New insights into the structure of cellulose in plant cell walls

Rosalie Cresswell[1], Parveen Kumar Deralia[2], Yoshihisa Yoshimi[2], Steven P. Brown[1], Ray Dupree[1], Paul Dupree[2]

1, Department of Physics, University of Warwick, Coventry CV4 7AL, UK;

2, Department of Biochemistry, University of Cambridge, Hopkins Building, The Downing Site, Tennis Court Road, Cambridge CB2 1QW, UK

Corresponding authors: Ray Dupree[1], Paul Dupree[2]

Abstract

The structure of native plant cellulose, despite its abundance and utility in industry, remains elusive. The cellulose structure of several species was studied using 2D solid-state Nuclear Magnetic Resonance (NMR) of ^{13}C labelled plants. Six major glucose environments were resolved which are common to the cellulose of poplar wood, spruce wood and grasses. The cellulose structure was maintained in isolated holo-cellulose nanofibrils, allowing more detailed characterisation. There are just two glucose environments within the fibril core which have the same NMR ^{13}C chemical shifts as tunicate cellulose Iβ. The third major glucose site with a carbon 4 shift near 89 ppm, previously assigned to the fibril interior, is one of four surface glucose environments. These advances allowed us to obtain a more accurate measure of the interior to surface ratio for poplar wood fibrils of 0.5, consistent with an 18 chain microfibril structure having 6 core and 12 surface chains.

Introduction

Cellulose is the most abundant natural polymer in the world.[1] It forms a major component in the cell walls of plants providing much of their mechanical strength.[2] Its properties are important for the pulp and lumber industries and more recently cellulose has become a major contender as a sustainable alternative to fossil fuels for producing materials, films and biofuels.[3–5] Despite the large industrial interest in cellulose, its structure in native plant cell walls remains elusive.

Cellulose is a β-1-4-linked polymer of D-glucose in long chains of perhaps 10,000 units.[6] These glucan chains have a 2-fold helical structure which allows them to crystallise via both stacking interactions and a network of hydrogen bonding to form long microfibrils in plant cell walls.[7] Estimates of the diameter of these cellulose microfibrils are between 3-5 nm.[8] In recent work, based on biophysical measurement and advances in understanding the biosynthesis, it is believed that a microfibril is formed from 18-24 glucan chains.[9–15] The arrangement of these chains in the microfibril into several stacked sheets also remains unknown, but computational studies have suggested some possible habits.[8,16]

Cellulose Iα and Iβ are well known forms of crystalline cellulose and have been characterised using both diffraction techniques and solid-state NMR.[17–21] Cellulose Iα has a P1 triclinic

structure and has non-equivalent glucosyl residues in a cellobiose unit within the same chain, whereas cellulose Iβ has a $P2_1$ monoclinic structure and its non-equivalent glucosyl residues are in alternating sheets of glucan chains, called centre and origin.[17,18] It is known that cellulose from bacteria and some algae have predominantly a cellulose Iα structure whilst tunicate cellulose, which consists of large crystals, is predominantly cellulose Iβ. Although it is possible to obtain high resolution X-ray and neutron diffraction patterns for these larger crystalline forms of cellulose, this is not achievable on native plant cellulose.[22–24] The very thin plant microfibrils result in low resolution diffraction patterns such that atomic resolution cannot be achieved.[24–26]

Solid-state NMR has been used for studying cellulose since the 1980s, mainly using 1D experiments.[27,28] It is particularly useful as, in contrast to diffraction, it does not require long-range order and samples can be analysed in situ. Nevertheless, the thin nature of the plant microfibrils means that the surface glucose residues contribute to the NMR spectra far more than in large crystals of tunicates or fibrils of bacteria, complicating attempts to fully assign the spectra. Furthermore, plant samples may contain non-cellulosic cell wall components that contribute to spectral peaks in the cellulose region, as well as a so-called amorphous cellulose component.[29,30] The cellulose fibrils also interact with other components in the cell wall which is likely to influence the chemical shifts of cellulose glucosyl residues.[31–33] Within these limitations, it is widely thought, based on NMR assignments and the fitting of X-ray diffraction patterns, that cellulose of higher plants is a mixture of mainly cellulose Iβ and some cellulose Iα[28] although there are also several studies that suggest plant cellulose could have its own distinct structure.[28,34,35]

In NMR spectra of cellulose, the carbon shifts of C4 and C6 of the different glucose units in cellulose are split into two main regions which, following Dupree et al.,[32] are defined as domain 1 glucosyl residues which have $C4^1$ and $C6^1$ shifts of ~89 ppm and ~65 ppm respectively, and domain 2 glucosyl residues which have $C4^2$ and $C6^2$ shifts of ~84 ppm and ~62 ppm. Domain 1 has been thought to reflect crystalline cellulose as its shifts are close to those of the crystalline cellulose Iα and Iβ[29,36], and these residues are normally also considered interior chains of the microfibril.[34] Domain 2 glucose residues are believed to be surface glucan chains of the cellulose microfibril and are also sometimes described as amorphous.[34,36] When restricted to 1D NMR it is only possible to resolve residues in these two domains using C4 and C6 shifts, as the C5 shifts overlap with those from C2 and C3.[32] The C4 cellulose region of the 1D NMR spectrum tends to be reasonably clear of contributions from hemicellulose and pectin etc. and so the ratio of the two domains ($C4^1$:$C4^2$) has been widely used as a measure of the crystallinity of cellulose and also to estimate the size of the cellulose microfibril.[37,38] It has been suggested [39] that the major source of this change in shift between the two domains arises from a change in the conformation of the hydroxymethyl group of the glucose unit from *tg* in domain 1 to *gt* or *gg* in domain 2 and this has since been supported by several studies using DFT or NMR.[16,40,41]

In recent years, developments in ^{13}C labelling of plant cell wall materials (and dynamic nuclear polarisation (DNP)) has enabled access to a wealth of 2D Magic Angle Spinning (MAS) NMR experiments that have allowed unparalleled insight into the structure of plant cell walls.[42] These studies have made significant strides in understanding hemicellulose structure and interactions with cellulose in both primary (PCW) and secondary cell walls (SCW).[43,44] NMR

has been used to identify the 2-fold helical structure of xylan in SCW, and the molecular architecture and water-polymer interactions in softwoods.[31,44,45] Several distinct glucosyl residues, each with differing chemical shifts due to differing environments within primary cell wall cellulose, were described by Wang et al. who found similar glucose environments in a range of different primary cell wall materials *(Brachypodium*, *Zea mays*, and *Arabidopsis thaliana*).[46] They reported that glucosyl residues of five environments are within the fibril, and two lie on the surface. They suggested several cellulose microfibril models based on the understanding that the domain 1 glucosyl residues (*tg* hydroxymethyl conformation) arise solely from sites in the interior chains of the microfibril.[41,47] Under this assumption, and using quantitative NMR, they found that cellulose had too much 'interior' relative to surface residues for the probable microfibril size of 18-24 chains. Additionally, some of the interior residues were more distant from water than others. As a result, the cellulose microfibrils bundle was suggested to produce more interior glucose residues.[37]

In this work we assign the NMR shifts of the main cellulosic glucose environments found in a range of plant secondary cell wall materials and begin to determine their position in the cellulose microfibril. To achieve this, we studied both poplar wood and xylanase-treated holo-cellulose nanofibrils (hCNFs) prepared from poplar wood using minimal processing to maintain the native cellulose structure.[48] By removing hemicellulose, lignin, and any amorphous cellulose present we are left with a clearer spectrum with improved resolution which provides more certainty in the determination of the NMR shifts for each distinct glucose environment. By comparing with the recently corrected NMR shift assignments for cellulose I we conclusively show that plant cellulose is not a mixture of Iα and Iβ, and that the core of the native cellulose microfibril has NMR shifts identical to tunicate cellulose 1β. We also show that the glucose residues in domain 1 and domain 2 cannot be split into crystalline and amorphous nor can they be considered to be exclusively interior and surface. In particular, we find that one of the domain 1 glucosyl residue environments is part of a surface glucan chain. This discovery allows a more realistic estimate of the interior to surface ratio of fibrils to be obtained than the many previous estimations using ^{13}C NMR.

Results

Assigning ^{13}C NMR chemical shifts for the glucosyl residue environments in cellulose within native plant cell walls

Throughout our work[31,45,49,50] never-dried native samples of many different plants have been studied using ^{13}C MAS NMR to understand the structure and interactions of cellulose hemicellulose, lignin and water in plant cell walls. The composition of the secondary cell wall of poplar is relatively simple with only 3 main components cellulose, xylan and lignin. Fig. 1 shows the 1D CP MAS NMR spectrum of the neutral carbohydrate region, which is dominated by the cellulose signal, of poplar wood. The carbon 1 (C1) peak is at ~105 ppm, C2, C3, C5 overlap in the region between 70-76 ppm and the C4 and C6 carbons are split into two main peaks. The cellulose glucose environments in domain 1 show $C4^1$ and $C6^1$ peaks at ~89 ppm and ~65 ppm respectively and the $C4^2$ and $C6^2$ domain 2 peaks are at ~84 ppm and ~62 ppm respectively. Interestingly the C4 peaks in the 1D CP MAS spectrum are not simple Lorentzian

line shapes indicating there are multiple glucose environments within both domains. The $C4^2$ region especially is clearly split into two peaks at ~83.5 and ~84.5 ppm.

Fully ^{13}C labelled poplar wood enabled the structure of plant cellulose to be explored using a wealth of 2D NMR experiments. 2D experiments such as CP INADEQUATE and 30 ms CP PDSD were used to resolve several different glucose environments within cellulose. The CP refocussed INADEQUATE is a double quantum experiment that correlates two covalently bonded carbons. The bonded carbons appear at the same DQ shift which is given by the sum of the respective SQ shift of the correlated carbons. Fig. 2 shows the neutral carbohydrate region of the CP refocussed INADEQUATE spectrum of poplar wood. The high resolution means that the carbons within each glucosyl residue can be followed through in the 2D spectrum. This is demonstrated in red for the glucose residue in domain 1 named environment 'c', and in black for the C4-C6 region for the residue in domain 2 environment 'j'. Our naming convention and assignments are based on those of Wang et al. [46] for cellulose in primary cell walls since many of the cellulose environments we observe have similar values, however glucose environment j, which is clearly visible for C4, C5 and C6, is newly reported. The C6 region of the spectrum (inset in Fig. 2) shows the main glucose environments identified within domain 1 and domain 2. This region also shows a minor domain 2 environment, named k. In total there are three major glucose environments a, b, c in domain 1 and three f, g, j in domain 2.

To make a more complete assignment of the NMR shifts of the glucose environments within cellulose, a 30 ms CP PDSD spectrum was analysed alongside the INADEQUATE spectrum. The CP PDSD experiment is a through space correlation experiment which, with a mixing time of 30 ms, shows cross peaks of carbons only if they are within the same glucose ring. SI Fig. 1 shows the 30 ms CP PDSD spectrum highlighting the key regions which are particularly useful for identifying different glucose environments. Two further minor glucose environments in domain 1 labelled d and e (which is very minor) have previously been assigned in Wang et al.[46]. Table 1 lists the ^{13}C NMR shifts for all the different glucose environments identified in cellulose, including the more minor ones.

Glucose environments in cellulose of cell walls of different plants

Having identified all the glucose environments in cellulose of poplar wood we were interested to compare these to cellulose of other plants. Over the years we have used high resolution 2D NMR to study a wide range of different plants including monocots, eudicots, and gymnosperms.[44,49,50] The cellulose of many of the plants investigated has remarkably similar NMR shifts such that it is possible to identify our assigned glucose environments in a wide range of plants. This is illustrated using a comparison of two different regions of the 30 ms CP PDSD spectra of a eudicot (poplar), a monocot (*Brachypodium*) and a gymnosperm (spruce) in Fig. 3 for the C1-C4 region and SI Fig. 2 for the C4-C6 region. The relative amount of each glucose environment varies between samples, for example, spruce wood appears to have relatively more b versus a or c whereas in poplar wood cellulose these are more similar in quantity. *Brachypodium* has significantly more of site d which is a minor cellulose site in the poplar wood secondary cell walls. Whilst not visible in Fig. 3 and SI Fig. 2, the cellulose environment e is a very minor environment in all three plants. There is one additional cellulose environment 's' seen in the spectra of spruce. There are some small differences in the shifts

for some of the more minor cellulose environments. For example, we found that environment d is generally broader than sites a, b, and c with its C4 NMR shift varying by ~ 0.3 ppm indicating that there are more variations in the d local environment. Generally, the NMR shifts remain similar between all the plants, indicating the glucose environments are constant in the cellulose fibrils.

Xylanase-treated holo-cellulose nanofibrils (hCNFs) maintain the native plant cellulose structure of poplar wood

Whilst studying the native plant cellulose in-situ is ideal for ensuring minimal disruption of the cell wall, the NMR spectrum can be crowded since signals from both hemicelluloses and lignin are present which tends to limit the resolution. The glucose environments in the fibrils may also be influenced by interactions of surface residues with hemicellulose or lignin. By removing the lignin during preparation of holo-cellulose nanofibrils (hCNFs) we can maintain the cellulose fibril structure whilst improving the resolution of the NMR spectrum.[48] To remove hemicellulose we treated hCNFs produced from the poplar wood with xylanase. TEM images of these hCNFs show long and thin cellulose fibrils which are loosely bundled, as observed with the 3nm width hCNFs prepared similarly from Arabidopsis see SI Fig. 3.[48] To investigate whether the xylanased hCNFs have maintained the cellulose environments we compared the NMR spectra of poplar wood with those of the xylanased hCNFs. Their 1D CP MAS spectra are shown in SI Fig. 4. There is a significant change in the total signal in the C4 region and in the ratio of domain 1 and domain 2 cellulose peaks. This change is due to the removal of both lignin and hemicelluloses as well perhaps as loss of a less ordered cellulose component.

The 2D 30 ms CP PDSD comparisons of the C1-C4 region and the C1-C6 region shown in Fig. 4 and SI Fig. 5 respectively show that the domain 1 glucose environments remain almost unchanged by the production of the hCNFs from wood. The domain 2 environments show some slight NMR shift changes, typically < 0.3 ppm, in both the f and g/j environments. As the domain 2 environments are surface chains of the fibril this may be due to the removal of both lignin and the hemicellulose xylan. The domain 2 cellulose region of the 30 ms CP PDSD spectrum shown in SI Fig. 5 exhibits distinctly narrower peaks for the xylanased hCNFs, i.e. a broad contribution of somewhat less ordered environments has been removed. Since there were no other changes, we are convinced that the production of the hCNFs causes relatively minimal disturbance to the different glucose environments in the poplar fibrils.

Plant cellulose fibril core environments are identical to tunicate cellulose Iβ

The improvement in the resolution of NMR spectra makes analysis of the xylanased hCNFs ideal for determining NMR shifts, quantities, and the relative location of the different glucose sites within the fibrils. There have been many 1D ^{13}C NMR studies of cellulose Iα and Iβ, but the use of different referencing has led to a spread in the published NMR shifts.[19,20,51,52] Recently Brouwer et.al. have resolved these discrepancies to give a consistent set of ^{13}C NMR values.[21] Using the same reference as Brouwer et al.[21] we can compare our shifts from the high-resolution spectra of xylanased hCNFs from poplar wood with those of cellulose Iα and Iβ. None of the domain 2 glucose environments are close to those of the NMR shifts of glucose in these cellulose allomorphs (SI Table 1). This means we only need to consider the ^{13}C NMR

shifts of domain 1 glucose residues, where there are only 3 major sites a, b and c. Fig. 5 shows the C4-C3/C5 region of the CP refocussed INADEQUATE spectrum of the xylanased hCNFs with the Iα and Iβ shift positions marked. It is evident that the shifts of cellulose Iα are different from those of native plant cellulose whereas the C3, C4 and C5 shifts of a and c are very close to those of Iβ. Indeed, all the shifts for sites a and c, apart from C1, are within ~0.2 ppm of those of cellulose Iβ (SI Table 1). Since the only substantial difference we observe from the Iβ shifts is in C1 where it seems likely that the Kono et al.[19,20] assignment was incorrect, as highlighted in SI Fig. 6 and SI Table 1. This misassignment by Kono et al.[20] presumably arose because a refocussed INADEQUATE spectrum was used for their assignment and both glucose environments in cellulose Iβ have nearly identical C2 NMR shifts making it difficult to distinguish the associated C1 shifts. Using the 30 ms CP PDSD spectrum (SI Fig. 6) together with the INADEQUATE spectrum allowed us to determine confidently the C1 shifts for environments a and c. In a later paper Brouwer and Mikolajewski also commented on the possibility that the Kono et al. C1 assignments could be swapped.[53] With this new NMR shift assignment of cellulose Iβ, all the shifts of fibril core glucose environments a and c match closely tunicate cellulose Iβ. We now assign environment a as origin chain, and environment c as centre chain, because the DFT calculations from the cellulose Iβ predict the C1 chemical shift of the residues in the origin chains is ~2 ppm higher than the C1 of residues in the centre chains, as seen here for environments a and c respectively.[46] Native plant cellulose is clearly not a mixture of cellulose Iα and Iβ since, despite environment b having similar shifts to one of the Iα glucose environments for both C1 and C4 (SI Table 2), there is no sign of the second Iα glucose environment which would also be visible in similar quantities within domain 1. This identification of glucose environments a and c now shows that the plant cellulose fibrils are cellulose Iβ.

Having determined that two of the main domain 1 environments correspond to glucose in the classical cellulose Iβ structure, we were interested to understand the origin of the third, remaining, domain 1 environment, b. Since b is from domain 1, it has been thought also to be interior to the fibril. To investigate this further, a 2D water-edited (w.e.) 30 ms CP PDSD spectrum was acquired to probe the glucose environments that are closest to water. Fig.6 shows the C1-C4 and C1-C6 regions of the w.e. spectrum compared with a standard 30 ms CP PDSD spectrum. As expected for surface glucose residues, the signal for the domain 2 sites f, g and j is enhanced indicating that they are more water accessible than the fibril core sites a and c. SI Fig. 7 shows similar proximity to water of these three major domain 2 glucose environments. Site j may be further from water than f and g, indicating these residues may be located on different surfaces of the fibril. Fig. 6a shows that the w.e. C1-C4 peak for glucose environment b is also significantly enhanced in a similar way to f, g, and j, indicating that b is also closer to water than the core sites a or c. Glucose environment b of domain 1 is therefore, unexpectedly, in a surface chain of the fibril. Interestingly, for the C1-C6 region shown in Fig. 6b the signal from b is still enhanced in the w.e. experiment over the signal for the core a and c environments, but to a lesser extent than for the domain 2 surface glucose environments. Being in domain 1, environment b likely reflects a glucose residue in a *tg* conformation. This conformation may be because the C6 is facing toward the interior of the cellulose fibril where the hydroxymethyl group hydrogen bonds with other residues.[16] Hence in the glucose residue of environment b the C6 is further from water than its C4, and further from water than the other surface environments where the C6 has a water-facing *gt* or *gg* conformation. Thus, in contrast to the widely accepted view, not all the domain 1 glucose residues are interior of the

fibril. A consequence of this finding is that the ratio of domain 1 and domain 2 signals cannot be used to estimate the size of a fibril since when site b is included in domain 1 the amount of interior will be overestimated, and surface underestimated. The ratio of interior vs surface of the fibril can however be estimated from the C1 region of the INADEQUATE spectrum where sites a and c are resolved from b, and interior a+c and surface D2+b can now be determined by integration (see SI Fig. 8). The value of ~ 0.5 in the poplar hCNFs is consistent with an 18-chain fibril with 6 interior chains and 12 surface chains. The relative amount of site b is ~ 0.7 of sites a or c (see SI Fig. 8).

To investigate the relative proximities of the different glucose environments, two longer (200 ms and 400 ms) mixing time CP PDSD experiments were acquired. As these experiments are probing distances up to ~ 5-8 Å, cross-peaks are between glucose residues in different sites within individual fibrils are additionally observed. Fig. 7 shows that at a mixing time of 200 ms the C1-C1 region of the CP PDSD spectrum gives clear cross peaks between a and c only. This shows that the residue environments a and c are closer to each other than either are to residue environment b. This is consistent with our finding that b is not situated with a and c in the core of the cellulose fibril. The first clear cross peaks between residues in the two cellulose domains can be seen in the C4-C4 region of the 400 ms CP PDSD (Fig. 8a). There are cross peaks from domain 2 glucose environments to the single $C4^1$ peak corresponding to glucose residue environments a plus b. Given we know that environments a and c are particularly close, and there are no cross peaks to c, the cross peaks from domain 2 observed here are likely to environment b only. This proximity of b and not c to domain 2 glucose residues is also observed in the C6-C6 region of the 400 ms CP PDSD spectrum (Fig. 8b). There are also cross peaks between the different domain 2 environments of cellulose. The proximities of b to the domain 2 environments provides further confirmation that b is one of the four major surface glucose residue environments.

Discussion

The structure of native plant cellulose fibrils has been the focus of many studies for decades. Although a wide range of techniques have been used, the structure remains to be fully elucidated.[30,54–56] The ^{13}C NMR chemical shift is very sensitive to the local environment such as differing conformation and hydrogen bonding. In principle, NMR therefore allows distinct and recurring glucose residues in the interior and surface of cellulose fibrils to be distinguishable. High resolution is needed to resolve signals from these residues, and was achieved here by studying never dried, hydrated, isolated ^{13}C labelled cellulose fibril samples with high field MAS NMR. We found plant cellulose is comprised of six major glucose environments, with four of these on the surface and just two in the core of the fibril. Importantly, these two core glucose residue sites have ^{13}C NMR shifts very similar to those of tunicate cellulose Iβ centre and origin chains and there is no detectable cellulose Iα, resolving longstanding questions about the composition of plant cellulose.

The core of the plant cellulose microfibril is comprised of two types of glucose residues, which lie in cellulose Iβ origin and centre chains. The plant fibril core glucose residue NMR shifts are essentially identical to those of tunicate cellulose Iβ. This finding relied in part on the unified and consistent NMR shifts of cellulose that have recently been determined by Brouwer et al.[21]

Second, we found that the C1 ^{13}C shift of the two glucose residues of the origin and centre chains of tunicate cellulose Iβ had previously been misassigned, as was also thought possible by Brouwer and Mikolajewski. [53] Our corrected assignment of the C1 shifts is also more consistent with DFT calculations of cellulose Iβ origin and centre chain environments.[57] We therefore propose the plant fibril glucose environments a and c correspond to cellulose Iβ origin and centre chain environments respectively (Table 1). Our NMR shifts for the plant cellulose fibril core are slightly smaller (~ 0.2 ppm) than the Brouwer values for tunicate cellulose Iβ.[21,53] Due to the thinness of the plant fibrils, all interior chains are next to surface chains, in contrast to the large crystals of tunicate cellulose that were studied by Kono et al.[20] Therefore the small NMR shift difference could result from these fibril core chains having neighbouring chains, similarly stacked as cellulose Iβ, but which are surface and hence they have different water interactions and bonding. [19–21,53]

Glucose residues with shifts in domain 1 (C4 around 89ppm) have long been considered to reside in the crystalline core of plant cellulose.[37,47] However, now we have shown that the third major glucose environment in domain 1, site b, is a residue on the surface of the fibril. We previously suggested[31] that there could be a surface component in the domain 1 region and recently this view has been supported by Addison et al.[33,58] Indeed, the NMR spectra of cotton of Kirui et al. have relatively less site b than found in the samples analysed here, consistent with a larger microfibril structure with lower surface to interior ratio.[34] Calculations have shown that a change in C6 hydroxymethyl conformation from *tg*, to *gt/gg* changes the C4 shift by ~ 5 ppm.[45] Adoption of the *tg* conformation is likely to occur where the C6 is facing towards the inside of the fibril and the orientation is fixed by hydrogen bonding to other glucosyl residues rather than water. This indicates that the glucose residue in environment b has the *tg* conformation and is likely to be a residue in a surface glucan chain but facing inwards towards another chain. It is now clearly more appropriate to consider domain 1 glucose residues as reflecting *tg* C6 hydroxymethyl confirmation rather than present only in crystalline cellulose and only in the core of fibril.

The earlier assignment of all types of domain 1 glucose residues as solely interior to the cellulose fibril has led to a number of debates concerning the core structure.[9,34,37,46,47] By comparing 1D NMR spectra in the domain 1 region, it was widely thought that the plant cellulose core is a mixture of cellulose Iα and cellulose Iβ.[51,54,59,60] This misinterpretation arose due to the glucose in environment b having similar C1 and C4 shifts as those of one of the glucose residues of cellulose Iα. Furthermore, the 1D spectrum appears significantly different to that of cellulose Iβ because the C4 and C6 signals from core residue a and surface residue b overlap. This means that the ratio of the core cellulose Iβ environments (now known to be a and c) appears not to be the expected 1:1. The mixture of cellulose of Iα and Iβ has been hypothesised to arise as faults in the fibril structure or by alternating sheets of cellulose Iα and Iβ.[29,54,55] There have also been suggestions that the core structure could be completely distinct from that of cellulose Iβ.[34,37] We have now shown the fibril is solely cellulose Iβ, and so cellulose modelling, computational studies and interpretation of diffraction patterns should be based on the Iβ structure.

The breadth of studies across plant and material sciences has given rise to terminology to describe cellulose which can be confusing, especially when the same terminology is applied to data acquired by different methods. For example, descriptions of cellulose as crystalline

and amorphous have different meanings between NMR studies and diffraction experiments which require order over a much longer distance.[61] Historically the glucose residues observed in the two domains of NMR spectra were assigned as crystalline and amorphous cellulose because the characterised crystalline forms of cellulose have ^{13}C shifts in the domain 1 (C4~89 ppm) region.[38,62,63] Surface chains are more mobile and are sometimes described as paracrystalline or amorphous.[25] In our work the line widths of the signals from the six glucose residue environments within the surface and core of the fibril are very similar (such as in the C6 region of the CP INADEQUATE spectrum of the xylanase treated hCNFs shown in SI Fig. 9 summarised in SI Table 2). NMR chemical shifts are sensitive only to short range structure (< ~ 5 A), and so the similar linewidths indicate that both domains have similar short-range order. Therefore, the core and surface glucose residues cannot be divided simply into crystalline and amorphous cellulose. On the other hand, a comparison of the 1D CPMAS spectrum of poplar wood versus the xylanase-treated hCNF sample in Fig. SI4 shows that a weak (<10%) broad signal underlying the domain 2 region, which may arise from some less ordered material, is removed in the production of the hCNF sample. This less ordered component includes lignin and xylan but also could be related to a less crystalline portion of cellulose that is known to be easily hydrolysed.[64] Therefore, whilst there might be a small component of less ordered cellulose giving rise to signal in the domain 2 region this term cannot describe all these glucose residues. The ratio of the $C4^1$ and $C4^2$ signals from 1D NMR spectra was initially called the crystallinity index (CI). The CI has been used to describe the crystallinity of cellulose and is regularly calculated alongside crystallinity measures by diffraction techniques.[38,62] This measure of crystallinity by NMR and diffraction has usually been significantly different, although the general trends between samples tend to be consistent.[38] The traditional CI as measured by NMR is a misnomer and does not measure crystallinity. It will be influenced by the surface to core ratio, as well as the presence of any amorphous material.

The better estimates of glucose residue environment proportions provide new constraints on plant cellulose fibril size and habit. Previous studies have used the ratio of $C4^1$ and $C4^2$ from 1D NMR spectra as a measure of microfibril size, which has resulted in variable estimations such as 24 chains in a fibril or in fibril bundling models.[9,37] As we now know domain 1 is not solely interior sites, thus the ratio of $C4^1$ and $C4^2$ will give an overestimate of the interior of the fibril and an overestimate of the fibril size. The ratio of the two domains could still be utilised for characterisation as it is still loosely related to the volume of the fibril, but care is needed in using this interpretation. In this work, we have provided a more accurate measure of the interior to surface ratio for the microfibril of poplar wood which was found to be ~0.5. We also provide new constraints on the proportions of the environments in the fibril, namely a and c origin and centre chains are similar in proportion, and surface environment b is ~0.7 of the amount of a or c chain environments (since the number of b is likely an integer, this ratio corresponds to 2/3 or 3/4). Given the further constraints on the fibril from the dimensions from biophysical measurements that limit the size to 18-24 chains, there are very few habits that fulfil these criteria. The core to surface value is consistent, for example, with an 18-chain fibril only in a habit that has 6 interior chains and 12 surface chains, or a 24 chain fibril with 8 interior and 12 surface chains. A schematic diagram of two potential 18 chain habits 234432 and 34443 that have the observed interior to surface ratio is shown in Fig. 9, and an alternative 333333 habit can be excluded. These habits fulfil our updated estimate of interior to core chain ratio, but it is difficult to see how the 34443 could have similar amounts of a and c origin and centre sheets. Currently therefore, 234432 is the habit currently that best

fits the data. The previously proposed concepts of multi-layered fibril environments[60] or bundling[37] to generate additional interior chains are not required to explain the data.

Whilst we have measured the relative amount of surface site b compared to core chain sites a and c for isolated fibrils from poplar wood, these values are likely to vary with any changes in the microfibril size, habit, and hemicellulose interactions in different plants. Hence the relative proportions of the six major glucose residue environments will be an important diagnostic tool for discovering cellulose variability, studying cellulose interactions, and for detecting changes during degradation or industrial processing of wood and other biomass. This work now provides a strong basis for understanding cellulose fibril surfaces as we discovered there are only four major types of surface residues. Understanding the origin of these NMR environments will be very instrumental in defining the hydrophobic and hydrophilic surfaces of fibrils. The exciting prospect now arises of a more complete description and understanding of cellulose fibrils surfaces and their interactions in biomaterials and in the plant cell wall.

Methods

Sample Production and Preparation

The poplar stem *Populus tremula* × *tremuloides* used in this work was grown from poplar shoot cuttings which were allowed to root for 2 weeks before being transferred into the growth chamber. Here the saplings were grown in a $^{13}CO_2$ atmosphere[31,65] for ~ 3 months to provide ~97% ^{13}C labelled material that is predominantly secondary cell wall. Only wood of lower (woody) never-dried stems was used after removing bark and cambium. The growth of spruce and *Brachypodium* was as described in Terrett et al. and Tryfona et al. respectively.[44,49]

To prepare holocellulose nanofibrils (hCNFs), the wood was first delignified four times with 3% peracetic acid (PAA; 0.35 g pure PAA/g dry material, pH 4.8) at 85 °C for 45 minutes without stirring. Between cycles, used PAA was decanted, one water wash was done, and new 3% PAA was added. Delignified material (holocellulose) was washed with water after the fourth cycle until the conductivity was below 10 S/cm. The holocellulose was blended for two minutes to form a homogenous slurry. Holocellulose dispersion (0.1 wt%) was blended for 30 minutes in a Vitamix A3500i blender to make a polydispersion. hCNFs were recovered as supernatant from the polydispersion by centrifugation at 5000 rpm for 15 minutes. The hCNFs suspensions were digested with xylanases as described previously.[48] The xylanase treated hCNFs were then freeze-dried and rewetted before being packed directly into the NMR rotor.

The never dried ^{13}C enriched poplar wood was debarked and cut into small pieces of ~ 1-2 mm size with a razor blade and then packed into a 3.2 mm MAS zirconia NMR rotor whilst removing excess water.

TEM

A ThermoFisher Scientific (FEI) TalosF200X G2 microscope operating in the scanning mode at 200 kV was used to obtain the TEM images.

Solid-state NMR

Solid-state NMR experiments of poplar wood and the xylanase treated hCNFs were acquired on a Bruker 1GHz AVANCE NEO solid-state NMR spectrometer operating at ^{1}H and ^{13}C Larmor frequencies of 1000.4 and 251.6 MHz respectively using a 3.2 mm E^{Free} triple resonance MAS probe. All experiments were conducted at an indicated temperature of 10°C and an MAS frequency of 12.5 kHz with a recycle delay of 2 s. The ^{13}C chemical shifts were determined using the carbonyl peak at 177.8 ppm of L-alanine as an external reference with respect to tetramethylsilane. This referencing was confirmed to correspond with referencing to the adamantane CH_2 peak to 38.48 ppm[66] to ensure direct comparison with Brouwer and Mikolajewski.[21,53] The 90° pulse lengths were typically 3.2 µs (^{1}H) and 4.0 µs (^{13}C). Cross polarization from ^{1}H to ^{13}C was achieved using ramped (70–100%) ^{1}H radiofrequency amplitude and a contact time of 1 ms[67]. SPINAL-64 decoupling was applied at a ^{1}H nutation frequency of 70-80 kHz during acquisition.[68] Sign discrimination in the indirect dimension of the 2D experiments was achieved using the States-TPPI method. For the principal assignments of cellulose environments, a CP refocused INADEQUATE ^{13}C double-quantum (DQ) ^{13}C single-quantum (SQ) correlation experiment was used whereby the evolution of DQ coherence for directly bonded carbons within the same glucan ring is correlated with directly observed SQ coherence.[69,70] The acquisition time in the indirect dimension was 6.67 ms with a spectral width of 37.5 kHz and 128 acquisitions per t_1 FID. The echo duration, τ, was 2.24 ms giving a total echo time of 8.96 ms. Both intra- and intermolecular contacts were probed using 2D ^{13}C-^{13}C proton driven spin diffusion (PDSD) experiments with mixing times of 30 to 400 ms.[71] The acquisition time in the indirect dimension (t_1) of the CP PDSD experiments was 5.5–8.1 ms. The spectral width in the indirect dimension was 37.5 kHz with at least 64 acquisitions per t_1 FID. The proximity of water to different cellulose environments was probed using a water-edited ^{13}C-^{13}C 30 ms CP PDSD experiment. This is based on the normal CP PDSD experiment, however before the CP contact time there is a ^{1}H T_2 filter to remove signal arising from directly attached protons followed by a delay to allow for diffusion from the water protons to its near neighbours.[72] The CP parameters were as stated above. The acquisition time in the indirect dimension was 4.8 ms with a spectral width of 37.5 kHz and 576 acquisitions per t_1 FID. The total ^{1}H T_2 filter was 320 µs followed by a diffusion delay of 2 ms. All 2D spectra were processed with Fourier transformation into 8 K (F_2) × 2 K (F_1) points with exponential line broadening of 20-50 Hz in F_2 and cubed sine bell processing in F_1 using Bruker Topspin v.3.6. Contour levels are x 1.1 throughout. The minimum contour is chosen to show the desired features.

Data availability

Unprocessed NMR data files will be available from http://wrap.warwick.ac.uk

References

1. Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science (1979)* **281**, 237–240 (1998).
2. Gibson, L. J. The hierarchical structure and mechanics of plant materials. *J R Soc Interface* **9**, 2749–2766 (2012).
3. Pauly, M. & Keegstra, K. Cell-wall carbohydrates and their modification as a resource for biofuels. *Plant Journal* **54**, 559–568 (2008).
4. Foroughi, F., Ghomi, E. R., Dehaghi, F. M., Borayek, R. & Ramakrishna, S. A review on the life cycle assessment of cellulose: From properties to the potential of making it a low carbon material. *Materials* **14**, 714 (2021).
5. Iglesias, M. C., Gomez-Maldonado, D., Via, B. K., Jiang, Z. & Peresin, M. S. Pulping processes and their effects on cellulose fibers and nanofibrillated cellulose properties: A review. *For Prod J* **70**, 10–21 (2020).
6. Klemm, D., Heublein, B., Fink, H. P. & Bohn, A. Cellulose: Fascinating biopolymer and sustainable raw material. *Angewandte Chemie - International Edition* **44**, 3358–3393 (2005).
7. Parthasarathi, R. *et al.* Insights into hydrogen bonding and stacking interactions in cellulose. *Journal of Physical Chemistry A* **115**, 14191–14202 (2011).
8. Kubicki, J. D. *et al.* The Shape of Native Plant Cellulose Microfibrils. *Sci Rep* **8**, 13983 (2018).
9. Tai, H. C. *et al.* Wood cellulose microfibrils have a 24-chain core–shell nanostructure in seed plants. *Nat Plants* **9**, 1154–1168 (2023).
10. Song, B., Zhao, S., Shen, W., Collings, C. & Ding, S. Y. Direct Measurement of Plant Cellulose Microfibril and Bundles in Native Cell Walls. *Front Plant Sci* **11**, 479 (2020).
11. Penttilä, P. A. & Paajanen, A. Critical comment on the assumptions leading to 24-chain microfibrils in wood. *Nature Plants* **10**, 1064–1066 (2024).
12. Jarvis, M. C. Cellulose biosynthesis: Counting the chains. *Plant Physiol* **163**, 1485–1486 (2013).
13. Purushotham, P. *et al.* A single heterologously expressed plant cellulose synthase isoform is sufficient for cellulose microfibril formation in vitro. *PNAS,* **113**, 11360–11365 (2016).
14. Nixon, B. T. *et al.* Comparative Structural and Computational Analysis Supports Eighteen Cellulose Synthases in the Plant Cellulose Synthesis Complex. *Sci Rep* **6**, 28696 (2016).
15. Cosgrove, D. J. *et al.* How many glucan chains form plant cellulose microfibrils? *Biomacromolecules* **in press**, (2024).
16. Yang, H. & Kubicki, J. D. A density functional theory study on the shape of the primary cellulose microfibril in plants: effects of C6 exocyclic group conformation and H-bonding. *Cellulose* **27**, 2389–2402 (2020).
17. Nishiyama, Y., Langan, P. & Chanzy, H. Crystal structure and hydrogen-bonding system in cellulose Iβ from synchrotron X-ray and neutron fiber diffraction. *J Am Chem Soc* **124**, 9074–9082 (2002).

18. Nishiyama, Y., Sugiyama, J., Chanzy, H. & Langan, P. Crystal Structure and Hydrogen Bonding System in Cellulose Iα from Synchrotron X-ray and Neutron Fiber Diffraction. *J Am Chem Soc* **125**, 14300–14306 (2003).
19. Kono, H. *et al.* CP/MAS ^{13}C NMR study of cellulose and cellulose derivatives. 1 . Complete assignment of the CP/MAS ^{13}C NMR spectrum of the native cellulose. *J Am Chem Soc* **124**, 7506–7511 (2002).
20. Kono, H., Erata, T. & Takai, M. Determination of the through-bond carbon-carbon and carbon-proton connectivities of the native celluloses in the solid state. *Macromolecules* **36**, 5131–5138 (2003).
21. Brouwer, D. H. & Mikolajewski, J. G. Resolving the discrepancies in reported ^{13}C solid state NMR chemical shifts for native celluloses. *Cellulose* **30**, 4827–4839 (2023).
22. Thomas, L. H. *et al.* Diffraction evidence for the structure of cellulose microfibrils in bamboo, a model for grass and cereal celluloses. *BMC Plant Biol* **15**, 158 (2015).
23. 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 Physiol* **163**, 1558–1567 (2013).
24. Šturcova, A., His, I., Apperley, D. C., Sugiyama, J. & Jarvis, M. C. Structural details of crystalline cellulose from higher plants. *Biomacromolecules* **5**, 1333–1339 (2004).
25. Thomas, L. H., Altaner, C. M. & Jarvis, M. C. Identifying multiple forms of lateral disorder in cellulose fibres. *J Appl Crystallogr* **46**, 972–979 (2013).
26. Thomas, L. H. *et al.* Structure and spacing of cellulose microfibrils in woody cell walls of dicots. *Cellulose* **21**, 3887–3895 (2014).
27. Deslandes, Y., Hamer, G. K. & Marchessault, R. H. Application of High-Resolution Solid-State NMR with Cross-Polarization/ Magic-Angle Spinning (CP/MAS) Techniques to Cellulose Chemistry. *Journal of Macromolecular Science, Part C* **23**, 187–216 (1983).
28. Atalla, R. H. & Vanderhart, D. L. Native Cellulose: A Composite of Two Distinct Crystalline Forms, *Science*, **223**, 283-285 (1984).
29. Newman, R. H. Evidence for Assignment of 13 C NMR Signals to Cellulose Crystallite Surfaces in Wood, Pulp and Isolated Celluloses, *Holzforschung,* **52**, 157-159 (1998).
30. Newman, R. H. Estimation of the lateral dimensions of cellulose crystallites using ^{13}C NMR signal strengths. *Solid State Nucl Magn Reson* **15**, 21–29 (1999).
31. Simmons, T. J. *et al.* Folding of xylan onto cellulose fibrils in plant cell walls revealed by solid-state NMR. *Nat Commun* **7**, 13902 (2016).
32. Dupree, R. *et al.* Probing the molecular architecture of *Arabidopsis thaliana* secondary cell walls using two- and three-dimensional ^{13}C solid state nuclear magnetic resonance spectroscopy. *Biochemistry* **54**, 2335–2345 (2015).
33. Addison, B. *et al.* Atomistic, macromolecular model of the Populus secondary cell wall informed by solid-state NMR. *Sci. Adv* **10**, 7965 (2024).
34. Kirui, A. *et al.* Atomic resolution of cotton cellulose structure enabled by dynamic nuclear polarization solid-state NMR. *Cellulose* **26**, 329–339 (2019).
35. Viëtor, R. J., Newman, R. H., Ha, M. A., Apperley, D. C. & Jarvis, M. C. Conformational features of crystal-surface cellulose from higher plants. *Plant Journal* **30**, 721–731 (2002).
36. Teeyaer, R. E., Zhbankov, R. G., Shipina, O. T., Marchenko, G. N. & Zabelin, L. V. NMR Spectroscopy of Cellulose and its Derivatives with Magic Angle Spinning of the sample (Review). *Zhurnal Prikladnoi Spektroskopii* **45**, 535–554 (1986).

37. Kirui, A. *et al.* Carbohydrate-aromatic interface and molecular architecture of lignocellulose. *Nat Commun* **13**, 538 (2022).
38. Park, S., Baker, J. O., Himmel, M. E., Parilla, P. A. & Johnson, D. K. Cellulose crystallinity index: measurement techniques and their impact on interpreting cellulase performance. *Biotechnol Biofuels* **3**, 10 (2010).
39. Fernandes, A. N. *et al.* Nanostructure of cellulose microfibrils in spruce wood. *PNAS* **108**, 1195–1203 (2011).
40. Phyo, P., Wang, T., Yang, Y., O'Neill, H. & Hong, M. Direct Determination of Hydroxymethyl Conformations of Plant Cell Wall Cellulose Using ^{1}H Polarization Transfer Solid-State NMR. *Biomacromolecules* **19**, 1485–1497 (2018).
41. Yang, H. *et al.* Structural factors affecting ^{13}C NMR chemical shifts of cellulose: a computational study. *Cellulose* **25**, 23–36 (2018).
42. Jarvis, M. C. NMR Studies on Cellulose and Plant Cell walls. in *Plant Cell Walls Research Milestones and Conceptual Insights* vol. 1 (CRC Press, 2023).
43. Wang, T., Zabotina, O. & Hong, M. Pectin-cellulose interactions in the arabidopsis primary cell wall from two-dimensional magic-angle-spinning solid-state nuclear magnetic resonance. *Biochemistry* **51**, 9846–9856 (2012).
44. Terrett, O. M. *et al.* Molecular architecture of softwood revealed by solid-state NMR. *Nat Commun* **10**, 4978 (2019).
45. Cresswell, R. *et al.* Importance of Water in Maintaining Softwood Secondary Cell Wall Nanostructure. *Biomacromolecules* **22**, 4669–4680 (2021).
46. 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* **17**, 2210–2222 (2016).
47. Kang, X. *et al.* Lignin-polysaccharide interactions in plant secondary cell walls revealed by solid-state NMR. *Nat Commun* **10**, 347 (2019).
48. Deralia, P. K. *et al.* Preserving native cellulose structure in isolated holocellulose nanofibrils and controlling their properties by tuning xylan structure for biomaterials. **In preparation**, (2024).
49. Tryfona, T. *et al.* Altering the substitution and cross-linking of glucuronoarabinoxylans affects cell wall architecture in Brachypodium distachyon. *New Phytologist* **242**, 524–543 (2024).
50. Bourdon, M. *et al.* Ectopic callose deposition into woody biomass modulates the nano-architecture of macrofibrils. *Nat Plants* **9**, 1530–1546 (2023).
51. Yamamoto, H. & Horii, F. CP/MAS ^{13}C NMR Analysis of the Crystal Transformation Induced for Valonia Cellulose by Annealing at High Temperatures. *Macromolecules* **26**, 1313–1317 (1993).
52. Larsson, T., Hult, E.-L., Wickholm, K., Pettersson, E. & Iversen, T. CPMAS ^{13}C-NMR spectroscopy applied to structure and interaction studies on cellulose I. *Solid State Nucl Magn Reson* **15**, 31–40 (1999).
53. Brouwer, D. H. & Mikolajewski, J. G. Solid-state NMR reveals a structural variation from the Iβ polymorphic form in bacterial cellulose. *Cellulose* **30**, 11341–11356 (2023).
54. Atalla, R. H. & Vanderhart, D. L. The role of solid state ^{13}C NMR spectroscopy in studies of the nature of native celluloses. *Solid State Nucl Magn Reson* **15**, 1–19 (1999).

55. Driemeier, C. & Francisco, L. H. X-ray diffraction from faulted cellulose I constructed with mixed Iα-Iβ stacking. *Cellulose* **21**, 3161–3169 (2014).
56. Hanley, S. J., Revol, J.-F. Ë. O., Godbout, L. & Gray, D. G. Atomic force microscopy and transmission electron microscopy of cellulose from Micrasterias denticulata; evidence for a chiral helical microfibril twist. *Cellulose,* **4**, 209-220 (1997)
57. Rehman, Z. *et al.* Organic NMR Crystallography: Enabling Progress for Applications to Pharmaceuticals and Plant Cell Walls. *Faraday Discuss* **Accepted Manuscript**, (2024).
58. Xue, Y., Li, H. & Kang, X. Molecular unraveling of polysaccharide digestion in wood-feeding termites: A solid-state NMR perspective. *Carbohydr Polym* **331**, 121843 (2024).
59. Smith, B. G., Harris, P. J., Melton, L. D. & Newman, R. H. Crystalline Cellulose in Hydrated Primary Cell Walls of Three Monocotyledons and One Dicotyledon. *Plant Cell Physiol* **39**, 711–720 (1998).
60. Larsson, T., Wickholm, K. & Iversen, T. A CP/MAS ^{13}C NMR Investigation of Molecular Ordering in Celluloses *Carbohydrate Research*, **302**, 19-25 (1997).
61. Bryce, D. L. NMR crystallography: Structure and properties of materials from solid-state nuclear magnetic resonance observables. *IUCrJ* **4**, 350–359 (2017).
62. Park, S., Johnson, D. K., Ishizawa, C. I., Parilla, P. A. & Davis, M. F. Measuring the crystallinity index of cellulose by solid state ^{13}C nuclear magnetic resonance. *Cellulose* **16**, 641–647 (2009).
63. Wickholm, K., Larsson, T. & Iversen, T. Assignment of non-crystalline forms in cellulose I by CP/MAS 13 C NMR spectroscopy. *Carbohydr Res* **312**, 123–129 (1998).
64. Chang, J. K. W., Duret, X., Berberi, V., Zahedi-Niaki, H. & Lavoie, J. M. Two-step thermochemical cellulose hydrolysis with partial neutralization for glucose production. *Front Chem* **6**, 117 (2018).
65. Chen, W. P. *et al.* An automated growth enclosure for metabolic labeling of Arabidopsis thaliana with ^{13}C-carbon dioxide - an in vivo labeling system for proteomics and metabolomics research. *Proteome Sci* **9**, 9 (2011).
66. Morcombe, C. R. & Zilm, K. W. Chemical shift referencing in MAS solid state NMR. *Journal of Magnetic Resonance* **162**, 479–486 (2003).
67. Metz Gunther, Wu Xiaoling & Smith Steven O. Ramped-Amplitude Cross polarization in Magic-angle-spinning NMR. *J Magn Reson A* **110**, 219–227 (1994).
68. Fung, B. M., Khitrin, A. K. & Ermolaev, K. An Improved Broadband Decoupling Sequence for Liquid Crystals and Solids, *J Magn Reson,* **142**, 97-101 (2000).
69. Fayon, F. *et al.* Through-space contributions to two-dimensional double-quantum J correlation NMR spectra of magic-angle-spinning solids. *J Chem Phys,* **122**, 194313 (2005).
70. Lesage, A., Bardet, M. & Emsley, L. Through-bond carbon-carbon connectivities in disordered solids by NMR. *J Am Chem Soc* **121**, 10987–10993 (1999).
71. Manolikas, T., Herrmann, T. & Meier, B. H. Protein structure determination from ^{13}C spin-diffusion solid-state NMR spectroscopy. *J Am Chem Soc* **130**, 3959–3966 (2008).
72. Ader, C. *et al.* Structural rearrangements of membrane proteins probed by water-edited solid-state NMR spectroscopy. *J Am Chem Soc* **131**, 170–176 (2009).

Author contributions

RC and RD conducted the NMR experiments and analysis. PKD prepared the hCNFs. YY generated models of cellulose fibril habits for conceptual analysis of glucose environments. PD and RD conceived and supervised the work. SJB contributed to experimental method development and interpretation. RC, RD and PD wrote the paper with comments from all authors.

Acknowledgements

Many thanks to Eva Hellmann from the Sainsbury lab at University of Cambridge for donating the poplar saplings that were used to grow the ^{13}C material used in this work. Many thanks to Theodora Tryfona for providing the mature leaves of *Brachypodium* as well as Alberto Echevarria Poza for all the help with growing and harvesting the ^{13}C labelled plants from the growth chamber, and Trent Franks for technical assistance. We wish to thank Mike Jarvis for giving us the ^{13}C NMR chemical shifts of his annealed flax and celery and for a draft version of his review[42.]. The UK High-Field Solid-State NMR Facility used in this research was funded by EPSRC and BBSRC (EP/T015063/1), as well as, for the 1 GHz instrument, EP/R029946/1. The work of RC, RD and PD was supported by the UKRI grant underwriting the ERC advanced grant EVOCATE, Function and evolution of plant cell wall architecture for sustainable technologies EP/X027120/1. PKD was supported by EPSRC grant Bio-derived and Bio-inspired Advanced Materials for Sustainable Industries (VALUED), EP/W031019/1. The work of YY generating models of cellulose fibril habits for conceptual analysis of glucose environments was supported as part of The Center for Lignocellulose Structure and Formation, an Energy Frontier Research Center funded by the US Department of Energy, Office of Science, Basic Energy Sciences, under Award number DE-SC0001090.

Competing interests.

The authors declare no competing interests

Main Figures

Cresswell et al.

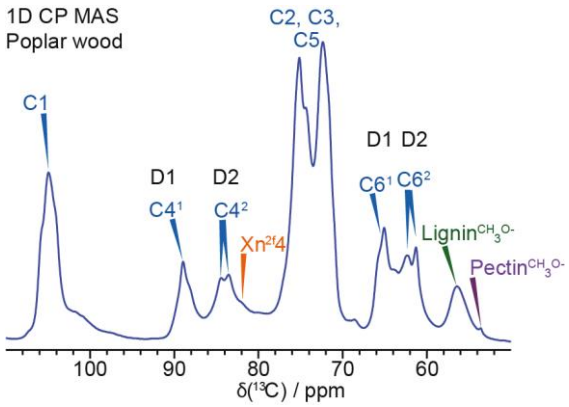

Figure 1: Neutral carbohydrate region of a ^{13}C 1D CP MAS NMR spectrum of poplar wood. The spectrum shows the assignments of cellulose (blue), xylan (orange), lignin (green) and pectin (purple). The glucose C4 and C6 peaks in cellulose are split into two regions named domain 1 (D1) and domain 2 (D2). The spectrum was recorded at a ^{13}C Larmor frequency of 251.4 MHz and a MAS frequency of 12.5 kHz.

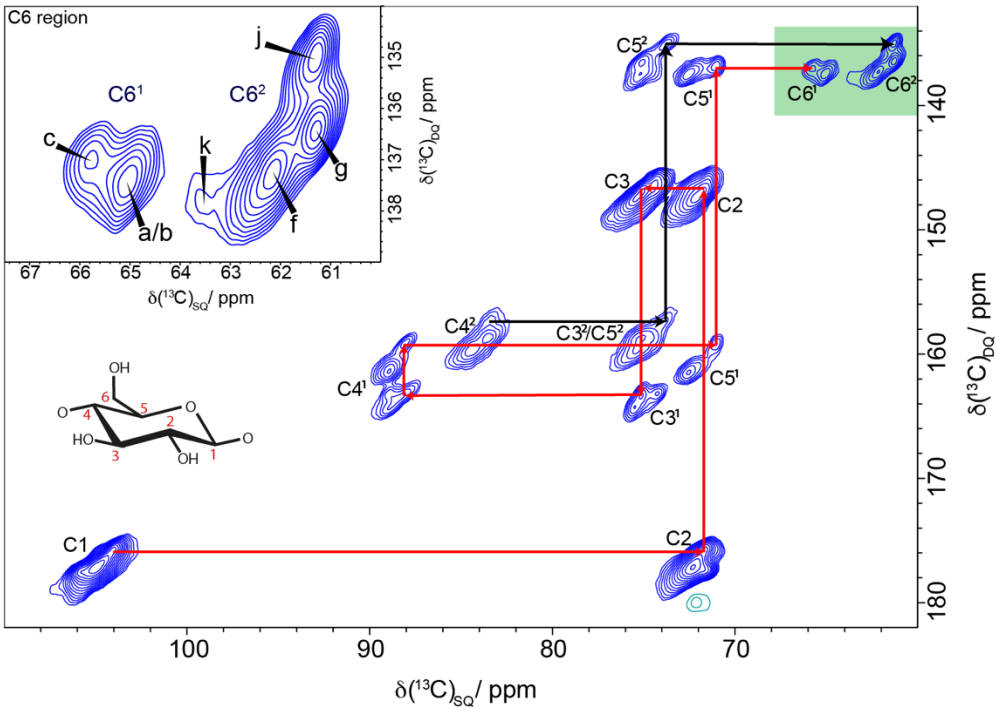

Figure 2: A ^{13}C 2D CP refocussed INADEQUATE MAS NMR spectrum of never-dried poplar wood showing multiple glucose environments in both domains 1 and 2 The path of ^{13}C carbon chemical shifts for cellulose glucose residue environments c (red) and j (black) are shown as example assignments. **Inset (top left):** The C6 region with the major glucose environments identified in poplar wood. The spectrum was as recorded in figure 1.

Domain 1						
Glucose Environment	**C1[1]**	**C2[1]**	**C3[1]**	**C4[1]**	**C5[1]**	**C6[1]**
a* (Cellulose Iβ origin)	105.8[$]	71.7	74.3	89.0	72.5	65.0
b	105.2	72.6	75.5	89.1	72.6	65.2
c (Cellulose Iβ centre)	104.1	71.8	75.2	88.1	71.2	65.8
d	*105.2*	*72.5*	*74.9*	*87.1*	*72.5*	*64.7*
e	*105.0*	-	*74.7*	*89.8*	*71.1*	*65.3*
Domain 2						
Glucose Environment	**C1[2]**	**C2[2]**	**C3[2]**	**C4[2]**	**C5[2]**	**C6[2]**
f	105.2	72.4	74.4	84.5	75.3	62.4
g	104.9	72.4	75.3	83.5	75.2	61.4
j	105.1	-	-	83.3	73.9	61.3
k	-	-	-	*83.8*	*74.3*	*63.5*

Table 1: NMR chemical shifts of all glucose environments assigned in cellulose of poplar wood. *The naming convention and assignments are based on those of Wang et al.[46] with additional environments. The minor glucose environments d, e and k are italicised. A comparison to Iα and Iβ cellulose NMR chemical shifts is shown in SI Table 1. [$]All ^{13}C NMR chemical shifts are in ppm and have an error of ± 0.1 ppm.

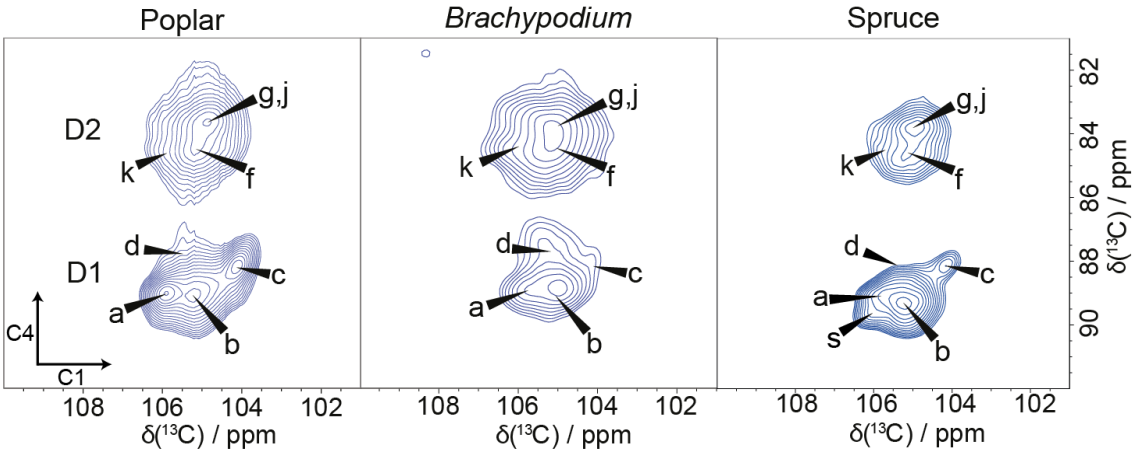

Figure 3: Glucose environments in cellulose are common across different plant lineages. A comparison of the C1-C4 region of 30 ms ^{13}C 2D CP PDSD MAS NMR spectra of poplar wood (a eudicot), *Brachypodium* mature leaves (a monocot) and spruce wood (a gymnosperm). Common glucose environments are labelled, noting that, whilst the relative amounts vary, the cellulose environments are common throughout. The spectra were recorded as in figure 1.

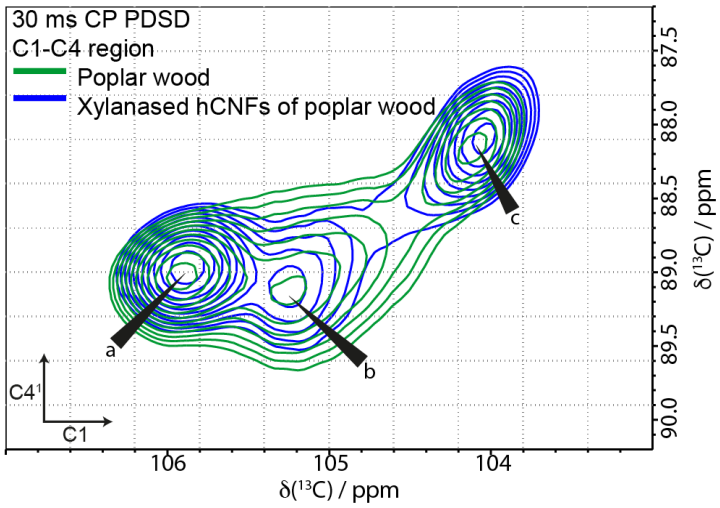

Figure 4: SCW cellulose of poplar wood maintains its structure in the fibrillation process to form hCNFs. Comparison of the cellulose C1-C4 domain 1 region of 30 ms ^{13}C 2D CP PDSD MAS NMR spectra of poplar wood (green) and xylanase treated hCNFs of poplar wood (blue). Spectra were recorded as in figure 1.

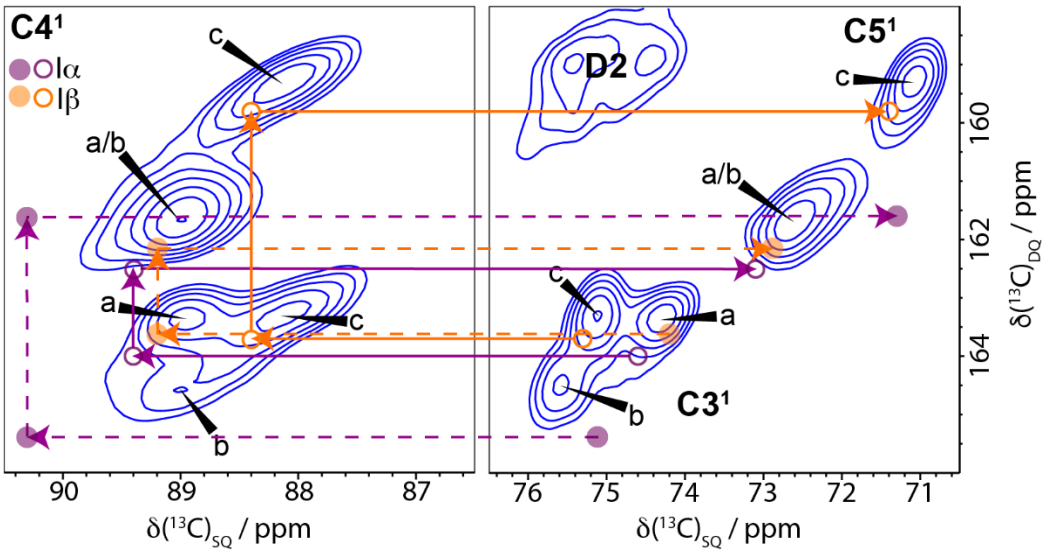

Figure 5: Plant secondary cell wall cellulose has a cellulose Iβ component but no cellulose Iα component. A comparison of the domain 1 C4-C3/C5 region of the ^{13}C 2D CP refocussed INADEQUATE MAS NMR spectrum of xylanase treated hCNFs of poplar wood with the ^{13}C chemical shifts of cellulose Iα and Iβ, as given by Brouwer and Mikolajewski[21] (see SI Table 1). The filled and empty circles correspond to two distinct ^{13}C chemical shifts in the asymmetric unit cells of cellulose Iα and Iβ.[17,18] The Iβ cellulose positions are within 0.2 ppm in the SQ dimension and 0.4 ppm in the DQ dimension of the native plant cellulose chemical shifts. The Iα cellulose positions do not match the ^{13}C NMR chemical shifts of any glucose environment acquired from the native never dried plant cell wall. Spectra were recorded as in figure 1.

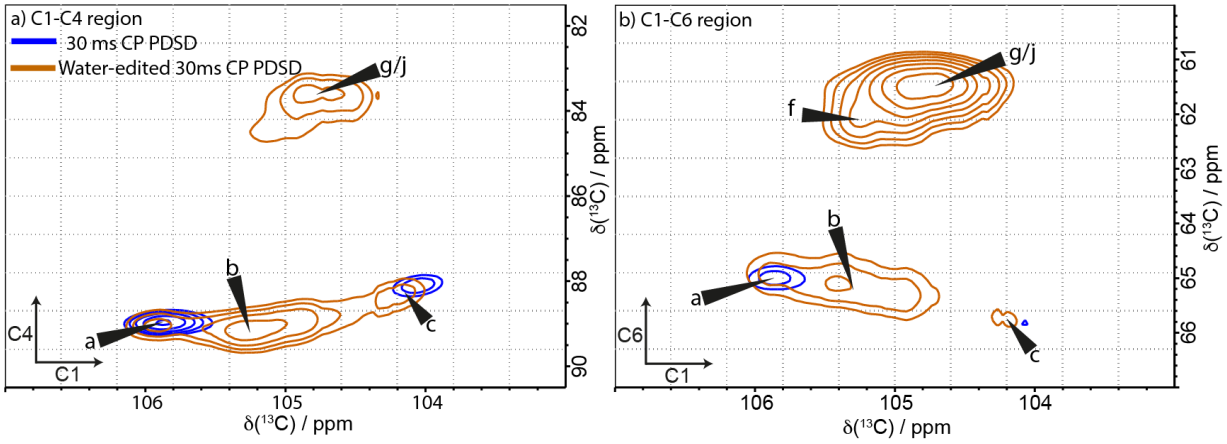

Figure 6: Domain 1 glucose environments in cellulose are not solely interior chains. Comparison of the standard (blue) and water edited (brown) 30 ms ^{13}C 2D CP PDSD MAS NMR spectra of xylanase treated hCNFs. a) the C1-C4 region, b) the C1-C6 region. Spectra were normalised to glucose environment a, and spectra were recorded as in figure 1.

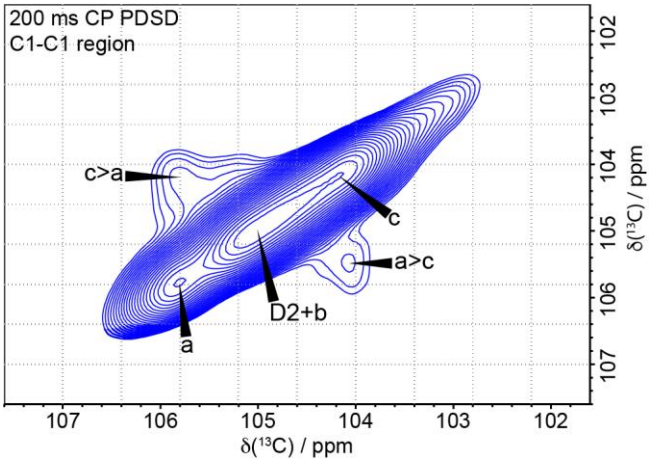

Figure 7: Domain 1 glucose environments a and c are in close proximity. The C1-C1 region of a 200 ms ^{13}C 2D CP PDSD MAS NMR spectrum of xylanase treated hCNFs of poplar wood shows cross-peaks between a and c. Spectra were recorded as in figure 1.

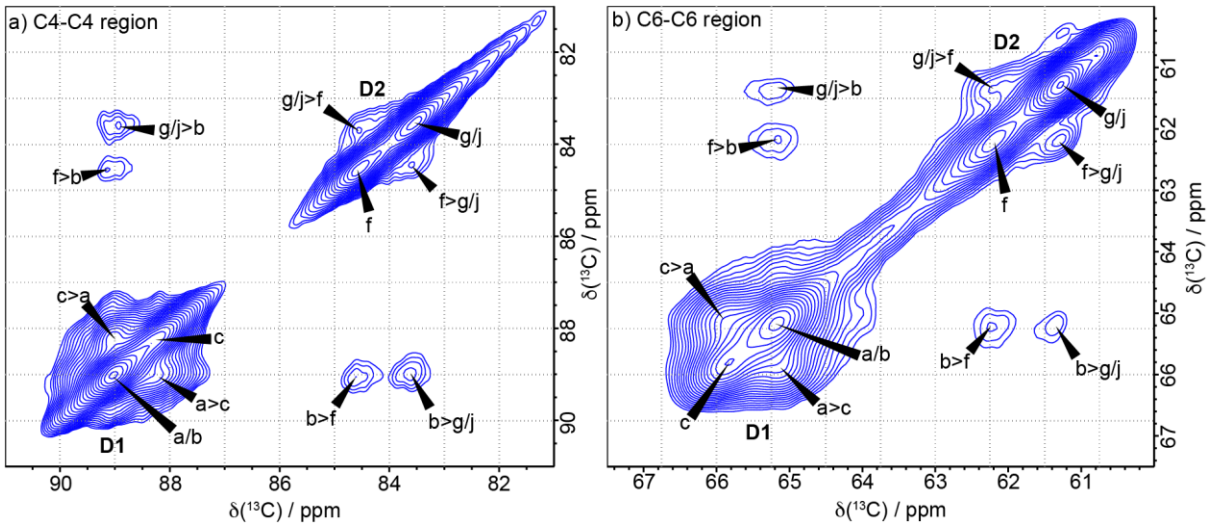

Figure 8: Domain 1 glucose environment b and the domain 2 glucose environments are in close proximity. 400 ms ^{13}C 2D CP PDSD MAS NMR spectrum of xylanase treated hCNFs of poplar wood. a) the C4-C4 region, b) the C6-C6 region. Spectra were recorded as in figure 1.

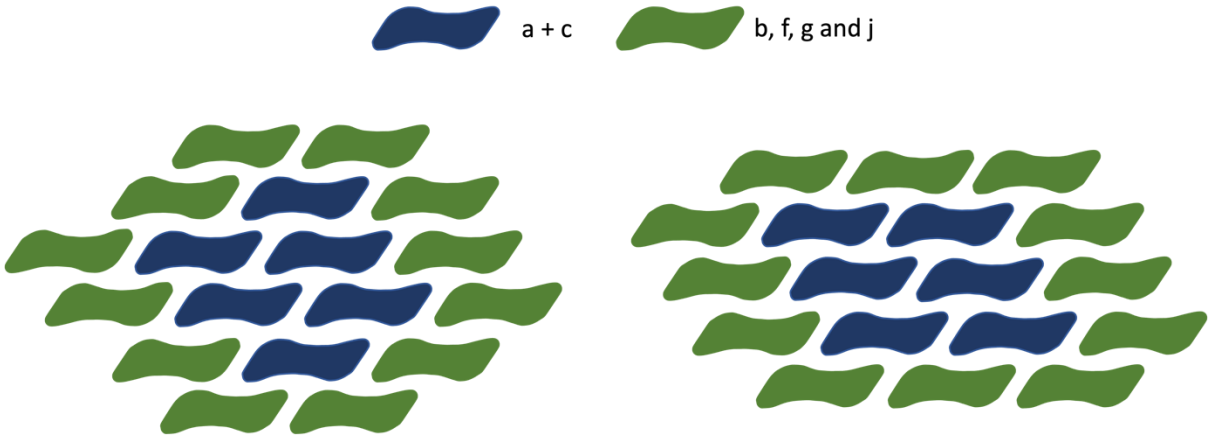

Figure 9: Schematic diagram of two potential 18 chain cellulose Iβ microfibrils, with 234432 and 34443 habits. The fibril core and the surface environments are shown in blue and green respectively. Although both habits fulfil the observation of core:surface ratio 0.5, only the 234432 habit would have a ratio of a (origin): c (centre) chain of 1.