A Revised Structure and Hydrogen-Bonding System in Cellulose II from a Neutron Fiber Diffraction Analysis

P. Langan,*,† Y. Nishiyama,‡ and H. Chanzy§

Life Sciences Division, Los Alamos National Laboratory, MS-M888, Los Alamos New Mexico 87545, Graduate School of Agricultural and Life Science, The University of Tokyo, Yayoi, Tokyo 113, 8657 Japan, and Centre de Recherches sur les Macromolécules Végétales, CNRS, affiliated with the Joseph Fourier University of Grenoble, BP 53, 38041 Grenoble Cedex 9, France

Received May 14, 1999

Abstract: The crystal and molecular structure and hydrogen bonding system in cellulose II have been revised using new neutron diffraction data extending to 1.2 Å resolution collected from two highly crystalline fiber samples of mercerized flax. Mercerization was achieved in $NaOH/H_2O$ for one sample and in $NaOD/D_2O$ for the other, corresponding to the labile hydroxymethyl moieties being hydrogenated and deuterated, respectively. Fourier difference maps were calculated in which neutron difference amplitudes were combined with phases calculated from two revised X-ray models of cellulose II, A and B′. The revised phasing models were determined by refinement against the X-ray data set of Kolpak and Blackwell,[8] using the LALS methodology.[37] Both models A and B′ have two antiparallel chains organized in a $P2_1$ space group and unit cell parameters: $a =$ 8.01 Å, $b =$ 9.04 Å, $c =$ 10.36 Å, and $\gamma = 117.1°$.[15] Model A has equivalent backbone conformations for both chains but different conformations for the hydroxymethyl moieties: *gt* for the origin chain and *tg* for the center chain. Model B′, based on the recent crystal structures of cellotetraose,[21−23] has different conformations for the two chains but nearly equivalent conformations for the hydroxymethyl moieties. On the basis of the X-ray data alone, model A and model B′ could not be differentiated. From the neutron Fourier difference maps, possible labile hydrogen atom positions were identified for each model and refined using LALS. We were able to eliminate model A in favor of model B′. The hydrogen-bonding scheme identified for model B′ is significantly different from previous proposals based on the crystal structures of cellotetraose,[21−23] MD simulations of cellulose II,[25] and any potential hydrogen-bonding network in the structure of cellulose II determined in earlier X-ray fiber diffraction studies.[7,8] The exact localization of the labile hydrogen atoms involved in this bonding, together with their donor and acceptor characteristics, is presented and discussed. This study provides, for the first time, the coordinates of all of the atoms in cellulose II.

Introduction

The crystalline nature of cellulose was revealed almost a century ago when Nishikawa and Ono recorded the first X-ray diffraction patterns from fiber bundles from various plants.[1] Following the success of these experiments, X-ray diffraction has become a standard tool for studying cellulose fibers either in their native state or after physical or chemical modifications.[2−4] These studies have not only allowed classification of the various celluloses into a number of crystalline allomorphs but also have triggered a strong interest in determining their crystal and molecular structure. The combination of X-ray diffraction with model building and conformational analyses in the 1970s provided “modern” crystal and molecular structures for most of the cellulose polymorphs and in particular for cellulose-I or native cellulose and cellulose-II or recrystallized cellulose.[5−10]

The development of high-resolution ^{13}C solid-state NMR techniques in the 1980s has brought a new dimension to the determination of the crystal structure of cellulose. Indeed the ^{13}C NMR spectra of highly crystalline cellulose such as that of *Valonia* showed unambiguously the presence of two crystalline allomorphs in cellulose I, namely, cellulose-I_α and $-I_\beta$.[11,12] On the other hand, only the I_β allomorph was found to be present in tunicin, another highly crystalline cellulose sample from animal origin.[13] These observations indicate that the crystal and molecular structure of cellulose I has to be revised in light of this dimorphism. This revision requires new diffraction data of pure I_α and I_β fibers. So far such data have not been obtained, and for this reason, the precise atomic coordinates of cellulose I and its polymorphs have not been obtained.

The X-ray structure of cellulose-II has defined the crystals of this polymorph as consisting of two antiparallel and crystallographically independent chains. The proposed structure has a monoclinic cell where the chains are aligned on the 2-fold screw axes of the cell. Both chains have equivalent backbone conformations but differ in the conformation of their hydroxy-

(1) Nishikawa, S.; Ono, S. *Proc. Tokyo Math. Phys. Soc*. **1913**, *7*, 131.
(2) Hermans, P. H. *Physics and Chemistry of Cellulose Fibres*; Elsevier Publishing Co: New York, 1949.
(3) Roelofsen, P. A. *The Plant Cell-Wall*; Gebrüder Borntraeger: Berlin, Germany, 1959.
(4) Krässig, H. A. *Cellulose, Structure, Accessibility and Reactivity*; Gordon and Breach Publishers: Yverdon, Switzerland, 1993.
(5) Gardner, K. H.; Blackwell, J. *Biopolymers* **1974**, 13, 1975.
(6) Sarko, A.; Muggli, R. *Macromolecules* **1974**, *7*, 486.
(7) Stipanovic, A.; Sarko, A. *Macromolecules* **1976**, *9*, 851.
(8) Kolpak, K. J.; Blackwell, J. *Macromolecules* **1976**, *9*, 273.
(9) Woodcock, C.; Sarko, A. *Macromolecules* **1980**, *13*, 1183.
(10) French, A. D.; Roughehad, W. A.; Miller, D. P. In *The Structure of Cellulose*; Atalla, R. H., Ed.; ACS Symposium Series 340, 1985; p 15.
(11) Atalla, R. H.; VanderHart, D. L. *Science* **1984**, *223*, 283.
(12) VanderHart, D. L.; Atalla, R. H. *Macromolecules* **1984**, *17*, 1465.
(13) Belton, P. S.; Tanner, S. F.; Cartier, N.; Chanzy, H. *Macromolecules* **1989**, *22*, 1615.

10.1021/ja9916254 CCC: $18.00 © 1999 American Chemical Society
Published on Web 10/16/1999

methyl groups.[7,8] These moieties are near the *gt* conformation[14] for the glycosyl residues located at the origin of the cell as opposed to the *tg* conformation for those of the center chain.[15]

The occurrence of two types of conformation for the hydroxymethyl groups of cellulose-II has been challenged by observations resulting from a number of ^{13}C NMR spectroscopy studies. In the cellulose II spectra, the C6 resonance occurs as a singlet near 64 ppm[16-19] and not as the expected doublet with resonance near 64 and 66 ppm if both *gt* and *tg* conformations are coexistent in the crystalline structure.[20]

The model of cellulose-II has been further challenged by the recent determination of the crystalline structure of β-cellotetraose hemihydrate, derived independently by Gessler et al. and Raymond et al.,[21-23] and that of methyl β-cellotrioside monohydrate 0.25 ethanolate.[24] Both cellodextrins adopt a crystalline packing almost equivalent to that of cellulose-II. Their molecular configuration is also similar to that of the cellulose-II model except in two main respects: all hydroxymethyl groups are in the *gt* conformation and the sugar pucker is different for the two chains. On the basis of these crystallographic observations and the aforementioned spectroscopic evidences, it is clear that the structure of cellulose-II should be re-examined.

Another important feature of crystalline cellulose that needs to be re-examined is that of the hydrogen-bond system. There are significant differences in the hydrogen-bond schemes proposed for β-D-cellotetraose by Raymond et al.[22] and Gessler et al.,[23] probably because the former group calculated H-atom positions with the AFIX routine of SHELX93 whereas the latter group used Fourier difference analyses. Although the Fourier difference analyses revealed less than half of the hydroxyl hydrogen atom positions, the observed positions were used as the basis for a complete 3D hydrogen-bonding network proposal. Molecular Dynamics simulations[25] on the crystal structures of cellulose-II confirm the hydrogen-bond network proposed by Gessler et al.[23] The hydrogen-bond networks of both Raymond et al.[22] and Gessler et al.[23] differ from any potential hydrogen-bond network in the structure of cellulose II determined in the earlier X-ray fiber diffraction studies.[7,8]

The power of neutron fiber diffraction for locating hydrogen atoms[26] and investigating hydrogen bonding[27,28] has already been demonstrated. Compared to oxygen and carbon, hydrogen is a weak scatterer of X-rays but not of neutrons. The scattering length of hydrogen for neutrons is negative, but it is positive and of comparable magnitude for deuterium, carbon, and oxygen (-0.37×10^{-12} cm for H, 0.667×10^{-12} cm for D, 0.665×10^{-12} cm for C, and 0.58×10^{-12} cm for O).[29] At the resolution of most fiber diffraction studies where individual atoms cannot be resolved, the scattering length of hydrocarbon and hydroxyl groups is small (0.291×10^{-12} cm for CH, -0.083×10^{-12} cm for CH_2, and 0.206×10^{-12} cm for OH) but of deuterioxyl groups is large (1.245×10^{-12} cm).

An earlier study has shown that, in cellulose-II, a partial replacement of the OH moieties by OD can lead to meaningful neutron fiber diffraction patterns where a substantial contrast can be observed when comparing data from the deuterated and hydrogenated samples.[30] Unfortunately, in this earlier work, the substitution of OH by OD was only partial and the fibers were poorly oriented so that the exact positions of these moieties within the lattice could not be determined.

This present work follows a preliminary report in which we described how an improved mercerization of flax fibers in NaOD/D_2O could lead to cellulose II fibers where the 6 independent H atoms involved in hydrogen bonding are replaced by 6 deuterium atoms, without any loss of crystalline perfection.[31] These deuterated fibers give high-resolution (1.2 Å) neutron fiber diffraction patterns with intensities that are substantially different from the intensities observed in neutron fiber diffraction patterns obtained from fibers prepared by conventional mercerization, as described in our preliminary report. In this work, we describe how measured intensities from these diffraction patterns have been used to compute Fourier difference maps leading to a full description of the hydrogen-bond system of cellulose-II. In the Fourier analyses, neutron diffraction amplitudes are combined with phases calculated from two improved models of cellulose-II determined from X-ray studies. We could not differentiate between these two models on the basis of the X-ray data alone. The subsequent refinement of the hydrogen-bond system against the neutron data has allowed us to eliminate one of these models. In the final model the sugar puckers were allowed to change and the hydroxymethyl moieties were allowed to adopt an all *gt* conformation.

Results

The Amplitudes from Neutron Studies. (A) Sample preparation. Two cellulose-II samples were prepared: one hydrogenous, designated H-cellulose-II, and the other specifically deuterated, designated D-cellulose-II. Purified flax fibers were swollen overnight and aligned by combing. Fiber bundles were clamped into stretchers during mercerization, washing, stretching, and relaxing. Highly crystalline samples were obtained using a method adapted from the work of Manjunath

(14) 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). Thus, an ideal *gt* conformation would be defined as the set of two angles: 60°, 180°.

(15) Throughout this work, we have used the unit cell of cellulose-II defined in ref 8 with $a = 8.01$ Å, $b = 9.04$ Å, $c = 10.36$ Å, and $\gamma = 117.1°$. This is a nonstandard setting of space group $P2_1$ with the c axis as the unique axis. When the z coordinate of O5 is greater than that of C6, the chain is defined as "up", otherwise as "down". We refer to the "up" and "down" chain as the origin and center chain, respectively, and atom labels are correspondingly post-fixed with an "o" or "c".

(16) Dudley, R. L.; Fyfe, C. A.; Stephenson, P. J.; Deslandes, Y.; Hamer, G. K.; Marchessault R. H. *J. Am. Chem. Soc*. **1983**, *105*, 2469.

(17) Fyfe, C. A.; Stephenson, P. J.; Veregin, R. P.; Hamer, G.; Marchessault, R. H. *Carbohydr. Chem*. **1984**, *3*, 663.

(18) Isogai, A.; Usuda, M.; Kato, T.; Uryu, T.; Atalla, R. H. *Macromolecules* **1989**, *22*, 3168.

(19) Horii, F.; Hirai, A.; Kitamaru, R.; Sakurada, I. *Cellul. Chem. Technol*. **1985**, *19*, 513.

(20) Horii, F.; Hirai, A.; Kitamaru, R.; Sakurada, I. *Polym. Bull*. **1983**, *10*, 3168.

(21) Gessler, K.; Krauss, N.; Steiner, T.; Betzel, C.; Sandman, C.; Saenger, W. *Science* **1994**, *266*, 1027.

(22) Raymond, S.; Heyraud, A.; Tran Qui, D.; Kvick, Å.; Chanzy, H. *Macromolecules* **1995**, *28*, 2096.

(23) Gessler, K.; Krauss, N.; Steiner, T.; Betzl, C.; Sarko, A.; Saenger, W. *J. Am. Chem. Soc*. **1995**, *117*, 11397.

(24) Raymond, S.; Henrissat, B.; Tran Qui, D.; Kvick, Å.; Chanzy, H. *Carbohydr. Res*. **1995**, *277*, 209.

(25) Kroon-Batenburg, L. M. J.; Kroon, J. *Glycoconjugate J*. **1997**, *14*, 677.

(26) Mahendrasingam, A.; Al-Hayalee, Y.; Forsyth, V. T.; Langan, P.; Fuller, W.; Oldman, R. J.; Blundell, D. J.; Mason, S. A. *Physica B* **1992**, *180 & 181*, 528.

(27) Shotton, M. W.; Pope, L. H.; Forsyth, V. T.; Langan, P.; Denny, R. C.; Giesen, U.; Dauvergne, M.-T.; Fuller, W. *Biophys. Chem*. **1997**, *69*, 85.

(28) Langan, P.; Forysth, V. T.; Mahendrasingam, A.; Pigram, W. J.; Mason, S. A.; Fuller, W. *J. Biomol. Struct. Dyn*. **1992**, *10*, 489.

(29) Koester, L.; Rauch, H. *Summary of Neutron Scattering Lengths*; IAEA contract 2517/Rb; International Atomic Energy Agency, Vienna, 1981.

(30) Langan, P.; Denny, R. C.; Mahendrasingam, A.; Mason, S. A.; Jaber, A. *J. Appl. Crystallogr*. **1996**, *29*, 383.

(31) Nishiyama, Y.; Okano, T.; Langan, P.; Chanzy, H. *Int. J. Biol. Macromol*. In press.

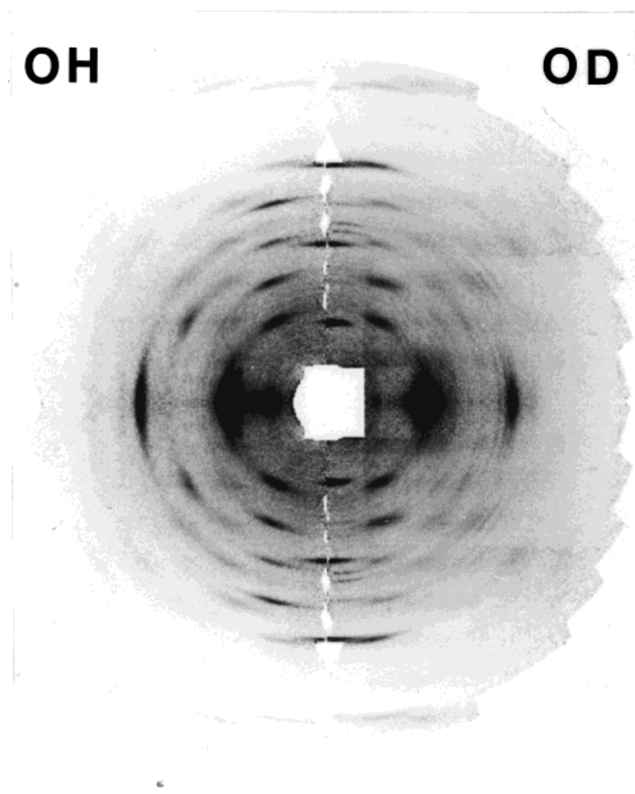

Figure 1. Neutron fiber diffraction patterns collected from two flax samples, one mercerized in NaOH/H_2O (left-hand side) and the other mercerized in NaOD/D_2O (right-hand side). The fiber axis is vertical and the two patterns have been displayed together by joining equivalent halves along the meridian.

and Venkataraman[32] and described in detail elsewhere.[31] The deuterated samples were prepared under conditions similar to those of the hydrogenated samples, except that NaOH was replaced by NaOD and H_2O by D_2O. The final samples occurred as 7 × 1 × 40 mm stiff thick ribbons weighing about 200 mg. IR spectra were recorded to confirm that OH moieties had been completely replaced by OD moieties.

(B) Data Collection. Neutron data were collected from both samples on diffractometer D19 at the Institut Laue Langevin, Grenoble. D19 is a 4-circle single-crystal diffractometer equipped with a large 4° × 64° position sensitive detector.[33] Data collection strategies have been developed for fibers that allow all reciprocal space to be covered out to ~1 Å resolution by stepping the sample in omega, chi, and phi and the detector in gamma at $\lambda = 1.5290$A.[30] The actual neutron diffraction patterns and the generic data collection strategy used have been described in detail in a preliminary report.[31] A complete scan took 36 h for each sample. The data were treated for the effects of attenuation with an effective linear absorption coefficient, μ, dominated by the effective absorption cross-section of hydrogen, σ_h. The values of μ for D-cellulose-II and H-cellulose-II were calculated to be 2.8 and 1.9 cm^{-1}, respectively, using a value of 49.3 barns for σ_h calculated from the linear fit to wavelength determined by Koetzle and McMullan.[34] Figure 1 shows the data after binning into reciprocal space using the D19 suite of programs called MCFIBRE.[35] There are clear differences in the intensities diffracted from D-cellulose-II and H-cellulose-II, particularly in the meridional regions. Diffraction features extend beyond 1.2 Å resolution. A 2D fit of the background and Bragg intensities in reciprocal space was made using the least-squares program LSQINT, part of the BBSRC funded CCP13 project.[36] The refined unit cell parameters agree (to within error) with the values obtained from X-ray studies.[8,15] We therefore assume that the samples are isomorphous.

The Phasing Model from X-ray Studies. Two 2-fold helical models were built and refined against the X-ray data, collected by Kolpak and Blackwell,[8] using the linked-atom-least-squares (LALS) procedure.[37] Hamilton's statistical test[38] was used to determine the significance of changes in agreement with the data, as represented by R'',[39] after introducing or removing parameters. The X-ray data consisted of 44 observed and 41 estimated amplitudes. The estimated amplitudes were included in the refinement only if their value was less than the corresponding calculated amplitude. Unlikely stereochemical contacts were taken as oxygen−oxygen distances of less than 2.6 Å and oxygen−carbon distances of less than 2.8 Å although oxygen−oxygen distances of less than 2.6 Å have been observed for hydrogen bonds involving water donors in biological molecules, they are unfavorable and involve disorder.[40,41] Over-short contacts were relieved by applying a constraint term in the minimization function.

The first model, designated A, was built from the published coordinates of cellulose-II.[8] The second, designated B, was identical to A except that all hydroxymethyl groups were in the *gt* conformation. Hydrogen atoms were explicitly included when covalently bound to carbon but not to oxygen so that hydrogen bonding was not initially incorporated into the refinement. In the first stage, χ' for each chain and the three packing parameters were varied; the relative displacement, z, of the origin and center chains with crystallographic positions given by (0,0,0) and (0.5,0.5,z) respectively, and the orientations of the chains about their helix axes. Along with a scale factor, K, and an isotropic temperature factor, B, this adds up to 7 parameters.

The parameters of both models remained close to their starting values during refinement. After relieving over-short contacts the value of R'' was 0.158 and 0.238 for models A and B, respectively. Model B can be rejected at a confidence level of 99.5% on the basis of the X-ray data alone. The refined parameters for model A and the corresponding value of R'' differ only slightly from those obtained by Kolpak and Blackwell.[8]

To allow the sugar puckers and glycosidic bonds to change conformation, we then included for each chain 2 glycosidic torsion angles, ϕ and ψ,[42] the 6 endocyclic torsion and 6 conformation angles for each glucose ring, 3 Eulerian angles, and the displacement of the helix axis from the symmetry axis,

(32) Manjunath, B. R.; Venkataraman, A. *J. Polym. Sci., Polym. Chem. Ed.* **1980**, *18*, 1407.

(33) Thomas, M.; Stansfield, R. F. D.; Berneron, M.; Filhol, A.; Greenwood, G.; Jacobe, J.; Feltin, D.; Mason, S. A. *Position Sensitive Detection of Thermal Neutrons*; Convert, P., Forsyth, J. B., Eds.; Academic Press: New York, 1983.

(34) Koetzle, T. F.; McMullan, R. K. *Research Memo C4*; Brookhaven National Laboratory, 1980.

(35) Shotton, M.; Langan, P. *ILL Report ILL95SH12T*, 1995.

(36) Denny, R. C.; Langan, P.; Forsyth, V. T.; Mant, G. R.; Squire, J. M. Manuscript in preparation.

(37) Campbell Smith, P. J.; Arnott, S. *Acta Crystallogr., Sect. A* **1978**, *34*, 3.

(38) Hamilton, W. C. *Acta Crystallogr.* **1965**, *18*, 502.

(39) $R'' = [\sum_i w_i(|F_i|_o - |F_i|_c)^2/\sum_i w_i|F_i|o^2]^{1/2}$, where F_o, F_c, and w are the observed and calculated structure factor amplitudes and their weights, respectively.

(40) Bouquiere, J. P.; Finney, J. L.; Savage, H. F. L. *Acta Crystallogr., Sect. B* **1994**, *50*, 566.

(41) Langan, P.; Lehmann, M. S.; Wilkinson, C.; Jogl, G.; Kratky, C. *Acta Crystallogr., Sect. D* **1999**, *55*, 51.

(42) The glycosidic torsion angles ϕ and ψ, which describe the relative orientation of adjacent glycosyl residues in the same chain, are defined as (O5−C1−O1−C4) and (C1−O1−C4−C3), respectively.

Table 1. Conformational Parameters of Models for Cellulose II: A Comparison between the Models Reported in This Study, A and B′ and Those of Kolpak and Blackwell (KB),[8] Stipanovic and Sarko (SS),[7] Raymond et al. (R),[46] Gessler et al. (G),[23] and Kroon-Batenburg and Kroon (KK)[48 a]

	origin chain				center chain				
	ϕ	ψ	χ'	θ	ϕ	ψ	χ'	θ	z
A	−95.4	92.3	−168.7	3.9	−95.1	92.3	−79.2	3.9	0.22
B′	−95.4	92.3	−172.4	3.9	−91.3	89.4	148.4	10.8	0.23
KB	−95.9	93.6	173.8	1.9	−96.4	94.0	−69.7	2.3	0.22
SS	−97.6	95.8	176.2	3.3	−97.6	95.8	−76.6	3.3	0.21
R	−93.2	96.4	−176.9	1.2	−92.0	86.4	171.2	11.6	0.23
G	−95.0	97.0	−176.0	1.5	−94.0	90.0	170.0	12.5	0.24
KK	−97.0	102.0	−169.0	5.9	−97.0	90.0	−179.0	17.6	0.23

[a] Our Final Model Is B′. The origin and center chains of SS have been inverted to follow the convention of KB used here. The parameters are defined as ϕ(O5−C1−O1−C4), ψ(C1−O1−C4−C3), χ'(C4−C5−C6−O6). θ is the puckering parameter defined by Cremer and Pople[43] (averaged for KK), and z is the relative fractional displacement of the origin and center chains in the c axis direction.

S. To maintain chain continuity and sugar ring closure, we included 24 independent constraint terms in the refinement, leaving19 net degrees of freedom. Refinements in which only the sugar ring of one chain was allowed to change conformation had 13 net degrees of freedom.

Model A refined to model A′ with only the sugar ring of the central chain allowed to change conformation; model A″ with only the sugar ring of the origin chain allowed to change conformation; model A‴ with the sugar rings of both chains allowed to change conformation. The corresponding values of R'' for models A′, A″, and A‴ were 0.154, 0.156, and 0.164, respectively, after relieving over-short contacts. The additional degrees of freedom involved in models A′, A″, and A‴ do not significantly improve the agreement with the X-ray data compared to model A, and we retain only model A for future consideration. Under the same conditions model B refined to models B′, B″, and B‴ with values of 0.146, 0.207, and 0.156 for R'', respectively, after relieving over-short contacts. Models B and B″ can be rejected with respect to model B′ at a confidence level of 99.5%. The additional degrees of freedom involved in model B‴ do not significantly improve the agreement with the X-ray data compared to model B′, and we retain only model B′ for future consideration. Although model B′ is in slightly better agreement with the X-ray data than model A, this is not significant even at a confidence level of 75%. We therefore cannot distinguish between models A and B′ on the basis of the X-ray data alone. The final parameters of models A and B′ are given in Table 1.

Hydrogen Bonding Refinement. Labile deuterium atom positions on the OD moieties were identified using Fourier difference synthesis. Two Fourier syntheses were calculated using coefficients $(F_d - F_h)\exp i\alpha_c$, where F_d and F_h are the observed structure factor amplitudes from D-cellulose-II and H-cellulose-II, respectively, and α_c are phases calculated from the X-ray models A and B′. The F_h amplitudes were weighted by the ratio of the spherically averaged amplitudes $\langle F_d\rangle/\langle F_h\rangle$. In both maps, all difference density peaks above a threshold of 1.3 times the root-mean-square density could be associated with possible deuterium atom positions, except for density features around the O6 groups, which we discuss later. The identified deuterium atom positions were then refined against the neutron diffraction data collected from D-cellulose-II using LALS.

A series of refinements was carried out in which the deuterium atom positions were added to the varied parameters in an incremental way. A deuterium atom position was explicitly included in the refinement by adding the corresponding torsion angle to the varied parameter list, while keeping its bond length fixed at 0.98 Å. The C−O−D bond angles were constrained to a global value, and this value was allowed to refine in a restrained manner around the standard value of 110°. The scattering power of deuterium atoms that were not explicitly included in the refinement was added to that of the corresponding deuterioxyl oxygen atom. The positions of the non-deuterium atoms were those corresponding to models A and B′. No constraint terms were used to relieve short contacts. No attractive hydrogen-bonding terms were included in the refinement. When no deuterium atoms were explicitly refined, the refinement had only two parameters, B and K. Including the deuterium atom positions associated with oxygen atoms O3, O2, and O6 increased the number of parameters to 5, 7, and 9, respectively.

The values of R'' for model A refined with 2, 5, 7, and 9 parameters were 0.307, 0.260, 0.236, and 0.222, respectively. The values of R'' for model B′ refined with 2, 5, 7, and 9 parameters were 0.294, 0.236, 0.222, and 0.194. At each step of the refinement for both models A and B′, adding deuterium atom positions has significantly improved the agreement of the model with the neutron data at a confidence level greater than 95%. Model B′ is in significantly better agreement with the neutron data than model A, and we can reject model A at a confidence level of greater than 95%. The coordinates for the final model B′ are given in Table 2. Values for the hydrogen-bonding parameters are given in Table 3. The final $(2F_d - F_c)$-exp(iαc) Fourier map is shown Figure 2. A schematic representation of the hydrogen bond network is given in Figure 3.

Table 2. The Atomic Coordinates (in Å) of the Final Model, B′, for Cellulose-II[a]

Origin Chain				Center Chain			
C1	−0.370	0.032	3.989	C1	1.683	4.310	−1.631
C2	−1.316	0.680	2.983	C2	0.508	4.129	−0.674
C3	−1.176	0.020	1.621	C3	0.834	4.517	0.759
O2	−2.650	0.577	3.451	O2	−0.592	4.900	−1.122
C4	0.285	0.001	1.186	C4	2.168	3.856	1.087
O3	−1.952	0.717	0.658	O3	−0.199	4.081	1.630
C5	1.158	−0.594	2.283	C5	3.259	4.160	0.068
O1	0.451	−0.779	0.000	O1	2.651	4.291	2.359
O5	0.971	0.120	3.511	O5	2.824	3.571	−1.163
C6	2.636	−0.531	1.963	C6	4.604	3.576	0.449
O6	3.420	−1.243	2.916	O6	5.373	3.219	−0.694
H1	−0.730	−0.942	4.148	H1	1.831	5.337	−1.789
H2	−1.052	1.685	2.832	H2	0.248	3.114	−0.606
H3	−1.468	−0.986	1.699	H3	0.968	5.557	0.808
H4	0.623	0.984	1.042	H4	2.060	2.812	1.061
H5	0.861	−1.584	2.472	H5	3.332	5.198	−0.071
H6b	2.775	−0.835	0.968	H6b	5.080	4.240	1.109
H6a	2.971	0.461	2.042	H6a	4.462	2.670	0.959
D6	3.408	−2.224	2.913	D6	6.288	3.561	−0.785
D3	−1.546	0.968	−0.198	D3	−0.037	4.059	2.596
D2	−3.292	0.028	2.956	D2	−0.894	5.665	−0.588

[a] The unit cell parameters are a = 8.01 Å, b = 9.04 Å, c = 10.36 Å, and γ = 117.1°.[15] The temperature factor, B, is 6.5 for the neutron data and 24.1 for the X-ray data. The C−O-D angle refined to a value of 119.4°.

Discussion

The Revised Crystal Structure of Cellulose II. In the crystal structure for cellulose-II presented here, model B′, the D-glucopyranoses of each chain are in the $^4C_1$ chair conformation with endocylic bond angles that do not deviate significantly from standard values. The calculated Cremer and Pople[43] puckering parameters indicate that, although the sugar of the origin chain

(43) Cremer, D.; Pople, J. A. *J. Am. Chem. Soc.* **1975**, *97*, 1354.

Table 3. The Hydrogen Bonding Parameters for Cellulose-II Identified in This Study[a]

D−H	*d* (D−H)	*d* (H..A)	<DHA	d(D..A)	*A* [symmetry generator]
O6o−H	0.981	2.015	119.75	2.643	O6c [-x+1, -y, z+1/2]
O6o−H	0.981	2.223	141.49	3.053	O3c [x, y-1, z]
O6o−H	0.981	2.489	122.08	3.123	O5c [-x+1, -y, z+1/2]
O3o−H	0.980	1.918	130.45	2.660	O5o [-x, -y, z-1/2]
O3o−H	0.980	2.803	113.07	3.312	O6o [-x, -y, z-1/2]
O2o−H	0.979	1.817	150.67	2.713	O6o [x-1, y, z]
O6c−H	0.981	1.784	150.44	2.682	O2c [x+1, y, z]
O3c−H	0.980	1.848	148.43	2.731	O5c [-x+1, -y+1, z+1/2]
O3c−H	0.980	2.500	130.02	3.219	O6c [-x+1, -y+1, z+1/2]
O2c−H	0.981	2.212	115.84	2.783	O2o [-x, -y+1, z-1/2]

[a] Acceptor and donor atoms are denoted as A and D, respectively. Although deuterium atom positions were identified and refined in this study, we denote the positions by H. Distances are in Angstroms and angles in degrees.

is conformationally unstrained, the sugar of the center chain is conformationally strained.[44] The relative orientation of adjacent glycosyl residues, described by the glycosidic torsion angles ϕ and ψ, is also different for the center and origin chains. The hydroxymethyl groups of both chains are near the *gt* conformation. The center and origin chains have a relative displacement along the *c* axis direction of 2.38 Å.

A comparison of selected parameters of the structure presented here with those of other models proposed for cellulose II is given in Table 1. Both Gessler et al.[23] and Raymond et al.[46] have proposed models for cellulose-II, designated G and R, respectively, which are based on the conformational features of the single crystal β-D-cellotetraose structures. Models G and R display the same alternation between center and origin chains of glycosidic linkage and sugar pucker as our model, although the exact values differ slightly. Models B′, G, and R were refined against the same set of X-ray data, and the differences are probably a reflection of the different refinement programs used, the different β-D-cellotetraose structures used to define the sugar conformations, and the different number of parameters allowed to refine in each case. Model R was refined with the sugar puckers constrained to the average conformations observed for the origin and center chains in β-D-cellotetraose, using the same refinement program, LALS, as used in this work. The parameters of R are in closest agreement with ours. Model G was refined using the program PS79 [47] with all structural parameters of both chains allowed to vary. The model published recently by Kroon-Batenburg and Kroon from MD simulations of cellulose II,[25,48] designated KK, is also in good agreement with our model showing the same alternation of glycosidic linkage and sugar pucker conformation between chains. Although B′, R, G, and KK have the hydroxymethyl groups of both chains near the *gt* conformation, the exact orientation of the hydroxymethyl group of the center chain is somewhat different in B′, having a value of χ' that is about 20° smaller than in R, G, and KK.

Model B′ differs significantly from the original fiber diffraction models of Kolpak and Blackwell[8] and Stipanovic and Sarko,[7] designated KB and SS, respectively. In particular the origin and center chains have similar sugar puckers and glycosidic linkages in KB and SS. In KB and SS the hydroxymethyl group is near the *gt* conformation for the origin chain and the *tg* conformation for the center chain. In KB and SS the

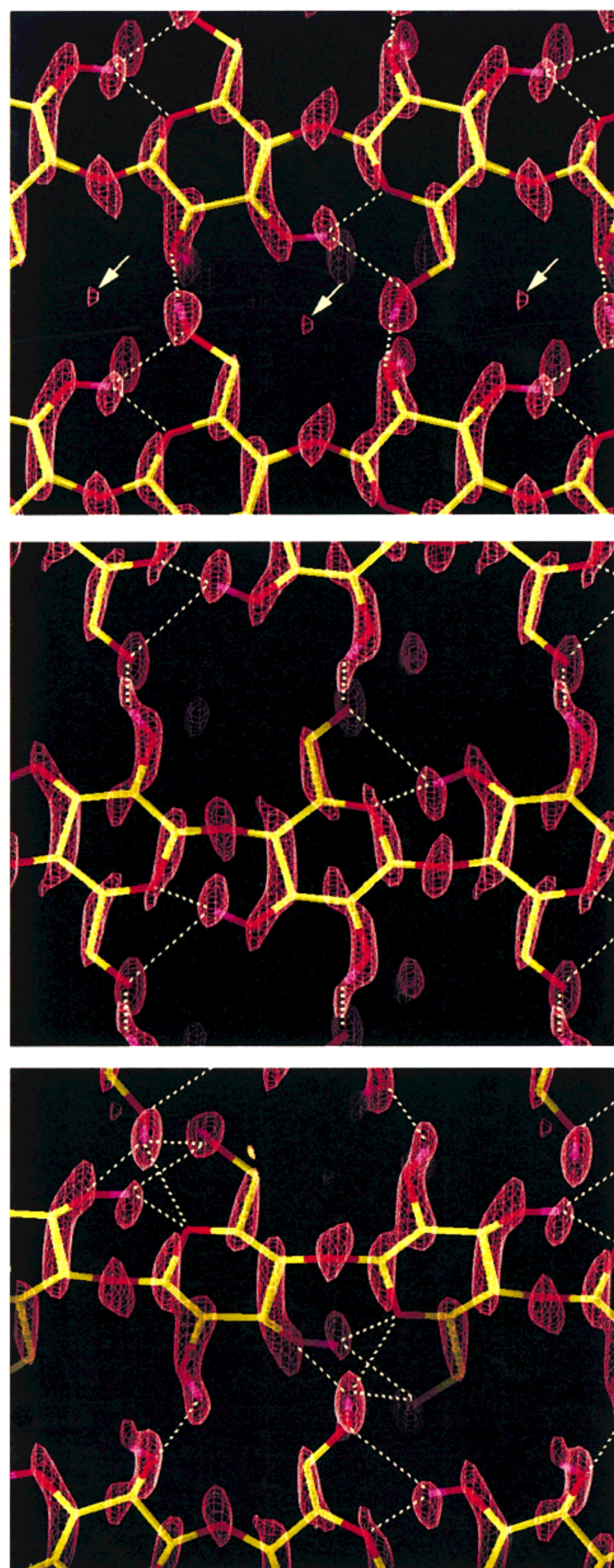

Figure 2. The final $2F_d - F_c$ map (red density) for model B′, showing views of the planes containing (a) the center chains (b) the origin chains and (c) origin and center chains. Cellulose chains are represented by a skeletal model in which carbon atoms are yellow, oxygen atoms are red, and labile hydrogen atoms are pink. Hydrogen atoms covalently bonded to carbon are not depicted. The arrows in (a) indicate density peaks that could not be accounted for by labile hydrogen atom positions. The potential hydrogen bonds are represented by broken lines.

(44) $\theta = 2.7°$ is the ideal value for an unstrained α-D-glucopyranose.[23]

(45) Dowd, M. K.; French, A. D.; Reilly, P. J. *Carbohydr. Res.* **1994**, *264*, 1.

(46) Raymond, S.; Kvick, Å.; Chanzy, H. *Macromolecules* **1995**, *28*, 8422.

(47) Zugenmaier, P.; Sarko, A. *Biopolymers* **1976**, *15*, 2121.

(48) The coordinates of the model resulting from the MD simulation of cellulose II carried out by Kroon-Batenburg and Kroon[25] were kindly supplied to us as a personal communication.

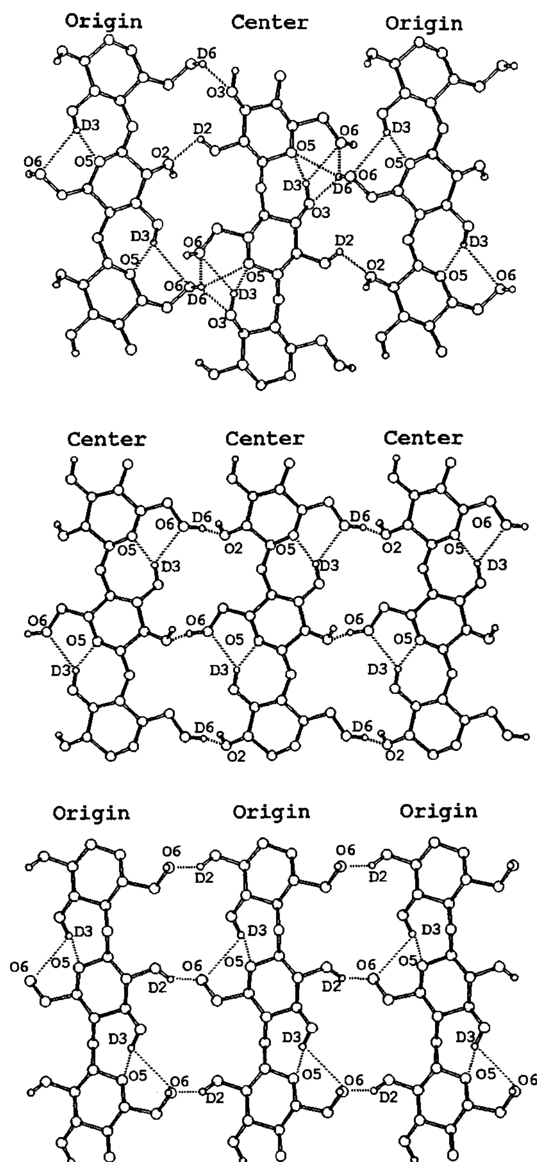

Figure 3. A schematic representation of the hydrogen bonds in cellulose II. Only atoms involved in hydrogen bonding are labeled. Hydrogen bonds are represented by dotted lines. Intermolecular hydrogen bonds are O2-D···O6 in sheets containing only origin molecules and O6-D···O2 in sheets containing only center molecules. In the sheet containing both center and origin molecules there are O6-D···O6 and O2-D···O2 intermolecular hydrogen bonds. The former has minor components involving O5 and O3 as acceptors. Intramolecular hydrogen bonds are O3-D···O5 in each molecule with a minor component involving O6 as acceptor.

relative displacement of the chains in the c axis direction is ~2.2 Å, whereas for B′ the value is ~2.4 Å.

As pointed out by Raymond et al.,[46] this increase in chain displacement from 2.2 to 2.4 Å means that the origin and center chains are staggered by almost half the length of a glycosyl residue. In model B′, because the hydroxymethyl moieties of both origin and center chains are near the *gt* conformation, there is an intermolecular hydrogen bond between O6 atoms. Reducing the relative displacement of the chains to 2.2 Å would introduce a steric clash between these O6 atoms. In models SS and KB, because the hydroxymethyl moieties of the origin and center chains have different conformations, there is no such interaction and the relative displacement of the chains is not restricted in this way. Thus the interaction between hydroxymethyl moieties in our structure can be related to the conformational strain observed on the sugar pucker of the central chain because of the different van der Waals, steric, and hydrogen-bond requirements at a chain displacement of 2.4 Å compared to 2.2 Å.

Despite the large conformational differences between models KK and SS on one hand and model B′ on the other, they are in similar agreement with the X-ray data. The analysis presented here has allowed us, for the first time, to differentiate between the *gt/tg* models represented by SS and KB in preference for the all *gt/gt* models represented by B′, G, R, and KK, on the basis of fiber diffraction data alone.

In both the $F_d - F_h$ and $2F_d - F_c$ Fourier maps, there is a difference density peak that cannot be assigned to a deuterium atom. This peak, which is identified by an arrow in Figure 2, is in a position that would be occupied by a hydroxymethyl group near the *tg* position. Allowing the occupancy of the hydroxymethyl group of the central chain to be shared between the *tg* and *gt* positions in the refinement of model B′ against the X-ray data involved two additional parameters and reduced the value of R'' to 0.133. This improvement is significant at a confidence level of greater than 97.5% and indicates that the *gt* position has an occupancy of ~70% and the *tg* position, with $\chi = 23.1°$, has an occupancy of ~30%. Refining a statistical mixture of models A and B′ against the X-ray data reduced R'' to 0.137 and indicated occupancies of ~40% and ~60%, respectively. This does not represent a significant improvement in the agreement with the X-ray data because of the large number of extra parameters involved. It is not clear to what extent an incomplete mercerization process could be responsible for 15% of all of the hydroxymethyl groups having a *tg* conformation. It is clear from the neutron data that there are no cellulose I crystallites present in the sample. It is interesting to note that, in the MD simulations of cellulose II by Kroon-Batenburg and Kroon,[25] the hydroxymethyl groups are not exclusively in the *gt* conformation. Some 5% of *tg* conformation occurs. The displacement parameters of the hydroxymethyl group of the central chain are also significantly larger than the thermal parameter of the other atoms in the MD simulations.

Hydrogen Bonding. An analysis of the hydrogen bonds in model B′ was carried out using the HTAB instruction in SHELX-97.[49] To facilitate the following discussion, we refer to the deuterium atoms in model B′ as H. The criteria used for identifying hydrogen bonds were that the A−H distance should be less than 2.8 Å and that the D−H−A angle should be greater than 110°, where D and A designate donor and acceptor atoms, respectively. A systematic intramolecular three-center hydrogen bond[50] is observed in both chains involving O3 as donor and O5 and O6 as acceptors. Intermolecular hydrogen bonds are observed between origin chains involving O2o as donor and O6o as acceptor and between center chains involving O6c as donor and O2c as acceptor. Intermolecular hydrogen bonds are observed between origin and center chains involving O2c as donor and O2o as acceptor and O6o as donor in a four-center hydrogen bond,[50] with O6c, O5c, and O3c as possible acceptors.

The intramolecular three-center hydrogen bond has a major component between O3 and O5 (O3o−O5o is 2.66 Å and O3c−O5c is 2.73 Å) and a minor component between O3 and O6 (O3o−O6o is 3.31 Å and O3c−O6c is 3.22 Å). These values are in good agreement with the average values reported by Gessler et al.[23] for β-D-cellotetraose (O3o−O5o is 2.84 Å, O3c−O5c is 2.91 Å, O3o−O6o is 3.32 Å, and O3c−O6c is 3.09 Å).

(49) Sheldrick, G. M. *SHELX-97* **1997**, a program for the Refinement of Single-Crystal Diffraction Data, University of Gottingen, Germany.

(50) Jeffrey, G. A.; Saenger, W. *Hydrogen Bonding in Biological Structures*; Springer-Verlag: Berlin, Germany, 1991.

In the MD simulation of cellulose-II, KK,[25,48] only the center chain, is reported to have a three-center hydrogen bond.

The intermolecular hydrogen bonding differs substantially from that observed in β-D-cellotetraose[23] and in the MD simulations.[25] Whereas in this structure O2o and O6c act as donors in hydrogen bonds to similar chains, in β-D-cellotetraose they act as donors in hydrogen bonds between origin and center chains. Similarly, in this structure O6o and O6c act as donors in hydrogen bonds between origin and center chains, whereas in β-D-cellotetraose they act as donors in hydrogen bonds between similar chains. Interestingly the D−A distances reported here and by Gessler et al.[23] for β-D-cellotetraose are very similar. O6o−O2o and O6c−O2c are 2.71 and 2.68 Å in this study and 2.72 and 2.65 Å for β-D-cellotetraose. O6c−O6o and O2c−O2o are 2.64 and 2.78 Å in this study and 2.75 and 2.76 Å for β-D-cellotetraose. The O6−H−O6 and O2−H−O2 bond angles are smaller in this study than for β-D-cellotetraose.

One consequence of the different intermolecular hydrogen-bonding arrangement reported here is that O6o can donate a hydrogen bond to three possible acceptors, the major component being between O6o and O6c (O6o−O6c is 2.64 Å) and the two minor components being between O6o and O3c (O6o−O3c is 3.05 Å) and O6o and O5c (06o−O5c is 3.12 Å). These three acceptors already interact with each other through a three-center hydrogen bond. The intricate hydrogen-bonding network involving O6c, O3c, O5c, and O6c is further extended by hydrogen bonds that O6o and O6c make to other atoms. It is not clear to what extent the observed disorder of the O6c group influences this hydrogen-bonding arrangement.

Model B′ was refined against the neutron data again, this time with the O6 and O2 deuterium atom starting positions corresponding to the hydrogen-bonding scheme observed for β-D-cellotetraose. The deuterium atom torsion angles had to be restrained in order to maintain this hydrogen-bonding scheme, the tendency being for them to adopt values corresponding to the hydrogen-bonding scheme reported in this study. The resulting value of R'' was 0.215, and this solution could be rejected at a 93% confidence level. We have used a rigid hydrogen-bond donor to hydrogen atom bond length and an overall hydrogen-bond angle parameter to reduce the number of parameters in our refinements. We have also used an overall thermal parameter rather than individual thermal parameters for the same reason. The exact values of the parameters determined in this work do not represent accurate values for acceptor and donor bond distances and angles. However the neutron data clearly support the hydrogen-bonding scheme reported in this work.

Conclusion

This study has provided, for the first time, a reliable set of coordinates for all of the atoms, including hydrogen, in the crystal structure of cellulose-II. The results presented here confirm that, in crystalline fibers of cellulose-II, a 3D network of hydrogen bonds exists. The hydrogen-bond network is substantially different from previous proposals based on X-ray fiber diffraction studies of cellulose-II and X-ray single-crystal diffraction studies of β-D-cellotetraose. The conformation of the cellulose chains is similar to those found in the crystal structure of β-D-cellotetraose and differs significantly from the original model for cellulose-II.

The structure of cellulose should be considered as a complete system of cellulose chains that interact through forces that are to a large extent due to hydrogen bonds. The identification of these bonds is necessary for an understanding of the structure, reactivity, and properties of cellulose in its various forms and the processes involved in conversion from one form to another. This study is the first in a series of fiber diffraction studies of the cellulose allomorphs, which exploit the power of neutron diffraction for locating hydrogen.

Acknowledgment. We thank the Institut Laue Langevin for the provision of neutron beamtime. Y.N. thanks the French Government and the Japanese Society for the Promotion of Science for financial support.

JA9916254