Unique Aspects of the Structure and Dynamics of Elementary Iβ Cellulose Microfibrils Revealed by Computational Simulations[1][OPEN]

Daniel P. Oehme, Matthew T. Downton, Monika S. Doblin, John Wagner, Michael J. Gidley*, and Antony Bacic*

IBM Research Collaboratory for Life Sciences-Melbourne, Victorian Life Sciences Computation Initiative, Carlton, Victoria 3010, Australia (D.P.O., M.T.D., J.W.); IBM Research-Australia, Carlton, Victoria 3010, Australia (D.P.O., M.T.D., J.W.); Australian Research Council Centre of Excellence in Plant Cell Walls, School of Botany (M.S.D., A.B.) and Bio21 Molecular Science and Biotechnology Institute (A.B.), University of Melbourne, Parkville, Victoria 3010, Australia; and Australian Research Council Centre of Excellence in Plant Cell Walls and Centre for Nutrition and Food Sciences, Queensland Alliance for Agriculture and Food Innovation, University of Queensland, St. Lucia 4072, Australia (M.J.G.)

The question of how many chains an elementary cellulose microfibril contains is critical to understanding the molecular mechanism(s) of cellulose biosynthesis and regulation. Given the hexagonal nature of the cellulose synthase rosette, it is assumed that the number of chains must be a multiple of six. We present molecular dynamics simulations on three different models of Iβ cellulose microfibrils, 18, 24, and 36 chains, to investigate their structure and dynamics in a hydrated environment. The 36-chain model stays in a conformational space that is very similar to the initial crystalline phase, while the 18- and 24-chain models sample a conformational space different from the crystalline structure yet similar to conformations observed in recent high-temperature molecular dynamics simulations. Major differences in the conformations sampled between the different models result from changes to the tilt of chains in different layers, specifically a second stage of tilt, increased rotation about the O2-C2 dihedral, and a greater sampling of non-TG exocyclic conformations, particularly the GG conformation in center layers and GT conformation in solvent-exposed exocyclic groups. With a reinterpretation of nuclear magnetic resonance data, specifically for contributions made to the C6 peak, data from the simulations suggest that the 18- and 24-chain structures are more viable models for an elementary cellulose microfibril, which also correlates with recent scattering and diffraction experimental data. These data inform biochemical and molecular studies that must explain how a six-particle cellulose synthase complex rosette synthesizes microfibrils likely comprised of either 18 or 24 chains.

Cellulose is the main structural polysaccharide in higher plant cell walls and has been studied intensively since it was first isolated by Anselme Payen in 1838 (Payen, 1838; Zugenmaier, 2001). As a result, there is significant accumulated knowledge of the most abundant biological macromolecule on earth. Cellulose plays an important role in defining the physical and mechanical properties of plant cells, such as their growth, flexibility, structural support, and responses to stresses (biotic and abiotic). These properties vary depending on whether the cell/tissue is undergoing rapid growth and expansion or is undergoing differentiation following the cessation of growth. Thus, the primary wall laid down at division must yield to stress produced by turgor pressure, leading to cell expansion. The shape of the cell is determined by differential yielding of walls in specific locations. At the completion of expansion, the mechanical properties of the wall can change so that it no longer yields to turgor pressure, but is rigidified so that it can withstand large compressive forces. Despite this, the mechanism(s) involved in how cellulose confers these properties are not well understood. One reason stems from the lack of consensus on the fine structure and dynamics of the elementary microfibril, the minimal functional unit of cellulose.

Produced at the plasma membrane via a rosette of cellulose synthase (CesA) catalytic subunits (in plants) utilizing cytoplasmic UDP-Glc, cellulose is made up of layers of planar (1,4)-β-glucan chains, stacked upon each other to form microfibrils, whose directionality is controlled initially by the underlying cytoskeleton of microtubules on which the cellulose synthase tracks and which are usually transverse to the direction of turgor pressure-driven growth. These microfibrils can also form larger order aggregates that differ in their

size depending upon the species/tissue/developmental state. X-ray and neutron diffraction experiments performed on cellulose from algae, such as *Valonia* spp., and in tunicates have identified single crystalline aggregates being up to 40 nm in diameter (Thomas et al., 2013), with those from higher plants generally ranging between 2 and 5 nm (Newman, 1999; Fernandes et al., 2011; Newman et al., 2013) and thought to be composed of single microfibrils.

A number of different polymorphs have been identified for crystalline phases of cellulose. The native state is defined as cellulose I, with two allomorphs, β and α, that differ slightly in their unit cells. The β form dominates in plants and the α form in bacteria and algae, although both the β and α forms can be found from the same source and possibly in the same microfibril (Atalla and Vanderhart, 1984; VanderHart and Atalla, 1984; Horii et al., 1987; Sugiyama et al., 1991a, 1991b).

The crystal and molecular structure of cellulose has been determined at atomic resolution for both the Iβ (Nishiyama et al., 2002) and Iα allomorphs (Nishiyama et al., 2003) using x-ray and neutron diffraction. Iβ cellulose crystals were found to have the monoclinic P2$_1$ space group, with two nonidentical chains arranged in a parallel manner. Iα cellulose crystals were found to have the triclinic P1 space group, with parallel chains of identical conformation. For both allomorphs, hydrogen bonding was reported within chains and between chains of the same layer, but not between layers. Additionally, the Iβ form is proposed to have two differing H-bond patterns, A and B, which are dependent on the position of hydrogens HO2 and HO6 (Fig. 1). The conformation of the exocyclic group at carbon C6 was also reported to reside in the TG conformation (Fig. 2).

Numerous ^{13}C-NMR studies have been performed to investigate the structure of cellulose microfibrils and
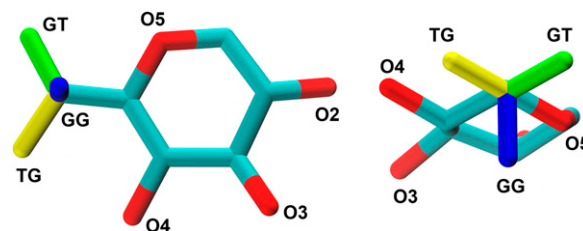
Figure 2. Rotation about the O5-C5-C6-O6 dihedral allows for three different conformations of the C6 exocyclic group of each Glc monomer. The C6-O6 bond is colored yellow for the TG conformation (dihedral angle of 180°), green for GT (60°), and blue for GG (–60°). Hydrogens are not shown.

aggregates (Earl and VanderHart, 1981; Horii et al., 1983, 1987; Isogai and Usuda, 1989; Newman and Hemmingson, 1994; Larsson et al., 1997; Ha et al., 1998; Wickholm et al., 1998; Viëtor et al., 2002; Larsson and Westlund, 2005; Park et al., 2009; Malm et al., 2010). Peaks attributed to the C4 and C6 carbons suggest that surface (amorphous) and interior (crystalline) chains have differing chemical shifts due to differences in either their environment, structure, or a combination of both (Viëtor et al., 2002). Data from the C6 peaks have also suggested that the exocyclic group for interior chains reside in the TG conformation, with this preference changing at the surface of the microfibril, where the GT and GG conformations are preferred (Viëtor et al., 2002). These conclusions were reached as a result of the earlier work of Horii et al. (1983) in which NMR studies on a range of oligosaccharides whose crystal structures had previously been solved identified three separate C6 peaks at 60 to 62, 62.5 to 64.5, and 65.6 to 66.6 ppm that corresponded to the GG, GT, and TG exocyclic conformations, respectively. Spectral fitting of NMR spectra has also been performed to deconvolute peaks and extract ultrastructural information about microfibril aggregates, such as crystalline, paracrystalline, surface, and allomorph proportions (Larsson et al., 1997; Wickholm et al., 1998; Larsson and Westlund, 2005).

Based upon these physical and ultrastructural measurements, a number of different models have been proposed for an elementary cellulose microfibril. A 36-chain, hexagonal-shaped model was suggested by Ding and Himmel (2006) and thought to be an optimal model from transmission electron microscopy and atomic force microscopy studies (Ding et al., 2014), because the cellulose synthase rosette complex from which cellulose is produced in plants is thought to be composed of 36 cellulose synthase catalytic proteins (Herth, 1983). Other 36-chain complexes have been suggested that have a square shape, in which the hydrophobic (100) and hydrophilic (010) faces are exposed to solvent (Mazeau, 2005; Matthews et al., 2006; Bergenstråhle et al., 2007), and a diamond-like shape where two hydrophilic (110 and 1-10) faces are exposed (Matthews et al., 2006; Zhang et al., 2011; Chen et al., 2014). However, there is now evidence from
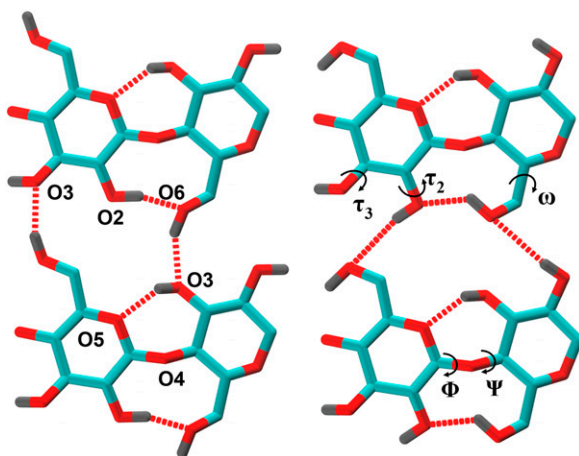

Figure 1. Cellobiose dimers in the A (left) and B (right) hydrogen bonding patterns. Oxygen (red) atoms are labeled in the left dimer, while key dihedral angles are labeled in the right dimer. Hydrogens (gray) are displayed if involved in hydrogen bonds. Otherwise, they are omitted.

wide-angle x-ray scattering (WAXS) and small-angle neutron scattering studies to suggest that 36 chains in a microfibril may be an overestimation (Fernandes et al., 2011; Thomas et al., 2013). Computer-simulated diffractograms for 18- and 24-chain models, with and without constant or regular shape, show a much better fit to the experimental data than the 36-chain model. Newman et al. (2013) also suggest that there are only six layers in the 100 plane of microfibrils, which differs from eight layers in the Ding and Himmel (2006) 36-chain model.

Thus, the structure of an elementary cellulose microfibril is yet to be determined unequivocally. This is related to the significant experimental challenge of characterizing small-diameter cellulose (Matthews et al., 2012). X-ray and neutron diffraction studies have given us models of crystalline cellulose, but these are generally acquired at nonnative temperatures and environments. NMR gives information on microfibril aggregates in a more natural environment through solid-state techniques. These techniques allow spectra to be produced with water contents similar to those found in the primary cell wall of plants (Brett and Hillman, 1985) and provide a bridge between structures determined by diffraction techniques and those which exist in solution (Zugenmaier, 2001). However, to understand the natural structure and dynamics of microfibrils, how they interact with other microfibrils and components of the cell wall, and their chemical reactivity, we must be able to investigate single microfibrils at the atomic level in an environment as close to solution as possible, considering that in primary

walls the aqueous phase can comprise 60% to 70% by weight (Doblin et al., 2010).

One alternative is to use computational approaches, with numerous studies published in which molecular dynamics (MD) simulations have been performed on cellulose microfibrils, generally with a 36-chain model (Mazeau, 2005; Matthews et al., 2006, 2011a, 2011b, 2012; Bergenstråhle et al., 2007; Gross and Chu, 2010; Zhang et al., 2011; Chen et al., 2014). These simulations have investigated the structure and dynamics of the microfibrils, temperature dependence, H-bonding patterns, and effect of solvent water, and compared different carbohydrate force fields. Simulations have also been performed in different environments, whether that be purely crystalline or fully solvated. Additionally, different-shaped microfibrils have been simulated, as have microfibrils of finite or infinite length. We are only aware of one study where MD simulations have been performed on 18- or 24-chain models of cellulose (Busse-Wicher et al., 2014).

In this study, we investigate the effect that different numbers of chains have on the structure and dynamics of a cellulose microfibril with the aim of obtaining a better understanding of how many chains make up an elementary cellulose microfibril. Considering the Iβ allomorph, we perform MD simulations on infinite-length microfibrils for long time scales using the Charmm carbohydrate force field. All microfibrils are built with the hexagonal shape suggested by Ding and Himmel (2006) such that the 1-10, 110, and 100/200 faces are exposed (Fig. 3A). This shape was chosen, as NMR data suggests that two hydrophilic faces (1-10

Figure 3. Initial conformations for the 36-chain (A), 24-chain (B), and 18-chain (C) microfibrils, viewed from the nonreducing end. In A, the different hydrophobic (100/200) and hydrophilic (110/1-10) surfaces are labeled, as are the spacings between the different layers and the unit cell vectors. Note that the c vector runs into the page, along the polymerization axis. Atoms are colored red for oxygen, cyan for carbon, and gray for hydrogen. C6 and C4 carbons are highlighted in green and purple, respectively, for one surface chain. The inset to A shows the environment directly surrounding this chain, with an interior chain to the left and water solvent to the right. Viewed from above the microfibril, this highlights that the two C4 carbons of a cellobiose repeat are in the same environment, while the two C6 carbons are not.

and 110) are exposed (Wickholm et al., 1998), while hydrophobic faces (100/200) have been included to correlate with the hexagonal shape of cellulose synthase rosettes. This also satisfies the requirement of the microfibril having a hydrophobic face with multiple chains for cellulose binding domains to interact with (Tormo et al., 1996; Lehtiö et al., 2003). From this work, we can define the dynamic characteristics of a microfibril containing 18, 24, or 36 chains in a natural environment with regards to exocyclic group conformation, O2 dihedral, H-bond patterns, tilt of chain with respect to the polymerization axis, and unit cell dimensions. We also obtain an understanding of how the different number of chains and, as a result, different surfaces affect the structuring of water around the microfibril. These insights will be useful for further investigating the role that cellulose plays in the structure and dynamics of the plant cell wall, particularly how it interacts with other microfibrils and the matrix phase polysaccharides during growth and development. This knowledge may also lead to insights into the underlying causes of the recalcitrance of plant walls to biotechnological utilization and also the mechanism(s) of cellulose synthesis.

RESULTS

MD simulations were performed to investigate the effect that differing numbers of chains had on the dynamics and structure of cellulose microfibrils in a hydrated environment. Simulations were performed on models of Iβ cellulose containing 18, 24, and 36 chains. The initial structures for each microfibril looking along the polymerization axis from the nonreducing end are shown in Figure 3, and the final simulation structures are shown in Figure 4. The work of Matthews et al. (2012) has shown that significant changes to the microfibril can occur over long time scales, and thus long MD simulations up to 640 ns were performed. Limited changes to microfibril structure and dynamics

were observed after 200 ns of simulation (Supplemental Fig. S1), so analysis was limited to this time scale given the computational requirements of such simulations.

The largest system considered in this work contained 36 chains in a 3, 4, 5, 6/6, 5, 4, 3 configuration (Fig. 4, A and B). This microfibril had 50% of its chains in the interior and 50% at the surface. Despite this, only 33% of exocyclic groups could be considered solvent exposed. This value was significantly less than 50% because, except for solvent-exposed chains on hydrophobic surfaces, every second exocyclic group in solvent-exposed chains points toward the center of the microfibril instead of out into the solvent. With an initial cross-sectional area of just under 1,200 Å^2, the initial water content of the system was 77%. After equilibration, the water content came down to 73% due to a decrease in the total volume of the system and expansion of the microfibril (Table I). The next system considered consisted of 24 chains in a 3, 4, 5/5, 4, 3 arrangement (Fig. 4C). With 58% of its chains at the surface and 42% of exocyclic groups considered solvent exposed, the equilibrated water content was 71%. The smallest system simulated consisted of 18 chains in a 2, 3, 4/4, 3, 2 configuration, with an initial cross-sectional area of 590 Å^2 (Fig. 4D). This system had 44% of its exocyclic groups solvent exposed (Table I), and its equilibrated water content was 74%.

A characteristic feature of the simulations was the tilting of the glucan chains in the hexagonal structures. Chain tilt occurred from the beginning of our simulations, consistent with observations in other MD simulations of cellulose using different force fields, crystalline models, and finite models (Bergenståhle et al., 2007; Matthews et al., 2011a, 2012; Zhang et al., 2011; Chen et al., 2014). When microfibrils were viewed from the nonreducing end, origin layers had a clockwise tilt (positive) compared with the crystal structure, while the center layers had an anticlockwise tilt (negative; Fig. 4, blue and red colored, respectively). The 36-chain model experienced tilts of −4° and 7.5° for the center and origin layers, respectively, while
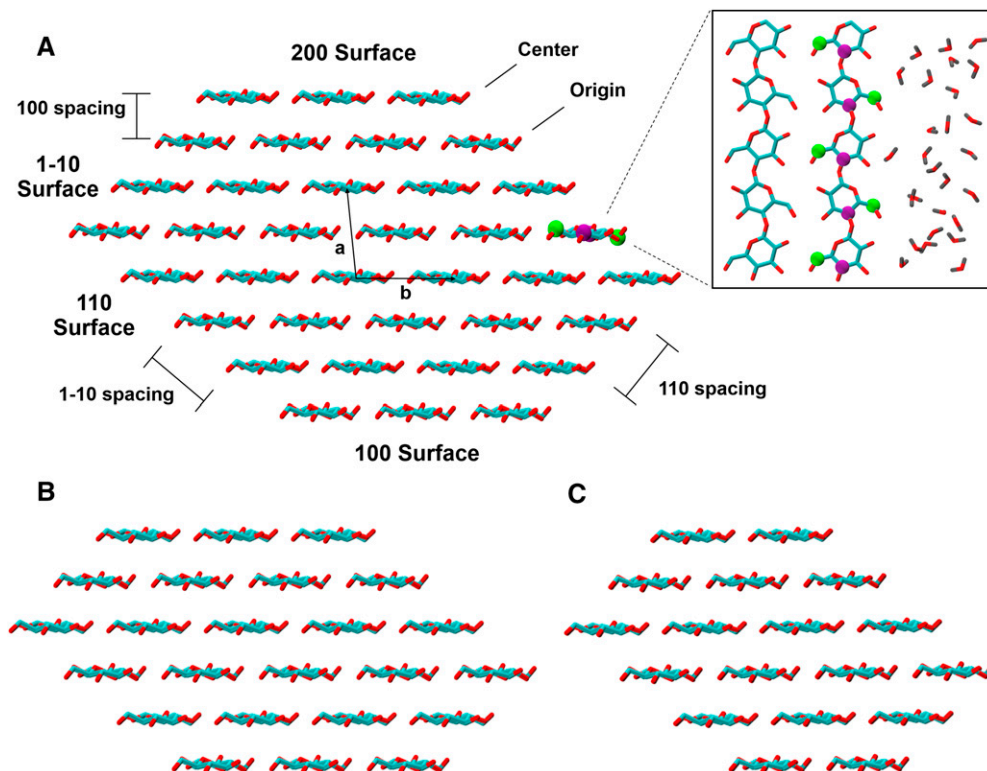
Figure 4. Final conformations for the 36-chain (A), 36-chain (B), 24-chain (C), and 18-chain (D) microfibrils, viewed from the nonreducing end. Layers are colored red for center and blue for origin.
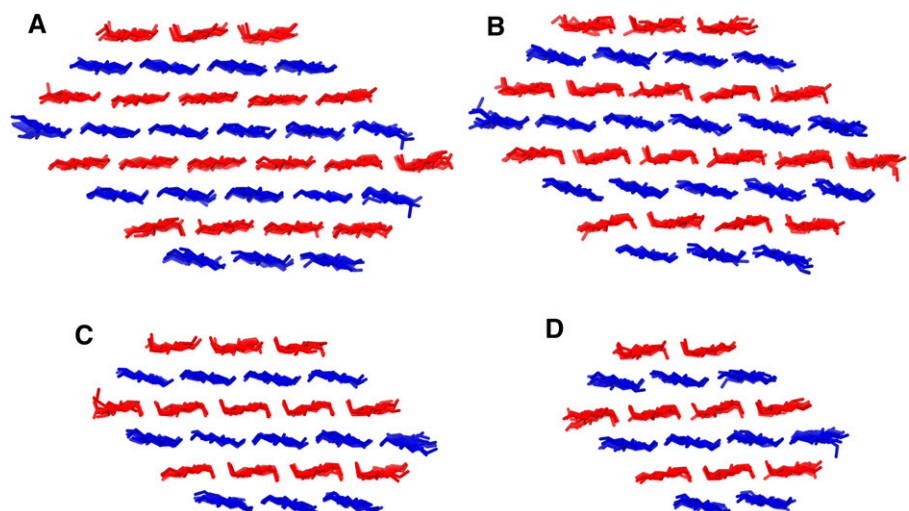
Table I. *Microfibril properties*

Simulation	Interior	Surface	Solvent[a] Exposed	Initial		Equilibrated	
				Cross[b] Section	Water Content	Cross[b] Section	Water Content
		%		A^2	v/v%	A^2	v/v%
18	33	67	44	588.2	80.2	680.2	74.0
24	42	58	42	769.8	76.6	858.1	71.1
36 (A)	50	50	33	1,188.6	76.7	1,261.3	73.3
36 (B)						1,279.4	72.9

[a]Solvent-exposed exocyclic groups, assuming that all exocyclic groups on hydrophobic surfaces are solvent exposed. [b]Cross-sectional area in the *xy* plane.

the 18- and 24-chain models experience larger changes in tilt of approximately ± 10° (Table II).

In addition to the tilting of the chains within the microfibrils, significant changes also occurred to the conformation of exocyclic groups. This is highlighted in Figure 5, which separately tracks the changes to the exocyclic conformation occupancy for groups pointing to the left and right, when looking down the polymerization axis from the nonreducing end, over the length of the simulations for each chain. Additionally, Supplemental Figure S2 shows the exocyclic conformation for all Glc residues of the microfibril at different time points throughout the simulations. It was apparent that the first changes occur to exocyclic groups that were exposed to solvent, with the changes occurring during equilibration. For all microfibrils, there was a preference for the GT conformation to dominate for solvent-exposed groups, although there was a high degree of variation (Table III). For the 36-chain model, the GT:TG:GG ratio for solvent exposed groups was 50:35:15. This was similar for the 24- and 18-chain models, except that the GG proportion increases and TG decreases. Interestingly, it was found that there were different exocyclic conformation ratios for the two different hydrophilic surfaces, with the 110 surface having a higher occupancy for GT and lower occupancy for GG (Supplemental Table S1). The TG conformations were of similar occupancy for the two surfaces.

Analysis of the exocyclic groups that were not solvent exposed indicates that the differing layers can sample different conformations. In origin chains, the TG conformation dominates, with occupancies greater than 80%. However, it was in the center chains where the relationship between an increase in GG sampling (and a corresponding decrease in TG) with decreasing numbers of chains was most evident. Simulations with the 36-chain model had a TG occupancy of 83% for center chains, while in the 18- and 24-chain simulations, significant changes were observed and the GG conformation now dominated with occupancies greater than 90% (Table III). The variability in interior chains was significantly reduced compared with the surface-exposed groups; however, it does appear to be smaller for chains closer to the center of the microfibril and when there were more chains in the microfibril.

Apart from the dihedral angles that dictate the conformation of exocyclic groups, there were four other key

dihedral angles. The φ and ψ angles determine the planarity of the β-glucan chains, while the O2-C2 (τ_2) and O3-C3 (τ_3) dihedrals affect possible interactions between the chains in the same layer of microfibrils (Fig. 1). In the crystal structure, the φ and ψ angles differed for origin and center chains, with φ and ψ being –88.7° and –147.1° for center chains and –98.5° and –142.3° for origin chains, respectively (Supplemental Tables S2 and S3). From the simulations, φ and ψ converge for the different layers to an average of –93° and –150°, respectively. This was consistent irrespective of whether a chain was in the origin or center layer or was solvent exposed or located in the interior. Given that we simulated an infinite microfibril, it was expected that the φ and ψ angles would not change markedly to keep the planar arrangement of chains. However, it was unexpected that they would equilibrate to the same value considering the differences that had been observed in the crystal structures.

From neutron diffraction studies (Nishiyama et al., 2002), it was suggested that hydrogens bonded to O2 and O3 were oriented in opposite directions (trans compared with the C2-C3 bond), and thus the initial dihedral angles about these O-C bonds were close to 180°. This was in contrast to a more cis-like conformation where the angle would be approximately 60° (Fig. 1). Similarly to the φ and ψ dihedrals, there was little variation about the τ_3 dihedral when compared with the crystal structure, with an average angle of approximately 161° for all simulations (Supplemental Table S4). This lack of variation resulted from the stabilizing O3-to-O5 intrachain H bond (see below).

Table II. *Degree of change of tilt for center and origin chains averaged over the last 50 ns of duplicate simulations*

Positive number indicates clockwise tilt, and negative indicates anticlockwise tilt.

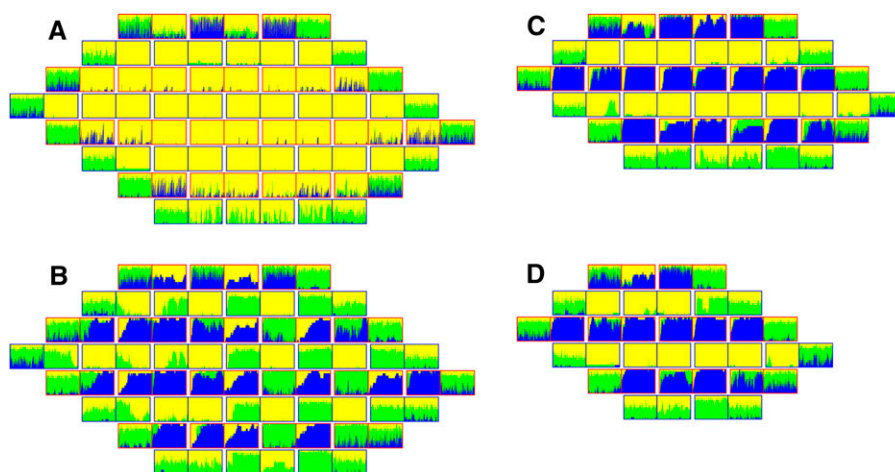| Simulation	Center	Origin
		°
18	−9.29	10.01
24	−10.43	9.06
36 (A)	−4.26	7.67
36 (B)	−7.92	10.73
18 (O2-C2)	−2.27	6.26

Figure 5. Plots of the exocyclic group conformation occupancies (TG in yellow, GG in blue, and GT in green) of each chain in the 36-chain (A), 36-chain (B), 24-chain (C), and 18-chain (D) models over the length of 200 ns simulations. Chains from center layers have plots with a red border, and origin layers have blue borders. Due to the different environments for the two exocyclic groups of each cellobiose unit, each chain has two plots; the left plot details those exocyclic groups that point to the left when looking down the microfibril polymerization axis from the nonreducing end, and the right plot details those that point to the right. Plots on the outside for each model represent exocyclic groups exposed to solvent, and these plots have high variance (as noted by the spiky nature of the plots) and sample all three conformations (all three colors are distinctly visible). For all other exocyclic groups, including those from surface chains but not exposed to the solvent, little variation is observed and generally one conformation (one color) dominates.

There were far greater changes observed for the τ_2 dihedral angle (Table IV). For the 36-chain simulations, the interior chains maintained their trans-like conformation (approximately 158°), while there was a drop in the angle for surface exposed chains (approximately 100°), suggesting that cis conformations were being sampled more regularly. For the 18- and 24-chain simulations, there were further reductions in the dihedral angle, most significantly for center chains. The dihedrals here were less than 60°, which suggests that only cis conformations were being sampled. Reductions in the dihedral angle of origin chains, though to a lesser extent for the 24-chain model, suggest more cis-like conformations were being sampled.

To understand how the changes in the τ_2 dihedral allow for the sampling of different exocyclic conformations, we also analyzed changes to the H-bonding patterns. For all these systems, the simulations were started with the microfibrils in the A H-bonding pattern. In this pattern, intrachain H bonds are formed where O2 donates its hydrogen to O6 and O3 to O5

(Fig. 1). Intralayer H bonds exist from O6 to O3, with those from O6 to O2 only found for center layers, while there were no interlayer H bonds.

The 36-chain microfibril showed limited changes to its H bonding throughout the simulations and can thus be thought to only sample the A H-bonding pattern (Table V). The O3-to-O5 intrachain H bonds continue to have strong occupancy, as does the O2-to-O6 intrachain H bonds. The O6-to-O3 intralayer H bond was observed to have high occupancy, coexisting with the O6-to-O2 intralayer H bond that formed in origin chains to a similar moderate level as those in center chains. The O6-to-O2 intrachain and O2-to-O6 intralayer H bonds are representative of the B H-bond pattern and were sampled to a low extent, generally due to chains at the surface of the microfibril (Supplemental Fig. S3). There were only sporadic interlayer H bonds, mainly from surface chains to interior chains (Supplemental Table S5).

The 18- and 24-chain models sample very similar H-bonding patterns to each other that can be classified

Table III. *Exocyclic conformation occupancies, expressed as a percentage, averaged over the last 50 ns of duplicate simulations*

Chain	18			24			36 (A)			36 (B)			18 (O2-C2)		
	GG	GT	TG	GG	GT	TG	GG	GT	TG	GG	GT	TG	GG	GT	TG
Surface	23	54	23	24	52	24	17	49	34	21	52	27	8	19	72
Interior	46	12	42	47	7	47	7	2	91	39	28	33	0	2	98
Center	92	7	2	93	6	1	14	3	83	77	18	4	1	3	96
Origin	0	17	83	0	8	92	0	1	99	0	38	62	0	0	100
Total	36	30	34	37	26	37	10	18	72	33	36	31	4	9	87

Table IV. *O2 dihedral angles from diffraction data and averaged over the last 50 ns of duplicate simulations*

Simulation	Center	Origin	Surface	Crystal	Total
			°		
Iβ (Crystal)	166.1	165.3			165.7
18	58.9	111.8	85.3	90.9	85.4
24	56.3	124.7	81.6	100.6	90.5
36 (A)	117.3	140.5	100.2	157.6	128.9
36 (B)	52.8	83.9	76.4	60.2	68.3
18 (O2-C2)	172.3	173.2	172.5	173.2	172.7

as being in both the A and B H-bonding patterns (Table V). As was expected, the O3-to-O5 intrachain H bond has very high occupancy. With regard to H bonds that were representative of the A pattern, the O2-to-O6 H bond had similar occupancy to the 36-chain model for origin chains but had almost no occupancy in center chains. This relationship was also observed for other A intralayer H bonds where the O6-to-O3-and-O2 occupancies for the origin chains were similar to those of the 36-chain microfibril, while those in center chains were nonexistent. Instead, in the center chains, occupancy was greater for those H bonds that represent the B pattern, such as the intrachain O2-to-O6 and intralayer O2-to-O6 H bonds. Interestingly, there was some occupancy in the origin chains for these B-pattern H bonds. In what was a significant difference to the 36-chain simulations, there were significant interlayer H bonds for the 18- and 24-chain models, specifically for the O6-to-O2 1-10 layer, with O6 in a center chain and O2 in an origin chain (Supplemental Table S5). This type of H bond was only found to occur when the exocyclic group was in the GG conformation.

Initial expansion of the microfibrils can be evidenced by limited increases to the unit cell parameters (Table VI). The majority of changes were below 2%, with the exception being the a-unit cell length (Fig. 3), which could differ by almost 5%. This discrepancy results from a change in the distance between 200/100 layers. In the crystal form, this distance was 7.78 Å, while from the simulations, the distance increases to 7.94 Å

for the 36-chain model and 8.15 Å for the 18- and 24-chain models. This increase correlates with WAXS data (Newman et al., 2013), where the lattice spacing between (200) planes increased to 0.41 nm (8.2 Å between layers). The largest change in unit cell angle parameters occurs for the γ angle and represents a slight slippage between 200 layers; the top layer moves to the left and the bottom layer to the right when viewed from the nonreducing end.

More information about the changes that occur to the structure of the microfibril was extracted from the simulations by calculating changes in the distance between hydrophobic (100/200) and hydrophilic (1-10/110) layers (Table VII). The average 100/200 distance increased by 0.1 Å (2.6%) for the 36-chain model and by 0.2 Å (4.7%) for the 18- and 24-chain models, whereas distances between 110 layers increased by under 0.1 Å, with the greatest change (>0.2 Å) occurring for the 1-10 layers. The extra increase for the 1-10-layer distances was not unexpected given the increase in the γ unit cell angle. These average layer spacings correlate as well, if not better, than the crystal structure with WAXS (Newman et al., 2013) and x-ray diffraction (Nishiyama, 2009) data, where spacings were calculated to be 4.1, 5.6, and 5.6/6.0 Å for the 100/200,110, and 1-10 planes, respectively. Overall, the changes in these distances, as well as the changes to the unit cell parameters, present themselves as an anticlockwise rotation of the microfibril about the polymerization axis (looking from the nonreducing end) compared with the crystal structure. Additionally, we have measured the change in the distance along the polymerization axis of an origin chain compared with a center chain. This gives a measure of slippage of chains along the polymerization axis, and from these simulations, an average change of 0.17 Å for the 36-chain simulation was observed, while this increased to 0.26 Å for the 18- and 24-chain simulations. Given that there was little change in the α- and β-unit cell angles, the slippage described here results solely from slippage between origin and center chains and not from slippage across the entire microfibril.

Table V. *Occupancy of hydrogen bonds averaged over the last 50 ns of duplicate simulations*

Simulation	Chain	Intralayer			Intrachain		
		O6→O3[a]	O6→O2[a]	O2→O6[b]	O3→O5	O2→O6[a]	O6→O2[b]
				%			
18	Center	3	2	89	86	1	2
	Origin	58	24	39	89	35	25
24	Center	4	2	87	89	1	19
	Origin	68	3	29	90	49	18
36 (A)	Center	58	28	2	87	6	1
	Origin	87	41	1	91	7	13
36 (B)	Center	1	0	91	88	1	23
	Origin	25	6	8	91	12	4
18 (O2-C2)	Center	71	37	0	8	84	0
	Origin	98	44	0	88	9	0

[a]A-type H bond. [b] B-type H bond.

Table VI. *Unit cell distance and angles from diffraction studies and averaged over the last 50 ns of duplicate simulations*

Chains	a	b	c	α	β	γ
			Å		°	
Iβ (Crystal)	7.78	8.20	10.38	90.00	90.00	96.50
Iβ (203°C)[a]	8.19	8.18	10.37	90.00	90.00	96.40
18	8.15	8.25	10.44	91.52	89.21	98.47
24	8.14	8.25	10.44	90.70	90.20	97.83
36 (A)	7.94	8.35	10.44	90.41	90.35	98.41
36 (B)	8.06	8.40	10.43	93.77	90.00	97.79
18 (O2-C2)	7.95	8.33	10.45	90.24	90.43	98.46

[a]Wada et al. (2010).

The final characteristic measured was the degree of water structuring surrounding the microfibrils. Densities were measured above all faces as a function of distance away from the layer. The water structuring was very similar for each model, and for the most part, there appeared to be two peaks in each layer's density plots. However, it was evident that water was structured differently above the different layers (Fig. 6A; Supplemental Figs. S4A–S6A). Water was least structured above the 1-10 surface, as evidenced by the much smaller initial peak in the density curves. Increased structuring was observed over the 110 surface, with the most significant structuring occurring over the hydrophobic (100/200) surfaces. With the densities measured starting from the heavy atoms that were most exposed to the surface (Fig. 6B), it was apparent that there was water density preceding these heavy atoms for the hydrophilic surfaces. This was due to the apparent tilting of the chains at these surfaces, which allowed water molecules to position themselves in the gaps between the center and origin layers (Supplemental Fig. S7). By analyzing the densities starting at the middle of the surface layers, we can deduce that the initial primary peaks in density start between 3.5 and 4.5 Å away from the middle of the layers. Additionally, minima in density between the primary and secondary peaks occur between 5 to 6 Å.

Another approach to analyze the water structuring is to measure the density as a function of distance along the polymerization axis (Fig. 6C; Supplemental Figs. S4C–S6C). From these plots, it was notable that, for the hydrophobic surfaces, there were two peaks per cellobiose repeat, while there were four for the hydrophilic surfaces. These peaks coincide with glycosidic linkages exposed on the surface. The reason that four peaks were found for the hydrophilic surfaces was that these surfaces were composed of both center and origin layers, and because the center layers were displaced by approximately 2.6 Å compared with the origin layers, we see peaks with this separation.

From the analysis of these simulations it was clear that the 18- and 24-chain microfibrils were sampling a different conformational space to the 36-chain microfibril. This conformational space appeared to be similar to high-temperature structures (I-HT) simulated by

Matthews et al. (2011a). Subsequently, a question that needed to be answered was what is causing these changes in the conformation of the microfibrils. While it appeared that the number of chains in the microfibril was having an effect on the simulations, it was thought it would be informative to pinpoint the changes that occurred in the 18- and 24-chain systems yet did not occur in the 36-chain model. To investigate this, two further sets of simulations were performed; one with a different initial H-bonding scheme and the other with the τ_2 dihedral restrained to its initial value.

36-Chain B

As noted above, a major difference between surface-exposed and interior chains for the 36-chain simulations and throughout the microfibril for simulations with smaller numbers of chains was the τ_2 dihedral angle. There was a dramatic change such that the O2 hydroxyl no longer sat in the crystal-like conformation pointing away from O3 of the same residue but now pointed toward it. This conformation was similar to one sampled by Matthews et al. (2011a), who found that at high temperature, and when simulating for long time scales, this τ_2 dihedral would change. Moreover, this also aligns with some uncertainty in the diffraction data of Nishiyama et al. (2002) for which two potential H-bond schemes were suggested.

We therefore performed simulations on the 36-chain model with the B H-bond scheme (Fig. 1), based on the proposed structures from the x-ray diffraction work (Nishiyama et al., 2002). In these simulations, we changed the τ_2 dihedral such that the simulations started with the O2 hydroxyl group pointed toward the O3 atom of the same Glc residue (i.e. the cis instead of the trans orientation). We also rotated the position of the HO6 (hydrogen of exocyclic hydroxyl group) so that there were no steric clashes, while the exocyclic group was kept in the TG conformation. These changes resulted in the O2-to-O6 H bond being the only interchain H bonds existing at the beginning of the simulation.

Many of the dynamic characteristics of this model were similar to those of the 18- and 24-chain models. The tilt of chains was very similar (Table II), as were the unit cell parameters (Table VI) and the distances between layers (Table VII). The τ_2 dihedral from center and solvent-exposed chains was similar to those seen

Table VII. *Distance between layers and chain shift along the polymerization axis from diffraction data and averaged over the last 50 ns of duplicate simulations*

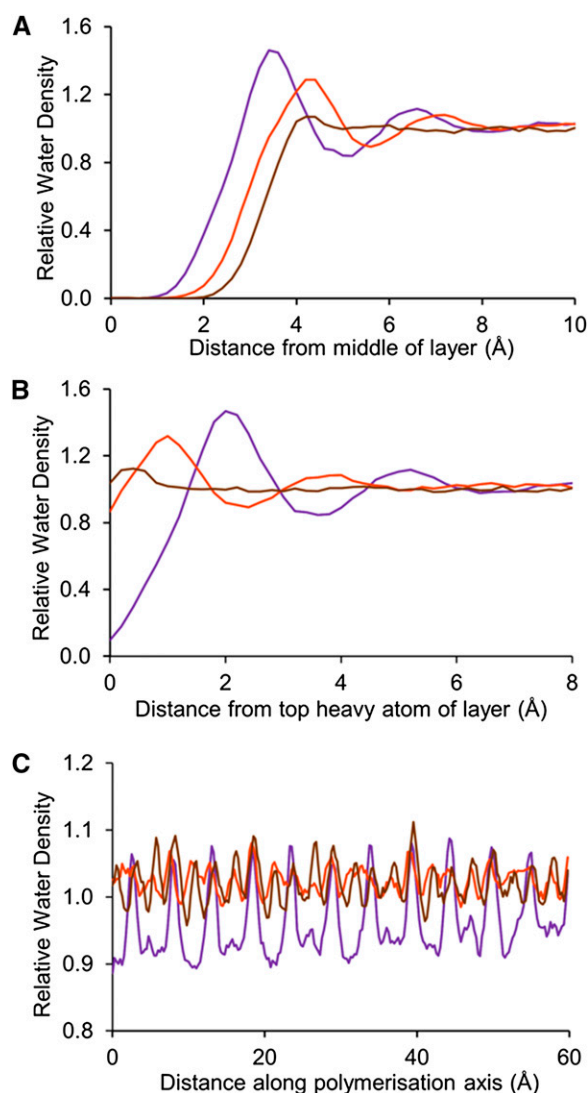| Chains	100	110	1-10	Z
			Å	
Iβ (Crystal)	3.84	5.31	5.97	2.71
18	4.01	5.39	6.23	2.96
24	4.04	5.40	6.18	2.98
36 (A)	3.94	5.36	6.22	2.88
36 (B)	4.00	5.42	6.21	2.93
18 (O2-C2)	3.83	5.34	6.18	2.89

Figure 6. Relative water densities about the 100 (purple), 110 (orange), and 1-10 (brown) layers for the 36 A simulations as a function of distance above the top heavy atom of the layer of interest in A, above the middle of the layer in B, and along the polymerization axis in C.

for previous simulations; however, the origin and interior chains had greatly reduced dihedrals and were therefore much more cis like (Table IV). These chains did not revert to the trans-like conformation of the A H-bonding pattern model as could be expected given the stability shown in the 36 A H-bond simulations.

As expected, by starting these simulations in a different H-bond pattern, an entirely different sampling of H bonds and exocyclic group conformations was observed when compared with the 36 A simulations (Table III; Fig. 5). However, the conformations sampled did resemble those sampled with the 18- and 24-chain models, though with some small differences. There was still some A character to origin chains for the O2-to-O6 intrachain and O6-to-O3 intralayer H bonds, but this was significantly reduced. It was

also noted that the occupancies for H bonds that are typical of the B pattern were far higher. Importantly, interlayer H bonds were observed between O6 and O2 for 1-10 layers at a similar occupancy to the 18- and 24-chain models. With regards to exocyclic group conformation, center chains had a greater likelihood to sample GG, while solvent-exposed chains more often sampled GT. Interestingly, origin chains were able to sample the GT conformation to a greater extent, although TG was still the preferred exocyclic conformation for these chains.

18-Chain τ_2 Restrained

From the simulations performed thus far, three major events took place to move the microfibril from the crystal phase to the so-called I-HT phase in the 18, 24, and 36 B simulations. These were the opposing tilts of center and origin layers, reduced sampling of the TG exocyclic conformation, and the rotation of the τ_2 dihedral away from the trans conformation. The effect of this final event can be investigated by restraining the dihedral angle to the trans conformation as observed in the crystal structure, and this was performed for the 18-chain model.

Unsurprisingly, there were limited changes that took place to the structural characteristics of the microfibril compared with the initial structure. There were limited changes to dihedral angles, which was to be expected given that the only dihedral angle to show significant variation in the other simulations was the τ_2 dihedral, and this was now restrained. Exocyclic groups not exposed to solvent were found to reside in the TG conformation for 98% of the simulation. There was also limited change in the exocyclic conformations of surface-exposed groups, with 72% still in the TG conformation. This suggested a link between τ_2 dihedral rotation and variation in exocyclic conformation.

Characteristics that showed slight changes, but only such that they sample values as observed in the 36 A simulations, were the unit cell parameters, chain shifts, and H-bond occupancies. One specific characteristic that showed significant changes compared with the initial structure, but still similar to values observed in the 36 A simulations, was the tilt of center and origin layers (Table II). As this tilt occurs, we can suggest that the tilt is not dependent on τ_2 rotation. However, as the tilt was reduced compared with the other simulations, it can be used as further evidence to suggest that tilt is a two-stage process, with the first stage innate to all structures and the second stage being dependent on rotation of the τ_2 dihedral.

DISCUSSION

We have performed MD simulations on three different models of Iβ cellulose microfibrils to investigate the structural and dynamical dependence on the number of chains in the microfibril. Models with 18, 24, and 36 chains

were produced, and the smaller 18- and 24-chain models were shown to sample a conformational space different to that of the initial crystalline phase and any phase identified experimentally, but similar to that observed in high-temperature MD simulations. Here, we discuss why particular exocyclic group conformations are sampled, why this depends on a second stage of tilt and dihedral τ_2 rotation, and why it only occurs for smaller microfibrils. We also compare the results of our simulations with other computational studies and the available experimental data.

Differences in Microfibril Structure

It was clearly evident that the simulations performed with 18- and 24-chain microfibrils, and those with 36 chains in the B H-bonding pattern, quickly moved from the crystalline conformation to one more reminiscent of the I-HT structures produced in the simulations of Matthews et al. (2011a). This was, however, different from the 36-chain simulations with the A H-bonding pattern in which the microfibrils stayed close to the crystalline phase. One of the major differences between the simulations was the sampling of the exocyclic group conformations. The 36-chain A simulations predominantly stayed in the TG conformation for interior chains, while the other simulations sampled the GG conformation if the chain belonged to a center layer and the traditional TG if in the origin layer. Exocyclic groups that were exposed to solvent behave in a similar manner no matter the number of chains in the microfibril. The predominant conformation for these groups was GT, with the next favored conformation dependent on the chains layer and following the same pattern as observed for interior chains.

The ability of the 18- and 24-chain simulations to sample different exocyclic conformations compared with the 36-chain A model was dependent on extra tilt and the rotation of the τ_2 dihedral. In all simulations, an initial stage of tilt was observed, with chains in the origin layer tilting in a clockwise manner and those in center layers exhibiting an anticlockwise tilt when looking down the microfibril from the nonreducing end. Tilt was able to occur in two different directions as a result of the shift of cellobiose repeat units along the polymerization axis between origin and center chains that positions the chains in different environments. This becomes apparent by analyzing the environment surrounding a cellobiose repeat from the different chains (Fig. 7). Firstly, it was noted that the chains above and below a center chain are oriented differently. For the first (Fig. 7A, bottom) residue of a center chain, the glycosidic linkage in the origin chains above and below the exocyclic group point down. This means that the hydrogens attached to the C1 and C4 carbons of the glycosidic linkage point upwards. This produces a shallow cavity above the exocyclic group and extra volume below it. For the second (Fig. 7A, top) residue of the cellobiose repeat, the shallow cavity is below the exocyclic group and extra volume above
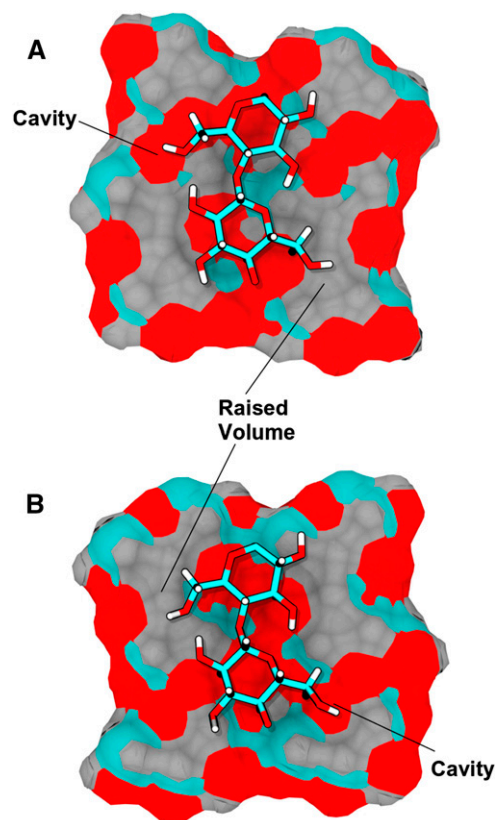

Figure 7. Initial structures of center chain (in licorice) above an origin layer (surface) in A and origin chain (licorice) above a center layer in B, colored by atom type (hydrogen, gray; oxygen, red; and carbon, cyan). This figure highlights the slight cavity below the exocyclic group in the top Glc of the center chain in A and the bottom Glc of the origin chain in B, while also highlighting the raised volume below the exocyclic group in the bottom Glc of the center chain in A and the top Glc of the origin chain in B. It is this steric environment that is responsible for the clockwise tilt in the origin chains and the anticlockwise tilt in the center chains that occurs early.

it. The extra volume of the hydrogens places steric strain on the alternating exocyclic groups, which is alleviated by a slight anticlockwise tilt to the chain. For the origin chain cellobiose repeat, the opposite orientation of glycosidic linkages occurs in the center chains above and below, and as a result, we see a clockwise tilt to alleviate the strain due to the hydrogens.

This explains why, in all simulations, an initial tilt of approximately 5° in either direction was observed and sets up each system for a second stage of tilt. For this second stage to occur, the intrachain H bond from O2 to O6 must break. In the 36-chain A and 18-chain O2-C2 restrained simulations, this H bond did not break, and as a result, the second stage of tilt was not observed. As this H bond was the major interaction that kept the exocyclic group in the TG conformation, these two simulations seldom sample conformations other than TG in interior chains. Additionally, the a-unit cell distance was smaller for these simulations, suggesting

that the layers must increase their distance between each other to allow for the second stage of tilt.

The link between τ_2 dihedral and exocyclic conformation was quite apparent. From the τ_2 restrained simulations, there was no rotation of the τ_2 dihedral, and as a result, almost no change to the exocyclic conformation. However, in all other simulations, coinciding with the rotation of τ_2 dihedral into a cis-like conformation, the O2-to-O6 intrachain H bond broke. As a result, the exocyclic group was free to rotate, and a second stage of tilt was observed. Despite this, it was predominantly only in the center chains where changes in the exocyclic group conformation were observed for interior chains. We suggest this was due to the steric restraints/freedoms placed on the exocyclic groups by the initial tilt of chains, causing steric clashes with exocyclic rotations of certain directions. For center chains, the TG-GG rotation was made easier by the anticlockwise tilt. With the τ_2 dihedral in a cis-like conformation and, as a result, the O2-to-O6 intrachain H bond being broken, the exocyclic groups of center chains could easily rotate from TG to GG. This rotation was further stabilized by the second stage of tilt and the forming of interlayer H bonds, specifically the O6-to-O2 1-10 layer H bonds. Stabilization was so strong that conversion back to TG was not observed. With the clockwise tilt of origin chains, the TG-to-GG rotation was hindered, while the TG-to-GT rotation was more favored. However, unlike the TG and GG conformation, there are no other interactions to stabilize the GT conformation in the interior chains and that is why the TG conformation still dominates. It should be noted that solvent-exposed exocyclic groups were not hindered in any way and could sample all conformations, with GT being of lowest energy.

Rotation of the τ_2 dihedral was observed to a high degree for the 18- and 24-chain simulations but was limited for 36 A chain simulations. The only difference between the simulations was the size of the microfibrils and, as a result, the greater the influence that the surrounding solvent had on the dynamics of each chain. It was clear from the plots of the exocyclic conformation (Fig. 5) that there was far more variation the closer a chain was to the solvent. The solvent must affect the dynamics of the exocyclic group and the τ_2 dihedral. Increased fluctuations led to increased chances of the O2-to-O6 intrachain H bond breaking and eventual freeing of the exocyclic group to sample different conformations. It was evident from Supplemental Figure S2 that these changes filter from the outside chains inwards, and as there are less interior chains in the 18- and 24-chain models, the changes filter throughout the entire microfibril. The 36-chain microfibril has enough interior chains that these changes were not able to infiltrate the microfibril. This is shown in Figure 5A for the center chains of the 36-chain A microfibril, where there was greater variation in the exocyclic conformation with a consistent GG conformation but

at low levels, preventing adoption of a 100% TG exocyclic conformation.

Comparison with Previous Experimental and in Silico Data

The simulations presented here resemble those performed on 36-chain microfibrils with three different force fields (Matthews et al., 2012) and those that heated the microfibrils to 500 K, which resulted in a so-called I-HT conformation (Matthews et al., 2011a). Though their simulations were performed on diamond-shaped finite microfibrils, our 18- and 24-chain simulations with infinite hexagonally shaped microfibrils were able to sample a similar I-HT conformation at simulated ambient temperature with regards to the tilt, exocyclic group conformation, and H-bond patterns. By contrast, our 36-chain model simulations sampled the more crystalline-like Iβ conformation and not the I-HT conformation at ambient temperature. The microfibrils in the Matthews et al. (2012) work were modeled as finite microfibrils and, as a result, were solvated in all directions, including at the reducing and nonreducing end, and thus had significantly more solvent exposure than our 36-chain model. Given the reduced degree of polymerization of these microfibrils, we suggest these models will be oversolvated compared with natural microfibrils, and this is a potential limitation of their simulations. As mentioned above, we believe the reason our 18/24- and 36-chain simulations sampled a different conformational space was the difference in degree of solvent exposure, and we propose that this was the reason that the Matthews et al. (2012) 36-chain simulations also sampled the I-HT conformation. From Matthews et al. (2012), specifically the simulations performed with the Charmm and Glycam force fields, we note that changes away from the crystalline TG conformation occur from the outside of the microfibril inwards. This correlates with what was observed in our 18- and 24-chain simulations, suggesting that changes to structure occur initially at solvent-exposed residues and then filter through the microfibril. With our 36-chain simulations, the decreased solvent exposure and higher number of interior chains means that the changes were not able infiltrate the entire microfibril.

It was evident from the exocyclic conformation occupancy plots that the favored conformation for solvent-exposed groups was predicted to be GT. This goes against both experimental and computational studies of Glc and cello-oligosaccharides in solution where GG is reported to be the favored form (Nishida et al., 1984; Cramer and Truhlar, 1993; Bock and Duus, 1994; Kirschner and Woods, 2001; Kuttel et al., 2002; Pereira et al., 2006; Shen et al., 2009). In these studies, the general consensus is that exocyclic groups have a distribution of about 55:40:5 for the GG:GT:TG conformations, respectively. With the NMR and x-ray data suggesting that TG was the only conformation in a cellulose microfibril, we further questioned this preference for GT. However, there is some evidence to

suggest that the simulations presented here are sampling the correct exocyclic conformation for these solvent-exposed groups. Recent experimental data suggests that GT may be more populated than GG for Glc in solution (Hansen and Hünenberger, 2011). X-ray data for cellobiose has shown that the two exocyclic groups in the crystal structure adopt the GT conformation (Chu and Jeffrey, 1968). Additionally, atomic force microscopy images and NMR data suggest that exocyclic groups at surfaces have the GT conformation (Baker et al., 2000). Finally, other MD simulations have also seen the GT conformation favored at the surface (Matthews et al., 2012). Thus, it appears that when Glc is formed into a polymer, and does not reside in an environment where it can interact with other chains, GT is its favored conformation, though with low-energy barriers such that significant variation between conformers is likely.

Experimental work suggests that the majority of the exocyclic groups of cellulose reside in the TG conformation. X-ray and neutron diffraction data suggest that this is the case for all crystalline chains. NMR data has generally been interpreted to suggest that chains in the interior of microfibril aggregates have a TG exocyclic conformation, while those that were solvent exposed were either GG or GT. The simulations performed here with the 36-chain model correlate well with this interpretation of the experimental data. However, as mentioned previously, simulations with the 18- and 24-chain models move away from the Iβ phase to a I-HT conformational space, which does not fit with the experimental data.

However, we suggest that it is possible that the NMR data are being interpreted incorrectly. It is well documented that the dual C4 peaks are separately due to the C4 carbons from solvent-exposed and interior chains (Ha et al., 1998; Newman, 1998). Given the almost identical sizes of the double peaks for C4 and C6 in cellulose, it is commonly thought that the upfield peaks result from the contribution of carbons in surface chains and the downfield peaks from interior chains (Viëtor et al., 2002). To make these conclusions, two assumptions must be made: all C4 and C6 carbons on surface chains will have the same chemical shifts, and exocyclic groups are restricted to GG and GT for surface and TG for interior. We have doubts about the validity of these assumptions. Given the 2-fold axis of cellulose, there is potential for the equivalent carbons on each cellobiose repeat to have two different chemical shifts. If we focus on surface chains, the environment around the two C4s is very similar, with comparable exposure to solvent (Fig. 3A, inset). However, the two C6 carbons and their attached hydroxyl groups exist in different environments. One points directly out into the solvent, while the other is positioned away from the solvent, directly interacting with another cellulose chain, and can therefore be thought of as actually being an interior exocyclic group. Thus, the two different C6 carbons should have different chemical shifts, with the nonsolvent-exposed group having a similar shift to the interior C6 and thus contributing to the downfield peak.

As a result, we should characterize these chains, and the atoms of interest, with regards to their solvent exposure instead of whether they are on surface or interior chains.

As we have identified that it is solvent exposure that we need to focus on to analyze NMR spectra, the second assumption should now be that exocyclic groups that are solvent exposed are restricted to the GG and GT conformations, while all other exocyclic groups are TG. If this is true, then for the 36-chain microfibril, the upfield C6 peak that represents all exocyclic groups that are solvent exposed should be 33% of the total C6 peak (Table I). However, the C4 peak will have a 50% contribution to the total C4 peak, and this does not correlate with the fact that the two peaks should be of similar size. Using the restriction assumption, the C6 upfield peak sizes for the 18- and 24-chain models are not similar to the C4 upfield peaks either (44% versus 67% and 42% versus 58% for 18 and 24 chains, respectively). Under the second assumption, none of the models fit the NMR data, suggesting that this assumption is also invalid.

Our simulations provide more evidence that this assumption is invalid and also suggest that the solvent-exposed exocyclic groups are not restricted to GT or GG, while interior groups are not restricted to TG. Therefore, we suggest that the percentage of exocyclic groups in the GT and GG conformation be a measure of the size of the upfield C6 peak, as is suggested by Horii et al. (1983). Using this criterion, we find that both the 18- and 24-chain models give C6 upfield peaks that would be of a very similar size to the C4 upfield peak. For the 24-chain microfibril, the C6 peak would be 63%, which compares favorably with the C4 surface peak of 58%, with the 18-chain microfibrils giving even better correlation with the C6 peak predicted to be 66% and the C4 peak predicted to be 67%. Additionally, the data for the 18- and 24-chain microfibrils correlate well with an NMR spectrum of sugar beet (*Beta vulgaris*) cellulose from parenchymatous primary cell walls, which was interpreted as having 54% of its chains located at the surface (Viëtor et al., 2002). Thus, using this new interpretation of the NMR data, it appears that the size of the C6 upfield peak is dictated by the percentage of exocyclic groups in the GT and GG conformations, and these conformations are not restricted to just surface chains or solvent-exposed groups.

Further evidence to support the theory that the elementary microfibril consists of either 18 or 24 chains comes from scattering and diffraction work (Fernandes et al., 2011; Newman et al., 2013). Data from these recent WAXS, wide-angle neutron scattering, and small-angle neutron scattering studies on mung bean (*Vigna radiata*) primary cell walls and spruce (*Picea sitchensis*) wood secondary cell walls suggest that microfibril diameters should be about 3 nm and that there should be six layers of 100 planes per microfibril. As a result, the microfibril should have a cross-sectional area of around 7 nm^2. After the expansion that was observed by performing these simulations at 300 K in a hydrated environment, the microfibrils had cross-sectional areas of 6.8, 8.6,

and 12.6 nm^2 for the 18-, 24-, and 36-chain microfibrils, respectively. This suggests that the 36-chain microfibril is too big and that 18 or 24 chains are more credible fits to the recent experimental data.

Given the results with the 36-chain model and all other experimental x-ray diffraction work, we still believe that cellulose fibers of larger diameter have a crystalline core. However, the results with the 18- and 24-chain models suggest that we should question the model of a purely crystalline cellulose core for small-diameter microfibrils. Smaller fibers are likely to have a very different structure, which we believe is a result of the greater solvent exposure that these chains experience.

Comparisons with experimental structural data are dependent on the systems that we are simulating being closely related to the systems used in the experiments. With the NMR, diffraction, and scattering work, this is an assumption that may not hold true. Due to issues with experimental isolation protocols, it is likely that the experimental data are produced with microfibril aggregates instead of small-diameter microfibrils, in an environment with significantly less solvent than what was modeled in this work. This may not be as great an issue for NMR studies, as it has been shown that water content of systems studied are close to the in vivo water content and that changes to spectra are observed when working on cellulose samples from different sources and different aggregate sizes (Malm et al., 2010). There is also a potential issue with the mechanical and chemical pretreatments applied to cellulose that allow them to be investigated with the differing experimental methods, which could lead to an altered cellulose microfibril structure.

It is not only the experiments that could affect our comparisons but also the computational methods, specifically the force field used. It could be possible that the C36 Charmm force field does not give a good account of Glc when in a cellulose microfibril. This force field was parameterized based on the structure and dynamics of Glc in solution and, as a result, may not be able to sample the correct energy profiles when a residue is in an interior chain, without any solvent exposure and interacting with many other Glc residues. It would be of interest to perform some ab initio calculations on small cellulose systems, such as those performed by Watts et al. (2013), to investigate the structure and energetics of Glc in a cellulose microfibril and how this compares with the structures and dynamics observed in these simulations.

CONCLUSION

We have performed MD simulations on 18-, 24-, and 36-chain models of Iβ cellulose and found that the smaller 18- and 24-chain models sample a conformational space different from that of the initial crystalline phase and any phase identified experimentally but similar to that observed in previous high-temperature MD simulations of a 36-chain model. The chains of all

microfibrils became tilted, with the origin chains experiencing a clockwise tilt and center chains an anticlockwise tilt due to the different steric environments in which the two chains were situated. As a result of increased variation in the τ_2 dihedral and the steric limitations of tilt, the 18- and 24-chain microfibrils were able to change exocyclic conformations in center layers, virtually exclusively to GG, while there were sporadic changes to GT in the origin layers. To accommodate these changes, a further stage of tilt was required. For all simulations, solvent-exposed exocyclic groups preferred to reside in a GT conformation; however, there was significant variability in the conformations sampled. As a result of these simulations and a reinterpretation of NMR data, it was concluded that the 18- and 24-chain models were more parsimonious than the 36-chain model, in agreement with recent scattering and diffraction data.

Determination of the detailed structure of cellulose microfibrils in a hydrated environment will allow for more rigorous investigations of cellulose in larger systems. Aggregation of microfibrils can be studied, which can, in turn, be used to understand more about the recalcitrance to enzymic and chemical hydrolysis of plant-derived cellulose. Simulation studies with atomic detail of the interaction of cellulose with noncellulosic polymers are now possible, leading to a better understanding of atomic interactions underlying the structure, dynamics, and physical properties of the plant cell wall matrix. Identifying the number of chains in a microfibril also gives insight into the structure of the CesA rosette complex. The interpretation of simulation and NMR data to suggest 18 to 24 chains in a microfibril means that either the 36 proposed CesA proteins in a rosette are regulated such that not all subunits are simultaneously active or that there are less than 36 CesA proteins in a rosette. There have been suggestions that each six elementary particles of the rosette could contain less than six CesA proteins (Newman et al., 2013) or that the particle could contain both initiating and extending CesA proteins (Read and Bacic, 2002). It would also be of interest to run simulations similar to the ones reported in this study on cellulose aggregates made up of many small-diameter microfibrils or microfibrils of much larger diameter to further investigate the effect of chain numbers on crystallinity, chain conformation, surface features, and dynamic characteristics of cellulose.

MATERIALS AND METHODS

Models of cellulose microfibrils were constructed with Cellulose Builder (Gomes and Skaf, 2012). Initial models containing 36 chains (eight layers) were created for the Iβ form with a hexagonal shape (Fig. 3). These models were then used as templates to produce the smaller 18- and 24-chain models (six layers in each; Fig. 3, B and C). Models in which the 36-chain model had the B H-bond pattern of Nishiyama et al. (2002) were produced by rotating the position of HO2 of every residue such that the C3-C2-O2-HO2 dihedral angle was –14° and rotation of HO6 such that the C5-C6-O6-HO6 dihedral angle was 26°. Each chain contained 20 Glc residues (i.e. 10 cellobiose repeat units) so that the microfibril had a length of approximately 102 Å. Periodic boundary conditions were applied along all three directions. The bonding between Glc

residues was arranged so that residues at one end of the simulation cell were covalently bonded to Glc residues on the other side of the periodic boundary. This creates a microfibril that is unaffected by boundary effects and is effectively infinite in length. Each microfibril was aligned so that its polymerization axis ran along the z axis of the simulation cell and was solvated in a box of transferable intermolecular potential 3P water so that the microfibril was padded in the x and y directions by at least 12 Å, with no padding in the z direction.

All simulations were performed in duplicate using NAMD 2.9 (Kale et al., 1999) with the Charmm 36 carbohydrate force field (Guvench et al., 2008) at 300 K. An initial simulation step was performed to optimize the position of the added waters. With a time step of 1 fs, a nonbonded cutoff of 10 Å, and the RATTLE/SETTLE algorithms (Andersen, 1983; Miyamoto and Kollman, 1992) used to constrain the bond lengths involving hydrogens in water, 5,000 steps of minimization were followed by 500,000 steps of constant pressure-temperature Langevin dynamics. Pressure was kept constant using a Langevin piston barostat, and the particle mesh Ewald method was used to calculate long-range electrostatic interactions (Darden et al., 1993). Production phase simulations were performed in the constant volume-temperature ensemble with a time step of 2 fs and all bonds from hydrogen to heavy atoms constrained.

To analyze the structure and dynamics of the microfibrils, multiple characteristics were calculated for the initial structure of each simulation and the last 50 ns. Unit cell parameters were calculated by splitting each microfibril into discrete unit cells and then averaging over all unit cells. The exocyclic group around C6 can take on three different orientations that can be described by the dihedral angle about the atoms O5-C5-C6-O6. With values of 0° to 120°, 120° to 240°, and 240° to 360°, these angles correspond to the GT, TG, and GG conformations, respectively (Fig. 2). The color scheme used in Matthews et al. (2012) has been used in this work to color cellulose chains and individual Glc residues. Intrachain, intralayer, and interlayer H-bond occupancies were calculated using the Visual Molecular Dynamics h_bonds command with a heavy atom cutoff of 3.4 Å and angle cutoff of 60°. The internal motion of the microfibrils was tracked by calculating the change in distance between 100/200 layers, 110 layers, 1-10 layers, and the shift of chains along the z axis (Fig. 1). The tilt in each chain was calculated by determining how much the plane of each chain rotates around its initial position along the xz plane. A positive value means there has been a clockwise rotation and a negative value anticlockwise rotation, when viewed from the nonreducing end. Average φ, ψ, τ_2, and τ_3 dihedral angles for each chain have also been calculated. The density of water above each layer was calculated as a function of distance from the microfibril and as a function of distance along the polymerization axis.

Supplemental Data

The following supplemental materials are available.

Supplemental Figure S1. Plot of exocyclic conformation occupancies for 640-ns simulation of 36 A model.

Supplemental Figure S2. Exocyclic conformations at different time points up to 200 ns.

Supplemental Figure S3. O6-to-O2 intrachain and O2-to-O6 intralayer H-bond occupancies.

Supplemental Figure S4. Relative water densities for 36 B chain model.

Supplemental Figure S5. Relative water densities for 24-chain model.

Supplemental Figure S6. Relative water densities for 18-chain model.

Supplemental Figure S7. Water densities above cellulose surfaces.

Supplemental Table S1. Exocyclic occupancies for 110 and 1-10 hydrophilic surfaces.

Supplemental Table S2. Average φ dihedral angles.

Supplemental Table S3. Average ψ dihedral angles.

Supplemental Table S4. Average τ_3 dihedral angles.

Supplemental Table S5. Occupancy of interlayer hydrogen bonds.

LITERATURE CITED

Andersen C (1983) Rattle: A "velocity" molecular version of the shake dynamics calculations for molecular dynamics calculations. J Comput Phys **52:** 24–34

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

Baker AA, Helbert W, Sugiyama J, Miles MJ (2000) New insight into cellulose structure by atomic force microscopy shows the i(α) crystal phase at near-atomic resolution. Biophys J **79:** 1139–1145

Bergenstråhle M, Berglund LA, Mazeau K (2007) Thermal response in crystalline Iβ cellulose: a molecular dynamics study. J Phys Chem B **111:** 9138–9145

Bock K, Duus JØ (1994) A conformational study of hydroxymethyl groups in carbohydrates investigated by 1H NMR spectroscopy. J Carbohydr Chem **13:** 513–543

Brett CT, Hillman JR (1985) Biochemistry of Plant Cell Walls, Ed 70. Cambridge University Press, Cambridge, UK

Busse-Wicher M, Gomes TCF, Tryfona T, Nikolovski N, Stott K, Grantham NJ, Bolam DN, Skaf MS, Dupree P (2014) The pattern of xylan acetylation suggests xylan may interact with cellulose microfibrils as a twofold helical screw in the secondary plant cell wall of *Arabidopsis thaliana*. Plant J **79:** 492–506

Chen P, Nishiyama Y, Putaux JL, Mazeau K (2014) Diversity of potential hydrogen bonds in cellulose I revealed by molecular dynamics simulation. Cellulose **21:** 897–908

Chu SSC, Jeffrey GA (1968) The refinement of the crystal structures of β-D-glucose and cellobiose. Acta Crystallogr B **24:** 830–838

Cramer CJ, Truhlar DG (1993) Quantum chemical conformational analysis of glucose in aqueous solution. J Am Chem Soc **115:** 5745–5753

Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An N.log(N) method for Ewald sums in large systems. J Chem Phys **98:** 10089–10092

Ding SY, Himmel ME (2006) The maize primary cell wall microfibril: a new model derived from Direct visualization. J Agric Food Chem **54:** 597–606

Ding SY, Zhao S, Zeng Y (2014) Size, shape, and arrangement of native cellulose fibrils in maize cell walls. Cellulose **21:** 863–871

Doblin MS, Pettolino F, Bacic A (2010) Evans Review: plant cell walls: the skeleton of the plant world. Funct Plant Biol **37:** 357–381

Earl WL, VanderHart DL (1981) Observations by high-resolution carbon-13 nuclear magnetic resonance of cellulose I related to morphology and crystal structure. Macromolecules **14:** 570–574

Fernandes AN, Thomas LH, Altaner CM, Callow P, Forsyth VT, Apperley DC, Kennedy CJ, Jarvis MC (2011) Nanostructure of cellulose microfibrils in spruce wood. Proc Natl Acad Sci USA **108:** E1195–E1203

Gomes TCF, Skaf MS (2012) Cellulose-builder: a toolkit for building crystalline structures of cellulose. J Comput Chem **33:** 1338–1346

Gross AS, Chu JW (2010) On the molecular origins of biomass recalcitrance: the interaction network and solvation structures of cellulose microfibrils. J Phys Chem B **114:** 13333–13341

Guvench O, Greene SN, Kamath G, Brady JW, Venable RM, Pastor RW, Mackerell AD Jr (2008) Additive empirical force field for hexopyranose monosaccharides. J Comput Chem **29:** 2543–2564

Ha MA, Apperley DC, Evans BW, Huxham IM, Jardine WG, Viëtor RJ, Reis D, Vian B, Jarvis MC (1998) Fine structure in cellulose microfibrils: NMR evidence from onion and quince. Plant J **16:** 183–190

Hansen HS, Hünenberger PH (2011) A reoptimized GROMOS force field for hexopyranose-based carbohydrates accounting for the relative free energies of ring conformers, anomers, epimers, hydroxymethyl rotamers, and glycosidic linkage conformers. J Comput Chem **32:** 998–1032

Herth W (1983) Arrays of plasma-membrane "rosettes" involved in cellulose microfibril formation of *Spirogyra*. Planta **159:** 347–356

Horii F, Hirai A, Kitamaru R (1983) Solid-state ^{13}C-NMR study of conformations of oligosaccharides and cellulose conformation of CH_2OH group about the exo-cyclic C-C Bond. Polym Bull **10:** 357–361

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

Isogai A, Usuda M (1989) Solid-state CP/MAS ^{13}C NMR study of cellulose polymorphs. Macromolecules **22:** 3168–3172

Kale L, Skeel R, Bhandarkar M, Brunner R, Gursoy A, Krawetz N, Phillips J, Shinozaki A, Varadarajan K, Schulten K (1999) NAMD2: greater scalability for parallel molecular dynamics. J Comput Phys **312:** 283–312

Kirschner KN, Woods RJ (2001) Solvent interactions determine carbohydrate conformation. Proc Natl Acad Sci USA 98: 10541–10545

Kuttel M, Brady JW, Naidoo KJ (2002) Carbohydrate solution simulations: producing a force field with experimentally consistent primary alcohol rotational frequencies and populations. J Comput Chem 23: 1236–1243

Larsson PT, Westlund PO (2005) Line shapes in CP/MAS ^{13}C NMR spectra of cellulose I. Spectrochim Acta A Mol Biomol Spectrosc 62: 539–546

Larsson PT, Wickholm K, Iversen T (1997) A CP/MAS ^{13}C NMR investigation of molecular ordering in celluloses. Carbohydr Res 302: 19–25

Lehtiö J, Sugiyama J, Gustavsson M, Fransson L, Linder M, Teeri TT (2003) The binding specificity and affinity determinants of family 1 and family 3 cellulose binding modules. Proc Natl Acad Sci USA 100: 484–489

Malm E, Bulone V, Wickholm K, Larsson PT, Iversen T (2010) The surface structure of well-ordered native cellulose fibrils in contact with water. Carbohydr Res 345: 97–100

Matthews JF, Beckham GT, Bergenstrahle-Wohlert M, Brady JW, Himmel ME, Crowley MF (2012) Comparison of cellulose Iβ simulations with three carbohydrate force fields. J Chem Theory Comput 8: 735–748

Matthews JF, Bergenstråhle M, Beckham GT, Himmel ME, Nimlos MR, Brady JW, Crowley MF (2011a) High-temperature behavior of cellulose I. J Phys Chem B 115: 2155–2166

Matthews JF, Himmel ME, Crowley MF (2011b) Conversion of cellulose Iα to Iβ via a high temperature intermediate (I-HT) and other cellulose phase transformations. Cellulose 19: 297–306

Matthews JF, Skopec CE, Mason PE, Zuccato P, Torget RW, Sugiyama J, Himmel ME, Brady JW (2006) Computer simulation studies of microcrystalline cellulose Iβ. Carbohydr Res 341: 138–152

Mazeau K (2005) Structural micro-heterogeneities of crystalline Iβ-cellulose. Cellulose 12: 339–349

Miyamoto S, Kollman PA (1992) SETTLE: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. J Comput Chem 13: 952–962

Newman RH (1998) Evidence for assignment of ^{13}C NMR signals to cellulose crystallite surfaces in wood, pulp and isolated celluloses. Holzforschung 52: 157–159

Newman RH (1999) Estimation of the lateral dimensions of cellulose crystallites using ^{13}C NMR signal strengths. Solid State Nucl Magn Reson 15: 21–29

Newman RH, Hemmingson JA (1994) Carbon-13 NMR distinction between categories of molecular order and disorder in cellulose. Cellulose 2: 95–110

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

Nishida Y, Ohrui H, Meguro H (1984) ^{1}H-NMR studies of (6R)- and (6S)-deuterated D-hexoses: assignment of the preferred rotamers about C5-C6 bond of D-glucose and D-galactose derivatives in solutions. Tetrahedron Lett 25: 1575–1578

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

Nishiyama Y, Langan P, Chanzy H (2002) Crystal structure and hydrogen-bonding system in cellulose Iβ from synchrotron x-ray and neutron fiber diffraction. J Am Chem Soc 124: 9074–9082

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

Park S, Johnson DK, Ishizawa CI, Parilla PA, Davis MF (2009) Measuring the crystallinity index of cellulose by solid state ^{13}C nuclear magnetic resonance. Cellulose 16: 641–647

Payen A (1838) Memoir on the composition of the tissue of plants and of woody material. C R Biol 7: 1052–1056

Pereira CS, Kony D, Baron R, Müller M, van Gunsteren WF, Hünenberger PH (2006) Conformational and dynamical properties of disaccharides in water: a molecular dynamics study. Biophys J 90: 4337–4344

Read SM, Bacic T (2002) Plant biology. Prime time for cellulose. Science 295: 59–60

Shen T, Langan P, French AD, Johnson GP, Gnanakaran S (2009) Conformational flexibility of soluble cellulose oligomers: chain length and temperature dependence. J Am Chem Soc 131: 14786–14794

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

Sugiyama J, Vuong R, Chanzy H (1991b) Electron diffraction study on the two crystalline phases occurring in native cellulose from an algal cell wall. Macromolecules 24: 4168–4175

Thomas LH, Forsyth VT, Sturcová A, Kennedy CJ, May RP, Altaner CM, Apperley DC, Wess TJ, Jarvis MC (2013) Structure of cellulose microfibrils in primary cell walls from collenchyma. Plant Physiol 161: 465–476

Tormo J, Lamed R, Chirino AJ, Morag E, Bayer EA, Shoham Y, Steitz TA (1996) Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose. EMBO J 15: 5739–5751

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

Viëtor RJ, Newman RH, Ha MA, Apperley DC, Jarvis MC (2002) Conformational features of crystal-surface cellulose from higher plants. Plant J 30: 721–731

Wada M, Hori R, Kim UJ, Sasaki S (2010) X-ray diffraction study on the thermal expansion behavior of cellulose Iβ and its high-temperature phase. Polym Degrad Stabil 95: 1330–1334

Watts HD, Mohamed MNA, Kubicki JD (2013) A DFT study of vibrational frequencies and ^{13}C NMR chemical shifts of model cellulosic fragments as a function of size. Cellulose 21: 53–70

Wickholm K, Larsson PT, Iversen T (1998) Assignment of non-crystalline forms in cellulose I by CP/MAS ^{13}C NMR spectroscopy. Carbohydr Res 312: 123–129

Zhang Q, Bulone V, Ågren H, Tu Y (2011) A molecular dynamics study of the thermal response of crystalline cellulose Iβ. Cellulose 18: 207–221

Zugenmaier P (2001) Conformation and packing of various crystalline cellulose fibers. Prog Polym Sci 26: 1341–1417