

SCIENTIFIC REPORTS



OPEN

The Shape of Native Plant Cellulose Microfibrils

James D. Kubicki¹, Hui Yang², Daisuke Sawada³, Hugh O'Neill⁴, Daniel Oehme¹ & Daniel Cosgrove²

Determining the shape of plant cellulose microfibrils is critical for understanding plant cell wall molecular architecture and conversion of cellulose into biofuels. Only recently has it been determined that these cellulose microfibrils are composed of 18 cellulose chains rather than 36 polymers arranged in a diamond-shaped pattern. This study uses density functional theory calculations to model three possible habits for the 18-chain microfibril and compares the calculated energies, structures, ¹³C NMR chemical shifts and WAXS diffractograms of each to evaluate which shape is most probable. Each model is capable of reproducing experimentally-observed data to some extent, but based on relative theoretical energies and reasonable reproduction of all variables considered, a microfibril based on 5 layers in a 34443 arrangement is predicted to be the most probable. A habit based on a 234432 arrangement is slightly less favored, and a 6 × 3 arrangement is considered improbable.

Cellulose is one of the most important materials on Earth. As the foundation of plant cell walls, it has a critical role in the biosphere. Because it has evolved to provide strength and resist biodegradation, cellulose is commonly used in paper, clothing, and myriad materials. Efforts to convert cellulose into fermentable compounds have become commercialized recently in order to produce ethanol from recyclable feedstocks that minimize impact on the food supply.

Landmark papers by Nishiyama and coworkers^{1–3} utilized X-ray and neutron diffraction to determine the crystalline structure of cellulose. This structure was used as the basis for density functional theory (DFT) calculations that subsequently reproduced the infrared, Raman, NMR and sum-frequency generation spectra of cellulose^{4,5}. However, much of the research on cellulose has been performed on bacterial and/or plant-extracted cellulose neither of which represents the state of plant cellulose *in muro*^{6,7}. Consequently, recent studies on non-extracted cellulose have been critical in helping to define the native state of cellulose within plant cell walls^{8–10}.

Although the atomic structure of cellulose has been characterized, an outstanding question is the shape of cellulose microfibrils in plant cell walls. For decades, the cellulose microfibril (CMF) was thought to be comprised of 36 chains of cellulose polymers (i.e., 1,4 β-linked glucan monomers) based on the observation of “rosettes” formed by the cellulose synthase complex (CSC)¹¹. Furthermore, a diamond-shaped habit based upon cellulose polymer layers arranged in a 1 × 2 × 3 × 4 × 5 × 6 × 5 × 4 × 3 × 2 × 1 pattern was assumed to be the habit of this 36-polymer CMF (e.g., Matthews *et al.*, 2012). This view of the CMF was challenged by papers based on wide-angle X-ray scattering (WAXS)^{12,13}. Based on diameter arguments, the 36-polymer CMF was judged to be too large, so the CMF would have to be smaller. This conclusion has been justified recently by the work of Nixon *et al.*¹⁴, Hill *et al.*¹⁵ and Venu *et al.*¹⁶ who have demonstrated that the CSC rosette is most consistent with a hexamer of trimers. Assuming only one cellulose polymer can be extruded from each of the Cesa enzymes in the CSC^{17–19}, the CMF should be made of 18 cellulose polymer chains rather than 36.

The difference between 18 and 36 polymers in the fundamental plant CMF will be reflected in the habit of the CMF as well as the size. The shape of the CMF is far from purely academic because the surfaces displayed by the CMF will depend upon this arrangement of cellulose chains. In turn, the interactions of other plant cell wall components (i.e., water, hemicellulose, pectin, lignin) will depend upon the nature of the exposed CMF surfaces. Biodegradation mechanisms of the cellulose will also depend heavily upon the nature of the CMF surfaces present. Hence, this study sought to determine the most probably arrangement of cellulose chains within the

¹Department of Geological Sciences, University of Texas at El Paso, El Paso, TX, USA. ²Department of Biology, The Pennsylvania State University, University Park, PA, USA. ³Department of Bioproducts and Biosystems, School of Chemical Engineering, Aalto University, Espoo, Finland. ⁴Neutron Scattering Division, Oak Ridge National Laboratory, Oak Ridge, USA. Correspondence and requests for materials should be addressed to J.D.K. (email: jdkubicki@utep.edu)

CMF in order to use this crystal habit as a basis for future studies of plant cell wall architecture, mechanics and biochemistry.

Based upon success in modeling cellulose atomic structure with density functional theory (DFT) calculations^{4,20,21}, we chose to use a similar methodology⁴ to predict the structure, energetics and spectroscopic properties of finite CMFs with various arrangements of 18 cellulose chains. The periodic DFT calculations in VASP were carried out in order to obtain atomic structures that could be used for comparison against observed X-ray and neutron diffraction studies^{1,2}. Furthermore, a key prediction is the relative energies of each CMF habit. Although one cannot expect CMFs formed in plant cell walls to necessarily be at thermodynamic equilibrium, I β cellulose is known to be thermodynamically stable and only slightly higher in Gibbs free energy than the extracted cellulose II and III forms²². Our assumption is that it does not make biochemical sense for plants to produce a less thermodynamically stable form both in terms of assembly kinetics and in regards to ultimate strength and resistance to biodegradation.

To supplement the atomic structure and thermodynamic criteria, this study also calculates the WAXS diffractograms and the ¹³C NMR chemical shifts ($\delta^{13}\text{C}$) of the model CMFs. The WAXS diffractograms were simulated using the CMF models taken from the structures as energy-minimized in VASP using the program WAXSiS^{23,24}. $\delta^{13}\text{C}$ values were calculated with Gaussian 09²⁵ using the same model structures terminated after 4 glucan monomers along the c-axis with methyl groups to avoid unsatisfied valence at the terminal O atoms⁸ and following methods used in previous studies²⁶.

Results and Discussion

Three arrangements of the 18 cellulose polymers were created based upon the criteria that the CMF should attempt to minimize surface area, be reasonably symmetric and maximize the van der Waals forces between layers and H-bonding within layers. These three crystal habits are shown in Fig. 1. The model CMF habits have the following arrangements of cellulose polymers 6×3 , 234432, and 34443. (A 3×6 structure is possible but the diameter of this shape is ≈ 5 nm – significantly wider than the 3 to 3.8 nm measured^{12,13}. The first two have been suggested previously as shapes of the CMF (see ref.¹³), and the third is based upon a structure created for coarse-grained simulations of CMFs by Vin Crespi (The Pennsylvania State University).

Comparison of the model atomic structures with the parameters reported for I β cellulose in Nishiyama *et al.*^{1,2} reveals that all three CMFs accurately reproduce observed interatomic distances and torsion angles (Tables 1 and 2). Structural differences between the center and origin chains of I β cellulose are larger than the differences among the three CMF habits. Significant discrepancies exist between the model and experimentally-reported center chain H-bond distances and angles (Table 1), but we note that the model structures produce vibrational frequencies in agreement with observation^{4,5} and better agreement than the original Nishiyama *et al.*¹ structure with observed $\delta^{13}\text{C}$ values⁴. Because the same model and method reproduces the reported structure of the origin chain, we conclude that the discrepancy is not the result of inaccuracies in the model but instead originate from the difficulties arising from predicting H atom positions in cellulose¹. The Φ (O5-C1-O4-C4), Ψ (C1-O4-C4-C5), χ (O5-C5-C6-O6) and χ' (C4-C5-C6-O6) torsion angles calculated are in reasonable agreement with the experimental values (Table 2; refs^{1,2}). There is no model that performs significantly better with respect to reproducing structural parameters, so we conclude that crystal habit has a minimal effect on these structural parameters.

Our second criterion for evaluating the probability of forming each CMF habit is the potential energy calculated for each. The periodic DFT calculations should better represent actual CMFs in that they are continuous rather than truncated after four glucan monomers. However, DFT has notable issues with estimating van der Waals forces²⁷, so although the D3 dispersion correction of Grimme *et al.*²⁸ was employed, this empirically-derived estimate is subject to a significant level of uncertainty. In order to test how robust the DFT-predicted energies might be, single-point energy calculations were also performed on the tetramer CMFs in Gaussian 09²⁵ using the Perdew-Wang²⁹, M05³⁰, and B3LYP^{31,32} exchange-correlation functions in both vacuum and aqueous solution conditions (IEF-PCM)³³. Table 3 lists the energies for each model and method. Both the VASP and Gaussian 09 results argue against the 6×3 model because it is higher in energy than the other two options. The +276 to +371 kJ discrepancy is large, but when normalized on a per glucan basis, the difference is 4 to 5 kJ/mol/glucan. However, this small energy difference is still not negligible, because an energy difference of 4 kJ/mol between two states would result in a Boltzmann distribution ratio of approximately 5 between the two states. The VASP energies indicate that the 234432 model is most stable whereas the Gaussian 09 calculations predict that the 34443 model is lower in energy in vacuum and aqueous solution. In the former case, the difference is less than 1 kJ/mol/glucan which is well within computational accuracy. Even the +4 kJ/mol/glucan difference favoring 34443 from Gaussian 09 is relatively small, so we cannot definitively distinguish between 234432 and 34443 based on energetics.

High-resolution ¹³C NMR spectra of cellulose *in vivo* have been reported recently⁸ that can be used as a basis for evaluating the proposed CMF habits. Table 4 compares observed and calculated values for all types of C atoms within the center and origin chains. For “interior” C atoms, all three model CMFs result in good agreement with observation with the 34443 model having a slightly better fit based on the root-mean-squared (RMSE) and maximum errors (MaxE). ¹³C NMR spectra of cellulose have observed C4, C5 and C6 peak doublets that have been ascribed to the changes in chemical environment and/or structure that occur on the surface of cellulose compared to the interior (see Harris *et al.*⁹, and references therein; Wang *et al.*⁸). This interior to exterior or ordered to disordered shift is especially pronounced for C4 which changes from 88.5 to 91.5 ppm on the interior (iC4) to 84 to 86 ppm on the surface (sC4). A similar shift may occur for C6 from iC6 of 66 ppm to sC6 of 62 ppm, but the sC6 peak is harder to resolve due to overlap with other plant cell wall polymers (Harris *et al.*⁹). Our results (Supplementary Table 1) indicate that surficial $\delta^{13}\text{C}$ C4 values (85.4 ± 0.9 – 87.5 ± 2.4 ppm) can match both the “iC4” and “sC4” peaks in Harris *et al.*⁹, and the surficial $\delta^{13}\text{C}$ C4 values are slightly higher than the interior $\delta^{13}\text{C}$ C4 values

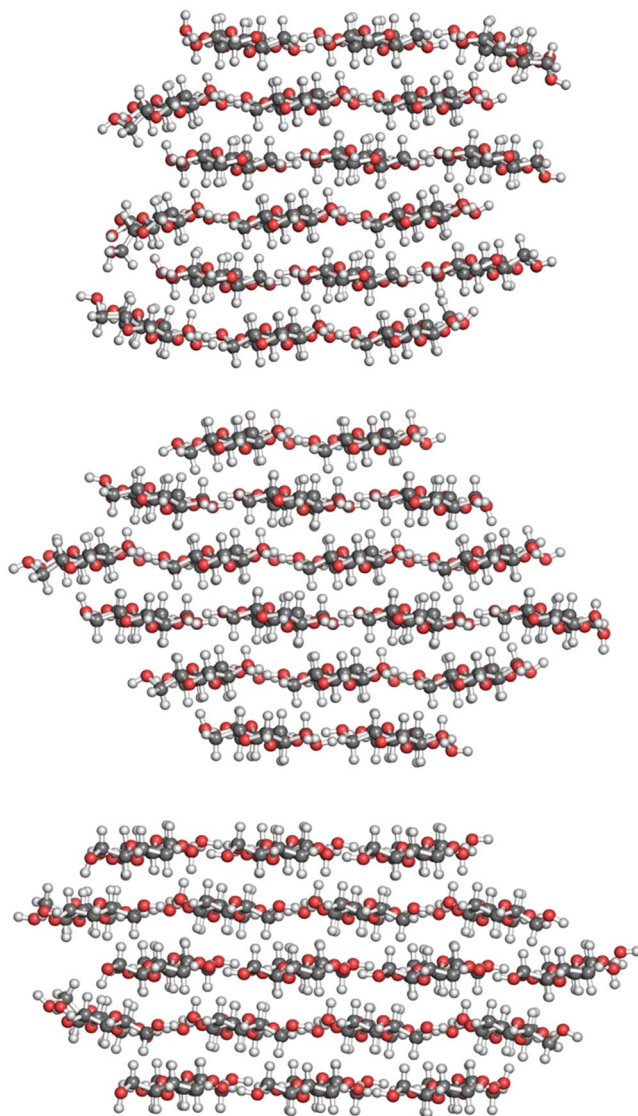


Figure 1. Potential crystal habits of native cellulose microfibrils: 6×3 , 234432 and 34443 models (from top to bottom).

| | Origin | | | | Center | | | |
|--------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| | Intra-molecular | | intermolecular | | Intra-molecular | | intermolecular | |
| | O3H-O5 | O2H-O6 | O6H-O3 | O6H-O2 | O3H-O5 | O2H-O6 | O6H-O3 | O6H-O2 |
| Expt. | 1.75 2.70 162 | 1.90 2.86 165 | 1.78 2.71 157 | 2.54 3.21 125 | 1.97 2.76 137 | 1.83 2.76 159 | 2.04 2.89 144 | 3.63 3.62 81 |
| 6×3 | 1.79 2.74 161 | 1.81 2.80 175 | 1.75 2.71 162 | 2.77 3.41 123 | 1.77 2.74 163 | 1.74 2.73 169 | 1.84 2.80 165 | 2.75 3.36 120 |
| 34443 | 1.79 2.74 161 | 1.82 2.81 174 | 1.74 2.70 163 | 2.77 3.39 121 | 1.77 2.74 163 | 1.74 2.73 169 | 1.83 2.80 163 | 2.75 3.36 121 |
| 234432 | 1.77 2.73 162 | 1.80 2.79 175 | 1.72 2.69 163 | 2.80 3.43 122 | 1.77 2.74 164 | 1.75 2.73 168 | 1.86 2.82 164 | 2.76 3.38 121 |

Table 1. H-bonding and selected O–O distances (Å) and H-bond angles (O–H–O in degrees) in 6×3 , 234432 and 34443 models. Experiment values are from refs^{1,2}.

| | Origin | | | | Center | | | |
|--------|--------|--------|--------|---------|--------|--------|--------|---------|
| | Φ | Ψ | χ | χ' | Φ | Ψ | χ | χ' |
| Expt. | -88.7 | -147.1 | 158.0 | -83.0 | -98.5 | -142.3 | 170.0 | -70.0 |
| 333333 | -93.3 | -146.1 | 168.9 | -73.4 | -90.3 | -144.1 | 175.2 | -67.1 |
| 34443 | -93.2 | -146.3 | 168.0 | -74.2 | -90.3 | -144.1 | 176.2 | -65.7 |
| 234432 | -93.4 | -146.5 | 170.6 | -71.6 | -90.0 | -143.7 | 176.5 | -65.0 |

Table 2. $I\beta$ cellulose torsion angles (degrees) in 6×3 , 234432 and 34443 models. Φ : O5C1O4C4, Ψ : C1O4C4C5, χ : O5C5C6O6, χ' : C4C5C6O6. Experiment values are from refs^{1,2}.

| | | 34443 | 234432 | 6×3 |
|---------------------------|---------------|-------|--------|--------------|
| <i>In vacuum</i> | | | | |
| VASP | Rel. E | 52 | 0 | 371 |
| | Rel. E/glucan | 0.7 | 0 | 5.1 |
| M05-2X/6-31 G(d) | Rel. E | 0 | 244.8 | 356.3 |
| | Rel. E/glucan | 0 | 3.4 | 4.9 |
| B3LYP-D3/6-31 G(d) | Rel. E | 0 | 260.4 | 435.8 |
| | Rel. E/glucan | 0 | 3.6 | 6.1 |
| B3LYP/6-31 G(d) | Rel. E | 0 | 335.6 | 291.4 |
| | Rel. E/glucan | 0 | 4.7 | 4.0 |
| mpw1pw91/6-31 G(d) | Rel. E | 0 | 316.6 | 276.6 |
| | Rel. E/glucan | 0 | 4.4 | 3.9 |
| <i>In water (IEF-PCM)</i> | | | | |
| M05-2X/6-31 G(d) | Rel. E | 0 | 134.4 | 280.3 |
| | Rel. E/glucan | 0 | 1.9 | 3.9 |
| B3LYP-D3/6-31 G(d) | Rel. E | 0 | 155.0 | 363.9 |
| | Rel. E/glucan | 0 | 2.2 | 5.1 |
| B3LYP/6-31 G(d) | Rel. E | 0 | 230.3 | 219.6 |
| | Rel. E/glucan | 0 | 3.2 | 3.1 |

Table 3. Calculated energies of 6×3 , 234432 and 34443 models with various methods (kJ/mol). Rel. E: relative potential energy, Rel. E/glucan: relative potential energy normalized on a per glucan unit base. The lowest energy obtained from each method was set to 0 kJ/mol.

| | Observed | | 333333 | | 34443 | | 234432 | |
|-------------------------|----------|-------|--------|--------|--------|--------|--------|--------|
| | c | b | Origin | Center | Origin | Center | Origin | Center |
| C1 | 104.1 | 104.8 | 106.0 | 102.8 | 105.4 | 102.5 | 105.2 | 102.4 |
| C2 | 71.8 | 72.6 | 71.1 | 71.0 | 70.3 | 69.8 | 70.8 | 70.8 |
| C3 | 75.2 | 75.4 | 72.5 | 72.7 | 73.2 | 73.9 | 72.8 | 72.6 |
| C4 | 88.1 | 88.8 | 85.8 | 84.3 | 85.7 | 84.7 | 84.7 | 83.9 |
| C5 | 71.2 | 72.6 | 72.4 | 71.6 | 73.0 | 70.9 | 72.5 | 71.4 |
| C6 | 65.9 | 65.5 | 64.0 | 66.7 | 65.5 | 65.9 | 64.3 | 66.6 |
| Slope | | | 1.06 | 0.93 | 1.02 | 0.94 | 1.03 | 0.92 |
| Intercept | | | -5.41 | 4.07 | -2.23 | 2.59 | -3.05 | 4.51 |
| Correlation coefficient | | | 0.993 | 0.993 | 0.993 | 0.996 | 0.992 | 0.993 |
| MAE | | | 1.8 | 2.2 | 1.6 | 2.1 | 1.8 | 2.4 |
| RMSE | | | 1.9 | 2.5 | 1.7 | 2.4 | 2.0 | 2.7 |
| MaxE | | | 2.7 | 4.5 | 2.4 | 4.1 | 3.4 | 4.9 |

Table 4. Observed and calculated ^{13}C NMR chemical shifts (ppm) for interior cellulose chains. Observed = Wang *et al.*⁸. Ideal slope, intercept and correlation coefficient values are 1.0, 0.0 and 1.0, respectively. MAE = Mean Absolute Error, RMSE = Root-Mean-Squared Error, and MaxE = Maximum Error.

(83.9 \pm 0.4–85.8 \pm 0.6 ppm). Hence, the shift from 88.5 to 91.5 ppm to 84 to 86 ppm may not indicate interior versus exterior hydroxymethyl groups³⁴. Furthermore, the shift is not necessarily the result of *tg* to *gt* rotations on the surface because none of the hydroxymethyl groups in our models are in a *gt* configuration. Instead, the difference may be caused by H-bonding or the lack thereof as well as other structural changes^{34,35}.

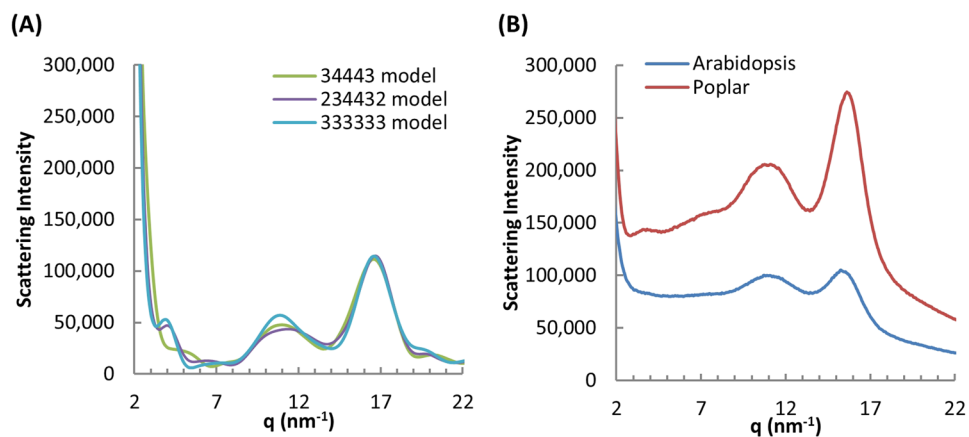


Figure 2. Simulated and experimental WAXS diffractograms for cellulose I. **(A)** All three CMF habits predict similar peaks near 10 to 11 and 16 nm^{-1} . The 34443 model is distinct however in not producing a WAXS peak near 4 nm^{-1} . **(B)** Wide-angle X-ray scattering from *Arabidopsis* and *Poplar*.

Wang *et al.*³⁶ have found that some interior glucan chains in CMFs are more than one chain away from the nearest surface chains. Hence, the CMF of primary cell wall must have two types of interior cellulose chains: a core fraction not directly contacting with the surface and a bound fraction directly contacting with the surface. As shown in Supplementary Fig. 1, only the 34443 model has the core fraction, because chain A in 34443 is an interior chain, based on the calculated NMR chemical shifts as shown in Supplementary Table 2. 234432 and 333333 models do not have the core fraction. In addition, since chain A in 34443 is an interior chain, the ratio of surface chain and interior chain (s:i) for model 34443 is ~ 1.57 , which is in good agreement with the s:i ratios of CMF in primary cell walls that Wang *et al.* found using solid-state NMR techniques. Therefore, based on the core domain and s:i ratio, the 34443 model is preferred.

WAXS data were instrumental in questioning the original 36-chain CMF model, so the models in this study were used to generate synthetic diffractograms for direct comparison with experimental observations. The theoretical diffraction patterns of the three CMF habits had peaks at ~ 10 to 11 nm^{-1} and 16 nm^{-1} similar to the experimental data for *Arabidopsis* and *Poplar* (Fig. 2). The 34443 model is distinct however in not producing a WAXS peak near 4 nm^{-1} (Fig. 2A). In addition, the 4 nm^{-1} peak is independent of the length of the cellulose microfibril models applied in WAXSiS calculations (as shown in Supplementary Fig. 2). It is also independent of the surrounding environment (water or vacuum) of CMF, as shown in Supplementary Fig. 4. Interestingly, when comparing to experimental WAXS results, we found that the WAXS diffractogram of *Poplar* cell walls demonstrated a weak peak around 4 nm^{-1} , whereas that of *Arabidopsis* did not show any peaks around that area. This is probably due to the difference in average electron density in real sample and inter-microfibril correlation. However, due to the complexity of plant cell walls, it is challenging to compare the simulation results with the experiments *in planta* especially in the q -range approaching the size of microfibrils. The peak around 4 nm^{-1} pointed us a possible way to differentiate CMF structures. However, more experiment and simulation work would be expected to further explore the origin of the peak.

Conclusion

In conclusion, DFT calculations on the structures, energies and spectroscopic parameters of three proposed CMF habits all predict accurate structures and $\delta^{13}\text{C}$ NMR values. Based comparison of sC4 and sC6 calculated versus observed $\delta^{13}\text{C}$, the 34443 model is slightly favored. Energetic considerations discount the 6×3 model as least favorable with periodic and molecular cluster calculations providing indistinguishable results on the 234432 and 34443 models; however, the calculated energy results favor 34443. WAXS diffractograms of the 34443 model are more consistent with observed patterns. We conclude the preponderance of evidence suggests that the 34443 habit is most probable because it cannot be eliminated from consideration based on comparisons to structural and ^{13}C NMR data nor energy estimations via DFT. The most distinguishing feature is the lack of a WAXS peak near 4 nm^{-1} which is predicted to exist for the other two models considered.

Methods

Density functional theory calculations in this study were performed with the Vienna Ab-initio Simulation Package (VASP^{37–40}) and Gaussian 09²⁵ codes. Initial atomic structures were based upon the experimental model of Nishiyama *et al.*¹ as supplied by Yoshi Nishiyama. The unit cell of cellulose was doubled to include 4 glucan units along the c -dimension, and 18 polymer CMFs were constructed in Materials Studio 2016. The simulation cell was expanded in the x - and y -dimensions to allow for space between the images of the CMFs in order to minimize any self-interactions. The model structures were converted into VASP input files (POSCAR) via a perl script written by Andrei V. Bandura (Saint Petersburg State University, Russia). All atomic positions were relaxed during energy minimizations using the VASP code. The simulation cell lattice parameters were also relaxed until the c -dimension converged to a stable state in order to the cellulose chains to stretch or contract. After convergence

of the chain lengths, the lattice parameters were then held constant in order to prevent artificial aggregation of the CMFs with their periodic images (i.e., contraction of the a and b cell dimensions until the CMF would sense images of itself.)

NMR calculations used the multi-standard method of⁴¹ with tetramethylsilane (TMS) and methanol as the primary and secondary standards^{4,26,34,42,43} in order to increase accuracy in reproducing experimental values. The modified Perdue–Wang exchange–correlation functional mPW1PW91⁴⁴ along with the 6–31 G(d) basis set⁴⁵ were used in the calculation of isotropic chemical shieldings for all nuclei. Only the added terminal methyl groups were relaxed in Gaussian 09 while all other atoms were fixed in the positions determined in the periodic energy minimizations performed with VASP. The 18-chain CMF structures (each chain contains 4 glucose units) were solvated in a box of TIP3P water⁴⁶ followed the protocol detailed out in previous work³⁵ using the Impact module of Maestro⁴⁷, followed by 1000 steps of conjugate gradient minimization performed before 10 ps molecular dynamics simulations at 298 K with a 1 fs time step in the NVT ensemble using the OPLS_2005 force field⁴⁸ keeping the atomic positions of cellulose molecules fixed. Only H₂O molecules within 3 Å of the cellulose clusters were included for NMR calculations as it was found that H₂O molecules further than 3 Å from the clusters had negligible effects on the calculated $\delta^{13}\text{C}$. An empirical correction of 49.5 ppm⁴³ was used for the difference between the $\delta^{13}\text{C}$ in methanol and TMS commonly used as an experimental ¹³C NMR standard⁴¹. This gives an isotropic chemical shielding of 193.0 ppm. To compute the $\delta^{13}\text{C}$ for any C nucleus i in cellulose, we used: $\delta^{13}\text{C}_i = 193.0 \text{ ppm} - \delta^{13}\text{C}_i$.

Small- and Wide-Angle X-ray Scattering curves of CMFs in solution were calculated based on explicit-solvent all-atom molecular dynamics (MD) simulations using WAXSiS^{23,24}.

The 18-chain CMF structures were obtained from the periodic energy minimizations performed with VASP. Each chain contained 19 glucose residues. Each CMF structure was solvated in a box of TIP3P water so that the CMF was surrounded in all directions by at least 20 Å. Short (50 ps) MD simulations were performed using AMBER⁴⁹ and the Glycam06 carbohydrate force field⁵⁰ at 300 K, at 1 bar, in the NPT ensemble, with 2 fs time step, 10 Å non-bonded interaction cutoff, and SHAKE-constrained hydrogen bonds. The positions of CMF were constrained using a force constant of 250 kcal/mol/Å². Simulation frames are written every 0.5 ps so that the solvent configurations are uncorrelated²³. From the MD trajectories, the theoretical scattering curves were then calculated using WAXSiS webserver. Two buffer subtraction methods have been used to calculate the scattering intensity: (i) total buffer intensity is subtracted, $I(q) = I_{\text{sample}}(q) - I_{\text{buffer}}(q)$; (ii) the buffer intensity is reduced by the volume fraction v of the solute, $I(q) = I_{\text{sample}}(q) - (1 - v) I_{\text{buffer}}(q)$ ²³. The two subtraction methods only slightly differ at small angles. As shown in Supplementary Fig. 3, total buffer subtraction method predicted 200 peaks at 15.8–16 nm⁻¹, and buffer scattering reduced by solute volume resulted in the 200 peaks at 16.5 nm⁻¹, very close to the widely accepted value for 200 of cellulose I (around 16 nm⁻¹)^{13,51–53}. But they highly differ at wide angles around the water scattering peak (20 nm⁻¹)²³. As shown in Supplementary Fig. 3, total buffer subtraction method resulted in negative intensities around 20 nm⁻¹. However, both predicted the peak around 4 nm⁻¹. The CRYSol⁵⁴ program was also used for calculation of theoretical diffraction patterns of the three CMF habits in water (with solvent density of 0.334 e/Å³) and in vacuum based on the atomic coordinate of the CMF structures obtained from the periodic energy minimizations performed with VASP.

The *Poplar* tension wood sample was obtained as described in Sawada *et al.*⁵⁵. The bottom 1 cm section of a seven week old stem from *Arabidopsis thaliana cold* was dried at 80 °C overnight before measurement. The XRD profiles were recorded as described in Sawada *et al.*⁵⁵.

References

- Nishiyama, Y., Langan, P. & Chanzy, H. Crystal structure and hydrogen-bonding system in cellulose I β from synchrotron X-ray and neutron fiber diffraction. *Journal of the American Chemical Society* **124**, 9074–9082 (2002).
- Nishiyama, Y., Johnson, G. P., French, A. D., Forsyth, V. T. & Langan, P. Neutron crystallography, molecular dynamics, and quantum mechanics studies of the nature of hydrogen bonding in cellulose I β . *Biomacromolecules* **9**, 3133–3140 (2008).
- Nishiyama, Y., Sugiyama, J., Chanzy, H. & Langan, P. Crystal structure and hydrogen bonding system in cellulose I α from synchrotron X-ray and neutron fiber diffraction. *Journal of the American Chemical Society* **125**, 14300–14306 (2003).
- Kubicki, J. D., Mohamed, M. N.-A. & Watts, H. D. Quantum mechanical modeling of the structures, energetics and spectral properties of I α and I β cellulose. *Cellulose* **20**, 9–23 (2013).
- Lee, C. M. *et al.* Hydrogen-bonding network and OH stretch vibration of cellulose: comparison of computational modeling with polarized IR and SFG spectra. *The Journal of Physical Chemistry B* **119**, 15138–15149 (2015).
- Cosgrove, D. J. Re-constructing our models of cellulose and primary cell wall assembly. *Current opinion in plant biology* **22**, 122–131 (2014).
- Jarvis, M. C. Structure of native cellulose microfibrils, the starting point for nanocellulose manufacture. *Phil. Trans. R. Soc. A* **376**, 20170045 (2018).
- Wang, T., Yang, H., Kubicki, J. D. & Hong, M. Cellulose Structural Polymorphism in Plant Primary Cell Walls Investigated by High-Field 2D Solid-State NMR Spectroscopy and Density Functional Theory Calculations. *Biomacromolecules* (2016).
- Harris, D. M. *et al.* Cellulose microfibril crystallinity is reduced by mutating C-terminal transmembrane region residues CESA1A903V and CESA3T942I of cellulose synthase. *Proceedings of the National Academy of Sciences* **109**, 4098–4103 (2012).
- Lee, C. M., Kaffle, K., Huang, S. & Kim, S. H. Multimodal broadband vibrational sum frequency generation (MM-BB-V-SFG) spectrometer and microscope. *The Journal of Physical Chemistry B* **120**, 102–116 (2015).
- Herth, W. & Weber, G. Occurrence of the putative cellulose-synthesizing “rosettes” in the plasma membrane of Glycine max suspension culture cells. *Naturwissenschaften* **71**, 153–154 (1984).
- Newman, R. H., Hill, S. J. & Harris, P. J. 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 physiology* **163**, 1558–1567 (2013).
- Fernandes, A. N. *et al.* Nanostructure of cellulose microfibrils in spruce wood. *Proceedings of the National Academy of Sciences* **108**, E1195–E1203 (2011).
- Nixon, B. T. *et al.* Comparative structural and computational analysis supports eighteen cellulose synthases in the plant cellulose synthesis complex. *Scientific reports* **6**, 28696 (2016).
- Hill, J. L., Hammudi, M. B. & Tien, M. The Arabidopsis cellulose synthase complex: a proposed hexamer of CESA trimers in an equimolar stoichiometry. *The Plant Cell* **26**, 4834–4842 (2014).

16. Vandavasi, V. G. *et al.* A structural study of CESA1 catalytic domain of Arabidopsis cellulose synthesis complex: evidence for CESA trimers. *Plant physiology* **170**, 123–135 (2016).
17. Yang, H., Zimmer, J., Yingling, Y. G. & Kubicki, J. D. How Cellulose Elongates - A QM/MM Study of the Molecular Mechanism of Cellulose Polymerization in Bacterial CESA. *The Journal of Physical Chemistry B* **119**, 6525–6535 (2015).
18. Sethaphong, L. *et al.* Tertiary model of a plant cellulose synthase. *Proceedings of the National Academy of Sciences* **110**, 7512–7517 (2013).
19. Morgan, J. L., Strumillo, J. & Zimmer, J. Crystallographic snapshot of cellulose synthesis and membrane translocation. *Nature* **493**, 181–186 (2013).
20. Li, Y., Lin, M. & Davenport, J. W. Ab initio studies of cellulose I: crystal structure, intermolecular forces, and interactions with water. *The Journal of Physical Chemistry C* **115**, 11533–11539 (2011).
21. Bučko, T., Tunega, D., Ángyán, J. G. & Hafner, J. Ab initio study of structure and interconversion of native cellulose phases. *The Journal of Physical Chemistry A* **115**, 10097–10105 (2011).
22. Goldberg, R. N. *et al.* A thermodynamic investigation of the cellulose allomorphs: Cellulose (am), cellulose I β (cr), cellulose II (cr), and cellulose III (cr). *The Journal of Chemical Thermodynamics* **81**, 184–226 (2015).
23. Knight, C. J. & Hub, J. S. WAXSiS: a web server for the calculation of SAXS/WAXS curves based on explicit-solvent molecular dynamics. *Nucleic acids research* **43**, W225–W230 (2015).
24. Chen, P.-c. & Hub, J. S. Validating solution ensembles from molecular dynamics simulation by wide-angle X-ray scattering data. *Biophysical journal* **107**, 435–447 (2014).
25. Frisch, M. *et al.* Gaussian 09, revision B. 01. Gaussian, Inc., Wallingford, CT (2010).
26. Watts, H. D., Mohamed, M. N. A. & Kubicki, J. D. A. DFT study of vibrational frequencies and ¹³C NMR chemical shifts of model cellulosic fragments as a function of size. *Cellulose* **21**, 53–70 (2014).
27. Grimme, S., Hansen, A., Brandenburg, J. G. & Bannwarth, C. Dispersion-corrected mean-field electronic structure methods. *Chemical reviews* **116**, 5105–5154 (2016).
28. Grimme, S., Ehrlich, S. & Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *Journal of computational chemistry* **32**, 1456–1465 (2011).
29. Perdew, J. P. & Wang, Y. Accurate and simple analytic representation of the electron-gas correlation energy. *Physical Review B* **45**, 13244 (1992).
30. Zhao, Y., Schultz, N. E. & Truhlar, D. G. Design of density functionals by combining the method of constraint satisfaction with parametrization for thermochemistry, thermochemical kinetics, and noncovalent interactions. *Journal of Chemical Theory and Computation* **2**, 364–382 (2006).
31. Becke, A. D. A new mixing of Hartree–Fock and local density-functional theories. *The Journal of chemical physics* **98**, 1372–1377 (1993).
32. Lee, C., Yang, W. & Parr, R. G. Development of the Colle–Salvetti correlation-energy formula into a functional of the electron density. *Physical review B* **37**, 785 (1988).
33. Cancès, E., Mennucci, B. & Tomasi, J. A new integral equation formalism for the polarizable continuum model: Theoretical background and applications to isotropic and anisotropic dielectrics. *The Journal of chemical physics* **107**, 3032–3041 (1997).
34. Kubicki, J. D., Watts, H. D., Zhao, Z. & Zhong, L. Quantum mechanical calculations on cellulose–water interactions: structures, energetics, vibrational frequencies and NMR chemical shifts for surfaces of I α and I β cellulose. *Cellulose* **21**, 909–926 (2014).
35. Yang, H. *et al.* Structural factors affecting ¹³C NMR chemical shifts of cellulose: a computational study. *Cellulose*, 1–14 (2017).
36. Wang, T. & Hong, M. Solid-state NMR investigations of cellulose structure and interactions with matrix polysaccharides in plant primary cell walls. *Journal of experimental botany* **67**, 503–514 (2015).
37. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical review B* **54**, 11169 (1996).
38. Kresse, G., Furthmüller, J. & Hafner, J. Theory of the crystal structures of selenium and tellurium: the effect of generalized-gradient corrections to the local-density approximation. *Physical Review B* **50**, 13181 (1994).
39. Kresse, G. & Hafner, J. Ab initio molecular dynamics for open-shell transition metals. *Physical Review B* **48**, 13115 (1993).
40. Kresse, G. & Hafner, J. Ab initio molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium. *Physical Review B* **49**, 14251 (1994).
41. Sarotti, A. M. & Pellegrinet, S. C. A multi-standard approach for GIAO ¹³C NMR calculations. *The Journal of organic chemistry* **74**, 7254–7260 (2009).
42. Watts, H. D., Mohamed, M. N. A. & Kubicki, J. D. Comparison of multistandard and TMS-standard calculated NMR shifts for coniferyl alcohol and application of the multistandard method to lignin dimers. *The Journal of Physical Chemistry B* **115**, 1958–1970 (2011).
43. Gottlieb, H. E., Kotlyar, V. & Nudelman, A. NMR chemical shifts of common laboratory solvents as trace impurities. *The Journal of organic chemistry* **62**, 7512–7515 (1997).
44. Adamo, C. & Barone, V. Exchange functionals with improved long-range behavior and adiabatic connection methods without adjustable parameters: The mPW and mPW1PW models. *The Journal of Chemical Physics* **108**, 664–675 (1998).
45. Rassolov, V. A., Ratner, M. A., Pople, J. A., Redfern, P. C. & Curtiss, L. A. 6-31G* basis set for third-row atoms. *Journal of Computational Chemistry* **22**, 976–984 (2001).
46. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* **79**, 926–935 (1983).
47. Schrödinger Release 2014-1: Maestro, version 9.7 (2014).
48. Banks, J. L. *et al.* Integrated modeling program, applied chemical theory (IMPACT). *Journal of computational chemistry* **26**, 1752–1780 (2005).
49. Case, D. A. *et al.* Amber **14** (2014).
50. Kirschner, K. N. *et al.* GLYCAM06: a generalizable biomolecular force field. *Carbohydrates. Journal of computational chemistry* **29**, 622–655 (2008).
51. Thomas, L. H. *et al.* Diffraction evidence for the structure of cellulose microfibrils in bamboo, a model for grass and cereal celluloses. *BMC plant biology* **15**, 153 (2015).
52. Crawshaw, J., Bras, W., Mant, G. & Cameron, R. Simultaneous SAXS and WAXS investigations of changes in native cellulose fiber microstructure on swelling in aqueous sodium hydroxide. *Journal of applied polymer science* **83**, 1209–1218 (2002).
53. Gubitosi, M. *et al.* Stable, metastable and unstable cellulose solutions. *Royal Society open science* **4**, 170487 (2017).
54. Svergun, D., Barberato, C. & Koch, M. H. CRYSOLE—a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of applied crystallography* **28**, 768–773 (1995).
55. Sawada, D. *et al.* Tension wood structure and morphology conducive for better enzymatic digestion. *Biotechnology for biofuels* **11**, 44 (2018).

Acknowledgements

This work is supported by the Center for Lignocellulose Structure and Formation, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences under Award # DE-SC0001090. Portions of this research were conducted with Advanced Cyber Infrastructure computational resources provided by the Institute for Cyber Science at The Pennsylvania State University (<http://ics.psu.edu>). This research also used resources of NERSC, supported by the “Office of Science of DOE under Contract No. DE-AC02-05CH11231”. We also acknowledge that the poplar sample in Fig. 2 was provided by the BER-funded BioEnergy Research Center.

Author Contributions

J.D.K. and D.C. conceived the study. J.D.K. conducted the VASP calculations. H.Y. performed the NMR calculations, and structural analysis. H.Y. and D.O. conducted WAXSiS and CRY SOL calculations. H.O. and S.D. performed the XRD experiments. J.D.K. and H.Y. wrote the paper and prepared the figures. All authors contributed edits and ideas, and approved the final manuscript version.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-32211-w>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018