Korean J. Chem. Eng., **30**(12), 2127-2141 (2013)
DOI: 10.1007/s11814-013-0162-0
pISSN: 0256-1115
eISSN: 1975-7220

REVIEW PAPER

Characterization of crystalline cellulose in biomass: Basic principles, applications, and limitations of XRD, NMR, IR, Raman, and SFG

Seong H. Kim†, Christopher M. Lee, and Kabindra Kafle

Department of Chemical Engineering and Materials Research Institute,
Pennsylvania State University, University Park, PA 16802, U.S.A.
(*Received 18 June 2013 • accepted 26 August 2013*)

Abstract−Cellulose is among the most important and abundant biopolymers in biosphere. It is the main structural component of a vast number of plants that carries vital functions for plant growth. Cellulose-based materials have been used in a variety of human activities ranging from papers and fabrics to engineering applications including production of biofuels. However, our understanding of the cellulose structure in its native form is quite limited because the current experimental methods often require separation or purification processes and provide only partial information of the cellulose structure. This paper aims at providing a brief background of the cellulose structure and reviewing the basic principles, capabilities and limitations of the cellulose characterization methods that are widely used by engineers dealing with biomass. The analytical techniques covered in this paper include x-ray diffraction, nuclear magnetic resonance, and vibrational spectroscopy (infrared, Raman, and sum-frequency-generation). The scope of the paper is restricted to the application of these techniques to the structural analysis of cellulose.

Key words: Cellulose, Biomass, XRD, NMR, IR, Raman, SFG

INTRODUCTION

Cellulose is among the most important polysaccharides, yet native cellulose in intact biomass is quite challenging to characterize. Cellulose is a linear polymer of β-D-glucose linked by the 1,4-glycosidic bond, but this simple molecular description is not sufficient to describe the cellulose properties [1]. Cellulose is a crystalline material; thus, its mechanical and chemical properties cannot be fully understood without knowing the crystal structure and its surface interactions with surroundings. If high-quality single crystalline samples were available, then these properties could be analyzed accurately. However, such crystals large enough for full structural characterization are not available. Cellulose cannot be synthesized in laboratory. If glucose is polymerized, only short oligomers can be obtained since the oligomers longer than 6 glucose units are not soluble in aqueous solution [2,3]. High molecular weight cellulose is obtained only from natural resources [4].

Among various biological species that can synthesize cellulose, plants are by far the most prolific producers. In plants, cellulose is synthesized by protein complexes embedded and traveling in plasma membranes of cells and excreted as a fibrillar form consisting of 24-36 chains into extracellular matrices where the cellulose fibrils are deposited and mixed with other polysaccharides forming cell walls [5-7]. Historically, the cellulose fibrils are called 'microfibrils'; they would have been called 'nanofibrils' if they were discovered in the late twentieth or early twenty-first century since their diameters are in the range of a few to ~20 nm. The cellulose chains in microfibrils have intra- and inter-chain hydrogen bonding interactions which order them into crystalline lattices [8]. The crystal structure of cellulose in microfibrils, cellulose microfibril deposition patterns, and their interactions with other wall matrix polymers can vary drastically depending on plant species as well as plant growth stages and environments [9,10]. Some heterotrophic bacteria (such as *Gluconacetobacter xylinus*), slime molds (such as *Physarum polycephalum*, *Dictyostelium discoideum*), and one group of animals (the tunicates, e.g., *Halocynthia roretzi*) can also synthesize cellulose [11]. But, plants such as trees and grasses are by far the main producers of cellulose. Cell walls of these plants contain crystalline cellulose and incorporated matrix polymers such as hemicellulose and lignin and are called 'lignocellulosic biomass' [12].

Lignocellulosic biomass materials are of great interest. They have been used as a combustion fuel source for heat and raw materials for pulps, papers, fabrics, and additives for food and plastics [13-15]. Recently, lignocellulosic biomass has gained much attention for the production of biofuels that can be used for transportation vehicles [12,16,17]. The main valuable ingredient for these applications is cellulose. The structures of crystalline cellulose in these biomass materials play important roles in engineering applications as solid additives for mechanical property reinforcements and as sustainable resources for liquid fuel or chemical productions.

Analytical techniques widely used for characterization of cellulose structure include x-ray and electron diffraction (XRD and ED), nuclear magnetic resonance (NMR), and vibration spectroscopy (infrared (IR) and Raman) [18-25]. Recently, a nonlinear laser spectroscopy called sum-frequency-generation (SFG) has been shown to be capable of providing some structural information of cellulose in lignocellulose biomass [26-29]. It is very important to understand the principle through which each technique probes the "ordering" or "structural constraint" imposed through intra- and inter-chain hydrogen bonding interactions of cellulose molecules and their inter-

†To whom correspondence should be addressed.
E-mail: shkim@engr.psu.edu

actions with other matrix polysaccharides of plant cell walls.

This short review paper provides the basic principles of XRD, NMR, IR, Raman, and SFG spectroscopy and their applications to cellulose structure characterization for engineers and plant biologists who are not familiar with modern instrumental analyses and cellulose structures. XRD seeks for information of the atomic arrangement within the structural building block (called "unit cell") that repeats infinitely along all three Cartesian coordinate directions. This is done through mathematical fitting and regression of the diffraction data. NMR detects small differences in the degree of magnetic field shielding at a specific atom position imposed by surrounding atoms. The length and angle of the covalent bond could be altered by structural constraints imposed by the "crystalline" packing of molecules, which can alter the magnetic field experienced by the atoms involved in the distorted or strained chemical bond. Similarly, the structural constraints of the crystalline packing could induce some shifts in the frequency (energy) of specific vibration modes of molecules. This can make the IR and Raman spectra of crystalline solids deviate from those of the isolated molecule which does not have any interactions with surrounding molecules. SFG is sensitive to the coherency of the noncentrosymmetric ordering of molecular components within the optical wavelength scale (which is of the order of 100s nanometers) as well as the molecular-level symmetry (within the bond-length scale). Thus, SFG can be seen as a hybrid version of diffraction and vibration analyses with unique noncentrosymmetry requirements. The working principles, capabilities, and limitations of each analytical method, along with the current model of the cellulose structure, will be discussed in this review paper.

CELLULOSE STRUCTURES

Cellulose is a linear 1,4-linked homopolymer of β-D-glucopyranose. The β-D-glucopyranose unit has a six-membered arm-chair conformation with all methine (C-H) groups at the axial position and all hydroxyl (OH) and hydroxymethyl (CH_2OH) groups at the equatorial positions (Fig. 1) [6]. Along the cellulose chain, two glucose units are rotated alternatively about 180° from each other [30]. Given this chemical definition, one might anticipate that the nature and chemical properties of cellulose could be understood in terms of the chemistry of its monomeric constituents along with properties arising from its polymeric nature. The reality, however, is that the physical and chemical properties of particular cellulose samples are much more complicated; they vary drastically depending on plant sources, isolation and purification processes, and aggregation states. Probably the most unusual nature of cellulose is its insolubility in water, even though its monomeric constituent (glucose) is highly soluble in water [1,3,31]. This insolubility is the consequence of the crystallization of cellulose chains while they are produced in biological systems [32].

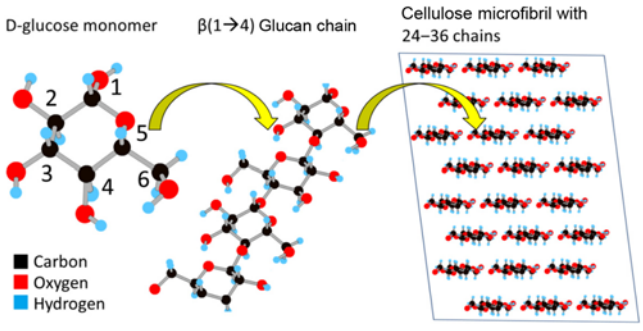

Fig. 1. Chemical structures of D-glucose, linear β(1→4) glucan chain and a 24-chain model for cellulose microfibrils [6]. For the cellulose microfibril, the cross section view is projected along the glucan chain.

As a result of crystalline packing, the naturally produced cellulose chains assume the shape of flat ribbons in cross section, which can be packed with high density (as high as 1.6 g/cm^3, typically ~1.55 g/cm^3 [33]. The flat sides of the ribbons (~0.9 nm wide) expose the C-H groups, while the thin edges of the ribbons (~0.4 nm thick) are terminated with the hydroxyl groups. Thus, the flat sides are hydrophobic and the edges are hydrophilic. The hydroxyl groups on the thin edges form inter-chain hydrogen bonds with those in the adjacent chains, producing an extended sheet. The flat sides of these sheets are hydrophobic and stacked with each other through van der Waals interactions between them. The high-density packing maximizes these intermolecular interactions, stabilizing the crystalline structure of cellulose. Thus, it is not surprising that cellulose is difficult to dissolve. A solvent system must disrupt both the hydrogen bonding and the van der Waals forces to break up and dissolve the crystalline solid [34-38].

Naturally produced cellulose has a broad molecular weight distribution [39], making the formation of high-crystalline lattice as a whole molecule impossible. The crystalline portion consists of small crystallites that often deviate from ideal crystals, by definition, having infinite lattice arrays. They expose significant fraction of cellulose chains at the crystal surface which will have different structures and dynamics than the chains inside the crystal [40-42]. The amorphous character is often thought to arise from the 'fringed micelle' model of the crystalline polymers [43,44]. In this model, the polymer molecules, much longer than the individual crystallites, pass through several crystalline domains (micelles). Amorphous regions are believed to exist where the polymer chains are not part of any crystallite. It is important to realize that this is only a hypothetical model; the reality could be much more intricate [1].

The naturally produced crystalline cellulose can have two polymorphic structures: Iα and Iβ [45]. Algae and lower plants produce mostly Iα allomorph, while higher plants produce mostly Iβ allomorph. Bacteria synthesize biofilms which contain mainly cellulose Iα, while the tunicates produce exclusively cellulose Iβ [24]. The crystalline cellulose can be isolated from the biomass through hydrolysis of other matrix components using strong acids (e.g., 50-72% HCl or H_2SO_4) at high temperatures [31]. The isolated cellulose crystals are often called nanowhiskers. Cellulose Iα can be converted to cellulose Iβ through annealing in weak alkaline solution (typically 0.1 N NaOH) at 260 °C [46]. The crystalline cellulose can be transformed to different polymorphs (Fig. 2). When cellulose Iα and Iβ are treated with high concentration alkaline solution (typically >23% NaOH) and then neutralized, cellulose II is produced [47] in a process called mercerization. Cellulose II can also be obtained through coagulation with water of a fully dissolved cellulose solution in strong ionic solvents [48]. This form is called regenerated cellulose. When cellulose I or II are treated with liquid ammonia or

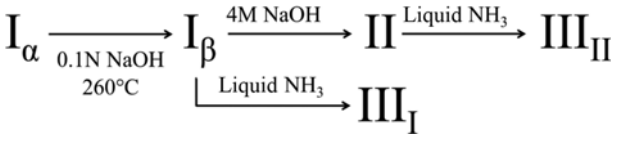

Fig. 2. Cellulose polymorphism. Conversion of native cellulose (Iα and Iβ) into mercerized cellulose (II) and ammonia treated cellulose III [28].

amines, then they convert to III_I and III_{II}, respectively [49-51]. Another polymorphic form that one can find in the literature is cellulose IV [52]. It is reported to be formed through high temperature treatments (about 260 °C) of cellulose III in glycerol [53]. Certain young primary cell walls of cotton were claimed to produce cellulose IV [54]. However, the exact structure of cellulose IV (or even its presence) remains controversial due to the poor reproducibility of cellulose IV synthesis and the poor analytical resolution for the artificially-produced and allegedly-natural cellulose IV form. Thus, cellulose IV will not be discussed in this paper.

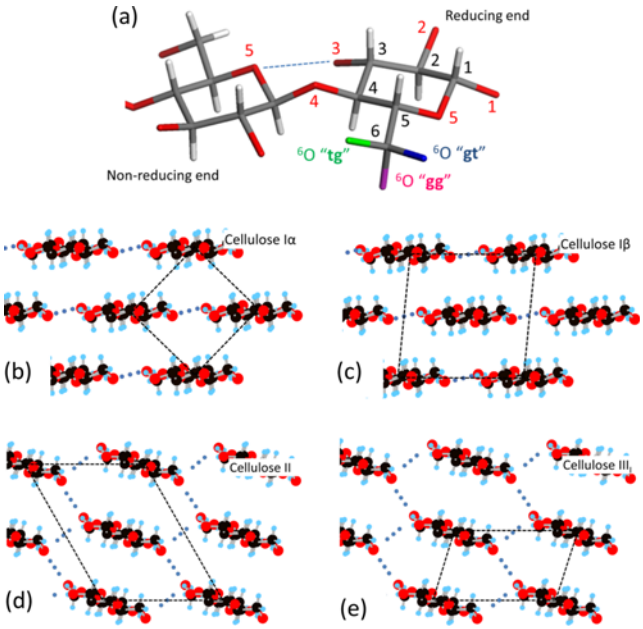

Fig. 3. (a) Schematic of the repeating unit along the cellulose chain (C=dark gray; O=red; H=light gray). Three conformations, *tg*, *gt* or *gg*, of the OH group in the exocyclic $C6H_2OH$ group are shown in the right glucose unit. The *t* and *g* characters stand for *trans* and *gauche* conformations, respectively. The first italic character refers to the relative position between O5 and O6 (O5-C5-C6-O6), and the second italic character refers to the relative position between C4 and O6 (C4-C5-C6-O6). In the left glucose unit, the $C6H_2OH$ group is in the *tg* conformation. The intra-chain hydrogen-bond (3O-H···O5) is shown with a blue dotted line. Hydrogen atoms of OH groups are omitted. Reprinted with permission from Ref. 28. Copyright 2013 Springer. (b)-(e) Cross-sectional view along the chain axis direction and the associated unit cell for each polymorph. Inter-chain hydrogen bonds shown in blue. Notice that native cellulose Iα and Iβ contain a two-dimensional network of hydrogen bonds where chains are packed into flat sheets ((b) and (c)). In cellulose II and III_I, chains are tilted slightly and a three-dimensional network of hydrogen bonds connects chains within and between sheets ((d) and (e)). The conformation of each chain is not accurate in this simplified schematic illustration ((b)-(e)).

A unique aspect of the naturally produced crystalline cellulose allomorphs is the orientation of the primary alcohol group. There are three possible orientations for the oxygen (O6) atom of the hydroxymethyl group [55]: (1) *trans* to O5 and *gauche* to C4, (2) *gauche* to O5 and *trans* to C4, and (3) *gauche* to O5 and *gauche* to C4. They are referred to as *tg*, *gt*, and *gg*, respectively. In the *tg* conformation, the O6H group is positioned in the plane of the six-membered ring; thus, the ribbon-shape cellulose chains connected through side-by-side inter-chain hydrogen bonds can form flat sheets, which then are stacked through van der Waals interactions forming microfibrils (Figs. 3(b) and 3(c)). The rotation of the hydroxymethyl group from the *tg* conformation to *gt* or *gg* can change the intra- and inter-chain hydrogen bonds and induce the tilting of the cellulose chains along the chain axis within the sheet (Figs. 3(d) and 3(e)). Only cellulose Iα and Iβ assume the *tg* conformation for the hydroxymethyl group [21,22]. All other crystals of cellulose polymorphs (II and III), amyloses, and glucose moieties exist in the *gg* or *gt* conformation [51, 55-57].

STRUCTURAL CHARACTERIZATION METHODS FOR CELLULOSE

1. X-ray Diffraction (XRD)

Diffraction occurs when a collimated beam of electromagnetic waves with a wavelength comparable to interatomic distances interacts with the periodic array of molecules in a crystal [58]. The incident beams with wavelengths of ~0.1 nm are generated by x-ray sources (photons), electron guns (electrons), and nuclear reactors (neutrons). Among these, x-ray diffraction (XRD) is most widely used. The electron diffraction of cellulose nanowhiskers was an essential tool to confirm the unit cell symmetry of cellulose [59,60]. However, the electron beam tends to severely damage the small cellulose crystals, making this experiment extremely difficult. The neutron diffraction data enabled the determination of the hydroxyl group positions [21,22], but this requires complete exchange of hydroxyl (OH) groups within the cellulose crystal with deuterated groups (OD). In contrast, XRD can be applied to structural analysis of cellulose, either isolated or in its native state, without severe irradiation damage or extensive OH/OD exchange.

The positions of the diffracted beams with respect to the incident x-ray beam are largely governed by the positions of atoms in the unit cell of the crystal, which is the smallest structural unit that can be repeated by simple translations to generate the entire crystal. The dimension of the unit cell is typically described with unit lengths a, b, c and interaxial angles α, β, γ (Fig. 4). When the unit cells in the sample have perfect three-dimensional periodicity, then the diffraction corresponds to a series of refracted beams that satisfy Bragg's law for constructive interference, $n\lambda=2d\times\sin\theta$, where n is a positive integer, λ is the wavelength of x-ray, d is the spacing between atomic planes with equal electron density, and θ is the incidence and reflection angle (Fig. 4). Thus, the diffraction peak measured at a specific angle (θ) corresponds to a d-spacing of the planes reflecting x-ray. The d-spacing values are then related to the unit cell dimensions (a, b, c and α, β, γ) through Miller indices [58]. Typical

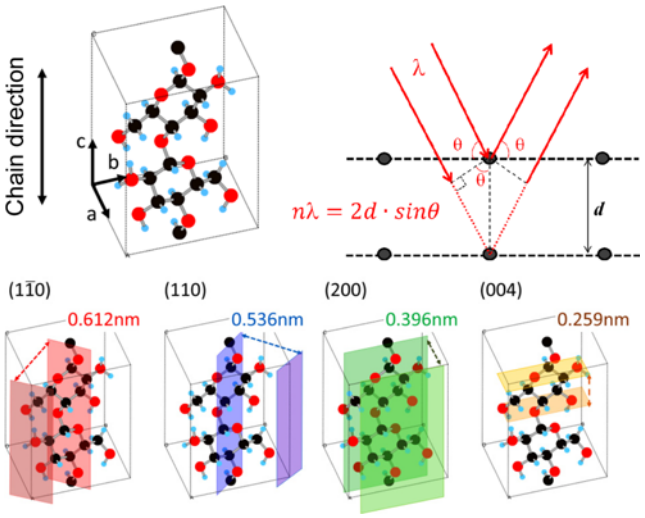

Fig. 4. Schematic illustration of the cellulose chain in a unit cell, Bragg's law, and a set of planes corresponding to the strongest reflections of powder XRD patterns for cellulose Iβ with their Miller indices and d-spacings [60].

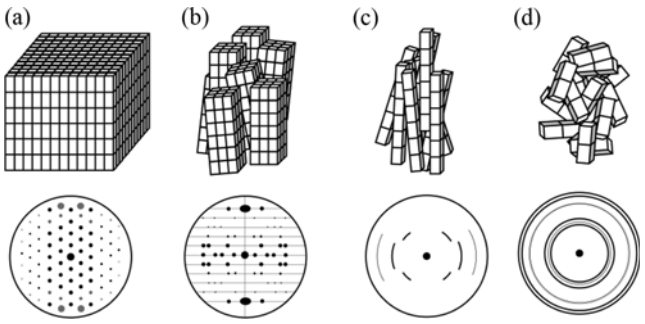

Fig. 5. Different distribution of crystallites and corresponding diffraction patterns within the Debye ring on the detector placed behind the sample [64].

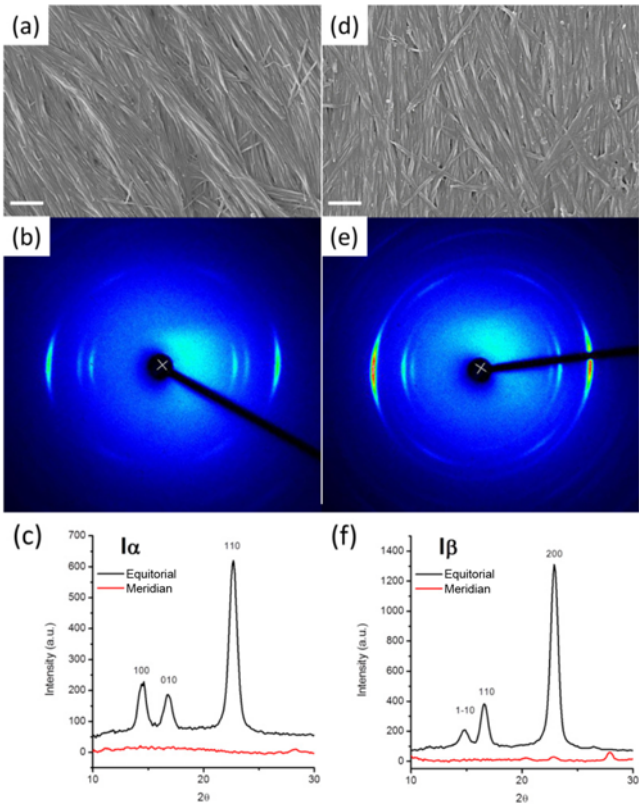

Fig. 6. Cellulose Iα: (a) SEM image, (b) 2D XRD diffraction pattern and (c) equatorial and meridian integrated plots. Cellulose Iβ: (d) SEM image, (e) 2D XRD diffraction pattern and (f) equatorial and meridian integrated plots. Scale bars in (a) and (b) =300 nm. In (b) and (e), the chain axes are approximately vertical. Reprinted with permission from Ref. 29. Copyright 2013 American Chemical Society.

Miller-index planes are shown for cellulose Iβ in Fig. 4. The intensities of the diffracted beams depend on the types and relative positions of the constituent atoms. The reciprocal of the width of the diffracted peak can be related to the size and degree of perfection of the crystallites [61].

To construct a structural model with a high precision, it is essential to use a large single crystal (usually >100 μm) to obtain a two-dimensional reciprocal array of sharp diffraction spots as schematically shown in Fig. 5(a). However, such single crystals are not available for cellulose. The largest cellulose crystals are ~20 nm wide and a few 100s nm long nanowhiskers; thus, the best situation for high quality XRD analysis is to use preferentially-oriented fibrous specimens (Fig. 5(b)) [62]. Individual crystals are rotated differently about their long axis which is nearly parallel to the cellulose chain axis. In this case, the diffraction spots are positioned along the meridian lines (parallel to the fiber axis) and a series of layer lines (perpendicular to the fiber axis; the middle one at the center is called an equatorial line). For thinner and shorter crystals, the degree of alignment along the chain axis would be much weaker (Fig. 5(c)). In this case, the diffraction patterns consist of a series of short arcs whose lengths vary with the degree of lateral orientation. The most commonly encountered cases would be disoriented short fibers (Fig. 5(d)); the diffractions from randomly oriented crystallites produce a series of concentric rings, each corresponding to a characteristic d-spacing. The intensity is uniform on a ring, but it varies among rings. This type of diffraction is commonly referred to as powder diffraction [63]. As an example, Fig. 6 shows the electron micrograph images of cellulose Iα and Iβ nanowhiskers packed preferentially along the long axis of the fibers, their two-dimensional diffraction patterns, and diffraction intensities along the equatorial and meridian lines [29].

Although analyzing these XRD patterns with mathematical models can provide crystallographic structures that have been the foundation of our current knowledge about the cellulose structure, it is important to note that the proposed structural models are based on the data set with limited accuracies. The main limitation stems from the fact that cellulose single crystals cannot be produced large enough for full, assumption-free analysis of diffraction data. In typical single crystal diffraction studies of proteins or small glucose-moiety molecules, crystals larger than 100 μm are used to obtain more than a thousand of reflections, which can then be fully analyzed for atomic coordinates of every single atom in the unit cell (except hydrogen which has very low x-ray scattering power) [65,66]. In theory, at

least three diffraction spots are needed for each atom (three degrees of freedom (x, y, z) per atom). Practically, far more than 10 diffractions per atom are needed for accurate determination of the atomic coordinates in the unit cell. However, this is not the case for cellulose (and most polysaccharides, including amylose). The cellulose crystals are far too small to yield substantial sets of diffraction data. Typical diffraction patterns of the aligned fibers contain far fewer reflections than the number of atomic coordinates in the unit cell. The best diffraction data reported in the literature contained ~250 reflection spots for cellulose Iα and ~300 for cellulose Iβ [21,22]. In the case of random cellulose crystals isolated from biomass originating from higher plants typical diffraction peaks are only ~5 or less.

To reduce the degree of freedom in solving the crystal structure, many known structural constraints are used [67]. These include bond lengths, bond angles, and some torsional angles, as derived from the standard geometry of a sugar ring. Then, only three conformational angles (two dihedral angles of the glycosidic bond and one dihedral angle of the exocyclic hydroxymethyl group) are needed to describe the geometry of the glucose unit along the chain, instead of 33 coordinates for 11 non-hydrogen atoms per glucose unit. After approximating these parameters, additional variables such as the number and packing arrangement (parallel vs. antiparallel) of chains in the unit cell must be optimized for the best fit model. These variables are adjusted and refined somewhat empirically until the smallest fitting error, called crystallographic *R*-value, is obtained [68]. For XRD analyses of aligned cellulose crystals, typical *R*-values are around 0.2 [21]. This is an order of magnitude larger than typical *R*-values (0.02 or smaller) reported for XRD analyses of single-crystal proteins and small glucose-moieties [69,70]. Thus, the crystallographic models proposed for cellulose are an order of magnitude less accurate than single crystals of proteins and carbohydrate molecules. Readers should keep in mind these structural assumptions and limited precisions involved in structural refinements of crystallographic models when they consider atomic positions within the cellulose crystal.

Table 1 compares the space group and unit cell parameters of cellulose polymorphs determined from XRD analyses. The cellulose chains are considered to have alternative residues with a ~180° rotation about the chain axis. This is close to the two-fold screw axis ($P2_1$) symmetry. Following the historic nomenclature of crystallography, early work proposed the unit cell for cellulose I with b as this symmetry axis and β as the monoclinic angle [71]. However, perhaps to facilitate comparative discussions of other polysaccharides, the current convention is to use c as the chain axis and γ as the obtuse monoclinic angle [72].

Cellulose polymorphs are consequences of slight variations in the shape of the glucose ring and more significantly in the dihedral angles of the β-(1,4) linkage and the hydroxymethyl exocyclic group. All chains in the cellulose Iα and Iβ crystals are parallel; in other words, the directions from the reducing ends to the non-reducing ends are all the same. But, there are subtle differences between these two polymorphs. Structural refinements of fiber XRD data suggested that all chains in cellulose Iα are identical and each successive unit along the chain has a slight difference in the amount of deviation from the ideal *tg* conformation of the hydroxymethyl group [22]. Thus, it does not fall into the $P2_1$ space group strictly; instead, it belongs to the P1 triclinic space group. The fiber XRD patterns of cellulose Iβ are fitted with the two-chain monoclinic unit cell in which all units in each chain are identical but two chains have slightly different dihedral angles for the hydroxymethyl groups, although both are close to the *tg* conformation [21]. Whether there is true symmetry in the cellulose crystallites has been questioned because some weak diffraction peaks that did not fit to the $P2_1$ space group were omitted in the structural refinement [1].

Cellulose II, produced by mercerization or regeneration processes, has the O6 atoms close to the *gt* position [56]. This allows the inter-sheet hydrogen bonding in addition to the inter-chain hydrogen bonding within the sheet (Fig. 3(d)). The unique aspect of cellulose II is that the two chains in the unit cell are antiparallel, which means that the direction from the non-reducing end to the reducing end is opposite for the two chains in one unit cell. In early days, there was speculation that mercerized cellulose II might be parallel while regenerated cellulose II was antiparallel. But, it is generally believed that cellulose II has antiparallel chain directions, although the detailed mechanisms for the conversion from the parallel to antiparallel directions are still debatable [73].

Cellulose III_I, obtained through amines or liquid ammonia treatments of cellulose I, has a single-chain unit cell [50]. Thus, all chains in the cellulose III_I crystal are parallel. Similar to cellulose II, the *gt* conformation of the hydroxymethyl group allows an extensive three-dimensional hydrogen bond network (Fig. 3(e)). In this regard, its structure is similar to cellulose II, except for its one-chain unit cell and therefore parallel chains in the crystal.

The crystal size and crystallinity of cellulose is of great interest to researchers handling lignocellulosic biomass and isolated cellulose materials. The full-width-at-half-maximum (FWHM) of a specific diffraction peak at 2θ can be related to the crystallite size along the d-spacing direction of that peak. The most widely-used method is the Scherrer formula assuming no structural distortions [74]:

$$T = \frac{K \cdot \lambda}{FWHM \times \cos\theta}$$

where T is the crystallite diameter and K is the Scherrer constant. The K value is derived to be about 0.9 assuming Gaussian line pro-

Table 1. Unit cell symmetry and dimension of cellulose polymorphs

Allomorph	Space group	# Chains in unit cell	Unit cell dimension (a, b, c in nm; α, β, γ in degree)						Refs.
			a	b	c	α	β	γ	
Cellulose Iα	P1	1	0.672	0.596	1.040	118.1	114.8	80.4	[22]
Cellulose Iβ	$P2_1$	2	0.784	0.820	1.038	90.0	90.0	96.5	[21]
Cellulose II	$P2_1$	2	0.810	0.903	1.031	90.0	90.0	117.1	[56]
Cellulose III_I	$P2_1$	1	0.445	0.785	1.031	90.0	90.0	105.1	[50]

files of XRD and small cubic crystals of uniform size; this value is widely used for the estimation of cellulose crystal size since it is not sensitive to the crystallite shape and symmetry [75]. Although this method is very convenient, it should not be ignored that this equation assumes no distortion of the lattice. Internal stresses due to bending or twisting of cellulose microfibrils can create compressive or tensile strains, which can also broaden the diffraction peak width. Even with these over-simplifications, the crystal size estimated from the XRD peak width appears to correlate reasonably with the surface area determined from water adsorption isotherms [76].

It has been proposed that the crystallinity index (CI) could be calculated by comparing the minimum intensity just before the largest diffraction peak with the main diffraction peak height [61]. In the case of cellulose, this is typically done by comparing the background intensity at 2θ=~18° with the peak height at 2θ=22.8° (110 peak for cellulose Iα and 200 peak for cellulose Iβ). Since the random scattering halo from the amorphous component has significant contribution in the region where the cellulose diffraction peaks are located, alternative methods such as peak deconvolution or amorphous subtraction methods were also suggested [77]. However, all these methods assume that the x-ray sensitivity to the crystalline and amorphous phases is the same, which is not true. Thus, depending on the calculation method used, the XRD CI value can vary over a wide range [77]. Especially, the CI value calculated from the peak height method proposed originally by Segal, which is the simplest, thus the most widely used, can be off by more than 100% from the true value [27]. More sophisticated methods based on diffraction theories can resolve these uncertainties in the CI calculation [78], although these advanced methods are not easy for non-XRD experts to use.

2. Nuclear Magnetic Resonance (NMR) Spectroscopy

NMR analysis is based on the fact that the resonance behaviors of the nuclei of atoms with magnetic moments under electromagnetic irradiation are very sensitive to tiny differences in magnetic field shielding imposed by surrounding atoms (most strongly influenced by the atoms covalently attached to that specific atom). For isolated molecules in liquid and gas, the positions of the surrounding atoms are determined by equilibrium bond angles and distances. But, in crystalline solids, they can vary due to the packing structure. The tiny differences in magnetic field shielding imposed by surrounding atoms can provide critical information needed to determine the structure of cellulose. First, let's briefly review the basic principle of the NMR process.

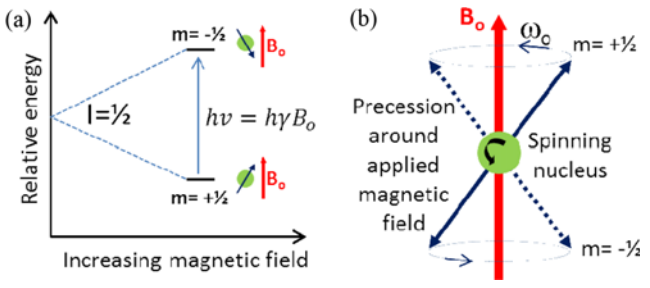

Fig. 7. Splitting of nucleus spin states (m=+½ and −½) under the influence of the external magnetic field (B_o) and schematic illustration of precession for spin-up and spin-down states at Larmor frequency (ω_o).

When an external magnetic field of strength B_o is applied, quantum mechanics dictates that the energy state of a nucleus with a magnetic spin number I=½ (such as ^{1}H and ^{13}C) will split into two states (2I+1): one has its magnetic moment along the B_o field and the other against the B_o field (Fig. 7(a)) [79]. The state with its magnetic moment aligned parallel with the applied field (denoted as m=+½) is energetically more favorable than the state with the antiparallel alignment (m=−½). The rotation axis of the spinning nucleus cannot be oriented exactly parallel or antiparallel with the applied B_o field; instead, it moves in precession about the B_o field (Fig. 7(b)). The angular velocity of the precessional orbit is defined by the Larmor frequency $\omega_o=\gamma B_o$, where γ is the gyromagnetic (or magnetogyric) ratio (for ^{1}H, $\gamma=2.674\times10^4$ gauss^{-1}sec^{-1}) [80]. If the system is irradiated with electromagnetic waves with frequency (υ) in the MHz range, the precessing nuclei in the energetically more favorable state (m=+½) will absorb the energy and be promoted to the less favorable higher energy state (m=−½). This transition process is called nuclear magnetic resonance (NMR) because the frequencies of the applied radiation and the precession resonate ($\upsilon=\omega_o$).

The energy separation between these spin states is quite small; the thermal energy at room temperature is high enough to populate significant portions of nuclei into the high energy spin state. The Boltzmann distribution predicts that for ^{1}H nuclei in an external field B_o=1.4 Tesla, the population difference between two energy states would be only ~30 for every one million spins [80,81]. Because of this nearly equal population distribution, the NMR absorption yields relatively weak signals that need to be averaged over a long period of time with a relatively large amount of sample (typically >100 mg) to obtain spectra with adequate signal to noise ratios.

Negatively charged electrons surrounding nuclei can produce magnetic fields that are antiparallel to the external magnetic field. This electronic magnetic field effect will cause nuclei with different chemical environments to yield resonance frequencies perturbed from the Larmor frequency: $\omega_o=\gamma(B_o-S)$ where S represents the shielding magnetic field caused by the electron magnetic moment [82]. The amounts of magnetic field shielding effects by electrons are measured with respect to a standard reference that is often added to the sample. In the case of ^{1}H NMR, tetramethylsilane (TMS) is chosen as a standard. Since the electronegativity of Si is lower than the carbon atom, the sigma bond electrons in the Si-C bond are shifted toward the carbon atoms in the four methyl groups. Consequently, the protons will have high shielding effects (large S), causing the resonance signal to be generated at a very high magnetic field (high B_o) at a fixed υ. Since all ^{1}H's in TMS are in the same environment, only one NMR peak is generated. This is the signal that all other ^{1}H NMR signals of a sample are referenced to. The difference from this reference signal (Δ) is expressed as $(\Delta\times10^6)/\upsilon$ in parts per million (ppm), which is called chemical shift (δ). The larger δ, the shielding effect by electrons is less, and thus the resonance can take place at a lower magnetic field; thus, the large chemical shift side is called a down-field side.

The application of NMR to cellulose has been advanced by the development of solid-state ^{13}C NMR technique. Although ^{13}C has the same nucleus spin (I=1/2) as ^{1}H, its natural abundance is only 1% and thus its signal is very weak. This problem was resolved by the proton-carbon cross polarization (CP) method to enhance the ^{13}C sensitivity and the proton decoupling method to eliminate the

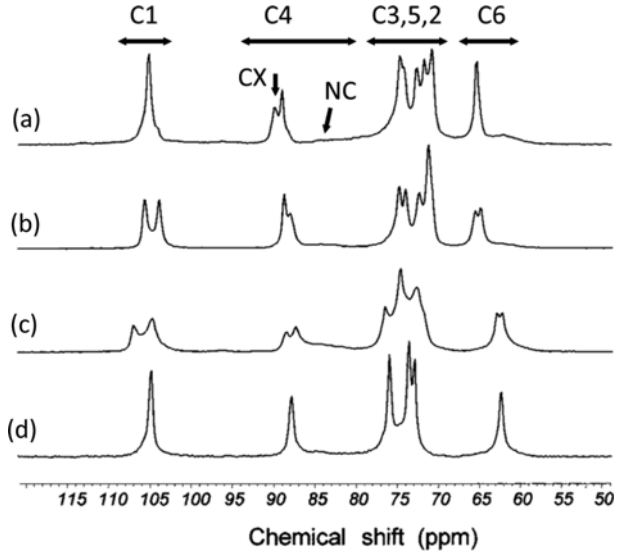

Fig. 8. ^{13}C CP/MAS NMR spectra of cellulose (a) Iα from *Glaucocystis*, (b) Iβ from tunicate, (c) II, and (d) III$_I$. The region of chemical shift for each carbon is labeled. For C4, the peaks attributed to crystalline and amorphous portions are marked as CX and NC, respectively. Modified with permission from Ref. 50. Copyright 2013 American Chemical Society.

dipolar interactions between the ^{13}C nuclei and neighboring protons [83,84]. In liquid, molecules are subject to fast thermal motions, which make them magnetically equivalent in the applied field. In the case of solid, the molecules are not mobile; thus, molecules at different locations in the solid are not magnetically isotropic in the field. This anisotropy of solid samples in the magnetic field can be annihilated by spinning the sample about the so-called magic angle [85]. This method is denoted as magic angle spinning (MAS) [86, 87]. One important attribute of the CP/MAS ^{13}C NMR analysis is that the relative intensities of the resonance peaks are linearly proportional to the carbon atom density in the sample. Thus, the relative intensities can be easily interpreted as the relative concentration of different carbon species.

Fig. 8 compares the NMR spectra of cellulose polymorphs - Iα, Iβ and II, and III$_I$ [50]. Starting from the down-field side, the 102-108 ppm region is attributed to C1, the 81-93 ppm region is to C4, the clusters between 70 and 81 ppm are to C2, C3, and C5, and finally the peaks between 60 and 70 ppm are to C6 [88]. This order from the downfield follows the number of electronegative atoms in the nearest and next-nearest neighbor positions to that specific carbon. For example, among six carbons in the glucose unit, the C1 atom has the highest influence from oxygen atoms (O1 and O5 in the nearest positions and O2 in the next-nearest positions) and the C6 atom has the smallest influence (O6 in the nearest positions and O5 in the next-nearest positions). Thus, C1 is at the lowest field side and C6 is at the highest field side. All others (C2, C3, C4, and C5) are between these two extreme cases. The ^{13}C chemical shifts of cellulose Iα and Iβ have recently been computed from the first principles using density functional theory (DFT) [89].

The splitting of the C4 resonance peak near 88-90 ppm is of special interest in plant biomass characterization since it distinguishes cellulose Iα and Iβ. Both allomorphs show a peak at 88.7 ppm. Cellulose Iα exhibits a second peak in the down-field side (89.7 ppm), while cellulose Iβ has a second peak in the up-field side (88.0 ppm). Thus, deconvolution of the peaks in this region can provide the relative abundance of these allomorphs in the sample [46]. The broad featureless peak in the up-field region (81-86 ppm) is mostly associated with the amorphous phase [18]. A part of this component could also be correlated with chains at the surfaces of the cellulose crystals which assume a less ordered structure than the core of the crystal [90,91]. Thus, the relative intensity of this broad peak in the 81-86 ppm region and the sharp peaks in the 86-91 ppm region can be used to estimate the cellulose crystallinity or the surface-to-volume ratio of the cellulose microfibrils [77,92]. Sometimes, there could be a broad feature in the C4 region that overlaps with the sharp peaks of cellulose Iα and Iβ in the 86-91 ppm region, but cannot be deconvoluted into the Iα and Iβ components. This component is often attributed to a 'paracrystalline' phase [91,93], but the exact nature or origin of the paracrystalline phase is still in debate.

In the C1 peak region, cellulose Iβ and II show double peaks while cellulose Iα shows a single sharp peak (Fig. 8) [94]. The origin of this doublet splitting is not fully understood yet. It could be due to the presence of nonequivalent glycosidic linkages in a single chain or two chains with the same glycosidic units in each chain but slightly different from each other [18]. It is also interesting that the amorphous contribution found in the C4 resonance region (81-86 ppm) is not prominent in the C1 resonance region. The reason for the absence of the amorphous component in the C1 region is not well understood; it was speculated that the anomeric effect of the C1 position could make it less flexible than the C4 position [18].

The chemical shift of the C6 resonance is correlated well with the dihedral angle of the hydroxymethyl group. The *gg*, *gt*, and *tg* conformers give a peak at 60-62 ppm, 62-64 ppm, and 65-66 ppm, respectively [55]. These chemical shifts could be used to determine the relative population of three rotational conformers. The cellulose chains at the crystal surface can assume the *gt* or *gg* conformation, contributing to the peaks in the up-field region (60-64 ppm) [95]; however, the amorphous component can also have a broad peak in this up-field region overlapping with the *gt* or *gg* contributions [18]. The chemical shifts of the C1 and C4 are also correlated with the dihedral angle of C1-O and O-C4 bonds [18], but their variances are not discussed extensively in the literature partly due to the complication with the multiple peaks and splitting.

3. Infrared (IR) and Raman Spectroscopy

Two most widely used vibrational spectroscopic analyses are IR and Raman spectroscopy. The vibration spectrum is obtained as a consequence of the absorption or emission of electromagnetic radiation at frequencies that correlate to the vibration of specific sets of chemical bonds in a molecule. The vibration energy states are the quantum mechanical states of electron waves confined in specific chemical bond coordinates. They are mostly determined by the equilibrium bond strength and geometry as well as the reduced mass of the atoms involved in specific vibrations. In the case of solid materials, the equilibrium energy states are also affected by the packing structure of molecules or interactions with surrounding functional groups in the solid phase. Thus, small changes in vibration peak position or intensity can provide the structural information of crystalline cellulose. This principle will be reviewed, starting from the vibrational spectroscopy basics.

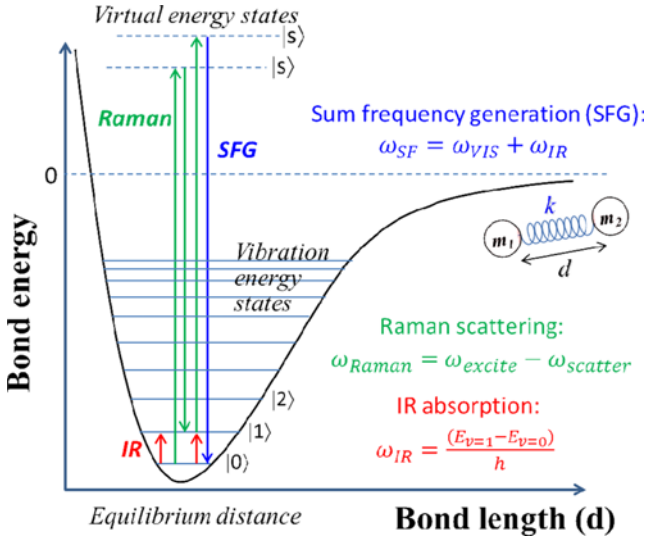

Fig. 9. Vibration energy states of diatomic molecules and electronic transitions involved in IR absorption (red arrow in the left), Raman scattering (a pair of green arrows in the middle), and SFG process (red, green, and blue arrows in the right). The diatomic molecules can be modeled as two masses (m_1 and m_2) connected with a spring with a spring constant k. The symbols $|0\rangle$, $|1\rangle$, $|2\rangle$, etc. denote the vibrational energy states.

The basic principle of vibrational spectroscopy can be derived with a simple harmonic oscillator model for a di-atomic chemical bond. A chemical bond can be viewed as a spring with a spring constant k holding two atomic species with masses m_1 and m_2 separated by a distance d (Fig. 9) [96]. The chemical potential near the equilibrium bond distance can be expressed with Hooke's law; solving the Schrödinger equation describing the wave functions of electrons involved in this chemical bond gives the discrete vibration energy states (Fig. 9) [96]. The vibration energy levels are expressed as $E_v = (h/2\pi)\sqrt{k/\mu}\,(v+(1/2))$ where h is the Planck constant, μ is the reduced mass ($m_1m_2/(m_1+m_2)$), and v is the vibration quantum number. This simple equation provides a link between the bond strength of the covalent bond between two atoms (or molecular fragments), the mass of the interacting atoms (molecular fragments) and the frequency of vibration.

IR spectroscopy is the measure of IR beam attenuation through the absorption of the IR photon whose energy is the same as the energy difference of two vibration energy states (red up-arrow in Fig. 9). At room temperature, most vibrational excitations corresponding to the IR absorption are from the v=0 state to v=1 state: $\omega_{IR}=E_{v=1}-E_{v=0}=(h/2\pi)\sqrt{k/\mu}$. The fundamental requirement for IR absorption is a net change in dipole moment during the vibration of a molecule or a functional group [97].

When electrons at the v=0 state absorb photons in the visible wavelength range, they can transition to a virtual electronic state (green up-arrow in Fig. 9). Upon the transition back to the ground electronic state, most of them come back to the vibrational ground state (v=0); but, a small fraction fall to the v=1 state (green down-arrow in Fig. 9). The former case is called Rayleigh scattering and the latter is Raman scattering. Thus, the energy difference between the originally absorbed photon (excitation) and the emitted photon (scattering) in the Raman scattering corresponds to the vibration energy level difference: $\omega_{Raman}=\omega_{excite}-\omega_{scatter}=(h/2\pi)\sqrt{k/\mu}$. The requirement for Raman activity is a net change in bond polarizability during the electronic transition upon absorption of electromagnetic wave [97]. Because of the different requirements in the IR (dipole change) and Raman (polarizability change) processes, they can provide spectral information complementary to each other.

When this simple concept of the vibration of a di-atomic bond is

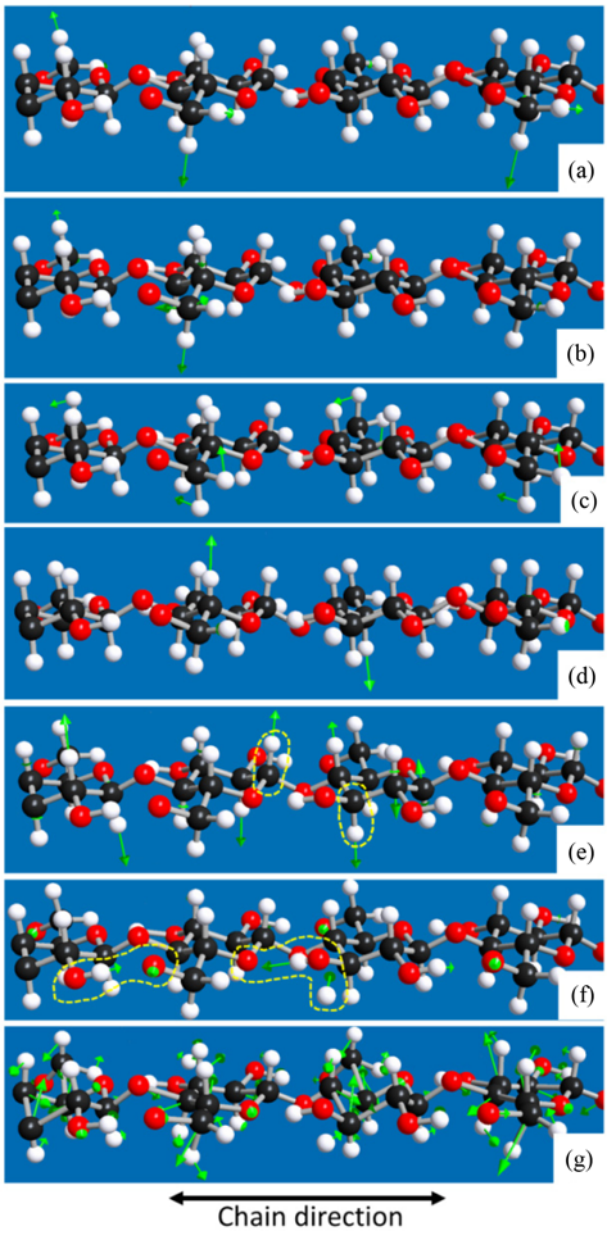

Fig. 10. Cellulose Iβ vibration modes. (a) CH_2 symmetric stretch at 2,850-2,860 cm^{-1}, (b) CH_2 asymmetric stretch at ~2,944 cm^{-1}, (c) CH_2 bend at 1,432-1,484 cm^{-1}, ((d) and (e)) localized and delocalized CH stretch at ~2,909 cm^{-1}, (f) stretch of intra- and inter-chain hydrogen-bonded OH groups at 3,240-3,400 cm^{-1}, and (g) skeletal stretch and bend at ~1,098 cm^{-1}. (C=dark gray; O=red; H=light gray; displacement vectors=green). Reprinted with permission from Ref. 29. Copyright 2013 American Chemical Society.

extended to the molecule as a whole, we need to consider concerted vibrations of atoms in the molecule. These vibrational motions of the molecule can be reduced by using group theory to a series of normal modes such as symmetric and asymmetric stretch vibrations of CH_2 and CH_3 groups, in-plane and out-of-plane bending vibrations of CH_2 groups, etc. [96]. The normal vibration modes of cellulose can be calculated theoretically. Some of the calculated vibration modes are shown in Fig. 10 [29]. The videos of these modes can be viewed at the Supporting Information of Ref. [29]. Each mode has its own energy (vibration frequency) and extinction coefficient (absorption cross-section). The overall vibration spectrum is essentially a composite of the group frequencies of these normal modes, with band intensities related to the contributions of individual functional groups.

Figs. 11 and 12 compare the infrared and Raman spectra, respectively, of uniaxially-aligned nanowhiskers of cellulose Iα and cellulose Iβ obtained from *Glaucocystis* and *Halocynthia* [29]. One significant difference between the IR and Raman spectra is that the O-H stretch vibrations in the 3,200-3,600 cm^{-1} region are strong in IR but very weak in Raman. This is because the dipole moment change of the O-H vibration is large, but its polarizability change is small. The O-H stretch band is inherently broad, reflecting the complexity of hydrogen bond networks in the crystal as well as water molecules present in the amorphous region [26]. The strong bands near 2,900 cm^{-1} observed in both IR and Raman can be attributed to

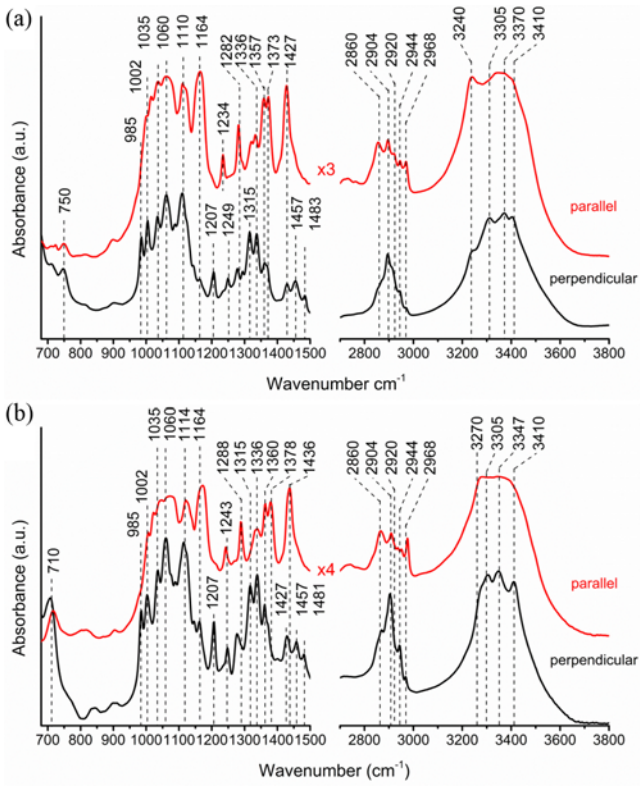

Fig. 11. Polarized FT-IR spectra of aligned nanowhiskers of (a) cellulose Iα (from *Glaucocystis*) and (b) cellulose Iβ (from *Halocynthia*). Spectra were taken with the IR polarization perpendicular (black) and parallel (red) to the nanowhisker alignment direction. Reprinted with permission from Ref. 29. Copyright 2013 American Chemical Society.

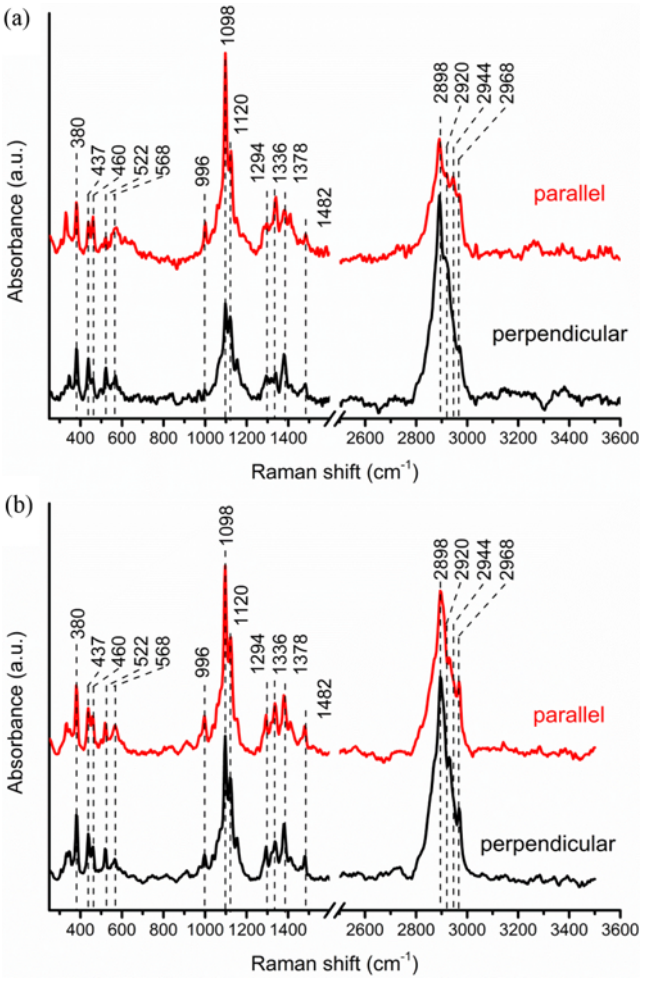

Fig. 12. Polarized FT-Raman spectra of aligned nanowhiskers of (a) cellulose Iα (from *Glaucocystis*) and (b) cellulose Iβ (from *Halocynthia*). Spectra were taken with the excitation laser polarization perpendicular (black) and parallel (red) to the nanowhisker alignment direction. Reprinted with permission from Ref. 29. Copyright 2013 American Chemical Society.

the C-H stretch vibrations. The shoulder peaks in the lower (2,840-2,880 cm^{-1}) and higher (2,920-2,970 cm^{-1}) energy sides of the C-H peak can be assigned to the symmetric and asymmetric CH_2 stretch vibrations, respectively, of the hydroxymethyl group of cellulose [29]. The deformation vibration, often called a scissoring mode, of the CH_2 group on C6 appears in the 1,450-1,480 cm^{-1} region. The bands between 1,200 and 1,450 cm^{-1} are due to C-C-H bending, CH_2 rocking and wagging, and C-O-H bending modes. The peaks in the region between 950 and 1,150 cm^{-1} originate from the skeletal ring stretches as well as the stretch modes of glycosidic C-O bonds and alcoholic C-O bonds. Below 950 cm^{-1}, the bending of skeletal C-C-C, C-O-C, O-C-C, and O-C-O bonds contributes. The region between 300 and 800 cm^{-1} could be due to the torsion or bending modes of the six-membered ring with respect to the glycosidic bonds [26,29,98].

In the case of small isolated molecules, the peak assignment to certain functional groups is relatively easy. In the case of cellulose, however, it is not that straightforward. This is because all bands,

especially at wavenumbers below 1,450 cm^{-1}, are highly coupled over several covalent bonds [98]. An example is shown for the 1,098 cm^{-1} mode in Fig. 10(g) [29]. So, the peaks in this region cannot be assigned to isolated vibrations of specific C-C or C-O bonds. For this reason, the interpretation of the polarization dependence of their vibration peak intensities is complicated. According to DFT calculations, the only local vibration modes are the stretch and bend modes of the CH_2 groups in the hydroxymethyl side chain [29]. A few examples of those modes are shown in Figs. 10(a), 10(b), and 10(c) [29].

When the vibration spectra of cellulose Iα and Iβ are compared to each other (for example, Fig. 11(a) versus Fig. 11(b); Fig. 12(a) versus Fig. 12(b)), it can be noticed their spectra are almost identical, except for minor differences in the relative intensities, in all regions associated with the skeletal vibrations as well as the regions associated with the bending and stretching vibrations of the CH and CH_2 groups. The main differences between the Iα and Iβ spectra are in the broad OH stretch region (3,200-3,600 cm^{-1}). The most distinct feature in the OH stretch region of the IR spectra is that cellulose Iα has a peak at 3,240 cm^{-1} and cellulose Iβ has a peak at 3,270 cm^{-1} [24]. Another characteristic difference in IR is the 750 cm^{-1} peak for cellulose Iα and the 710 cm^{-1} peak for cellulose Iβ [24, 99]. These results imply that the chain conformations in cellulose Iα and Iβ are very close to each other and the main structural difference is the hydrogen bonding networks and the dihedral angles at the glycosidic linkages [1,100].

Spectral resemblance is also observed when the vibration spectra of cellulose I, II, and III_I are compared. The peaks in the 800-3,000 cm^{-1} regions are similar to each other with only minor variations in the relative intensities [101]. The main differences are observed in the regions below 800 cm^{-1} and above 3,000 cm^{-1}. Cellulose II has two characteristic peaks at 3,450 and 3,480 cm^{-1}; cellulose III_I has a distinct peak at 3,480 cm^{-1}. These reflect that the basic structure of the glucose ring does not change significantly; the main differences are the dihedral angles at the glycosidic linkages and inter-chain hydrogen bond networks.

The cellulose crystallinity can be estimated from the vibration spectra of cellulose. In the case of Raman, the 380 cm^{-1} peak is observed only for crystalline cellulose I; it does not exist for amorphous cellulose [102]. The relative intensity of this peak with an internal standard peak, which is found in both crystalline and amorphous phases, can be used to calculate the crystallinity of cellulose [102,103]. There are several peaks in IR that can be correlated to the crystallinity index calculated from XRD [104,105]. However, the correlation of these IR peaks is weaker than the 380 cm^{-1} Raman peak. Water molecules in the amorphous phase can easily be exchanged with D_2O [106]. Upon H_2O/D_2O exchange, the OH groups of the cellulose chains in the amorphous phase are converted to the OD groups. Thus, the relative area of the OH and OD vibration peaks can be used to estimate the amorphous portion accessible by water [107,108]. In the case of highly crystalline samples, the OH/OD exchange can be used to probe the cellulose chains exposed at the crystal surface [109,110].

IR and Raman spectroscopy is routinely applied to cellulose analysis. However, when cellulose is imbedded in lignocellulosic biomass, other matrix polymers present in the plant cell walls (such as hemicellulose, pectins, and lignin) will interfere with cellulose vibration peaks [26]. Thus, IR and Raman analyses of cellulose present inside lignocellulosic biomass samples are not that simple.

4. Sum Frequency Generation (SFG) Vibrational Spectroscopy

SFG vibrational spectroscopy is a non-linear optical spectroscopy that can selectively detect molecules or functional groups arranged noncentrosymmetrically in a medium interacting with high-intensity laser beams [111,112]. SFG is well known as a surface-sensitive analytical tool [113], but it is even more sensitive to noncentrosymmetric crystalline materials such as piezoelectric crystals as well as cellulose [114,115]. Although the experimental set-up is much more complicated than typical IR and Raman spectroscopy, SFG can provide many advantages and complementary information that the conventional spectroscopy methods cannot offer. One is its intrinsic selectivity to crystalline cellulose present inside lignocellulose biomass. This allows the analysis of crystalline cellulose without spectral interferences from non-cellulosic components in the lignocellulose biomass [26]. This eliminates the need for chemical treatments of biomass to isolate or separate cellulose from biomass. The SFG process is also very sensitive to structural ordering over an optical coherence length (which is of the order of 100s nm) [116,117]. Thus, it can provide three-dimensional information about structural hierarchy of cellulose microfibrils inside biomass. The basic principles and applications of SFG to cellulose will be reviewed in this section.

The electric field of a light wave, $E(\omega)$, propagating through a medium exerts a force on the valence electrons of the molecules comprising the medium. The sum of the molecular electric dipoles gives rise to a dipole moment per unit volume. This is the polarization of the medium $P(\omega)$ [111]:

$$P(\omega)=\varepsilon_o(\chi^{(1)}E(\omega)+\chi^{(2)}E(\omega)^2+\chi^{(3)}E(\omega)^3+\cdots)$$

where ε_o is the dielectric constant of vacuum and χ is the susceptibility. The product of $\varepsilon_o\chi$ is the dielectric constant of a material. The high order χ terms are extremely small compared to $\chi^{(1)}$. When the material is irradiated with typical continuous irradiation sources, as in the case of IR and Raman spectroscopy, the high order polarization terms are negligible and cannot be probed. However, when high-intensity laser pulses are irradiated to the sample, then $E(\omega)^2$ and $E(\omega)^3$ terms become large enough to make the high order polarization terms significant.

Let's consider the second-order polarization term, $\chi^{(2)}E(\omega)^2$, when two laser beams with frequencies ω_1 and ω_2 are irradiated. The electric fields of the laser beams can be expressed as the sum of simple cosine waves, $E(\omega)=E_1\cos\omega_1 t+E_2\cos\omega_2 t$. Then, the second-order polarization term can be expanded using simple trigonometric conversions:

$$\begin{aligned}P^{(2)}&=\varepsilon_o\chi^{(2)}(E_1\cos\omega_1 t+E_2\cos\omega_2 t)^2=\varepsilon_o\chi^{(2)}(E_1^2\cos^2\omega_1 t\\&\quad+2E_1E_2\cos\omega_1 t\cos\omega_2 t+E_2^2\cos^2\omega_2 t)=\varepsilon_o\chi^{(2)}\Big(E_1^2\frac{(1+\cos2\omega_1 t)}{2}\\&\quad+2E_1E_2\frac{(\cos(\omega_1+\omega_2)t+\cos(\omega_1-\omega_2)t)}{2}+E_2^2\frac{(1+\cos2\omega_2 t)}{2}\Big)\\&=\varepsilon_o\chi^{(2)}\Big(\frac{1}{2}(E_1^2+E_2^2)+\frac{1}{2}(E_1^2\cos2\omega_1 t+E_2^2\cos2\omega_2 t)\\&\quad+E_1E_2\cos(\omega_1+\omega_2)t+E_1E_2\cos(\omega_1-\omega_2)t\Big)\end{aligned}$$

This simple mathematical manipulation explains that the second-order responses of the material polarization contain multiple terms

with frequencies equal to two times the input frequencies ($2\omega_1$ and $2\omega_2$), the sum of input frequencies ($\omega_1+\omega_2$), and the difference of input frequencies ($\omega_1-\omega_2$). The SFG spectroscopy focuses on the photon emitted from the sample with a frequency ($\omega_1+\omega_2$).

To probe vibration modes of molecules in the sample, one of the input laser frequencies is tuned to mid-IR range (ω_{IR}) and the other beam is typically in the visible range (ω_{VIS}). Then the output frequency, $\omega_{SFG}=\omega_{VIS}+\omega_{IR}$, is in the blue to ultraviolet range (blue down-arrow in Fig. 9). The intensity of the SFG output signal, $I(\omega_{SFG})$, is expressed as [118]:

$$I(\omega_{SFG}) \propto \left|\chi_{eff}^{(2)}\right|^2 I(\omega_{VIS}) I(\omega_{IR})$$

where $I(\omega_{VIS})$ and $I(\omega_{IR})$ are the intensities of the visible and infrared laser beams and $\chi_{eff}^{(2)}$ is the effective nonlinear susceptibility [118]:

$$\chi_{eff}^{(2)} = \frac{N \sum_{\alpha,\beta,\gamma} \langle M_{\alpha\beta} A_\gamma \rangle}{\varepsilon_o(\omega_{IR} - \omega_q - i\Gamma)}$$

Here, N is the number density of the molecules producing SFG signals, $\langle M_{\alpha\beta} A_\gamma \rangle$ is the angle-average of the product of the Raman and IR tensors, ω_q is the frequency of a normal vibration mode, and Γ is the damping constant. In the SFG analysis with a tunable IR beam, signals are resonantly enhanced as the frequency of the input IR beam (ω_{IR}) approaches ω_q in the sample. The peak width is determined by the damping constant Γ. The numerator of the $\chi_{eff}^{(2)}$ term states that the vibration mode must be both Raman- and IR-active to be SFG-active. Otherwise, either Raman or IR tensors will be zero, making $\chi_{eff}^{(2)}=0$. Another very important selection rule for SFG comes from the third-order tensor nature of $\chi_{eff}^{(2)}$, which dictates that the SFG process takes place only when the irradiated medium has no inversion symmetry, otherwise $\chi_{eff}^{(2)}$ is zero [111,119,120].

The noncentrosymmetry requirement is schematically illustrated in Fig. 13. When vibration modes (denoted with brown arrows) are arranged randomly (Fig. 13(a)) or with inversion symmetry (Fig. 13(b)), then $\chi_{eff}^{(2)}$ is zero. Or if the molecule itself has inversion symmetry (Fig. 13(c)), then $\chi_{eff}^{(2)}$ is zero. Only when the molecule does not have inversion symmetry and those molecules are arranged noncentrosymmetrically in space (Fig. 13(d)), $\chi_{eff}^{(2)}$ can have non-zero values and SFG signals can be generated [111]. In addition, the phase matching of the SFG process is very sensitive to the spatial distribution of vibration groups over the optical coherence length, which is on the order of hundreds of nanometers [116]. Thus, the SFG spectral features such as relative intensities of various vibration modes

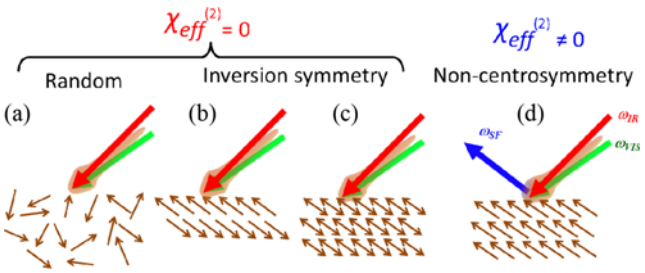

Fig. 13. Schematic illustration of the noncentrosymmetry requirement of $\chi_{eff}^{(2)}$ in the SFG process. The directionality of vibration modes is depicted by brown arrows.

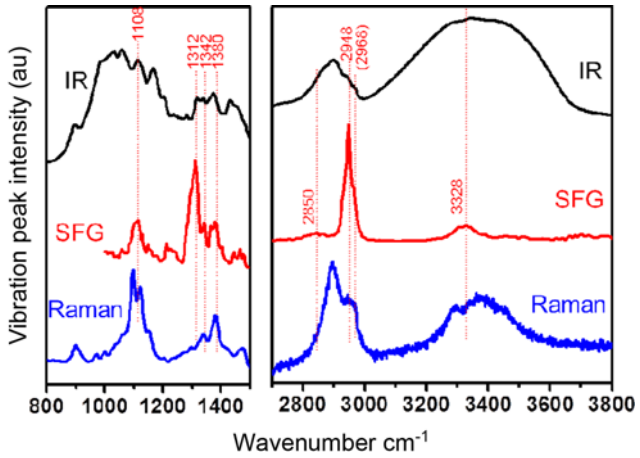

Fig. 14. Comparison of IR, Raman and SFG vibrational spectra of microcrystalline cellulose (Avicel® PH-101). Reprinted with permission from Ref. 26. Copyright 2013 American Chemical Society.

can provide information about the structural assembly of the vibration groups over this length scale.

Fig. 14 compares the IR, Raman, and SFG spectra of microcrystalline cellulose (commercial Avicel® PH-101) [26]. The SFG spectrum of cellulose shows much sharper and fewer vibration peaks compared with IR and Raman spectra, owing to the noncentrosymmetry selection rule of the SFG process. The absence of the peak at ~2,900 cm^{-1} in SFG is due to the symmetry cancellation of the axial C-H bonds [26]. In the cellulose I crystal, there are equal numbers of axial C-H groups on the opposite sides of the chain; thus, it creates a situation similar to Fig. 13(b). The peaks at 2,850 cm^{-1} (weak) and 2,945 cm^{-1} (strong) are attributed to the symmetric and asymmetric vibrations, respectively, of the CH_2 group at the C6 position with the *tg* conformation [26]. These peak positions are in good agreement with the vibration frequencies predicted from DFT calculations [29].

The SFG peak at ~3,325 cm^{-1} can be attributed to the intra-chain hydrogen-bonded OH groups such as O3-H and O2-H [26]. Although water molecules or hydroxyl groups in the amorphous region produce a broad interfering band in IR, they do not produce any SFG signal since they are randomly arranged in the sample (Fig. 13(a)). After proper H/D exchange treatments of the hydroxyl groups in the amorphous region accessible by D_2O, the cellulose SFG spectra showed neither discernible O-D peak growth nor O-H peak decrease [26,29]. This confirmed that SFG is sensitive to the crystalline cellulose chains, not the disordered chains in the amorphous region.

Hemicellulose and lignin in biomass do not produce SFG signals, because they are randomly arranged in the sample [26]. This allows direct and selective detection of crystalline cellulose inside plant cell walls without any extraction or separation of amorphous components [26]. In addition, the SFG signal intensity is sensitive to a wide range of crystallinity, although it varies non-linearly with the mass fraction of cellulose in the samples [27]. The non-linearity might be due to the matrix effects such as light scattering and absorption of IR by other carbohydrate polymers, as well as the nonlinear dependence of the SFG process on the density of the SFG-

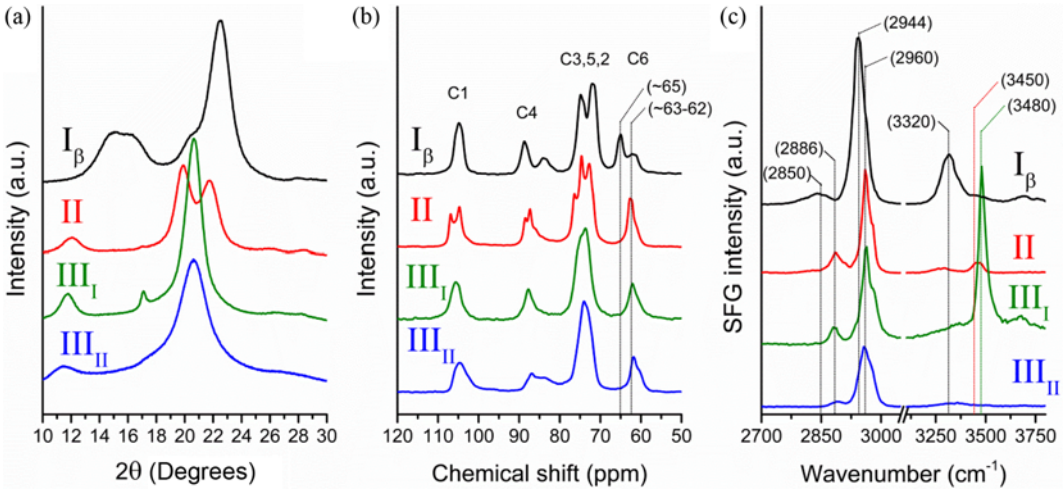

Fig. 15. Polymorphism of commercial cellulose Avicel PH-101. (a) XRD, (b) ^{13}C NMR, and (c) SFG spectra. Reprinted with permission from Ref. 28. Copyright 2013 Springer.

active molecules. As long as the cellulose crystal size and packing do not change, the SFG intensity can be correlated to the amount of crystalline cellulose in biomass using a proper calibration curve [27].

SFG can distinguish cellulose polymorphs. Cellulose Iα and Iβ show characteristic OH peaks at 3,240 cm^{-1} and 3,270 cm^{-1}, respectively [29]. Fig. 15 compares the XRD, NMR, and SFG spectra of cellulose Iβ, II, III_I, and III_{II}. Compared to cellulose Iβ, cellulose II, III_I, and III_{II} show a blue shift of the CH_2 peaks: from 2,850 cm^{-1} to 2,886 cm^{-1} for the symmetric stretch and from 2,944 cm^{-1} to 2,960 cm^{-1} for the asymmetric stretch [29]. The CH_2 peak positions in the cellulose SFG spectra can be correlated with the conformations of the hydroxymethyl groups of these polymorphs. The NMR spectra in Fig. 15(b) show the C6 signal at 64.9 ppm for cellulose Iβ, and 62.1-62.8 ppm for cellulose II, III_I, and III_{II}. These $^{13}C6$ chemical shifts in NMR correspond to *tg* and *gt* conformations, respectively (see Section 2). Thus, the CH_2 peak at 2,960 cm^{-1} can be attributed to the *gt* conformation of the hydroxymethyl groups in the cellulose II, III_I, and III_{II} polymorphs.

In the OH stretch region of SFG, drastic differences are observed among cellulose polymorphs (Fig. 15(c)). Cellulose Iβ shows a peak at 3,320 cm^{-1} with a weak shoulder at 3,270 cm^{-1}. Cellulose III_I has a very sharp and strong peak at 3,480 cm^{-1}; in contrast, cellulose II and III_{II} display very weak or barely noticeable peaks at 3,450-3,480 cm^{-1}. The OH peaks near 3,450 and 3,480 cm^{-1} of cellulose II and III_I are known to have dipole moments aligned along the cellulose chain direction, similar to the 3,320 cm^{-1} peak in cellulose Iβ [50, 121]. Thus, the relative SFG intensities of these intra-chain hydrogen-bonded O-H groups can be correlated to the chain directionality in each polymorph using the noncentrosymmetry rule of the SFG process. In cellulose Iβ, all cellulose chains are arranged in the same (parallel) direction. In this structure, the dipoles of the O3H groups add up over the entire crystal, becoming SFG-active. In contrast, cellulose II shows a negligible O-H peak in SFG compared to the CH_2 peak. The absence (or weakness) of the OH peak whose dipole is along the chain direction indicates that the cellulose chains are aligned in an antiparallel fashion in the unit cell [29]. The ammonia-treatment does not affect the chain directionality of the starting material [73]. Thus, cellulose III_I has a strong OH peak, but cellulose III_{II} does not show any OH peak.

PERSPECTIVES

The incomplete understanding of cellulose structure is rooted in the complexity of its structure in its native state and the different forms into which it is transformed during isolation or pretreatment processes. These complexities are further amplified by the inadequacy of conceptual models and methodologies used for cellulose characterization. It is important to recognize that different analytical techniques probe the cellulose structure through completely different principles. Thus, they observe different aspects of the cellulose structure. Sometimes, in the absence of better scientific proof in the past, an old empirical model became a *convenient dogma* of the field. An example is the XRD crystallinity index calculation based on the peak heights [27,77]. It was originally proposed as an easy, time-saving estimation method in 1959, but it is often treated as an absolute value in recent literatures.

There is still a debate on the crystal structure of native celluloses. In his recent review paper [1], Atalla argued that the diffraction and spectroscopy analyses arrived at different conclusions about the repeat unit in the cellulose Iα and Iβ unit cells and how each glucose monomer is oriented. In XRD analysis, the $P2_1$ (two-fold screw axis) symmetry was assumed for the structural models for cellulose I before the discovery of Iα and Iβ duality in native celluloses [122]. And some diffraction peaks that did not fit well to the $P2_1$ symmetry model were omitted in the structural refinement. These analyses ended up with three different chain conformations: one chain in the Iα unit cell and two chains in the Iβ unit cell. But, Atalla claimed that both cellulose Iα and Iβ must have the same ring conformation. This proposal was made based on the fact that Raman spectra of cellulose Iα and Iβ show essentially the same ring vibration modes (peak positions are the same; only the relative intensities are different) and there was no peak splitting for cellulose Iβ although the peak splitting is expected for the two-chain unit cell structure model. The origin of paracrystalline cellulose and the presence of cellulose IV are still unclear [6,90,91,93]. Knowing the controversy of the structure of purified cellulose, it is not surprising

that current understanding of the crystal structure and spatial assembly of cellulose microfibrils in native plant cell walls is even less [9,100].

To advance our understanding on the structure of cellulose, in either purified or native state, it is very important to first know the basic principle of the analytical technique and take cautious steps when the conclusion from one analysis is extrapolated or generalized. The main focus of conventional diffraction and spectroscopy analyses is on the issues of *"repeating unit cells" corresponding to domains that are of the order of sub-nm in dimension*. To study the organization of these repeating unit cells at the next level in the hierarchy of assembly over 1 nm and 100's nm within plant cell walls and biomass, there are a number of new issues that have not been considered. One issue is the departures from a linear lattice model, which have been generally regarded as measures of disorder; they must originate from the hierarchical organization in a biological structure, rather than purely random crystal packing. The fact that significant fractions of the cellulose molecules are exposed at the surface of the microfibrils of native celluloses, particularly in the case of higher plants, is another important issue. These issues arise in relation to their biogenesis since the cellulose chains at the surface of microfibrils are in intimate contact with hemicellulose and other small molecules (including water), which will affect the thermodynamics of the crystallization process. Finally, the packing and assembly of cellulose microfibrils into three-dimensional 'fibrous composite structure' of cell walls are closely related to the cell growth through cell wall expansion and mechanical support at various developmental stages or living environments of plants [123,124]. They also have huge impacts on cell wall degradability through chemical and enzymatic processes, which is the main focus of many engineers aiming to produce biofuels and other useful chemicals from lignocellulosic biomass. The new developments of *nondestructive* analytical tools that can *reveal hierarchical structures of cellulose microfibrils in the range between 1 nm and 100's nm* can provide critical insights needed to explore the relationships between structure and reactivity of cellulose and lignocellulosic biomass. SFG could provide such information due to its phase matching condition over the optical coherence length scale. The availability of detailed characterization of the cellulose aggregation states should permit more systematic exploration of the influence of biological source and treatment history of cellulosic samples on their reactivity in chemical and biological processes [125,126].

Clearly, computational analyses would play critical roles in all aspects of cellulose characterization [32,127-129]. Computer simulations can calculate and predict fundamental properties that have not been tested experimentally. They can also calculate parameters or properties that complement the experimental findings or assist the interpretation of experimental data. To catalyze more reliable and meaningful computations, it is needed to develop more accurate potentials and force fields describing chemical bonds and molecular interactions in cellulose.

ACKNOWLEDGEMENT

This work was supported by the National Science Foundation (Grant No. CBET-1152824) and by the Center for Lignocellulose Structure and Formation, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, and Office of Basic Energy Sciences under Award Number DE-SC0001090. SHK and CML were primarily supported by DOE BES (DE-SC000 1090) and KK was partially supported by NSF (CBET-1152824).

REFERENCES

1. R. H. Atalla and A. Isogai, *Celluloses in comprehensive natural products ii: Chemistry and biology*, Elsevier Science (2010).
2. Y. Yu and H. Wu, *Ind. Eng. Chem. Res.*, **48**, 10682 (2009).
3. M. Gray, A. Converse and C. Wyman, "Sugar monomer and oligomer solubility," *Appl. Biochem. Biotechnol.*, B. Davison, J. Lee, M. Finkelstein and J. McMillan, Eds., Humana Press, 179 (2003).
4. A. C. O'Sullivan, *Cellulose*, **4**, 173 (1997).
5. C. Somerville, *Annu. Rev. Cell Dev. Biol.*, **22**, 53 (2006).
6. A. N. Fernandes, L. H., Thomas, C. M., Altaner, P. Callow, V. T. Forsyth, D. C. Apperley, C. J. Kennedy and M. C. Jarvis, *PNAS*, **108**, E1195 (2011).
7. S. Y. Ding and M. E. Himmel, *J. Agric. Food Chem.*, **54**, 597 (2006).
8. Y. Nishiyama, *J. Wood Sci.*, **55**, 241 (2009).
9. P. Sarkar, E. Bosneaga and M. Auer, *J. Exp. Bot.*, **60**, 3615 (2009).
10. C. T. Brett, *Int. Rev. Cytol.*, **199**, 161 (2000).
11. I. M. Saxena and R. M. Brown, *Ann. Bot.*, **96**, 9 (2005).
12. M. E. Himmel, S. Y. Ding, D. K. Johnson, W. S. Adney, M. R. Nimlos, J. W. Brady and T. D. Foust, *Science*, **315**, 804 (2007).
13. W. Zhao and A. Berg, *Lab on a Chip*, **2008**, 1988 (2008).
14. S. Pérez and D. Samain, *Advances in Carbohydrate Chemistry and Biochemistry*, **64**, 25 (2010).
15. S. Park, D. K. Johnson, C. I. Ishizawa, P. A. Parilla and M. F. Davis, *Cellulose*, **16**, 641 (2009).
16. C. Somerville, H. Youngs, C. Taylor, S. C. Davis and S. P. Long, *Science*, **329**, 790 (2010).
17. A. Carroll and C. Somerville, *Ann. Rev. Plant Biol.*, **60**, 165 (2009).
18. R. Atalla and D. VanderHart, *Solid State Nucl. Magn. Reson.*, **15**, 1 (1999).
19. W. L. Earl and D. L. VanderHart, *Macromolecules*, **14**, 570 (1981).
20. F. Horii, A. Hirai and R. Kitamaru, *Polym. Bull.*, **10**, 357 (1983).
21. Y. Nishiyama, P. Langan and H. Chanzy, *J. Am. Chem. Soc.*, **124**, 9074 (2002).
22. Y. Nishiyama, J. Sugiyama, H. Chanzy and P. Langan, *J. Am. Chem. Soc.*, **125**, 14300 (2003).
23. H. Marrinan and J. Mann, *J. Polym. Sci.*, **21**, 301 (1956).
24. J. Sugiyama, J. Persson and H. Chanzy, *Macromolecules*, **24**, 2461 (1991).
25. J. Mann and H. Marrinan, *J. Polym. Sci.*, **27**, 595 (1958).
26. A. L. Barnette, L. C. Bradley, B. D. Veres, E. P. Schreiner, Y. B. Park, J. Park, S. Park and S. H. Kim, *Biomacromolecules*, **12**, 2434 (2011).
27. A. L. Barnette, C. Lee, L. C. Bradley, E. P. Schreiner, Y. B. Park, H. Shin, D. J. Cosgrove, S. Park and S. H. Kim, *Carbohydr. Polym.*, **89**, 802 (2012).
28. C. M. Lee, A. Mittal, A. L. Barnette, K. Kafle, Y. B. Park, H. Shin, D. K. Johnson, S. Park and S. H. Kim, *Cellulose*, **20**, 991 (2013).
29. C. M. Lee, N. M. A. Mohamed, H. D. Watts, J. D. Kubicki and S. H. Kim, *J. Phys. Chem. B*, **117**, 6681 (2013).
30. P. Zugenmaier, *Crystalline cellulose and cellulose derivatives: Characterization and structures*, Springer (2008).
31. Y. Habibi, L. A. Lucia and O. J. Rojas, *Chem. Rev.*, **110**, 3479 (2010).

32. S. K. Cousins and R. M. Brown Jr., *Polymer*, **36**, 3885 (1995).
33. P. Hermans and D. Vermaas, *J. Polym. Sci.*, **1**, 149 (1946).
34. R. P. Swatloski, S. K. Spear, J. D. Holbrey and R. D. Rogers, *J. Am. Chem. Soc.*, **124**, 4974 (2002).
35. M. Zavrel, D. Bross, M. Funke, J. Büchs and A. C. Spiess, *Bioresour. Technol.*, **100**, 2580 (2009).
36. P. Conte, A. Maccotta, C. De Pasquale, S. Bubici and G. Alonzo, *J. Agric. Food Chem.*, **57**, 8748 (2009).
37. A. Isogai and R. H. Atalla, *Cellulose*, **5**, 309 (1998).
38. Y. Wang and Y. Deng, *Biotechnol. Bioeng.*, **102**, 1398 (2009).
39. W. Blaschek, H. Koehler, U. Semler and G. Franz, *Planta*, **154**, 550 (1982).
40. R. J. Viëtor, R. H. Newman, M. A. Ha, D. C. Apperley and M. C. Jarvis, *Plant J.*, **30**, 721 (2002).
41. R. H. Newman and J. A. Hemmingson, *Cellulose*, **2**, 95 (1995).
42. J. F. Matthews, M. Bergenstrahle, G. T. Beckham, M. E. Himmel, M. R. Nimlos, J. W. Brady and M. F. Crowley, *J. Phys. Chem. B*, **115**, 2155 (2011).
43. J. Hearle, *J. Appl. Polym. Sci.*, **7**, 1175 (1963).
44. A. Scallan, *Text. Res. J.*, **41**, 647 (1971).
45. R. H. Atalla and D. L. Vanderhart, *Science*, **223**, 283 (1984).
46. H. Yamamoto and F. Horii, *Macromolecules*, **26**, 1313 (1993).
47. P. Langan, Y. Nishiyama and H. Chanzy, *Biomacromolecules*, **2**, 410 (2001).
48. D. Ruan, L. Zhang, J. Zhou, H. Jin and H. Chen, *Macromol. Biosci.*, **4**, 1105 (2004).
49. R. Hori and M. Wada, *Cellulose*, **13**, 281 (2006).
50. M. Wada, L. Heux, A. Isogai, Y. Nishiyama, H. Chanzy and J. Sugiyama, *Macromolecules*, **34**, 1237 (2001).
51. M. Wada, H. Chanzy, Y. Nishiyama and P. Langan, *Macromolecules*, **37**, 8548 (2004).
52. M. Wada, L. Heux and J. Sugiyama, *Biomacromolecules*, **5**, 1385 (2004).
53. E. S. Gardiner and A. Sarko, *Can. J. Chem.*, **63**, 173 (1985).
54. H. Chanzy, K. Imada and R. Vuong, *Protoplasma*, **94**, 299 (1978).
55. F. Horii, A. Hirai and R. Kitamaru, *Polym. Bull.*, **10**, 357 (1983).
56. P. Langan, Y. Nishiyama and H. Chanzy, *J. Am. Chem. Soc.*, **121**, 9940 (1999).
57. G. Rappenecker and P. Zugenmaier, *Carbohydr. Res.*, **89**, 11 (1981).
58. Y. Waseda, *Xray diffraction crystallography*, Springer Verlag Berlin Heidelberg (2011).
59. G. Honjo and M. Watanabe, *Nature*, **181**, 326 (1958).
60. J. Sugiyama, R. Vuong and H. Chanzy, *Macromolecules*, **24**, 4168 (1991).
61. L. Segal, J. Creely, A. Martin and C. Conrad, *Text. Res. J.*, **29**, 786 (1959).
62. Y. Nishiyama, S. Kuga, M. Wada and T. Okano, *Macromolecules*, **30**, 6395 (1997).
63. R. Jenkins and R. Snyder, *Introduction to x-ray powder diffractometry*, Wiley-Interscience (1996).
64. R. Chandrasekaran, *Adv. Carbohydr. Chem. Biochem.*, **52**, 311 (1997).
65. S. Chu and G. Jeffrey, *Acta Crystallogr.*, **B24**, 830 (1968).
66. J. T. Ham and D. G. Williams, *Acta Crystallogr.*, **B26**, 1373 (1970).
67. K. H. Gardner and J. Blackwell, *Biopolymers*, **13**, 1975 (1974).
68. A. T. Briinger, *Nature*, **355**, 472 (1992).
69. K. Gessler, N. Krauss, T. Steiner, C. Betzel, A. Sarko and W. Saenger, *J. Am. Chem. Soc.*, **117**, 11397 (1995).
70. D. Picot, P. J. Loll and R. M. Garavito, *Nature*, **367**, 243 (1994).
71. K. Meyer and L. Misch, *Helu Chim. Acta*, **20**, 232 (1937).
72. A. D. French, *Cellulose*, **In press** (2013).
73. N.-H. Kim, T. Imai, M. Wada and J. Sugiyama, *Biomacromolecules*, **7**, 274 (2006).
74. A. Patterson, *Phys. Rev.*, **56**, 978 (1939).
75. R. P. Oliveira and C. Driemeier, *J. Appl. Crystallogr.*, **46**, 1196 (2013).
76. C. Driemeier and J. Bragatto, *J. Phys. Chem. B*, **117**, 415421 (2012).
77. S. Park, J. O. Baker, M. E. Himmel, P. A. Parilla and D. K. Johnson, *Biotechnol. Biofuels*, **3**, 1 (2010).
78. C. Driemeier and G. A. Calligaris, *J. Appl. Crystallogr.*, **44**, 184 (2010).
79. R. Drago, *Physical methods in chemistry*, W.B. Saunders Company, Philadelphia (1977).
80. A. E. Derome, *Modern nmr techniques for chemistry research*, Pergamon Books Inc., New York (1987).
81. R. Pohmann, "Physical basics of nmr," *In vivo nmr imaging*, Schröder, L. and Faber, C., Eds., Humana Press, 3 (2011).
82. M. J. Duer, *Solid state nmr spectroscopy: Principles and applications*, Wiley-Blackwell (2008).
83. D. C. Apperley, *Solid state nmr: Basic principles & practice*, Momentum Press, New York (2012).
84. D. D. Laws, H.-M. L. Bitter and A. Jerschow, *Angew. Chem. Int. Ed.*, **41**, 3096 (2002).
85. R. Pettifer, C. Brouder, M. Benfatto, C. Natoli, C. Hermes and M. R. Lopez, *Physical Review B*, **42**, 37 (1990).
86. J. W. Hennel and J. Klinowski, "Magic-angle spinning: A historical perspective," *New techniques in solid-state nmr*, Springer, 1 (2005).
87. E. R. Andrew, *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, **299**, 505 (1981).
88. H. Kono, S. Yunoki, T. Shikano, M. Fujiwara, T. Erata and M. Takai, *J. Am. Chem. Soc.*, **124**, 7506 (2002).
89. J. D. Kubicki, M. N.-A. Mohamed and H. D. Watts, *Cellulose*, **20**, 9 (2013).
90. P. T. Larsson, K. Wickholm and T. Iversen, *Carbohydr. Res.*, **302**, 19 (1997).
91. K. Wickholm, P. T. Larsson and T. Iversen, *Carbohydr. Res.*, **312**, 123 (1998).
92. R. H. Newman, *Holzforschung*, **58**, 91 (2004).
93. R. H. Newman, *Solid State Nucl. Magn. Reson.*, **15**, 21 (1999).
94. H. Yamamoto and F. Horii, *Macromolecules*, **26**, 1313 (1993).
95. R. H. Newman and T. C. Davidson, *Cellulose*, **11**, 23 (2004).
96. C. N. Banwell, *Fundamentals of molecular spectroscopy: 4e*, Tata McGraw-Hill Education (1994).
97. B. H. Stuart, *Infrared spectroscopy: Fundamentals and applications*, Wiley (2004).
98. J. H. Wiley and R. H. Atalla, *Carbohydr. Res.*, **160**, 113 (1987).
99. T. Imai and J. Sugiyama, *Macromolecules*, **31**, 6275 (1998).
100. M. Jarvis, *Nature*, **426**, 611 (2003).
101. R. H. Atalla, *IPC Technical Paper Series*, **19**, 1 (1975).
102. U. Agarwal, R. Reiner and S. Ralph, *Cellulose*, **17**, 721 (2010).
103. K. Schenzel, S. Fischer and E. Brendler, *Cellulose*, **12**, 223 (2005).
104. M. Åkerholm, B. Hinterstoisser and L. Salmén, *Carbohydr. Res.*, **339**, 569 (2004).

105. M. L. Nelson and R. T. O'Connor, *J. Appl. Polym. Sci.*, **8**, 1325 (1964).
106. R. Jeffries, *Polymer*, **4**, 375 (1963).
107. Y. Hishikawa, E. Togawa, Y. Kataoka and T. Kondo, *Polymer*, **40**, 7117 (1999).
108. R. Jeffries, *J. Appl. Polym. Sci.*, **8**, 1213 (1964).
109. K. Hofstetter, B. Hinterstoisser and L. Salmén, *Cellulose*, **13**, 131 (2006).
110. Y. Horikawa and J. Sugiyama, *Cellulose*, **15**, 419 (2008).
111. A. G. Lambert, P. B. Davies and D. J. Neivandt, *Appl. Spectrosc. Rev.*, **40**, 103 (2005).
112. Y. Shen, *Nature*, **337**, 519 (1989).
113. H.-F. Wang, W. Gan, R. Lu, Y. Rao and B.-H. Wu, *Int. Rev. Phys. Chem.*, **24**, 191 (2005).
114. S. A. Denev, T. T. A. Lummen, E. Barnes, A. Kumar and V. Gopalan, *J. Am. Ceram. Soc.*, **94**, 2699 (2011).
115. H. C. Hieu, N. A. Tuan, H. Li, Y. Miyauchi and G. Mizutani, *Appl. Spectrosc.*, **65**, 1254 (2011).
116. R. LaComb, O. Nadiarnykh, S. S. Townsend and P. J. Campagnola, *Opt. Commun.*, **281**, 1823 (2008).
117. L. Tian, J. Qu, Z. Guo, Y. Jin, Y. Meng and X. Deng, *J. Appl. Phys.*, **108**, 054701 (2010).
118. F. Vidal and A. Tadjeddine, *Rep. Prog. Phys.*, **68**, 1095 (2005).
119. C. T. Williams and D. A. Beattie, *Surf. Sci.*, **500**, 545 (2002).
120. G. L. Richmond, *Chem. Rev.*, **102**, 2693 (2002).
121. A. Šturcová, I. His, T. J. Wess, G. Cameron and M. C. Jarvis, *Biomacromolecules*, **4**, 1589 (2003).
122. A. D. French and G. P. Johnson, *Cellulose*, **16**, 959 (2009).
123. A. Neville, *BioEssays*, **3**, 4 (1985).
124. U. Kutschera, *Ann. Bot.*, **101**, 615 (2008).
125. W. Wang, X. Chen, B. S. Donohoe, P. N. Ciesielski, A. Mittal, R. Katahira, E. M. Kuhn, K. Kafle, C. M. Lee, S. Park, S. H. Kim, M. P. Tucker, M. E. Himmel and D. K. Johnson, Submitted to Biomass and Bioenergy (2013).
126. Y. B. Park, C. M. Lee, T. Zhang, B.-W. Koo, S. Park, D. J. Cosgrove and S. H. Kim, *Plant Physiol.* (http://www.plantphysiol.org/content/early/2013/08/30/pp.113.225235.abstract, Published online before print August 2013, DOI:http://dx.doi.org/10.1104/pp.113.225235 Plant Physiology August 2013 pp.113.225235) (2013).
127. A. P. Heiner, J. Sugiyama and O. Teleman, *Carbohydr. Res.*, **273**, 207 (1995).
128. K. Mazeau and L. Heux, *J. Phys. Chem. B*, **107**, 2394 (2003).
129. J. F. Matthews, C. E. Skopec, P. E. Mason, P. Zuccato, R. W. Torget, J. Sugiyama, M. E. Himmel and J. W. Brady, *Carbohydr. Res.*, **341**, 138 (2006).