Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Looking at hydrogen bonds in cellulose

Yoshiharu Nishiyama,[a]* Paul Langan,[b] Masahisa Wada[c,e] and V. Trevor Forsyth[d,f]

[a]CERMAV–CNRS, BP 53, 38041 Grenoble CEDEX 9, France, [b]Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA, [c]School of Agricultural and Life Sciences, University of Tokyo, Tokyo 113-8657, Japan, [d]ILL, BP 156, 38042 Grenoble CEDEX 9, France, [e]College of Life Sciences, Kyung Hee University, Gyeonggi-do 446-701, Republic of Korea, and [f]EPSAM/ISTM, Keele University, Staffordshire ST5 5BG, England

Correspondence e-mail: yoshi@cermav.cnrs.fr

Received 6 May 2010
Accepted 11 August 2010

A series of cellulose crystal allomorphs has been studied using high-resolution X-ray and neutron fibre diffraction to locate the positions of H atoms involved in hydrogen bonding. One type of position was always clearly observed in the Fourier difference map ($F_d - F_h$), while the positions of other H atoms appeared to be less well established. Despite the high crystallinity of the chosen samples, neutron diffraction data favoured some hydrogen-bonding disorder in native cellulose. The presence of disorder and a comparison of hydrogen-bond geometries in different allomorphs suggests that although hydrogen bonding may not be the most important factor in the stabilization of cellulose I, it is essential for stabilizing cellulose III, which is the activated form, and preventing it from collapsing back to the more stable cellulose I.

1. Introduction

Cellulose is produced in nature as continuous crystalline filaments with nanometric lateral dimensions called microfibrils (Nishiyama, 2009) as a result of the simultaneous polymerization of many chains by synthase complexes. The native cellulose crystal is chemically stable and does not allow water penetration under ambient conditions, despite having three hydroxyl groups per glucose residue. The turnover of this 1,4-linked β-glucan in nature is much slower than that of chitin, its counterpart in the animal kingdom, or indeed of any other polysaccharide generated from glucose linked in different ways. This low turnover rate makes cellulose the most important biomass on earth and thus a key target for fossil energy replacement since it does not compete with food markets, in contrast to its α-glucan counterpart starch. On the other hand, the conversion of cellulose into sugars which can then be used to produce fuels and other products is inherently challenging owing to its recalcitrance.

The crystalline nature of cellulose is believed to be one of the most important factors responsible for its recalcitrance towards hydrolysis. Conventional native cellulose degrades rapidly to fragments corresponding to the length of the crystalline domains, which are about 150 nm in length (Nishiyama, Kim *et al.*, 2003). Severe conditions or a special set of enzymes are required to degrade cellulose further.

In the absence of the macroscopic single crystals required for conventional crystallographic analyses, the crystal structure of cellulose has mostly been studied using the technique of fibre diffraction. In fact, the native form of cellulose has never been crystallized *in vitro* and can be only found in native systems, in which it is generated by a unique process of crys-

tallization in the vicinity of the site of polymerization where all chains are pointing in the same direction.

Native cellulose in general is known to be composed of two similar but distinct crystal allomorphs, namely cellulose I_α and cellulose I_β. Cellulose I_α has a one-chain triclinic unit cell and is the major component of some green algal and bacterial celluloses, while cellulose I_β has a two-chain monoclinic unit cell and is the dominant form in terrestrial plant cellulose. These allomorphs can be transformed to another allomorph called cellulose III by treatment with various amines. When treated with supercritical ammonia, this transformation can take place without an appreciable loss in crystallinity (Isogai & Usuda, 1992). We have been studying the detailed crystal structures of these allomorphs using polycrystalline samples in which nanometric single crystals have been aligned amost uniaxially. High-resolution X-ray and neutron fibre diffraction data were collected to atomic resolution and the basic structures were determined without ambiguity (Langan *et al.*, 1999, 2001; Nishiyama *et al.*, 2002; Nishiyama, Sugiyama *et al.*, 2003; Wada *et al.*, 2004).

Recently, it was demonstrated that certain *Trichoderma viride* cellobiohydrolases degraded cellulose transformed to the cellulose III allomorph five times faster than native cellulose, even when normalized to the specific surface of the substrate (Igarashi *et al.*, 2007). Clearly, the crystal structure of cellulose has an important impact on its enzymatic reactivity and a detailed analysis of the crystal structures of cellulose should provide insight into its degradation behaviour. In this article, we will review the structural details of cellulose crystals and discuss the implications for their reactivities.

2. Methods

2.1. Sample preparation

2.1.1. Oriented films. The outer mantle of tunicate, *Halocynthia roretzi*, is composed of protein embedding highly crystalline cellulose I_β. The cellulose was purified by alternating treatment with 5% KOH solution and 1% sodium chlorite solution. The resulting cellulose was hydrolyzed using 60% sulfuric acid with about 2% solid content overnight with strong stirring to obtain a suspension of single crystals that formed an oriented gel when shear-aligned. This gel was washed free of acid with a mixture of ethanol and water and subsequently dried to produce oriented films (Nishiyama *et al.*, 1997).

Glaucocystis nostochinearum, a micro-alga whose cell wall contains cellulose of almost pure I_α phase, was cultured in an airlift fermenter under continous illumination. The oriented film was obtained in the same manner as for *H. roretzi*.

For cellulose III, *Chladophora wrightiana*, a green alga, was chosen as the starting material since it contains one of the thickest microfibrils found so far. The algae were collected on the coast of Chikura, Chiba, Japan and transformed into an oriented specimen as above. The oriented sample was subsequently treated with liquid ammonia: the sample was placed in a glass tube cooled in a dry-ice/methanol bath and ammonia gas was condensed on the sample. The system was brought to above the critical temperature of ammonia in an oil bath and the ammonia slowly leaked out. The product was washed free of ammonia with methanol and dried under vacuum.

Deuterated samples in which all intra-crystalline hydroxyl groups were deterium-exchanged were prepared by treating the cellulose I sample in heavy water with 0.1 *N* NaOD at 483 K (Nishiyama *et al.*, 1999). When brought back to room temperature the intra-crystalline hydroxyl groups did not back-exchange with hydrogen even when left in the atmosphere for more than a year. Deuterated cellulose III samples were prepared by swelling cellulose in deuterated ammonia and subsequently washing with deuterated methanol and vacuum drying.

2.1.2. Uniaxially oriented samples. During data analysis, we noticed that in addition to the axial orientation along the shear direction, there was a tendency for certain crystallographic planes to orient with respect to the film surface. Also, the axial orientation was not cylindrically symmetric, but rather was ellipsoidal owing to the film geometry. This added another complication to the data-reduction procedure, although in principle it provides us with additional information. We thus prepared samples with a uniaxial orientation and pure fibre cylindrical symmetry. This was performed by embedding the cellulose crystals in a polyvinyl alcohol (PVA) physical gel and drying it while elongating it with a suspended weight. Briefly, aqueous suspension of crystals were mixed with a high-molecular-weight PVA aqueous solution to make a liquid containing about 3.5% PVA and 3.5% cellulose. Saturated borax solution was then added dropwise to the mixture while mixing the solution to obtain a gel. After removing the bubbles by centrifugation, the gel was stretched, dried and washed with methanol to remove the borax.

2.2. Data collection

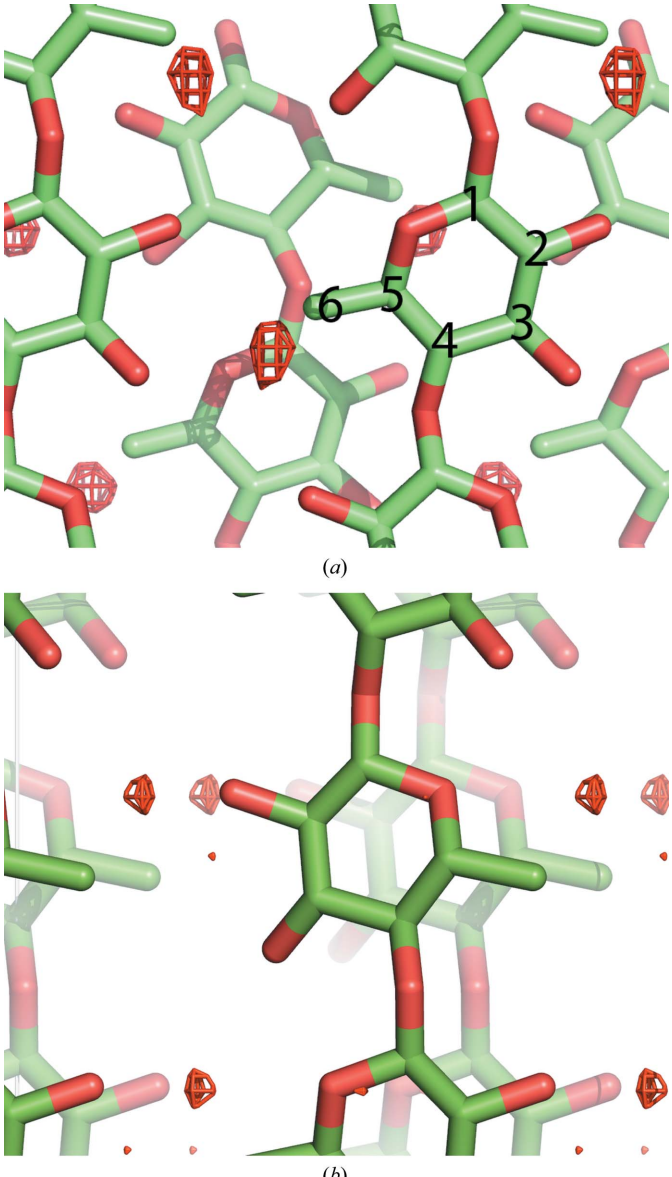
X-ray diffraction data from oriented films were recorded using a flat image-plate system and a goniometer to measure a series of pseudo-fibre diffraction patterns with different tilts and rotations around the main axis.

Neutron diffraction data were collected on the D19 diffractometer at the Institut Laue–Langevin. For data acquisition, 1 mm thick block samples were prepared from both hydrogenated and deuterated crystals by stacking the films on top of one another. The thickness was limited by the attenuation caused by the H atoms linked on the carbon. Typically, three slices of reciprocal space that were roughly parallel, perpendicular and at 45° to the film surface were swept using a four-circle goniometer and a $4 \times 64°$ position-sensitive detector (now replaced by a $30 \times 120°$ detector) at room temperature and were reconstituted into pseudo-fibre patterns by binning. The PVA composite fibre sample was measured using the new $30 \times 120°$ detector to reconstitute the longitudinal slice of the cylindrically averaged reciprocal space at room temperature and at 15 K.

The background of the neutron diffraction diagram is dominated by inelastic scattering from the H atoms in the

sample. This scattering should be constant over scattering vector and thus an absorption correction was applied so that the background became flat.

Firstly, the texture and the peak widths arising from the limited crystal dimensions were determined from the intensity distribution of a few diffraction spots. Lists of integrated peak intensities were then obtained from background-corrected files by a linear least-squares fit, as the peak profiles can be uniquely determined from the texture and crystallite sizes. The overlap of close peaks makes the least-squares problem underdetermined, so the minimum norm solution was obtained using the GELSD driver of the linear algebra package *LAPACK* (Anderson *et al.*, 1999). This solution splits the overlapping intensities without structural bias but is completely arbitrary. Composite intensity lists, generated by regrouping the intensities of reflections that overlap with each other in the same layer line, were also created.



(*a*)

(*b*)

Figure 1
Fourier OMIT maps of cellulose I_β (*a*) and cellulose III (*b*) showing the position of O6 omitted from the model.

One of the drawbacks of this method is the lack of error estimation for the intensity data, since the Jacobian matrix is always rank deficient. An arbitrary standard deviation proportional to the distance from the meridian was associated with each intensity.

2.3. Structure analysis

Structure refinement was carried out with *SHELX*-97 using restraints on bond distances and bond angles restricted to the mean value in the cellotetraose single-crystal structure. The C-atom and O-atom positions were refined against the composite X-ray intensity list using the 'HKLF 6' option intended for high-resolution powder diffraction analysis. Fourier OMIT maps and difference maps were calculated using the minimum norm intensity list distributing the intensity of overlapping peaks in a flat manner.

3. Results and discussion

3.1. X-ray structure

In the absence of software capable of performing Rietveld-like refinement using the entire data set, structure analysis was carried out based on the intensity lists obtained by fitting as described here. Cellulose is composed of anhydro-glucose residues that have fairly rigid pyranose rings. The torsion angles around the glycosidic bond have some rotational freedom, but like most polysaccharides they are limited to a small region of conformational space owing to steric hindrance (French *et al.*, 2001). Given the unit-cell parameters, and thus the polymer repeat distance, the values that the glycosidic torsion angles can take is very limited, resulting in a ribbon-shaped flat molecule. Owing to the high density and tight packing in cellulose crystals, the positional and rotational freedoms of the chains with respect to the unit cell are also quite limited, with two choices of orientation, pointing in opposite directions, and the possibility of sliding along the chain axis in some cases. With these considerations in mind, and neglecting H atoms, the only rotation torsion angle with a large amount of freedom is around the C5—C6 bond (refer to Fig. 1 for atom numbering) of a exocyclic hydroxymethyl group in each residue. Thus, in our approach to structure determination, the positions and orientations of the chains were first determined with the measured intensities, with the O6 atoms of the freely rotating hydroxymethyl groups omitted. In all allomorphs studied here, both the 'minimum norm' intensity list and the composite intensity list of X-ray diffraction data allowed easy discrimination between different chain-packing possibilities.

Fourier OMIT maps were then calculated based on the 'minimum norm' X-ray intensity data set and the model. O6 appeared in the OMIT map at a sensible position when the correct model was chosen (Fig. 1), although its clarity was dependent on the allomorph, probably owing to differing symmetry and data quality. In fact, in the fibre diffraction data not only the phase but also the intensity distribution in a

composite peak is uncertain, which makes the resulting maps quite noisy despite the small missing component in the model.

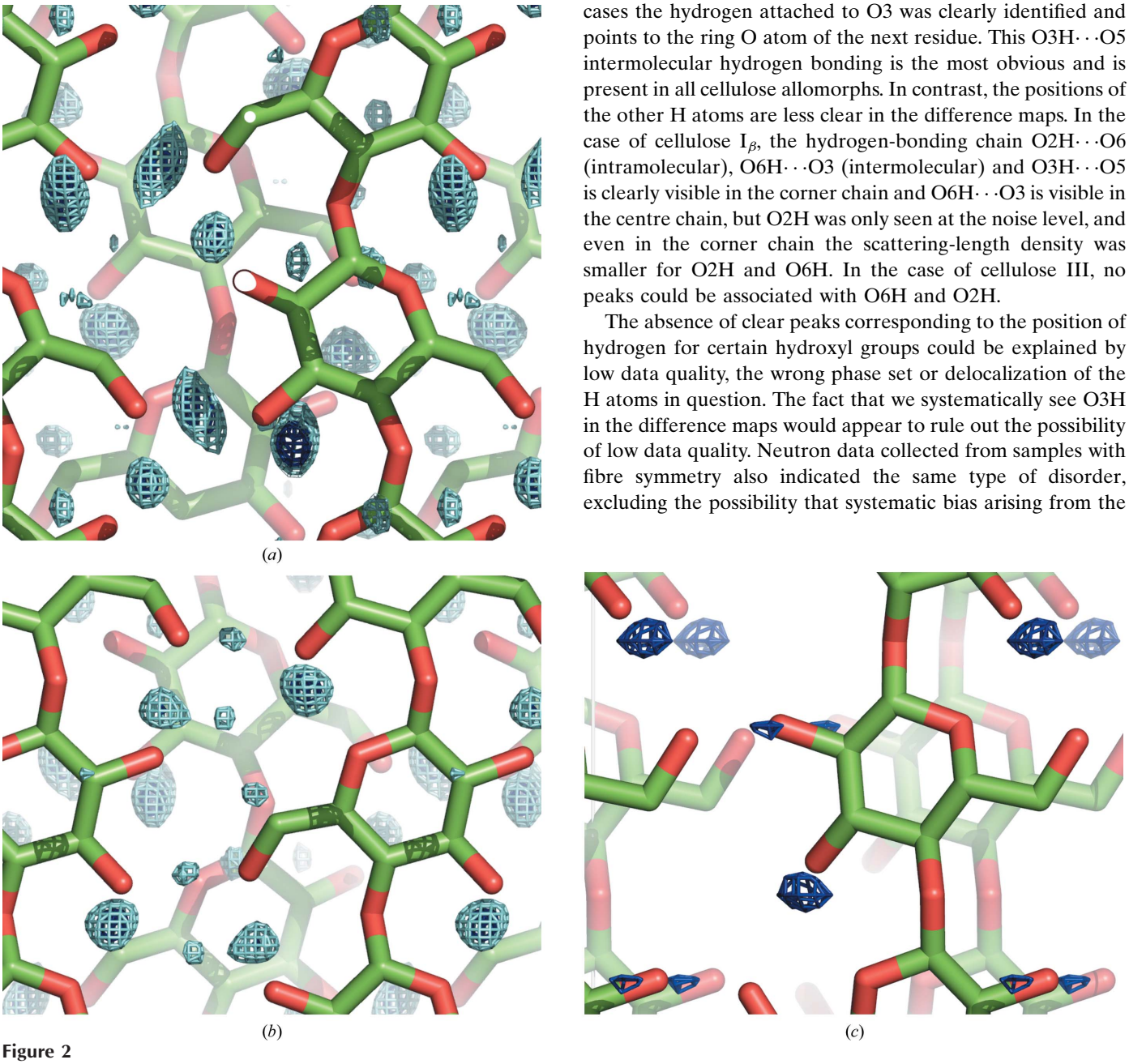
The fact that we see the omitted O6 atoms in the Fourier maps means that the quality of the intensity data obtained from the fibre or pseudo-fibre diffraction diagrams is sufficient to unambiguously determine the basic structure despite the extensive peak overlap arising from the limited crystal size and cylindrical averaging around the fibre axis. The coordinates of the C and O atoms were then refined against the composite intensity list in a similar way to powder diffraction.

3.2. Positions of H atoms

Once the positions of all of the C and O atoms have been determined and fixed, there are only a limited number of degrees of freedom for the undetermined H-atom positions: six rotations of independent hydroxyl groups for cellulose I and three for cellulose III. Again, the 'minimum norm' neutron intensity list was used to directly visualize the hydrogen positions. To enhance the contrast, $F_d - F_h$ was used as the amplitude, where F_d is the amplitude from the deuterated sample and F_h is the amplitude from the hydrogenated sample. The intensities were scaled using *SHELX*, with the positions of the C and O atoms determined in the X-ray studies fixed. The positions of H atoms covalently bonded to carbon were also fixed in standard positions. The hydroxyl H atoms were left out of the model at this point. The phase was thus calculated from seven H, five O and six C atoms per residue to locate the three labile hydroxyl H atoms.

Fig. 2 shows $F_d - F_h$ maps for the three allomorphs. In all cases the hydrogen attached to O3 was clearly identified and points to the ring O atom of the next residue. This O3H···O5 intermolecular hydrogen bonding is the most obvious and is present in all cellulose allomorphs. In contrast, the positions of the other H atoms are less clear in the difference maps. In the case of cellulose I_β, the hydrogen-bonding chain O2H···O6 (intramolecular), O6H···O3 (intermolecular) and O3H···O5 is clearly visible in the corner chain and O6H···O3 is visible in the centre chain, but O2H was only seen at the noise level, and even in the corner chain the scattering-length density was smaller for O2H and O6H. In the case of cellulose III, no peaks could be associated with O6H and O2H.

The absence of clear peaks corresponding to the position of hydrogen for certain hydroxyl groups could be explained by low data quality, the wrong phase set or delocalization of the H atoms in question. The fact that we systematically see O3H in the difference maps would appear to rule out the possibility of low data quality. Neutron data collected from samples with fibre symmetry also indicated the same type of disorder, excluding the possibility that systematic bias arising from the



(*a*)

(*b*)

(*c*)

Figure 2
$F_d - F_h$ Fourier maps. (*a*) and (*b*) show corner chains and centre chains of cellulose I_β; (*c*) shows cellulose III.

film texture influenced the result. The accuracy of the phasing of the structure in the region of O2H should be as well defined as that for O3H, as the positions of the atoms in its immediate surroundings are well established in the X-ray structure. Thus, the difference maps suggests inherent disorder of the hydrogens attached to O2 and O6.

The positions of H atoms were also analyzed by refining the hydrogen positions against the composite intensity list of the deuterated samples, as no software was available that could handle hydrogenated and deuterated fibre data at the same time. O3H has no other candidate acceptor atom and the positions were determined without any doubt. The O2 and O6 atoms have different potential acceptors. For cellulose I_α and I_β, in addition to the classical hydrogen-bonding schemes O2H$\cdots$O6 and O6H$\cdots$O3, the O6 hydrogen can tautomerize and be positioned at O6 instead of O2 (Gardner & Blackwell, 1974*a*,*b*; Sarko & Muggli, 1974; Woodcock & Sarko, 1980). In both cases, when two sets of coordinates were refined for each atom with partial occupancies, the agreement with the observed composite intensities was significantly improved. Cellulose III has an obvious intermolecular hydrogen-bonding chain O2H$\cdots$O6H$\cdots$O2H$\cdots$, but again has the possibility of tautomerizing. The mixed structure improved the agreement with the observed intensities, but the difference was not significant enough with respect to the additional number of degrees of freedom to allow us to decide conclusively that disorder was present.

These observations triggered theoretical work on the possibility of hydrogen-bonding disorder in native cellulose.

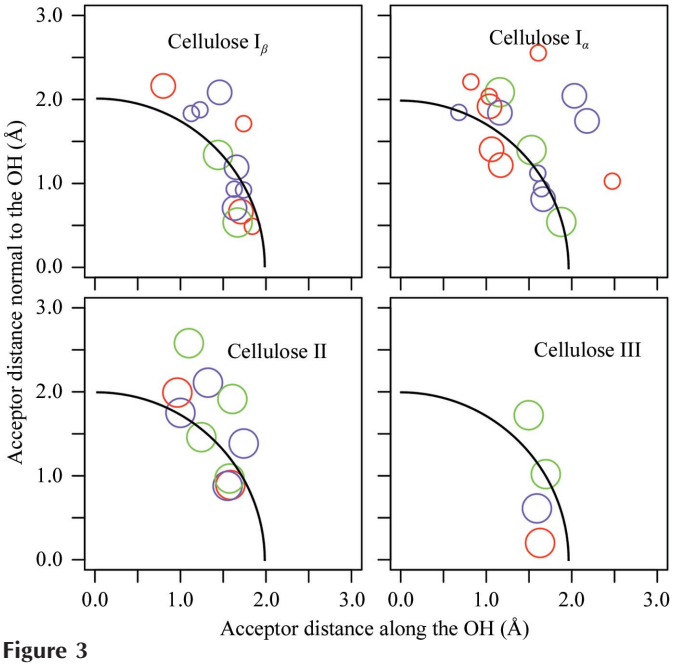


Figure 3
Hydrogen-bonding geometries in different allomorphs of cellulose. The positions of the acceptor with respect to the donor hydrogen are plotted, where an OH$\cdots$O angle of 180° lies on the horizontal axis. The distance from the origin is the hydrogen–acceptor distance; circles at 2 Å are shown as a guide to the eye. The areas of the circles are proportional to the relative occupancy in the refined structure. The different hydroxyl groups are indicated by colour: green, O3H; red, O2H; purple, O6H.

Mazeau (2005) pointed out based on molecular-dynamics calculations that the alternative hydrogen-bonding scheme in native cellulose was not stable, but resulted in a number of different hydrogen-bonding schemes organized in small domains which remained stable. A quantum-mechanical calculation also suggested that the minor hydrogen-bonding scheme was of much higher energy, except for the surface chains. In the case where O6H$\cdots$O3 hydrogen bonding was impossible owing to the absence of a neighbouring chain, the alternative hydrogen bonding was more stable, since O2H$\cdots$O6 hydrogen bonding has some penalty owing to lone-pair electron interference between O2 and O3 (Nishiyama *et al.*, 2008).

While neutron refinement clearly favoured a mixture of two coexisting hydrogen-bonding systems over a single position, it is certainly not the only solution that can fit the intensity data. The apparent disorder is not simply a consequence of high thermal motion, since the disorder persisted at temperatures as low as 15 K. Different hydrogen-bonding arrangements are frozen at room temperature, as suggested by the molecular-dynamics studies.

The idea of disordered hydrogen bonding has also been suggested independently by infrared spectroscopy, which, based on temperature-induced spectral changes, suggested that part of O2H was not hydrogen bonded to O6 but was almost free of hydrogen bonding (Maréchal & Chanzy, 2000)

This unexpected result of our neutron diffraction studies suggests that our perception of native cellulose as a very simple and regular crystal structure has to be revised. This perception was inspired particularly by the direct imaging of the crystalline lattice of cellulose by electron microscopy (Sugiyama *et al.*, 1984, 1985) and scanning probe microscopy (Baker *et al.*, 1997), as well as from the extremely sharp X-ray and neutron diffraction peaks obtained to atomic resolution from the highly crystalline samples used here. However, the hydrogen-bonding disorder can be reconciled with a fairly well ordered structure for C and O atoms in the chain backbone. Small deviations from this ordered backbone structure can completely change the hydrogen-bonding environment. These small deviations would not significantly affect the average structure observed in the crystallographic analysis. It may be significant that although diffraction data can extend to above 0.7 Å resolution in the chain direction (Honjo & Watanabe, 1958), diffraction intensity cannot be recognized above 1 Å resolution in the direction perpendicular to the chain direction.

3.3. Hydrogen-bond strength and structure stability

Hydrogen bonding is often considered to play an important role in the structure formation and stability of cellulose. However, when the hydrogen-bonding geometries were compared among different allomorphs of cellulose, cellulose III, which is considered to be less stable because it spontaneously converts to cellulose I on heating (Wada, 2002), is seen to have much stronger hydrogen bonding (Fig. 3). The difference in free energy between the different allomorphs is

Table 1
Crystal density of different allomorphs of cellulose calculated from the unit-cell parameters.

Allomorph	Density (g cm^{-3})
Cellulose I_β	1.615
Cellulose I_α	1.595
Cellulose II	1.58
Cellulose III	1.53

difficult to obtain experimentally, since the crystals generally have very limited size and their morphology depends on their preparation and hence on the allomorph. However, if cellulose III is of higher energy than cellulose I, the enthalpic contribution of hydrogen bonding to stabilization is clearly minor since the hydrogen-bonding geometry is more favourable in cellulose III. At higher temperature, the possibility of making different hydrogen bonds would contribute to a gain in entropy (Shen & Gnanakaran, 2009), favouring cellulose I.

One should nevertheless be careful not to overemphasize the role of hydrogen bonding, since the native crystallization process and also enzymatic hydrolysis take place in an aqueous environment in which the hydrogen-bonding energy between hydroxyl groups is comparable with the solvation energy. When the crystal densities of the different allomorphs are compared (Table 1), it can be seen that cellulose I has a significantly higher value compared with cellulose III, which corresponds to stronger van der Waals interactions. Taking the above observations into consideration, we propose that the high reactivity of cellulose III in enzymatic hydrolysis is partly a consequence of the presence of strong hydrogen bonding that helps to prevent its swollen structure from collapsing to the more stable and dense cellulose I.

Part of this study was supported by the French Agence National de la Recherche. VTF acknowledges support from the UK Engineering and Physical Sciences Research Council (EPSRC) for grant GR/R47950/01 to Durham, Keele and Bath Universities for the construction of the D19 diffractometer at ILL. We also thank Sax Mason, John Archer and John Allibon for assitance during the neutron experiments.

References

Anderson, E., Bai, Z., Bischof, C., Blackford, S., Demmel, J., Dongarra, J., Du Croz, J., Greenbaum, A., Hammarling, S., McKenney, A. & Sorensen, D. (1999). *LAPACK Users' Guide*, 3rd ed. Philadelphia: Society for Industrial and Applied Mathematics.
Baker, A. A., Helbert, W., Sugiyama, J. & Miles, M. J. (1997). *J. Struct. Biol.* **119**, 129–138.
French, A. D., Kelterer, A.-M., Johnson, G. P., Dowd, M. K. & Cramer, C. J. (2001). *J. Comput. Chem.* **22**, 65–78.
Gardner, K. H. & Blackwell, J. (1974*a*). *Biochim. Biophys. Acta*, **343**, 232–237.
Gardner, K. H. & Blackwell, J. (1974*b*). *Biopolymers*, **13**, 1975–2001.
Honjo, G. & Watanabe, M. (1958). *Nature (London)*, **181**, 326–328.
Igarashi, K., Wada, M. & Samejima, M. (2007). *FEBS J.* **274**, 1785–1792.
Isogai, A. & Usuda, M. (1992). *Mokuzai Gakkaishi*, **38**, 562–569.
Langan, P., Nishiyama, Y. & Chanzy, H. (1999). *J. Am. Chem. Soc.* **121**, 9940–9946.
Langan, P., Nishiyama, Y. & Chanzy, H. (2001). *Biomacromolecules*, **2**, 410–416.
Maréchal, Y. & Chanzy, H. (2000). *J. Mol. Struct.* **523**, 183–196.
Mazeau, K. (2005). *Cellulose*, **12**, 339–349.
Nishiyama, Y. (2009). *J. Wood Sci.* **55**, 241–249.
Nishiyama, Y., Isogai, A., Okano, T., Müller, M. & Chanzy, H. (1999). *Macromolecules*, **32**, 2078–2081.
Nishiyama, Y., Johnson, G. P., French, A. D., Forsyth, V. T. & Langan, P. (2008). *Biomacromolecules*, **9**, 3133–3140.
Nishiyama, Y., Kim, U. J., Kim, D. Y., Katsumata, K. S., May, R. P. & Langan, P. (2003). *Biomacromolecules*, **4**, 1013–1017.
Nishiyama, Y., Kuga, S., Wada, M. & Okano, T. (1997). *Macromolecules*, **30**, 6395–6397.
Nishiyama, Y., Langan, P. & Chanzy, H. (2002). *J. Am. Chem. Soc.* **124**, 9074–9082.
Nishiyama, Y., Sugiyama, J., Chanzy, H. & Langan, P. (2003). *J. Am. Chem. Soc.* **125**, 14300–14306.
Sarko, A. & Muggli, R. (1974). *Macromolecules*, **7**, 486–494.
Shen, T. & Gnanakaran, S. (2009). *Biophys. J.* **96**, 3032–3040.
Sugiyama, J., Harada, H., Fujiyoshi, Y. & Uyeda, N. (1984). *Mokuzai Gakkaishi*, **30**, 98–99.
Sugiyama, J., Harada, H., Fujiyoshi, Y. & Uyeda, N. (1985). *Planta*, **166**, 161–168.
Wada, M. (2002). *J. Polym. Sci. B Polym. Phys.* **40**, 1095–1102.
Wada, M., Chanzy, H., Nishiyama, Y. & Langan, P. (2004). *Macromolecules*, **37**, 8548–8555.
Woodcock, C. & Sarko, A. (1980). *Macromolecules*, **13**, 1183–1187.