J|A|C|S
ARTICLES
Published on Web 07/10/2002

Crystal Structure and Hydrogen-Bonding System in Cellulose Iβ from Synchrotron X-ray and Neutron Fiber Diffraction

Yoshiharu Nishiyama,[†] Paul Langan,*[,‡] and Henri Chanzy[§]

Contribution from the Department of Biomaterials Science, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo 113-8657, Japan, Biosciences Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, and Centre de Recherches sur les Macromolécules Végétales—CNRS, affiliated with the Joseph Fourier University of Grenoble, B.P. 53, 38041 Grenoble Cedex 9, France

Received January 28, 2002

Abstract: The crystal and molecular structure together with the hydrogen-bonding system in cellulose Iβ has been determined using synchrotron and neutron diffraction data recorded from oriented fibrous samples prepared by aligning cellulose microcrystals from tunicin. These samples diffracted both synchrotron X-rays and neutrons to better than 1 Å resolution (>300 unique reflections; $P2_1$). The X-ray data were used to determine the C and O atom positions. The resulting structure consisted of two parallel chains having slightly different conformations and organized in sheets packed in a "parallel-up" fashion, with all hydroxymethyl groups adopting the *tg* conformation. The positions of hydrogen atoms involved in hydrogen-bonding were determined from a Fourier-difference analysis using neutron diffraction data collected from hydrogenated and deuterated samples. The hydrogen atoms involved in the intramolecular O3···O5 hydrogen bonds have well-defined positions, whereas those corresponding to O2 and O6 covered a wider volume, indicative of multiple geometry with partial occupation. The observation of this disorder substantiates a recent infrared analysis and indicates that, despite their high crystallinity, crystals of cellulose Iβ have an inherent disorganization of the intermolecular H-bond network that maintains the cellulose chains in sheets.

Unraveling the crystalline structure of cellulose has been one of the most studied structural problems in polymer science. Close to one century ago, Nishikawa and Ono obtained the first X-ray diffraction traces from wood, hemp, and bamboo.[1] Subsequently, X-ray diffraction patterns from a number of cellulose fibers were published, and from them, several cellulose structures were proposed. Landmarks in this development were the molecular models built by Meyer and Mark,[2] later revised by Meyer and Misch.[3] The derived atomic coordinates provided a realistic molecular description of the glucosyl rings in an antiparallel two-chain monoclinic unit cell.

Various improved models were proposed[4−6] that incorporated intramolecular hydrogen-bonding after the glucosyl moieties were shown to adopt a "bent" $^4C_1$ chair conformation.[7] However, refined electron diffraction diagrams from highly crystalline *Valonia* cellulose[8] could not be resolved on the basis of a two-chain monoclinic unit cell, and it was not clear whether an eight-chain unit cell or a disordered structure might be more correct. From 1974 onward, with the development of computer methods in model building, the groups of Sarko and Blackwell refined structures that were in better agreement with the available data: for native cellulose (cellulose I) and for other allomorphs (cellulose II, III_I and IV_I).[9−14] According to these authors, cellulose I can be modeled by an approximate monoclinic unit cell containing two parallel and conformationally equivalent chains. One chain (the origin chain) is positioned at the corner of the unit cell parallel to the *c* axis direction, and a second chain (the center chain) passes through the center of the *a*/*b* plane and is translated in the *c* axes direction by about *c*/4 with respect to the origin chain. Whereas in the description given by Sarko and Muggli[10] the direction of the chains was determined to be "parallel-up", in the description given by Gardner and Blackwell[9] it was determined to be "parallel-down".[15]

Although many of the features of these structures are still accepted, certain details have been contradicted by results from the development of ^{13}C CP/MAS solid-state NMR. This technique revealed unexpected details in highly crystalline

* Address correspondence to this author. Tel.: +1 505 665 8125. Fax: +1 505 665 3024. E-mail: langan_paul@lanl.gov.
[†] The University of Tokyo.
[‡] Los Alamos National Laboratory.
[§] Centre de Recherches sur les Macromolécules Végétales−CNRS.

(1) Nishikawa, S.; Ono, S. *Proc. Tokyo Math. Phys. Soc.* **1913**, *7*, 131.
(2) Meyer, K. H.; Mark, H. *Ber. Dtsch. Chem. Ges.* **1928**, *61*, 593.
(3) Meyer, K. H.; Misch, L. *Helv. Chim. Acta* **1937**, *20*, 232.
(4) Hermans, P. H.; DeBooys, J.; Mann, C. *Kolloid Z.* **1943**, *102*, 169.
(5) Jones, D. W. *J. Polym. Sci.* **1958**, *32*, 371.
(6) Liang, C. Y.; Marchessault, R. H. *J. Polym. Sci.* **1959**, *37*, 385.
(7) Ferrier, W. G. *Acta Crystallogr.* **1963**, *16*, 1023.
(8) Honjo, G.; Watanabe, M. *Nature* **1958**, *181*, 326.
(9) Gardner, K. H.; Blackwell, J. *Biopolymers* **1974**, *13*, 1975.
(10) Sarko, A.; Muggli, R. *Macromolecules* **1974**, *7*, 486.
(11) Stipanovic, A. J.; Sarko, A. *Macromolecules* **1976**, *9*, 851.
(12) Kolpak, F. J.; Blackwell, J. *Macromolecules* **1976**, *9*, 273.
(13) Sarko, A.; Southwick, J.; Hayashi, J. *Macromolecules* **1976**, *9*, 857.
(14) Gardiner, E. S.; Sarko, A. *Can. J. Chem.* **1985**, *63*, 173.
(15) The "parallel-up" and "parallel-down" structures are defined according to the definition given in the following: French, A. D.; Howley, P. S. In *Cellulose and Wood, Chemistry and Technology*; Schuerch, C., Ed.; Wiley: New York, 1989; p 159.

10.1021/ja0257319 CCC: $22.00

cellulose I which could be explained only by a system consisting of two distinct crystal phases, designated Iα and Iβ.[16] It became clear that the relative amount of Iα and Iβ was dependent on cellulose origin. Whereas *Valonia* and bacterial cellulose are rich in Iα,[17] cellulose from tunicate (*Halocynthia roretzi*) or animal cellulose consists of predominantly Iβ.[18,19] Recently, samples of nearly pure Iα have been reported in the cell wall of *Glaucocystis*.[20] Clearly, the structure of cellulose I must be revised by two separate structures for cellulose Iα and Iβ. Conceivably, these structures could be determined from refinement against X-ray diffraction patterns collected from highly crystalline mixed-phase cellulose samples, such as *Valonia*, separated into Iα and Iβ subdiagrams. Another approach is to collect diffraction data from samples consisting of purely Iα or Iβ phase. The second approach has recently become feasible through advances in sample characterization and is much preferred to the first because of the uncertainties in separating overlaid diffraction features from mixed-phase cellulose.

In recent X-ray and electron diffraction studies, the space group and chain packing of the Iα and Iβ phases have been characterized.[21,22] Cellulose Iα has a triclinic unit cell containing one chain, whereas cellulose Iβ has a monoclinic unit cell containing two parallel chains, similar to the approximate unit cell proposed for cellulose I by Gardner and Blackwell and Sarko and Muggli.[9,10] The "parallel-up" chain-packing organization favored by Sarko and Muggli[10] has been confirmed by an elegant electron microscopy study.[23] These results have allowed a number of molecular descriptions for Iα and Iβ to be produced in modeling studies.[24−30] There has also been a re-examination of the cellulose Iβ structure determined from the X-ray patterns of *Valonia* cellulose.[31] However, there have been no reported atomic coordinates derived from diffraction data collected from samples of pure Iα or Iβ phase.

In this work, which follows a recent revised description of cellulose II,[32,33] we describe the crystal structure and hydrogen-bonding arrangement in cellulose Iβ, as determined by synchrotron X-ray and neutron diffraction studies. Oriented films consisting of a parallel assembly of microcrystals of nearly pure cellulose Iβ from tunicate have been prepared, and we report synchrotron X-ray diffraction data to nearly atomic resolution ($d < 1$ Å, >300 unique reflections). Previously reported X-ray data from *Valonia* cellulose extend to $d \approx 1.85$ Å and consist of around 40 diffraction spots.[9,10] Analysis of our new data is described, and a high-resolution crystal structure for cellulose Iβ is presented.

Despite the high resolution of the X-ray data, we were unable to reliably determine hydrogen atom positions. Hydrogen is a relatively weak scatterer of X-rays but not of neutrons. Neutrons are also scattered in a significantly different way by hydrogen and its isotope, deuterium. The power of neutron diffraction in combination with specific deuteration for locating hydrogen atoms and investigating hydrogen-bonding in cellulose fibers has already been demonstrated.[32] Previously, we reported a novel intracrystallization technique for replacing the six independent hydrogen atoms involved in hydrogen-bonding by six deuterium atoms in cellulose Iβ, without any loss in crystalline perfection.[34] We report here neutron diffraction data from both hydrogenated and deuterated cellulose Iβ samples to resolutions comparable to those obtained with synchrotron X-rays, but with markedly different relative intensities. A Fourier synthesis-based analysis of these data is described, and the hydrogen positions and the hydrogen-bonding arrangement are presented.

Our approach differs from earlier fiber diffraction analyses of cellulose. The exceptional resolution and wealth of data has allowed us to use crystallographic analysis methods usually reserved for small-molecule systems. In particular, the conformations of the hydroxymethyl groups have been determined by their direct location in X-ray Omit maps. We have combined X-ray and neutron diffraction in order to obtain the position of all atoms, including the hydrogen atoms. We have used recently developed data analysis software that, for the first time, takes account of textural effects occurring in fiber samples.[35]

X-ray Experimental Section

(A) Sample Preparation. Cellulosic mantles of tunicate (*Halocynthia roretzi*) were thoroughly purified and hydrolyzed into microcrystals with sulfuric acid, followed by reconstitution into oriented films as described previously.[36] Highly oriented films were selected under a polarization microscope and stacked parallel on top of one another up to a thickness of ~200 μm. The films readily stuck together by wetting and drying and became stiff enough to stand unsupported. The sample was finally trimmed with a sharp blade to a width of about 500 μm.

(B) Data Collection. The sample was fixed with modeling clay to a goniometer head and mounted on four-circle diffractometer ID2A at the European Synchrotron Radiation Facility (Grenoble), so that the fiber axis was parallel to the ϕ axis. Data were collected using an online MAR345 image plate with a specimen-to-detector distance of 175 mm, a wavelength of 0.7208 Å, and a beam size of 100 μm. During data collection, it was discovered that diffraction from the sample was not truly cylindrically symmetric about the fiber axis. The observed texture corresponded to the *c* axis having an ellipsoidal distribution about the fiber axis and the *b* axis having preferred orientation in the plane of the constituent films and perpendicular to the fiber axis. To account for this, a series of images was collected at 15° intervals in ϕ. In addition to this scan, images were collected with χ at 90° and stepping ω in 2° intervals in order to collect data in the near meridional region.

Each image was remapped into reciprocal space, where the background component was removed using a two-dimensional extension of the algorithm by Sonneveld and Visser.[37] An orientation function was measured from the intensity distribution of the (0,0,4) reflection in reciprocal space, and the ratios of the intensities of reflections

(16) Atalla, R. H.; VanderHart, D. L. *Science* **1984**, *223*, 283.
(17) VanderHart, D. L.; Atalla, R. *Macromolecules* **1984**, *17*, 1465.
(18) Belton, P. S.; Tanner, S. F.; Cartier, N.; Chanzy, H. *Macromolecules* **1989**, *22*, 1615.
(19) Larsson, P. T.; Westermark, U.; Iversen, T. *Carbohydr. Res.* **1995**, *278*, 339.
(20) Imai, T.; Sugiyama, J.; Itoh, T.; Horii, F. *J. Struct. Biol.* **1999**, *127*, 248.
(21) Sugiyama, J.; Okano, T.; Yamamoto, H.; Horii, F. *Macromolecules* **1990**, *23*, 3196.
(22) Sugiyama, J.; Vuong, R.; Chanzy H. *Macromolecules* **1991**, *24*, 4168.
(23) Koyama, M.; Helbert, W.; Imai, T.; Sugiyama, J.; Henrissat, B. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 9091.
(24) Heiner, A. P.; Sugiyama, J.; Teleman, O. *Carbohydr. Res.* **1995**, *273*, 207.
(25) Aabloo, A.; French, A. D.; Mikelsaar, R.-H.; Pertsin, A. J. *Cellulose* **1994**, *1*, 161.
(26) Aabloo, A.; French, A. D. *Macromol. Theory Simul.* **1994**, *3*, 185.
(27) Kroon-Batenburg, L. M. J.; Kroon, J. *Glycoconjugate J.* **1997**, *14*, 677.
(28) Viëtor, R. J.; Mazeau, K.; Lakin, M.; Pérez, S. *Biopolymers* **2000**, *54*, 342.
(29) Simon, I.; Glasser, L.; Sheraga, A.; St John Manley, R. *Macromolecules* **1988**, *21*, 990.
(30) Hardy, B. J.; Sarko, A. *Polymer* **1996**, *37*, 1833.
(31) Finkenstadt, V. L.; Millane, R. P. *Macromolecules* **1998**, *31*, 7776.
(32) Langan, P.; Nishiyama, H.; Chanzy, H. *J. Am. Chem. Soc.* **1999**, *121*, 9940.
(33) Langan, P.; Nishiyama, Y.; Chanzy, H. *Biomacromolecules* **2001**, *2*, 410.
(34) Nishiyama, Y.; Isogai, A.; Okano, T.; Müller, M.; Chanzy, H. *Macromolecules* **1999**, *32*, 2078.
(35) Nishiyama, Y.; Langan, P. *Fiber Diffr. Rev.* **2000**, *9*, 18.
(36) Nishiyama, Y.; Kuga, S.; Wada, M.; Okano, T. *Macromolecules* **1997**, *30*, 6395.
(37) Sonneveld, E. J.; Visser, J. W. *J. Appl. Cryst.* **1975**, *8*, 1.

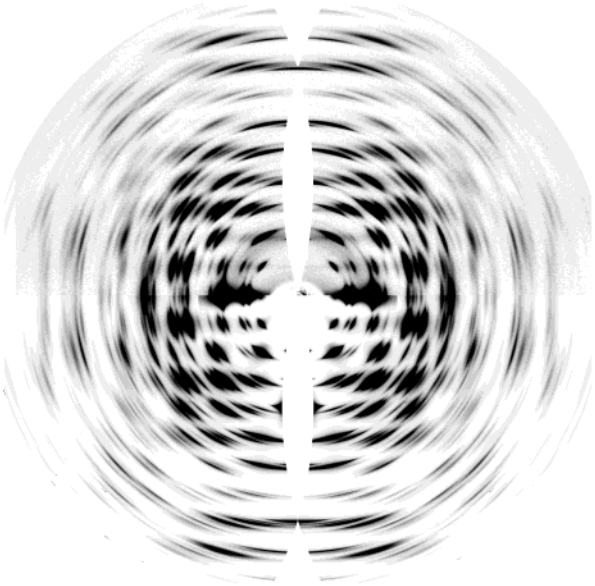

Figure 1. (Top) Synchrotron X-ray diffraction data collected on an online MAR image plate from fibers of *Halocynthia* cellulose Iβ on station ID2A at the ESRF, Grenoble, France. (Bottom) A 3D fit of the Bragg intensities, done using custom-written software that takes into account fiber texture. The images have been remapped into cylindrical reciprocal space with the fiber axis vertical.

Table 1. Experimental Details

	X-ray	neutron
	Crystal Data	
chemical formula	$C_{12}H_{20}O_{10}$	$C_{12}H_{14}D_6O_{10}$
cell setting, space group	monoclinic, $P2_1$	monoclinic, $P2_1$
a (Å)	7.784(8)	7.784(8)
b (Å)	8.201(8)	8.201 (8)
c (Å)	10.380(10)	10.380(10)
γ (°)	96.5	96.5
V (Å^3)	658.3(11)	658.3(11)
Z	2	
radiation type	synchrotron X-ray	neutron
λ (Å)	0.72080	1.30580
	Data Collection	
diffractometer	ID2A	D19
independent reflections	312	216
reflections $>2\sigma(I)$	298	
θ_{max} (°)	21.10	45.83
range of h	$0 \rightarrow 7$	$0 \rightarrow 4$
range of k	$-8 \rightarrow 8$	$-4 \rightarrow 4$
range of l	$0 \rightarrow 10$	$0 \rightarrow 11$
	Refinement	
refinement on	F^2	F^2
R [$F^2 > 2\sigma(F^2)$]	0.1857	0.2095
$\omega R(F^2)$	0.4242	0.4272
$\Delta\rho_{max}$	0.486	0.317
$\Delta\rho_{min}$	−0.597	−0.282

(1,−1,0) to (1,1,0) and (2,0,0) to (1,1,0) were used to calculate the diffraction profiles of the reflections in reciprocal space as described previously.[35] The intensity and position of each spot were determined by fitting the measured orientation function and diffraction spot profile to 31 images recorded at different sample orientations, using linear least-squares methods. An estimate of the error associated with each intensity was determined by repeating the least-squares fit using different subsets of the 31 frames and then taking the standard deviation. A comparison of the fit to the data is shown in Figure 1. A summary of the experimental parameters is given in Table 1.

(C) Structure Solution. X-ray structure refinement was carried out using SHELX-97.[38] The space group of cellulose Iβ was $P2_1$, and the

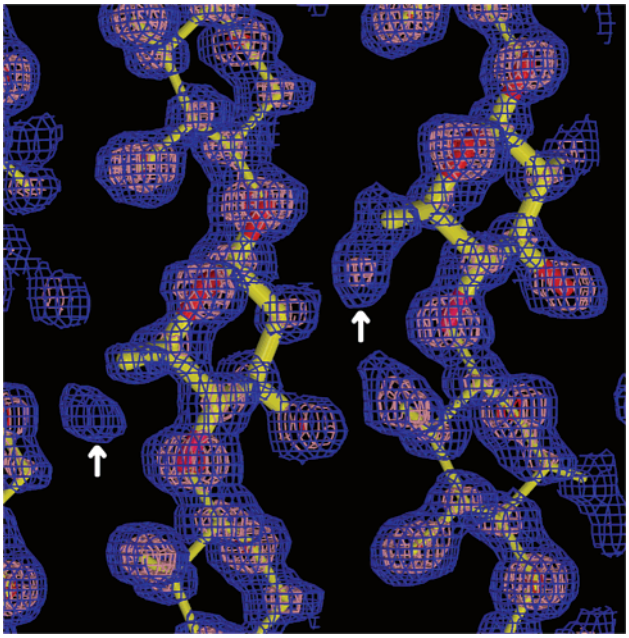

Figure 2. A section through an Omit map containing both an origin chain (left) and a center chain (right), calculated using the observed amplitudes and model phases but omitting the hydroxymethyl group oxygen atoms from the phase calculation. The skeletal model represents the cellulose chains, with one residue of both the origin and center chains highlighted by thicker lines. The map is represented at two contour levels in blue and pink. Density (indicated by arrows) can be clearly associated with the hydroxymethyl group in the *tg* position.

unit cell parameters were $a = 7.784$ Å, $b = 8.201$ Å, c (chain direction and unique axis) $= 10.38$ Å, and $\gamma = 96.5°$. Two residuals were monitored during refinement, a conventional F-based residual, designated R, and an F^2-based weighted residual, designated $R\omega$.[39] Since SHELX-97 refines against F^2, $R\omega$ was used in Hamilton's statistical test[40] to determine the significance of changes in agreement with the data. Initial atomic parameters were taken from the Iβ component of the structure of *Valonia* cellulose.[31] Refining only the scale factor and a global thermal parameter resulted in values of 0.2706 and 0.7244 for R and $R\omega$, respectively. To confirm that the starting model had the correct hydroxymethyl group conformations, a refinement was carried out with the hydroxymethyl group atoms removed. The calculated Omit map showed no sign of hydroxymethyl group disorder and clearly indicated that both groups were in the *tg* conformation (Figure 2).

All atomic positions were then allowed to refine with bond lengths and angles restrained to the mean of the values reported in the single-crystal structure of cellotetraose,[41,42] with weights proportional to the inverse of their standard deviation, using the DFIX and DANG options in SHELX-97. The coordinates of hydrogen atoms covalently bonded to carbon were determined using the HFIX option in SHELX-97 and were subsequently allowed to ride on the coordinates of the carbon atoms. Non-hydrogen atoms were constrained to have a global thermal parameter. Hydrogen atoms had thermal parameters tied to 1.2 times the global thermal parameter. Those hydrogen atoms covalently bonded to oxygen were not refined at this stage. The refinement involved 66 additional parameters and 186 restraints and resulted in values of 0.2186

(38) Sheldrick, G. M. *SHELX-97*, a program for the refinement of Single-Crystal Diffraction Data; University of Gottingen: Gottingen, Germany, 1997.
(39) R is calculated from $\sum(|F_o| - |F_c|)/\sum|F_o|$, with $F_o > 4\sigma$, where F_o and F_c are the observed and calculated amplitudes, respectively. $R\omega$ is calculated from $[\sum\omega(F_o^2 - F_c^2)^2/\sum\omega(F_o^2)^2]^{1/2}$, where ω is a weight $(1/\sigma^2)$ applied to each F^2 term in the least-squares refinement. $R\omega$ will be more than twice the size of the R.
(40) Hamilton, W. C. *Acta Crystallogr.* **1965**, *18*, 502.
(41) Gessler, K.; Krauss, N.; Steiner, T.; Betzel, C.; Sarko, A.; Saenger, W. *J. Am. Chem. Soc.* **1995**, *117*, 11397.
(42) Raymond, S.; Heyraud, A.; Tran Qui, D.; Kvick, A.; Chanzy, H. *Macromolecules* **1995**, *28*, 2096.

and 0.551 for R and $R\omega$, respectively. The resultant structure was in significantly better agreement with the data at a confidence level of 99.5%.

An analysis of variance showed that R was relatively flat, with a small rise at very high resolution, as might be expected, when calculated for reflections grouped in resolution shells. When K, defined as the mean F_o^2 over the mean F_c^2, was calculated for reflections grouped in shells of decreasing $F_c/F_c(\max)$, where $F_c(\max)$ is the maximum value of F_c, it started to rise below a value of about 0.03. This indicates that, on average, the measured weak intensities are too large, probably because of insufficient background subtraction. To restore a flat analysis of variance, and therefore to remove any statistical bias from the refinement, all reflections with $F_c/F_c(\max) < 0.03$ were excluded, as explained in a previous report.[33] This reduced the number of data from 344 to 312 and resulted in the structure, designated *A*, with values of 0.1857 and 0.425 for R and $R\omega$. Residual density in the Fourier difference map had a root-mean-square deviation from the mean of 0.14 e/Å^3, and maximum and minimum peak values of 0.49 and −0.6 e/Å^3. We were not able to reliably locate the hydroxyl group hydrogen atoms.

Allowing individual isotropic thermal parameters to refine resulted in values of 0.1747 and 0.405 for R and $R\omega$, respectively, and involved 22 additional parameters. This did not involve a large change in the atomic coordinates and was not a significant improvement in the agreement with the data at a high level of confidence. A refinement, designated *B*, with the sugar rings of each chain restrained to have the same stereochemical parameters, using the SAME command in SHELX-97, resulted in values of 0.1975 and 0.490 for R and $R\omega$. A refinement, designated *C*, with the glycosidic torsion angles, Φ and Ψ, and the sugar ring stereochemical parameters of each chain restrained to be the same, using the command SADI in SHELX-97, resulted in values of 0.2052 and 0.512 for R and $R\omega$. Refinements B and C could be rejected with respect to refinement A at a confidence level of >99.5%. The coordinates of the final model, A, are given in the Crystallographic Information File supplied as Supporting Information.

Neutron Experimental Section

(A) Sample Preparation. The neutron sample was prepared in a fashion similar to the X-ray sample, but with larger dimensions (3.5 mm × 7.0 mm × 10 mm) in order to exploit the relatively weak flux of the available neutron beam. Two samples were prepared, one hydrogenated, designated H-cellulose-Iβ, and the other with all OH moieties replaced by OD, designated D-cellulose-Iβ. Intracrystalline deuteration of the D-cellulose-Iβ sample was carried out by a hydrothermal treatment achieved in 0.1 N NaOD in D_2O at 210 °C for 30 min, as described previously.[34]

(B) Data Collection. Neutron diffraction data were collected from both samples on D19 at the Institute Laue-Langevin (Grenoble). D19 is a four-circle diffractometer equipped with a large 4° × 64° position-sensitive detector. Generic strategies for collecting fiber diffraction data from cellulose fibers on D19 have been described previously.[43,44] The goniometer angles were stepped in order to collect slices of reciprocal space that each contained the fiber axis but that were oriented about the fiber axis by approximately 0°, 45°, and 90° with respect to the constituent film surfaces. After remapping of the data into reciprocal space, the intensities of diffraction spots were determined for each slice of reciprocal space by fitting the orientation function and reflection profile determined from the X-ray analysis using linear least-squares methods. Measured intensities from each slice were then averaged, and their standard deviation was taken as representative of the uncertainty in their measurement. The unit cell parameters determined in the X-ray analysis were used in fitting the reflection positions. A representation of the data after binning into reciprocal space is shown in Figure 3. There are clear differences in the intensities diffracted from D-cellulose-Iβ and H-cellulose-Iβ, particularly in the meridional regions. The experimental parameters are summarized in Table 1.

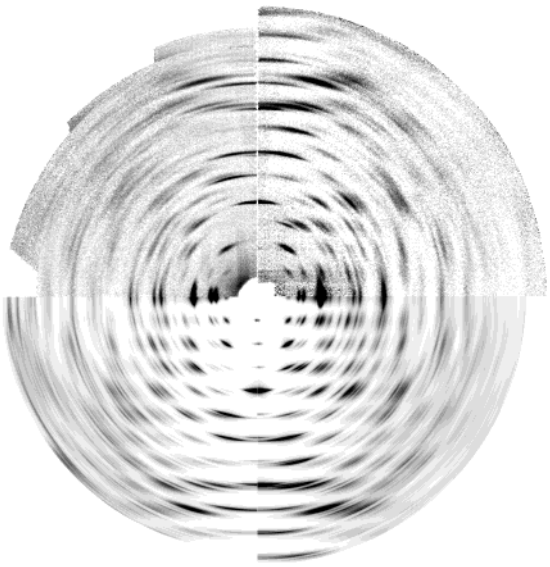

Figure 3. Neutron fiber diffraction patterns collected from two fibers of *Halocynthia* cellulose Iβ, one hydrogenated (top left-hand quadrant) and the other deuterated (top right-hand quadrant). The bottom quadrants show 3D fits of the Bragg intensities, done using custom-written software that takes into account fiber texture. The images have been remapped into cylindrical reciprocal space with the fiber axis vertical.

(C) Deuterium Atom Location. The positions of deuterium atoms were identified using the Fourier difference synthesis calculated with coefficients $(F_d - F_h) \exp i\alpha_c$, where F_d and F_h are the observed structure factor amplitudes from D-cellulose-Iβ and H-cellulose-Iβ, respectively, and α_c are phases calculated from the X-ray structure, A, reported here. Well-defined difference density peaks, shown in Figure 4, could be clearly identified with possible deuterium atom positions associated with the secondary alcohol O3 atoms. The atom numbering scheme is illustrated in Figure 6, and labels for atoms of the origin and center chains were postfixed with "o" and "c", respectively. Difference density peaks could also be associated with possible deuterium atom positions associated with the secondary alcohol O2 atoms and the primary alcohol O6 atoms, but they were weaker and less well-defined. The identified positions were refined against the neutron diffraction data collected from D-cellulose-Iβ using SHELX-97.[38]

Positions of non-deuterium atoms identified in the X-ray analysis were not refined. With no deuterium atoms included, the scale factor and a global thermal parameter were refined against 141 reflections with $d > 1.3$ Å, resulting in values of 0.3487 and 0.6761 for R and $R\omega$. Adding each deuterium atom increased the number of parameters by 3 and the number of restraints by 2. The O−D and C−O−D bond lengths and angles were restrained to 0.98 Å and 110°, respectively. The values of R after including the deuterium atoms bound to the O2, O3, and O6 atoms were 0.2744, 0.2668, and 0.2349, and the values of $R\omega$ were 0.5802, 0.5435, and 0.4565. At each step of the refinement, adding the deuterium atoms has significantly improved the agreement of the model with the neutron data at a confidence level of >99%. The higher resolution data were then included in the refinement in steps of 0.1 Å, with the final refinement corresponding to 216 data and values of 0.2228 and 0.4600 for R and $R\omega$, respectively.

A Sigma-A Fourier synthesis was carried out in order to detect any problems with the final model.[45] The map, shown in Figure 5, indicates

(43) Nishiyama, Y.; Okano, T.; Langan, P.; Chanzy, H. *Int. J. Biol. Macromol.* **1999**, *26*, 279.
(44) Langan, P.; Denny, R. C.; Mahendrasingam, A.; Mason S. A.; Jaber, A. *J. Appl. Crystallogr.* **1996**, *29*, 383.
(45) Read, R. J. *Acta Crystallogr.* **1986**, *A42*, 140.

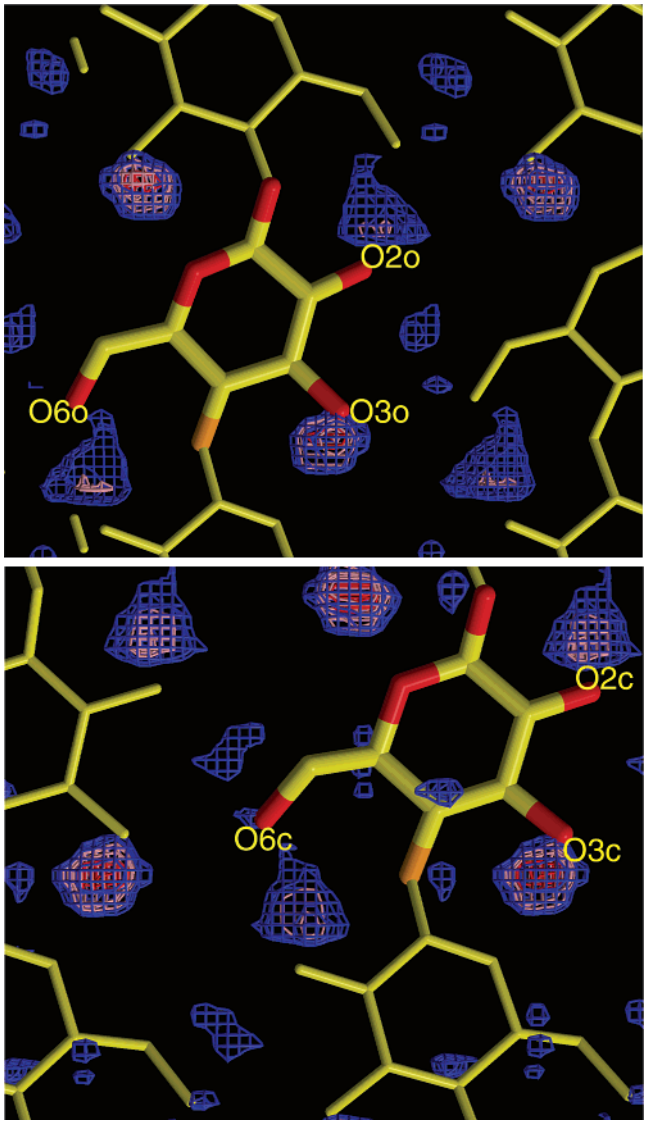

Figure 4. Sections containing the origin sheet (top) and center sheets (bottom) through the Fourier difference synthesis calculated with coefficients $(F_d - F_h) \exp i\alpha_c$, where F_d and F_h are the observed structure factor amplitudes from D-cellulose-Iβ and H-cellulose-Iβ, respectively, and α_c are phases calculated from the X-ray structure, A, reported here. The skeletal model represents the cellulose chains, with one residue of both the origin and center chains highlighted by thicker lines. The difference density is represented at two contour levels in blue and pink. The positions of deuterium atoms associated with the secondary alcohol oxygen O3 are clearly indicated by well-defined density peaks. Density indicative of the deuterium atoms associated with the secondary alcohol oxygen O2 and the primary alcohol oxygen O6 is also present but less well defined.

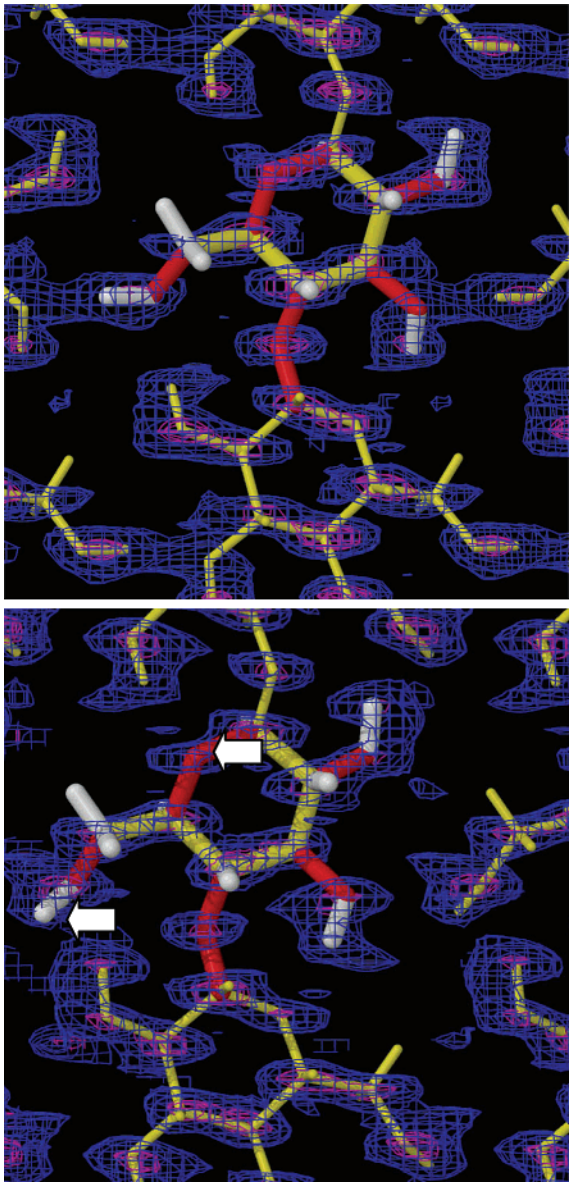

Figure 5. Sections through the center sheet (top) and origin sheet (bottom) of the Sigma-A Fourier synthesis of the structure of cellulose Iβ incorporating the deuterium atom positions. The Fourier synthesis was calculated using coefficients $2mF_o - F_c$, where m is calculated from F_o and F_c in order to incorporate an estimate of the phase error.[45] The skeletal model represents the cellulose chains, with one residue highlighted by thicker lines. The map is represented in blue. Negative density associated with hydrogen atoms bound to carbon is not shown for clarity. Careful inspection of the map indicates that the overall agreement of the model with the data is good, but that there is a discrepancy with the position of the D6o atom (arrow), and the intracyclic C1o−O5o−C5o (arrow) is slightly too small.

that the overall agreement of the model with the neutron data is excellent but that D6o attached to O6o should point more toward the O3o atom of the neighboring chain, and that the intracyclic C1o−O5o−C5o angle is slightly too small and the corresponding bond lengths are too long. The angle discrepancy is probably due to the use of over-restrictive restraints, and we took no further action. Changing the D6o position to agree with the Sigma-A map did not improve the agreement with the data, and the discrepancy suggests the presence of disorder at this position.

The presence of deuterium atom disorder was investigated further. For each deuterium atom position identified in the Fourier difference analysis, a second alternative deuterium atom was added if there existed the possibility of an alternative hydrogen bond. The total occupancy of the two deuterium components was constrained to unity. No reasonable alternatives existed for the D3 atom positions identified in the Fourier difference analysis. For the D2 and D6 atom positions, there were alternatives to those identified in the Fourier analysis with reasonable hydrogen-bonding geometry. Refining disordered D6 deuterium atoms resulted in values of 0.2113 and 0.4438 for R and $R\omega$. Further refinement with disordered D2 atoms resulted in values of 0.2094 and 0.4261 for R and $R\omega$, respectively. Each refinement step involved eight additional parameters and four additional restraints. These improvements in the agreement with the data are significant at a confidence level of >97.5%. Refining the second disordered deuterium atom components, designated D2oB, D2cB, D6oB, and D6cB, did not significantly change the position of the deuterium atoms identified in the Fourier difference analysis, designated D2oA, D2cA, D6oA, and

Table 2. Hydrogen Bonds with H···A < *r*(A) + 2.000 Å and ∠DHA > 110°

D–H	*d*(D–H)	*d*(H···A)	∠DHA	*d*(D··A)	A
O2o−D2oA	0.977	1.832	158.72	2.765	O6o [$-x,-y,z+1/2$]
O2o−D2oA	0.977	2.304	110.28	2.797	O1o [$-x,-y,z+1/2$]
O3o−D3o	0.979	1.966	137.08	2.764	O5o [$-x,-y,z-1/2$]
O6o−D6oA	0.979	2.040	144.26	2.892	O3o [$x,y+1,z$]
O6o−D6oB	0.974	1.876	150.23	2.765	O2o [$-x,-y,z-1/2$]
O6o−D6oB	0.974	2.152	121.59	2.789	O1o
O2c−D2cA	0.982	1.904	165.12	2.865	O6c [$-x+1,-y+1,z+1/2$]
O2c−D2cB	0.978	2.440	135.44	3.211	O6c [$x,y-1,z$]
O3c−D3c	0.983	1.752	162.23	2.705	O5c [$-x+1,-y+1,z-1/2$]
O6c−D6cA	0.985	1.779	156.61	2.711	O3c [$x,y+1,z$]
O6c−D6cA	0.985	2.544	124.98	3.211	O2c [$x,y+1,z$]
O6c−D6cB	0.975	1.967	152.06	2.865	O2c [$-x+1,-y+1,z-1/2$]
O62-D6cB	0.975	2.243	123.21	2.894	O1c

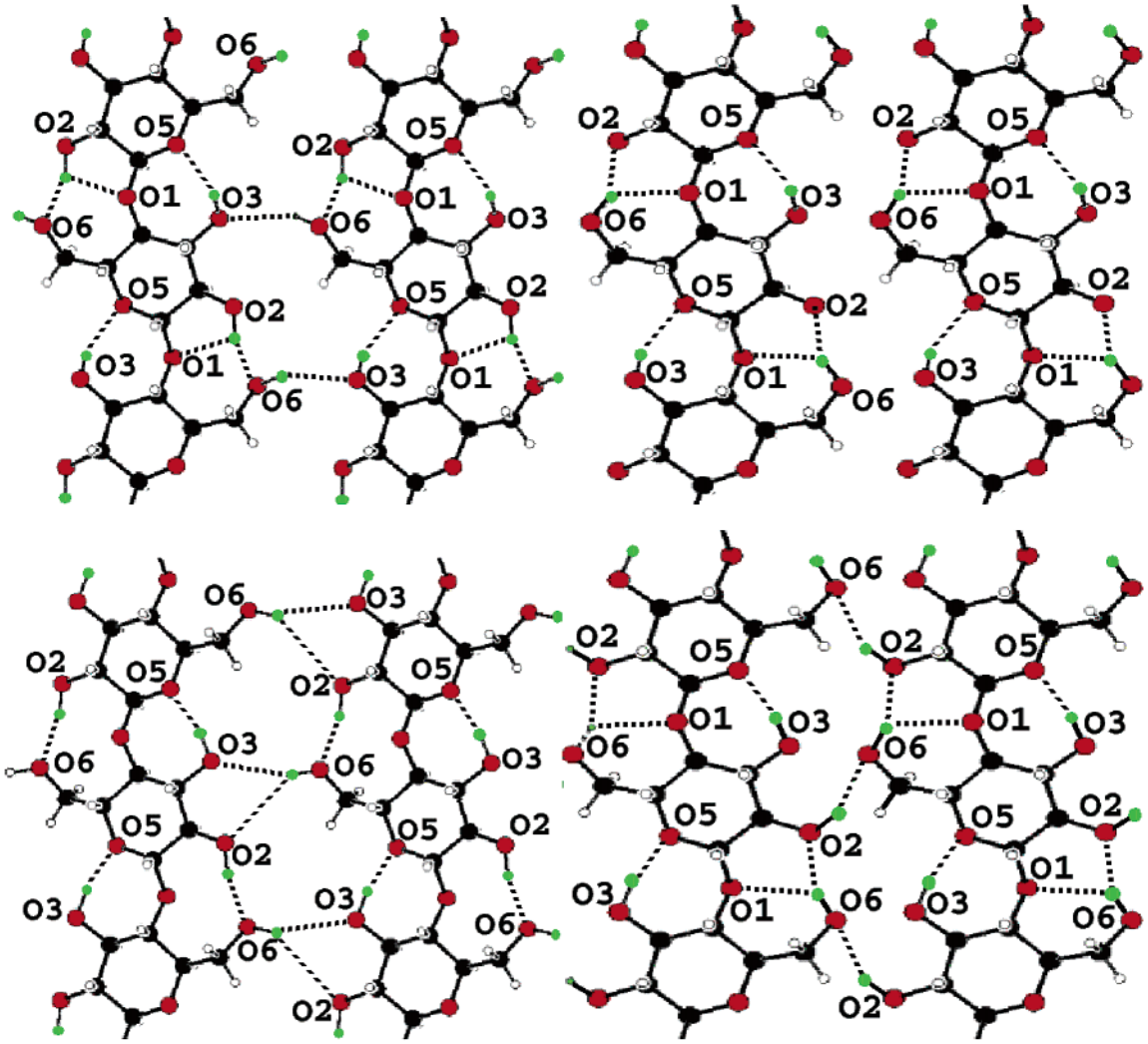

Figure 6. Schematic representation of the hydrogen bonds in the origin (top) and center (bottom) sheets of cellulose Iβ. Carbon, oxygen, hydrogen, and deuterium atoms are colored black, red, white, and green, respectively. Hydrogen bonds are represented by dotted lines. Only the oxygen atoms involved in hydrogen-bonding have been labeled for clarity. Deuterium atoms D2oA, D3o, and D6oA are included in the top left view (O2o−D2oA···O6o, O2o−D2oA···O1o, O3o−D3o···O5o, O6o−D6oA···O3o), D3o and D6oB in the top right view (O3o−D3o···O5o, O6o−D6oB···O2o, O6o−D6oB···O1o), D2cA, D3c, and D6cA in the bottom left view (O2c−D2cA···O6c, O3c−D3c···O5c, O6c−D6cA···O3c, O6c−D6cA···O2c), and D2cB, D3c, and D6cB in the bottom right view (O2c−D2cB···O6c, O3c−D3c···O5c, O6c−D6cB···O2c, O6c−D6cb···O1c).

D6cA, but did change their occupancies. The occupancy of the disordered D2oB atom went to zero, and it was dropped from the refinement. The hydrogen-bonding parameters are given in Table 2. The coordinates of the hydrogen atoms are given in the Crystallographic Information File supplied as Supporting Information. The hydrogen-bonding scheme is represented schematically in Figure 6.

Discussion

The unit cell parameters reported here for cellulose Iβ are similar to those reported previously for cellulose I, except in the *a* direction. As opposed to the present value of $a = 7.784$ Å, an earlier value of 8.01 Å was deduced from an electron

Table 3. Selected Model Parameters

	origin chain					center chain				
	Φ	Ψ	χ	χ′	θ	Φ	Ψ	χ	χ′	θ
A	−98.5(20)	−142.3(19)	170(3)	−70(3)	10.2	−88.7(20)	−147.1(16)	158(3)	−83(3)	6.7
B	−95.0(20)	−139.1(12)	161(2)	−80(2)	2.2	−89.96(20)	−143.8(13)	164(2)	−79(2)	2.2
C	−93.9(20)	−143.8(11)	164(2)	−77(2)	3.2	−92.2(20)	−145.5(14)	161(2)	−80(2)	3.2

diffraction analysis of the Iβ fraction of *Microdictyon* cellulose.[22] It is likely that this value was overestimated since the beam damage resulting from the electron beam irradiation has a tendency to swell the *a* parameter of cellulose.[46] Other values of *a*, namely 7.86 [9] and 7.88, [10] were obtained by X-ray for *Valonia* cellulose. These values are also overestimated since *Valonia* is dominated by cellulose Iα, which has a lower density[22] and therefore larger lattice parameters than those of cellulose Iβ. Selected molecular parameters for the structure presented here, A, and those of the rejected structures, B and C, are given in Table 3. Structure C is similar to that recently proposed for cellulose Iβ,[31] obtained by revising the refinement against the data published by Gardner and Blackwell[9] and Sarko and Muggli,[10] collected from fibers of *Valonia* cellulose. Allowing the relative orientation of adjacent glycosyl residues, described by the glycosidic torsion angles Φ and Ψ,[47] to refine in an unrestrained manner has introduced a slight difference in the conformation of the glycosidic linkages of the origin chain and the center chain in structure B. In structure A, where the sugars have also been refined without restraints, this difference in glycosidic linkage conformation is more significant (Φ and Ψ are −98.5(20)° and −142.3(19)° for the origin chain and −88.7(20)° and −147.1(16)° for the center chain), and the conformations of the hydroxymethyl groups of the different chains, described by the torsion angle χ and χ′,[47] are different (χ, χ′ are 170(3)°, −70(3)° for the origin chain and 158(3)°, −83(3)° for the center chain). The calculated Cremer and Pople[48] puckering parameters indicate that, relative to the ideal value for an unstrained α-D-glycopuranose ($\theta = 2.7°$),[49] the sugars of both origin ($\theta = 10.2°$) and center ($\theta = 6.7°$) chains are conformationally strained in structure A.

Differences between the origin and center chains in cellulose II and its related oligomers have been reported in X-ray studies of cellotetraose single crystals,[41,42] in X-ray and neutron studies of cellulose fibers,[32,33] and in modeling studies.[27] It has been suggested that the chains are different because sheets containing origin chains can pack snugly against sheets containing center chains if the center chains adopt a slightly unfavorable conformation to permit tight intersheet contacts.[41] Thus, the center chain in cellotetraose has a strained sugar conformation ($\theta = 13(1)°$), while the origin chain has an unstrained sugar conformation ($\theta = 2(1)°$). Intersheet O—H···O hydrogen-bonding imposes different hydrogen-bonding requirements on the two chains and may contribute significantly to the conformational differences. There are no O—H···O intersheet hydrogen bonds in cellulose Iβ. There are a number of C—H···O hydrogen bonds, but there is no great difference in the mean C···O bond length between structures A and C (the means of the six shortest C···O donor−acceptor distances are 3.38, 3.96, and 3.4 Å for structures A, B, and C). Perhaps more significantly, both structures B and C contain overshort hard contacts between the hydrogen atom attached to C1o of the origin chain and one of the hydrogen atoms attached to C6c of the center chain (the H−H distances are 2.2620 and 0.2138 Å in structures B and C). Allowing the sugar conformations to refine relaxes this contact while maintaining favorable C−H···O interactions in structure A.

The existence of nonequivalent chains provides a possible explanation for the fine detail observed in the ^{13}C CP-MAS spectra of cellulose Iβ.[50] The resonances assigned to the C1, C4, and C6 atoms all show distinct splitting, with the splitting at C1 being the most pronounced. Horii et al.[51,52] have shown that there is a correlation between the chemical shifts in CP-MAS spectra and the dihedral angles defined by the bonds associated with the particular carbon atom. Thus, the chemical shift of the C6 resonance can be correlated with χ and χ′, and the chemical shifts associated with C1 and C4 can be correlated with Φ and Ψ, respectively. The different values of χ, χ′, Φ, and Ψ for the nonequivalent chains reported here provide an explanation for these splittings.

One of the most striking results of this study is the atomic resolution of the geometry of the hydrogen bonds that are present in the crystals of cellulose Iβ. As noted in many previous reports, there is no hint of intersheet O—H···O hydrogen bonds in cellulose Iβ, and therefore the cellulose sheets are held together by only hydrophobic interactions and weak C—H···O bonds. Within each sheet, the difference maps calculated from our neutron data allow direct visualization of the hydrogen bonds O3—H···O5 from the origin sheet as well as those from the center sheet. Not only are these hydrogen bonds visualized, but their bond lengths and angles can be very precisely quantified. The spherical shape of the density peaks defining the location of the D3o and D3c atoms indicates that in both chains, the hydrogen atom linked to O3 occupies a well-defined localized position. In contrast with the situation at O3, the hydrogen atoms linked to O2 and O6 are associated with nonspherical density peaks and are split between more than one location. Thus, for these hydrogen bonds, there is a spectrum of possibilities. In particular, a close scrutiny of the maps indicates that each hydroxymethyl group of a given cellulose chain is partly engaged in an intramolecular hydrogen bond with the O2 atom

(46) Revol, J.-F. *J. Mater. Sci. Lett.* **1985**, *4*, 1347.
(47) The glycosidic torsion angles, Φ and Ψ, which describe the relative orientation of adjacent glycosyl residues in the same chain are defined by (O5−C1−O1−C4) and (C1−O1−C4−C5), respectively. The conformation of the hydroxymethyl group is defined by two letters, the first referring to the torsion angle χ (O5−C5−C6−O6) and the second to the torsion angle χ′ (C4−C5−C6−O6). An ideal *tg* conformation would be defined as the set of two angles 180°, −60°.
(48) Cremer, D.; Pople, J. A. *J. Am. Chem. Soc.* **1975**, *97*, 1354.
(49) Dowd, M. K.; French, A. D.; Reilly, P. *J. Carbohydr. Res.* **1994**, *264*, 1.
(50) Atalla, R. H. In *Comprehensive Natural Products Chemistry, Carbohydrates and their derivatives including tannins, cellulose and related lignins*; Pinto, B. M., Ed.; Elsevier: Cambridge, 1999; Vol. 3, p 529.
(51) Horii, F.; Hirai, A.; Kitamaru, R. *Polym. Bull.* **1983**, *10*, 357.
(52) Horii, F.; Hirai, A.; Kitamaru, R. In *Polymers for Fibers and Elastomers*; Arthur, J. C., Ed.; ACS Symposium Series 260; American Chemical Society: Washington, DC, 1984; p 27.

of the adjacent glucosyl moieties of the same chain and partly in an intermolecular hydrogen bond with another chain of the same sheet.

Both chains have intramolecular O2—H···O6 hydrogen bonds. However, in the origin sheet, D2oA is also in a position to form a weak hydrogen bond to O1o of the same chain in a three-centered hydrogen bond arrangement (O2o—D2oA···O6o/O1o). In the center sheet, the D2cA atom involved in the O2—H···O6 hydrogen bond has 62% occupancy. A disordered component, D2cB with 38% occupancy, is in a position to hydrogen bond to O6c of a neighboring chain. The hydrogen-bonding involving the O6 hydroxymethyl groups is more complicated. Both chains have O6—H···O3 interchain hydrogen bonds, with 81% and 74% occupancy for the D6oA and D6cA atoms. In the center sheet, the D6cA atom can also donate to O2c in a three-centered hydrogen bond (O6c—D6cA···O3c/O2c). The disordered components, D6oB with occupancy 19% and D6cB with occupancy 26%, are both in positions to hydrogen bond to O2 and O1 atoms of the same chain in a three-centered hydrogen bond arrangement (O6o—D6oB···O2o/O1o and O6c—D6cB···O2c/O1c).

We have modeled disorder by refining the deuterium atom positions so that the total occupancy associated with any hydroxyl group is constrained to unity. However, an alternative approach is to recognize that some positions are mutually exclusive. The total occupancy of D2oA/D6oB and D2cA/D6cB should not be greater than unity. From our refinements, the percentage occupancies of D2oA/D6oB and D2cA/D6cB are 84%/16% and 71%/29%, respectively, and the total occupancies are 0.88 and 1.19, respectively. When the total occupancies of D2oA/D6oB and D2cA/D6cB are constrained to be unity, the percentage occupancies of D2oA/D6oB and D2cA/D6cB refine to 86%/14% and 63%/37%, respectively. These two different approaches to treating the disorder give qualitatively consistent results.

At first glance, the multiple possibilities for the H-bonds at O2 and O6 is surprising, considering the perfection of the crystals of tunicate cellulose, and one would be tempted to correlate this uncertainty with some shortcoming in the structure determination that is based on the deconvolution of fiber diagrams. However, the multiple hydrogen-bonding scheme explains the complex O−C stretching bands observed in the infrared spectra of cellulose Iβ.[53] The multiple hydrogen-bonding model reported here is also in qualitative agreement with the hydrogen-bonding model deduced from the detailed assignment of O−H and C−O stretching bands in the infrared spectra. Taken together, our present diffraction data and the previous infrared observations substantiate the concept of intrinsic disorganization of the hydrogen bond pattern for O2 and O6 in cellulose Iβ, whereas the intramolecular hydrogen bond at O3 seems unambiguously well organized.

In cellulose Iβ, the exact nature of the hydrogen bond disorder is unclear. The principal deuterium atom positions identified in the neutron Fourier difference analysis can be regarded as a hydrogen-bonding network, designated I: intramolecular hydrogen bonds O3o—D3o···O5o, O3c—D3c···O5c, O2o—D2o···O6o/O1o, O2c—D2cA···O6cA and intermolecular hydrogen bonds O6o—D6oA···O3o, O6c—D6cA···O3c/O2c. The disordered deuterium atom positions can be regarded as a second hydrogen-bonding network, designated II: intramolecular hydrogen bonds O3o—D3o···O5o, O3c—D3c···O5c, O6o—D6oB···O2o, O6c—D6cB···O2c/O1c and intermolecular hydrogen bond O2c—D2cB···O6c. It is possible that these hydrogen-bonding networks exist exclusively at different positions in the sample, or they may reflect a dynamic balance between the two networks at any given position. However, it is also possible that the deuterium positions do not correspond to two exclusive hydrogen-bonding networks but rather to a more complicated situation where the local hydrogen-bonding geometry changes with both position and time. The apparent bifurcated hydrogen bonds probably reflect spatial and temporal averaging of two or more conventional two-centered hydrogen bonds, as observed in previous studies of cellulose II,[32,33] rather than true three-centered hydrogen bonds.

It is interesting to compare the hydrogen bond pattern reported here for cellulose Iβ with that reported in our previous neutron study of cellulose II.[32] In cellulose II, the location of hydrogen/deuterium atoms was such that it corresponded to a single hydrogen bond network. This hydrogen bond network accommodated a certain amount of disorder due to the presence of hydroxymethyl group disorder. In cellulose Iβ, there would appear to be no detectable hydroxymethyl group disorder, and yet the hydrogen-bonding pattern is more disordered than in cellulose II. It is possible that the higher resolution data obtained from the samples of tunicin, which are clearly more crystalline than those of cellulose II, have allowed us to identify the presence of disorder in greater detail. However, it is also possible that the more disordered hydrogen-bonding in cellulose Iβ has its origin in the pattern of intramolecular hydrogen bonds established during cellulose biosynthesis in order to accommodate parallel packing of cellulose chains. Whereas in cellulose II there is only one intramolecular hydrogen bond, the *tg* conformation of the hydroxymethyl group of cellulose Iβ is such that two intramolecular hydrogen bonds can be established: between O5 and O3, and between O6 and O2. It is likely that the formation of these two intramolecular hydrogen bonds through the glycosidic linkage brings an extensive constraint to the structure, resulting in an instability of the hydrogen bond pattern. At a given time, one of the intramolecular hydrogen bonds may be favored as opposed to the other. In this respect, it is interesting to note that, whereas in the origin chain the geometry of the O2—H···O6 hydrogen bond is clearly shorter than that of the O3—H···O5 hydrogen bond, in the center chain the reverse is true.

The unique possibility of the hydrogen bond at O3 as opposed to the multiple possibilities at O2 and O6 may have some importance for the description of the surface of cellulose crystals and microfibrils, where all the reactivity of cellulose takes place. In fact, the reactivity of the three hydroxyl groups located at the surface of cellulose crystals can be probed by so-called chemical microstructural analysis (CMA). In this method, devised by Rowland et al.,[54] surface derivatives resulting from a mild interaction of cellulose with *N*,*N*-diethylarizidinium chloride are analyzed, and the reactivity of O2, O3, and O6 is determined. Remarkably, with well-organized native cellulose crystals, the surface O3 atom is totally unreactive, suggesting that the precise geometry of the intramolecular hydrogen bond linking O3 to O5 persists at the surface of cellulose Iβ

(53) Marechal, Y.; Chanzy, H. *J. Mol. Struct.* **2000**, *523*, 183.

(54) Rowland, S. P.; Roberts, E. J.; Wade C. P. *Text. Res. J.* **1969**, *39*, 530.

crystals.[55,56] The O2 and O6 atoms at the crystal surfaces are substantially reactive, in keeping with the multiple possibilities for the hydrogen bonds, and therefore conformations, at O6 and O2 within the crystal. It remains to be seen whether the hydroxymethyl groups located at the crystals surfaces can adopt a *gt* conformation, as proposed for the interpretation of recent high-resolution AFM images of cellulose microfibrils.[57]

The values of the conventional crystallographic residual reported here for the X-ray refinement (0.1857) and the neutron refinement (0.2095) are high compared to those for single-crystal structures. In particular, the crystal structure of β-D-cellotetraose hemihydrate was refined to a residual of 0.085.[41,42] Whereas the intensity profile of reflections in single-crystal studies is usually dominated by instrument effects and is relatively easy to model, in this study it is dominated by the disordered and textured distribution of microcrystallites within the fiber and is difficult to model exactly. In addition, the disordered distribution of microcrystallites around the fiber axis results in the overlap of the intensity profiles of independent reflections, particularly at high resolution. These two factors increase the uncertainty in measuring the intensity of reflections. The crystallographic structure of microcrystallites within fibers might also be expected to be less well ordered than the structure of single crystals because of surface and interface effects. This structural disorder will affect the intensity of reflections but is difficult to model in the structure refinements. Finally, the structure reported here has been refined with isotropic thermal parameters due to a lack of sufficient data, whereas the single-crystal structure reported for β-D-cellotetraose hemihydrate was refined with anisotropic thermal parameters.

(55) Rowland, S. P.; Howley, P. S. *J. Polym. Sci. A* **1988**,*26*, 1769.
(56) Verlhac, C.; Dedier, J.; Chanzy, H. *J. Polym. Sci. A* **1990**, *28*, 1171.
(57) Baker, A. A.; Helbert, W.; Sugiyama, J.; Miles, M. J. *Biophys. J.* **2000**, *79*, 1139.

Conclusion

This study has provided, for the first time, a set of coordinates for all of the atoms, including hydrogen, in the crystal structure of cellulose Iβ. The results presented here indicate that the structure of cellulose Iβ differs from those of cellulose Iα and previous models for cellulose I in having two unique sheets containing conformationally distinct chains with different hydrogen-bonding. The hydrogen-bonding scheme is more complex and disordered than initially projected, but there is evidence for the presence of cooperative hydrogen-bonding networks. These results have important implications for our understanding of the factors that influence the structure and properties of cellulose.

Acknowledgment. The authors thank the Institut Laue Langevin and the European Synchrotron Radiation Facilities for the provision of beamtime. Y.N. thanks the French Government and the Japanese Society for the Promotion of Science for financial support. P.L. thanks the Office of Science and the Office of Biological and Environmental Research of the U.S. Department of Energy for financial support.

Supporting Information Available: Crystallographic data (CIF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA0257319