-4070
- 1571 Cael J, Gardner K, Koenig J, Blackwell J (1975) Infrared and Raman spectroscopy of
1572 carbohydrates. Paper V. Normal coordinate analysis of cellulose I *J Chem Phys* 62:1145-
1573 1153
- 1574 Cammarata R, Eby R (1991) Effects and measurement of internal surface stresses in materials with
1575 ultrafine microstructures *J Mater Res* 6:888-890
- 1576 Cammarata RC (1994) Surface and interface stress effects in thin films *Prog Surf Sci* 46:1-38
- 1577 Chen H, Ferrari C, Angiuli M, Yao J, Raspi C, Bramanti E (2010) Qualitative and quantitative
1578 analysis of wood samples by Fourier transform infrared spectroscopy and multivariate
1579 analysis *Carbohydr Polym* 82:772-778
- 1580 Chen L, Wilson RH, McCann MC (1997) Investigation of macromolecule orientation in dry and
1581 hydrated walls of single onion epidermal cells by FTIR microspectroscopy *J Mol Struct*
1582 408:257-260
- 1583 Chen X, Lee CM, Wang H-F, Jensen L, Kim SH (2017) Experimental and Theoretical Study of
1584 Azimuth Angle and Polarization Dependences of Sum-Frequency-Generation Vibrational
1585 Spectral Features of Uniaxially Aligned Cellulose Crystals *J Phys Chem C* 121:18876-
1586 18886
- 1587 Cho SH et al. (2015) In vitro synthesis of cellulose microfibrils by membrane protein from
1588 protoplasts of the non-vascular plant *Physcomitrella patens* *Biochem J* 470:195-205
- 1589 Cintrón MS, Hinchliffe DJ (2015) FT-IR examination of the development of secondary cell wall
1590 in cotton fibers *Fibers* 3:30-40

1591 Cosgrove DJ (2005) Growth of the plant cell wall *Nat Rev Mol Cell Biol* 6:850-861

1592 Cosgrove DJ (2014) Re-constructing our models of cellulose and primary cell wall assembly *Curr*

1593 *Opin Plant Biol* 22:122-131

1594 Cullity BD (1978) *Elements of X-ray Diffraction*. Addison-Wesley, New York

1595 de Beer AG, Campen RK, Roke S (2010) Separating surface structure and surface charge with

1596 second-harmonic and sum-frequency scattering *Phys Rev B* 82:235431

1597 de Beer AG, Roke S (2007) Sum frequency generation scattering from the interface of an isotropic

1598 particle: Geometrical and chiral effects *Phys Rev B* 75:245438

1599 de Beer AG, Roke S (2009) Nonlinear Mie theory for second-harmonic and sum-frequency

1600 scattering *Phys Rev B* 79:155420

1601 de Beer AG, Roke S, Dadap JI (2011) Theory of optical second-harmonic and sum-frequency

1602 scattering from arbitrarily shaped particles *J Opt Soc Am B* 28:1374-1384

1603 Déjardin A, Laurans F, Arnaud D, Breton C, Pilate G, Leplé J-C (2010) Wood formation in

1604 Angiosperms *C R Biol* 333:325-334

1605 Dion M, Rydberg H, Schröder E, Langreth DC, Lundqvist BI (2004) Van der Waals density

1606 functional for general geometries *Phys Rev Lett* 92:246401

1607 Donaldson L, Xu P (2005) Microfibril orientation across the secondary cell wall of Radiata pine

1608 tracheids *Trees* 19:644

1609 Fagard M et al. (2000) PROCUSTE1 encodes a cellulose synthase required for normal cell

1610 elongation specifically in roots and dark-grown hypocotyls of Arabidopsis *Plant Cell*

1611 12:2409-2423

1612 Fan M, Dai D, Huang B (2012) Fourier transform infrared spectroscopy for natural fibres. In:

1613 *Fourier transform-materials analysis*. InTech, pp 45-68

1614 Fernandes AN et al. (2011) Nanostructure of cellulose microfibrils in spruce wood *Proc Natl Acad*

1615 *Sci U S A* 108:E1195-E1203

1616 Frisch M et al. (2010) Gaussian 09, revision b. 01, Gaussian Inc, Wallingford, CT 6492

1617 Frisch MJ, Yamaguchi Y, Gaw JF, Schaefer III HF, Binkley JS (1986) Analytic Raman intensities

1618 from molecular electronic wave functions *J Chem Phys* 84:531-532

1619 Fujita M et al. (2013) The any1 D604N mutation in the Arabidopsis thaliana cellulose synthase 1

1620 catalytic domain reduces cell wall crystallinity and the velocity of cellulose synthase

1621 complexes *Plant Physiol* 162:74-85

1622 Fujita M, Wasteneys GO (2014) A survey of cellulose microfibril patterns in dividing, expanding,

1623 and differentiating cells of Arabidopsis thaliana *Protoplasma* 251:687-698

1624 George J, Sabapathi S (2015) Cellulose nanocrystals: synthesis, functional properties, and

1625 applications *Nanotechnol Sci Appl* 8:45

1626 Gierlinger N (2018) New insights into plant cell walls by vibrational microspectroscopy *Appl*

1627 *Spectrosc Rev* 53:517-551

1628 Goda K, Sreekala M, Gomes A, Kaji T, Ohgi J (2006) Improvement of plant based natural fibers

1629 for toughening green composites—Effect of load application during mercerization of ramie

1630 fibers *Composites, Part A* 37:2213-2220

1631 Goodell B et al. (2017) Modification of the nanostructure of lignocellulose cell walls via a non-

1632 enzymatic lignocellulose deconstruction system in brown rot wood-decay fungi *Biotechnol*

1633 *Biofuels* 10:179

1634 Gorman M (1957) The evidence from infrared spectroscopy for hydrogen bonding: A case history

1635 of the correlation and interpretation of data *J Chem Educ* 34:304

1636 Grimme S (2006) Semiempirical GGA - type density functional constructed with a long - range
1637 dispersion correction J Comput Chem 27:1787-1799
1638 Grimme S, Antony J, Ehrlich S, Krieg H (2010) A consistent and accurate ab initio parametrization
1639 of density functional dispersion correction (DFT-D) for the 94 elements H-Pu J Chem Phys
1640 132:154104
1641 Ha MA, Apperley DC, Jarvis MC (1997) Molecular rigidity in dry and hydrated onion cell walls
1642 Plant Physiol 115:593-598
1643 Habibi Y, Lucia LA, Rojas OJ (2010) Cellulose nanocrystals: chemistry, self-assembly, and
1644 applications Chem Rev 110:3479-3500
1645 Handakumbura PP et al. (2013) Perturbation of Brachypodium distachyon CELLULOSE
1646 SYNTHASE A4 or 7 results in abnormal cell walls BMC Plant Biol 13:131
1647 Hill JL, Hammudi MB, Tien M (2014) The Arabidopsis cellulose synthase complex: a proposed
1648 hexamer of CESA trimers in an equimolar stoichiometry Plant Cell 26:4834-4842
1649 Hoang Chi H, Nguyen Anh T, Hongyan L, Yoshihiro M, Goro M (2011) Sum Frequency
1650 Generation Microscopy Study of Cellulose Fibers Appl Spectrosc 65:1254-1259
1651 Hofstetter K, Hinterstoisser B, Salmén L (2006) Moisture uptake in native cellulose—the roles of
1652 different hydrogen bonds: a dynamic FT-IR study using deuterium exchange Cellulose
1653 13:131-145
1654 Hohm T et al. (2014) Plasma membrane H⁺ - ATPase regulation is required for auxin gradient
1655 formation preceding phototropic growth Mol Syst Biol 10:751
1656 Horii F, Hirai A, Kitamaru R (1983) Solid-state ¹³C-NMR study of conformations of
1657 oligosaccharides and cellulose Polym Bull 10:357-361
1658 Huang S, Makarem M, Kiemle SN, Hamed H, Sau M, Cosgrove DJ, Kim SH (2018a)
1659 Inhomogeneity of Cellulose Microfibril Assembly in Plant Cell Walls Revealed with Sum
1660 Frequency Generation Microscopy J Phys Chem B 122:5006-5019
1661 Huang S et al. (2018b) Dehydration-induced physical strains of cellulose microfibrils in plant cell
1662 walls Carbohydr Polym 197:337-348
1663 Imai T, Sugiyama J, Itoh T, Horii F (1999) Almost pure I α cellulose in the cell wall of Glaucocystis
1664 J Struct Biol 127:248-257
1665 Jarvis M (1992) Control of thickness of collenchyma cell walls by pectins Planta 187:218-220
1666 Jarvis MC (1984) Structure and properties of pectin gels in plant cell walls Plant, Cell Environ
1667 7:153-164
1668 Johnson III RD, Irikura KK, Kacker RN, Kessel Rd (2010) Scaling factors and uncertainties for
1669 ab initio anharmonic vibrational frequencies J Chem Theory Comput 6:2822-2828
1670 Jolliffe I (2011) Principal component analysis. In: International encyclopedia of statistical science.
1671 Springer, Berlin, pp 1094-1096
1672 Kacurakova M, Capek P, Sasinkova V, Wellner N, Ebringerova A (2000) FT-IR study of plant
1673 cell wall model compounds: pectic polysaccharides and hemicelluloses Carbohydr Polym
1674 43:195-203
1675 Kafle K, Greeson K, Lee C, Kim SH (2014a) Cellulose polymorphs and physical properties of
1676 cotton fabrics processed with commercial textile mills for mercerization and liquid
1677 ammonia treatments Text Res J 84:1692-1699
1678 Kafle K, Lee CM, Shin H, Zoppe J, Johnson DK, Kim SH, Park S (2015a) Effects of delignification
1679 on crystalline cellulose in lignocellulose biomass characterized by vibrational sum
1680 frequency generation spectroscopy and X-ray diffraction BioEnergy Res 8:1750-1758

1681 Kafle K, Park YB, Lee CM, Stapleton JJ, Kiemle SN, Cosgrove DJ, Kim SH (2017) Effects of
1682 mechanical stretching on average orientation of cellulose and pectin in onion epidermis
1683 cell wall: A polarized FT-IR study *Cellulose* 24:3145-3154

1684 Kafle K et al. (2014b) Vibrational sum-frequency-generation (SFG) spectroscopy study of the
1685 structural assembly of cellulose microfibrils in reaction woods *Cellulose* 21:2219-2231

1686 Kafle K, Shin H, Lee CM, Park S, Kim SH (2015b) Progressive structural changes of Avicel,
1687 bleached softwood, and bacterial cellulose during enzymatic hydrolysis *Sci Rep* 5:15102

1688 Kafle K, Xi X, Lee CM, Tittmann BR, Cosgrove DJ, Park YB, Kim SH (2014c) Cellulose
1689 microfibril orientation in onion (*Allium cepa* L.) epidermis studied by atomic force
1690 microscopy (AFM) and vibrational sum frequency generation (SFG) spectroscopy
1691 *Cellulose* 21:1075-1086

1692 Karimi K, Taherzadeh MJ (2016) A critical review of analytical methods in pretreatment of
1693 lignocelluloses: composition, imaging, and crystallinity *Bioresour Technol* 200:1008-1018

1694 Kim HJ et al. (2017) Comparative physical and chemical analyses of cotton fibers from two near
1695 isogenic upland lines differing in fiber wall thickness *Cellulose* 24:2385-2401

1696 Kim HJ, Liu Y, French AD, Lee CM, Kim SH (2018) Comparison and validation of Fourier
1697 transform infrared spectroscopic methods for monitoring secondary cell wall cellulose
1698 from cotton fibers *Cellulose* 25:49-64

1699 Kim N-H, Imai T, Wada M, Sugiyama J (2006) Molecular directionality in cellulose polymorphs
1700 *Biomacromolecules* 7:274-280

1701 Koch W, Holthausen MC (2015) A chemist's guide to density functional theory. John Wiley &
1702 Sons, Weinheim

1703 Kong L, Lee C, Kim SH, Ziegler GR (2014) Characterization of starch polymorphic structures
1704 using vibrational sum frequency generation spectroscopy *J Phys Chem B* 118:1775-1783

1705 Kouyama W, Ogawa A, Li H, Miyauchi Y, Mizutani G, Sano H (2014) Sum frequency generation
1706 confocal microscopy observation of a fish scale e-J *Surf Sci Nanotechnol* 12:259-262

1707 Kubicki JD, Mohamed MN-A, Watts HD (2013) Quantum mechanical modeling of the structures,
1708 energetics and spectral properties of I α and I β cellulose *Cellulose* 20:9-23

1709 Kumar M, Campbell L, Turner S (2015) Secondary cell walls: biosynthesis and manipulation *J*
1710 *Exp Bot* 67:515-531

1711 Lambert AG, Davies PB, Neivandt DJ (2005) Implementing the Theory of Sum Frequency
1712 Generation Vibrational Spectroscopy: A Tutorial Review *Appl Spectrosc Rev* 40:103-145

1713 Langan P, Nishiyama Y, Chanzy H (2001) X-ray structure of mercerized cellulose II at 1 Å
1714 resolution *Biomacromolecules* 2:410-416

1715 Lasch P, Naumann D (2006) Spatial resolution in infrared microspectroscopic imaging of tissues
1716 *Biochim Biophys Acta, Biomembr* 1758:814-829

1717 Laury ML, Carlson MJ, Wilson AK (2012) Vibrational frequency scale factors for density
1718 functional theory and the polarization consistent basis sets *J Comput Chem* 33:2380-2387

1719 Lee C, Dazen K, Kafle K, Moore A, Johnson DK, Park S, Kim SH (2015a) Correlations of apparent
1720 cellulose crystallinity determined by XRD, NMR, IR, Raman, and SFG methods. In:
1721 *Cellulose chemistry and properties: fibers, nanocelluloses and advanced materials.*
1722 Springer International Publishing, pp 115-131

1723 Lee CM, Gu J, Kafle K, Catchmark J, Kim SH (2015b) Cellulose produced by *Gluconacetobacter*
1724 *xylinus* strains ATCC 53524 and ATCC 23768: Pellicle formation, post-synthesis
1725 aggregation and fiber density *Carbohydr Polym* 133:270-276

1726 Lee CM, Kafle K, Belias DW, Park YB, Glick RE, Haigler CH, Kim SH (2015c) Comprehensive
1727 analysis of cellulose content, crystallinity, and lateral packing in *Gossypium hirsutum* and
1728 *Gossypium barbadense* cotton fibers using sum frequency generation, infrared and Raman
1729 spectroscopy, and X-ray diffraction *Cellulose* 22:971-989

1730 Lee CM, Kafle K, Huang S, Kim SH (2016) Multimodal broadband vibrational sum frequency
1731 generation (MM-BB-V-SFG) spectrometer and microscope *J Phys Chem B* 120:102-116

1732 Lee CM, Kafle K, Park YB, Kim SH (2014) Probing crystal structure and mesoscale assembly of
1733 cellulose microfibrils in plant cell walls, tunicate tests, and bacterial films using vibrational
1734 sum frequency generation (SFG) spectroscopy *Phys Chem Chem Phys* 16:10844-10853

1735 Lee CM, Kubicki JD, Fan B, Zhong L, Jarvis MC, Kim SH (2015d) Hydrogen-bonding network
1736 and OH stretch vibration of cellulose: Comparison of computational modeling with
1737 polarized IR and SFG spectra *J Phys Chem B* 119:15138-15149

1738 Lee CM et al. (2013a) Cellulose polymorphism study with sum-frequency-generation (SFG)
1739 vibration spectroscopy: identification of exocyclic CH₂OH conformation and chain
1740 orientation *Cellulose* 20:991-1000

1741 Lee CM, Mohamed NM, Watts HD, Kubicki JD, Kim SH (2013b) Sum-frequency-generation
1742 vibration spectroscopy and density functional theory calculations with dispersion
1743 corrections (DFT-D2) for cellulose I α and I β *J Phys Chem B* 117:6681-6692

1744 Lei L et al. (2014) The *jiaoyao1* mutant is an allele of *korriگان1* that abolishes endoglucanase
1745 activity and affects the organization of both cellulose microfibrils and microtubules in
1746 *Arabidopsis* *Plant Cell* 26:2601-2616

1747 Li S et al. (2016) Cellulose synthase complexes act in a concerted fashion to synthesize highly
1748 aggregated cellulose in secondary cell walls of plants *Proc Natl Acad Sci U S A* 40:11348–
1749 11135

1750 Li Y-H, Chen D-N, Niu H-B (2017) A method for achieving super resolution vibrational sum-
1751 frequency generation microscopy by structured illumination *IEEE Photonics J* 9:1-8

1752 Li Y, Lin M, Davenport JW (2011) Ab initio studies of cellulose I: crystal structure, intermolecular
1753 forces, and interactions with water *J Phys Chem C* 115:11533-11539

1754 Libowitzky E (1999) Correlation of O-H stretching frequencies and O-H...O hydrogen bond
1755 lengths in minerals *Monatsh Chem* 130:1047-1059

1756 Liu Y (2013) Recent progress in fourier transform infrared (FTIR) spectroscopy study of
1757 compositional, structural and physical attributes of developmental cotton fibers *Materials*
1758 6:299-313

1759 Ma D, Lee CM, Chen Y, Mehta N, Kim SH, Liu Z (2017) Vibrational sum frequency generation
1760 digital holography *Appl Phys Lett* 110:251601

1761 Makarem M, Lee CM, Sawada D, O'Neill HM, Kim SH (2018) Distinguishing surface versus bulk
1762 hydroxyl groups of cellulose nanocrystals using vibrational sum frequency generation
1763 spectroscopy *J Phys Chem Lett* 9:70-75

1764 Makarem M et al. (2017) Dependence of sum frequency generation (SFG) spectral features on the
1765 mesoscale arrangement of SFG-active crystalline domains interspersed in SFG-inactive
1766 matrix: A case study with cellulose in uniaxially aligned control samples and alkali-treated
1767 secondary cell walls of plants *J Phys Chem C* 121:10249-10257

1768 Maréchal Y, Chanzy H (2000) The hydrogen bond network in I β cellulose as observed by infrared
1769 spectrometry *J Mol Struct* 523:183-196

1770 Marques MA, Gross EK (2004) Time-dependent density functional theory *Annu Rev Phys Chem*
1771 55:427-455

1772 Matthäus C, Bird B, Miljković M, Chernenko T, Romeo M, Diem M (2008) Infrared and Raman
1773 microscopy in cell biology *Methods Cell Biol* 89:275-308

1774 McCann M, Wells B, Roberts K (1990) Direct visualization of cross-links in the primary plant cell
1775 wall *J Cell Sci* 96:323-334

1776 McCann MC, Hammouri M, Wilson R, Belton P, Roberts K (1992) Fourier transform infrared
1777 microspectroscopy is a new way to look at plant cell walls *Plant Physiol* 100:1940-1947

1778 McNamara JT, Morgan JL, Zimmer J (2015) A molecular description of cellulose biosynthesis
1779 *Annu Rev Biochem* 84:895-921

1780 Moon RJ, Martini A, Nairn J, Simonsen J, Youngblood J (2011) Cellulose nanomaterials review:
1781 structure, properties and nanocomposites *Chem Soc Rev* 40:3941-3994

1782 Newman RH, Hill SJ, Harris PJ (2013) Wide-angle x-ray scattering and solid-state nuclear
1783 magnetic resonance data combined to test models for cellulose microfibrils in mung bean
1784 cell walls *Plant Physiol* 163:1558-1567

1785 Nishiyama Y, Langan P, Chanzy H (2002) Crystal structure and hydrogen-bonding system in
1786 cellulose I β from synchrotron X-ray and neutron fiber diffraction *J Am Chem Soc*
1787 124:9074-9082

1788 Nishiyama Y, Sugiyama J, Chanzy H, Langan P (2003) Crystal structure and hydrogen bonding
1789 system in cellulose I α from synchrotron X-ray and neutron fiber diffraction *J Am Chem*
1790 *Soc* 125:14300-14306

1791 Nixon BT et al. (2016) Comparative structural and computational analysis supports eighteen
1792 cellulose synthases in the plant cellulose synthesis complex *Sci Rep* 6:28696

1793 Norris JH et al. (2017) Functional specialization of cellulose synthase isoforms in a moss shows
1794 parallels with seed plants *Plant Physiol* 175:210-222

1795 Northey RA, Glasser WG, Schultz TP (2000) Lignin: historical, biological, and materials
1796 perspectives vol 742. American Chemical Society, Washington

1797 Novak A (1974) Hydrogen bonding in solids correlation of spectroscopic and crystallographic
1798 data. In: *Large Molecules*. Springer, pp 177-216

1799 Oehme DP, Yang H, Kubicki JD (2018) An evaluation of the structures of cellulose generated by
1800 the CHARMM force field: comparisons to in planta cellulose *Cellulose* 25:3755-3777

1801 Ogawa Y, Lee CM, Nishiyama Y, Kim SH (2016) Absence of Sum Frequency Generation in
1802 Support of Orthorhombic Symmetry of α -Chitin *Macromolecules* 49:7025-7031

1803 Oh SY et al. (2005) Crystalline structure analysis of cellulose treated with sodium hydroxide and
1804 carbon dioxide by means of X-ray diffraction and FTIR spectroscopy *Carbohydr Res*
1805 340:2376-2391

1806 Park S, Baker JO, Himmel ME, Parilla PA, Johnson DK (2010) Cellulose crystallinity index:
1807 measurement techniques and their impact on interpreting cellulase performance *Biotechnol*
1808 *Biofuels* 3:10

1809 Park YB, Cosgrove DJ (2012) A revised architecture of primary cell walls based on biomechanical
1810 changes induced by substrate-specific endoglucanases *Plant Physiol* 158:1933-1943

1811 Park YB, Kafle K, Lee CM, Cosgrove DJ, Kim SH (2015) Does cellulose II exist in native alga
1812 cell walls? Cellulose structure of *Derbesia* cell walls studied with SFG, IR and XRD
1813 *Cellulose* 22:3531-3540

1814 Park YB, Lee CM, Kafle K, Park S, Cosgrove DJ, Kim SH (2014) Effects of plant cell wall matrix
1815 polysaccharides on bacterial cellulose structure studied with vibrational sum frequency
1816 generation spectroscopy and x-ray diffraction *Biomacromolecules* 15:2718-2724

- 1817 Park YB, Lee CM, Koo BW, Park S, Cosgrove DJ, Kim SH (2013) Monitoring meso-scale
 1818 ordering of cellulose in intact plant cell walls using sum frequency generation spectroscopy
 1819 *Plant Physiol* 163:907-913
- 1820 Popescu C-M, Popescu M-C, Singurel G, Vasile C, Argyropoulos DS, Willfor S (2007) Spectral
 1821 characterization of eucalyptus wood *Appl Spectrosc* 61:1168-1177
- 1822 Raghunathan V, Han Y, Korth O, Ge N-H, Potma EO (2011) Rapid vibrational imaging with sum
 1823 frequency generation microscopy *Opt Lett* 36:3891-3893
- 1824 Reiher M, Neugebauer J, Hess BA (2003) Quantum chemical calculation of Raman intensities for
 1825 large molecules: The photoisomerization of [$\text{Fe}(\text{S}^4\text{PR}_3)_2(\text{N}_2\text{H}_2)$]($\text{S}^4\text{S}^2=$ 1, 2-bis
 1826 (2-mercaptophenylthio)-ethane (2-)) *Z Phys Chem* 217:91-104
- 1827 Rivnay J, Noriega R, Kline RJ, Salleo A, Toney MF (2011) Quantitative analysis of lattice disorder
 1828 and crystallite size in organic semiconductor thin films *Phys Rev B* 84:045203
- 1829 Robinson JW (1996) Atomic spectroscopy. CRC Press,
- 1830 Roke S, Bonn M, Petukhov AV (2004) Nonlinear optical scattering: The concept of effective
 1831 susceptibility *Phys Rev B* 70:115106
- 1832 Roke S, Roeterdink WG, Wijnhoven JE, Petukhov AV, Kleyn AW, Bonn M (2003) Vibrational
 1833 sum frequency scattering from a submicron suspension *Phys Rev Lett* 91:258302
- 1834 Ruppin R, Englman R (1970) Optical phonons of small crystals *Reports on Progress in Physics*
 1835 33:149
- 1836 Sacui IA et al. (2014) Comparison of the properties of cellulose nanocrystals and cellulose
 1837 nanofibrils isolated from bacteria, tunicate, and wood processed using acid, enzymatic,
 1838 mechanical, and oxidative methods *ACS Appl Mater Interfaces* 6:6127-6138
- 1839 Saxe F, Eder M, Benecke G, Aichmayer B, Fratzl P, Burgert I, Rüggeberg M (2014) Measuring
 1840 the distribution of cellulose microfibril angles in primary cell walls by small angle X-ray
 1841 scattering *Plant methods* 10:25
- 1842 Schwanninger M, Rodrigues J, Pereira H, Hinterstoisser B (2004) Effects of short-time vibratory
 1843 ball milling on the shape of FT-IR spectra of wood and cellulose *Vib Spectrosc* 36:23-40
- 1844 Sene CF, McCann MC, Wilson RH, Grinter R (1994) Fourier-transform Raman and Fourier-
 1845 transform infrared spectroscopy (an investigation of five higher plant cell walls and their
 1846 components) *Plant Physiol* 106:1623-1631
- 1847 Shen Q, Zhong L, Hu J-F (2004) Characterization of the surface properties of xylan by FT-Raman
 1848 spectroscopy and wicking technique *Colloids Surf, B* 39:195-198
- 1849 Shen Y-R (1984) *The principles of nonlinear optics* New York, Wiley-Interscience, 1984, 575 p
- 1850 Siesler HW, Ozaki Y, Kawata S, Heise HM (2008) *Near-infrared spectroscopy: principles,*
 1851 *instruments, applications.* John Wiley & Sons, Weinheim, Germany
- 1852 Somerville C (2006) Cellulose synthesis in higher plants *Annu Rev Cell Dev Biol* 22:53-78
- 1853 Sugiyama J, Okano T, Yamamoto H, Horii F (1990) Transformation of Valonia cellulose crystals
 1854 by an alkaline hydrothermal treatment *Macromolecules* 23:3196-3198
- 1855 Suslov D, Verbelen J (2006) Cellulose orientation determines mechanical anisotropy in onion
 1856 epidermis cell walls *J Exp Bot* 57:2183-2192
- 1857 Suslov D, Verbelen JP, Vissenberg K (2009) Onion epidermis as a new model to study the control
 1858 of growth anisotropy in higher plants *J Exp Bot* 60:4175-4187
- 1859 Synytsya A, Čopíková J, Matějka P, Machovič V (2003) Fourier transform Raman and infrared
 1860 spectroscopy of pectins *Carbohydr Polym* 54:97-106

- 1861 Szymanska-Chargot M, Zdunek A (2013) Use of FT-IR spectra and PCA to the bulk
1862 characterization of cell wall residues of fruits and vegetables along a fraction process Food
1863 biophysics 8:29-42
- 1864 Terashima N, Awano T, Takabe K, Yoshida M (2004) Formation of macromolecular lignin in
1865 ginkgo xylem cell walls as observed by field emission scanning electron microscopy C R
1866 Biol 327:903-910
- 1867 Thomas LH et al. (2012) Structure of cellulose microfibrils in primary cell-walls from
1868 collenchyma Plant Physiol 161:465-476
- 1869 Tsuboi M (1957) Infrared spectrum and crystal structure of cellulose J Polym Sci 25:159-171
- 1870 Vandavasi VG et al. (2016) A structural study of CESA1 catalytic domain of Arabidopsis thaliana
1871 cellulose synthesis complex: evidence for CESA trimers Plant Physiol 170:123-135
- 1872 Velarde L, Wang H-f (2013) Capturing inhomogeneous broadening of the–CN stretch vibration in
1873 a Langmuir monolayer with high-resolution spectra and ultrafast vibrational dynamics in
1874 sum-frequency generation vibrational spectroscopy (SFG-VS) J Chem Phys 139:084204
- 1875 Wada M, Heux L, Sugiyama J (2004) Polymorphism of cellulose I family: reinvestigation of
1876 cellulose IVI Biomacromolecules 5:1385-1391
- 1877 Wada M, Kondo T, Okano T (2003) Thermally induced crystal transformation from cellulose I α
1878 to I β Polym J 35:155
- 1879 Wang H-W et al. (2014a) Vibrational density of states of strongly H-bonded interfacial water:
1880 Insights from inelastic neutron scattering and theory J Phys Chem C 118:10805-10813
- 1881 Wang HF, Gan W, Lu R, Rao Y, Wu BH (2005) Quantitative spectral and orientational analysis
1882 in surface sum frequency generation vibrational spectroscopy (SFG-VS) Int Rev Phys
1883 Chem 24:191-256
- 1884 Wang T, Hong M (2015) Solid-state NMR investigations of cellulose structure and interactions
1885 with matrix polysaccharides in plant primary cell walls J Exp Bot 67:503-514
- 1886 Wang T, Park YB, Daniel JC, Hong M (2015) Cellulose–pectin spatial contacts are inherent to
1887 never-dried Arabidopsis thaliana primary cell walls: evidence from solid-state NMR Plant
1888 Physiol 168:871-884
- 1889 Wang T, Zabolina O, Hong M (2012) Pectin–cellulose interactions in the Arabidopsis primary cell
1890 wall from two-dimensional magic-angle-spinning solid-state nuclear magnetic resonance
1891 Biochemistry 51:9846-9856
- 1892 Wang W et al. (2014b) Effect of mechanical disruption on the effectiveness of three reactors used
1893 for dilute acid pretreatment of corn stover Part 1: chemical and physical substrate analysis
1894 Biotechnol Biofuels 7:57
- 1895 Ward I (1985) Determination of molecular orientation by spectroscopic techniques. In:
1896 Characterization of Polymers in the Solid State I: Part A: NMR and Other Spectroscopic
1897 Methods Part B: Mechanical Methods. Springer, pp 81-115
- 1898 Watanabe A, Morita S, Ozaki Y (2007) Temperature-dependent changes in hydrogen bonds in
1899 cellulose I α studied by infrared spectroscopy in combination with perturbation-correlation
1900 moving-window two-dimensional correlation spectroscopy: comparison with cellulose I β
1901 Biomacromolecules 8:2969-2975
- 1902 Wickholm K, Hult E-L, Larsson PT, Iversen T, Lennholm H (2001) Quantification of cellulose
1903 forms in complex cellulose materials: a chemometric model Cellulose 8:139-148
- 1904 Wiley JH, Atalla RH (1987) Band assignments in the Raman spectra of celluloses Carbohydr Res
1905 160:113-129

- 1906 Wilkes G (1971) The measurement of molecular orientation in polymeric solids. In: Fortschritte
1907 der Hochpolymeren-Forschung. Springer, pp 91-136
- 1908 Williamson G, Hall W (1953) X-ray line broadening from fided aluminium and wolfram Acta
1909 Metall 1:22-31
- 1910 Wilson EB, Decius JC, Cross PC (1955) Molecular Vibrations: The Theory of Infrared and Raman
1911 Vibrational Spectra. Dover, New York
- 1912 Wilson RH, Smith AC, Kačuráková M, Saunders PK, Wellner N, Waldron KW (2000) The
1913 mechanical properties and molecular dynamics of plant cell wall polysaccharides studied
1914 by Fourier-transform infrared spectroscopy Plant Physiol 124:397-406
- 1915 Withana-Gamage TS, Hegedus DD, Qiu X, Yu P, May T, Lydiate D, Wanasundara JP (2013)
1916 Characterization of Arabidopsis thaliana lines with altered seed storage protein profiles
1917 using synchrotron-powered FT-IR spectromicroscopy J Agric Food Chem 61:901-912
- 1918 Yamaguchi Y, Frisch M, Gaw J, Schaefer III HF, Binkley JS (1986) Analytic evaluation and basis
1919 set dependence of intensities of infrared spectra J Chem Phys 84:2262-2278
- 1920 Ye D et al. (2018) Resonant soft X-ray scattering reveals cellulose microfibril spacing in plant
1921 primary cell walls Sci Rep 8:12449
- 1922 Zeng Y, Yarbrough JM, Mittal A, Tucker MP, Vinzant TB, Decker SR, Himmel ME (2016) In
1923 situ label-free imaging of hemicellulose in plant cell walls using stimulated Raman
1924 scattering microscopy Biotechnol Biofuels 9:256
- 1925 Zhang T, Mahgoudy-Louyeh S, Tittmann B, Cosgrove DJ (2013) Visualization of the nanoscale
1926 pattern of recently-deposited cellulose microfibrils and matrix materials in never-dried
1927 primary walls of the onion epidermis Cellulose 21:853-862
- 1928 Zhang T, Vavylonis D, Durachko DM, Cosgrove DJ (2017) Nanoscale movements of cellulose
1929 microfibrils in primary cell walls Nat Plants 3:17056
- 1930 Zhang T, Zheng Y, Cosgrove DJ (2016) Spatial organization of cellulose microfibrils and matrix
1931 polysaccharides in primary plant cell walls as imaged by multichannel atomic force
1932 microscopy Plant J 85:179-192
- 1933 Zheng Y, Cosgrove DJ, Ning G (2017) High-resolution field emission scanning electron
1934 microscopy (FESEM) imaging of cellulose microfibril organization in plant primary cell
1935 walls Microsc Microanal 23:1048-1054
- 1936 Zhong R, Burk DH, Morrison WH, Ye Z-H (2002) A kinesin-like protein is essential for oriented
1937 deposition of cellulose microfibrils and cell wall strength Plant Cell 14:3101-3117
- 1938 Zugenmaier P (2001) Conformation and packing of various crystalline cellulose fibers Prog Polym
1939 Sci 26:1341-1417
- 1940