

The Crystal Structure and Hydrogen Bonding System in Cellulose from Neutron Fibre Diffraction Data

P. Langan^{1*}, Y. Nishiyama² and H. Chanzy³

* Author for correspondence.

¹ B Division, Los Alamos National Laboratory, MS-M888, Los Alamos NM 87545 USA. (<http://lsdiv.lanl.gov/nsb>)

² Graduate School of Agricultural and Life Science, The University of Tokyo, Yayoi, Tokyo 113, 8657 Japan.

³ Centre de Recherches sur les Macromolécules Végétales, CNRS, affiliated with the Joseph Fourier University of Grenoble, BP 53, 38041 Grenoble Cedex 9, France.

Cellulose is a linear poly (1-4) β -D glucan which is normally biosynthesised as slender rod-like crystalline microfibrils acting as structural elements in plant cell walls and various other living organisms. The chemical and chain structure of cellulose was established by the 1930s using classical organic chemistry and developing polymer studies. A key feature of this structure is the presence of three hydroxyl groups on each glycosyl monomer; two secondary and one primary alcohol. The potential cohesive interchain hydrogen bonding of these hydroxyl groups provided an explanation for some surprising chemical and physical properties specific to cellulosic materials. It was also realised that one of the secondary alcohols, and possibly the primary alcohol, could form intrachain hydrogen bonds between adjacent glycosyl monomers, producing a planar chain conformation with 2_1 symmetry [1].

The crystalline nature of cellulose was revealed nearly 90 years ago when Nishikawa and Ono recorded the first X-ray patterns from fibre bundles from various plants [2]. X-ray diffraction has become a standard tool for studying cellulose fibres, allowing classification of the various celluloses into a number of crystalline allomorphs [3,4]. Cellulose is crystallised into a metastable form during biosynthesis. Native cellulose, or cellulose I, can be made to undergo an irreversible transition to a stable form, cellulose II, by two distinct processes: regeneration and mercerisation. Regeneration involves either preparing a solution of cellulose in an appropriate solvent or of an intermediate derivative followed by coagulation and recrystallisation. Mercerisation involves intra-crystalline swelling of cellulose in concentrated aqueous NaOH followed by washing and recrystallisation. A number of other cellulose polymorphs can be obtained by physical or chemical modification, in particular cellulose III and

cellulose IV [5,6]. These processes have been commonly used in industrial preparations in order to improve mechanical properties and reception to dye.

One of the first detailed molecular models for cellulose I was proposed in 1937 and consisted of antiparallel chains packed into a monoclinic cell [7]. However subsequent studies on cellulose from a variety of sources using different diffraction techniques have produced a number of different unit cells and ways of packing cellulose chains into these cells. In particular, electron diffraction from *Valonia* cellulose taken by Honjo and Watanabe contained spots that could not be indexed by the conventional two-chain unit cell [8]. They proposed an eight-chain unit cell to explain the diffraction pattern. 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. 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 cellulose I β [9,10]. On the other hand, the I β allomorph was found to be the predominant form in tunicin, another highly crystalline cellulose sample from animal origin [11]. Cellulose from *Glaucocystis* has been shown to consist of essentially cellulose I α [12, 13].

It is now recognised that cellulose has characteristics that allow the formation of unique cellulosic structures within specific species and often within different tissues of the same organism [14]. The species specific compositional ratio of cellulose I α and I β and also the relative distribution of cellulose I α and I β domains in biosynthesised microfibrils is thought to affect structure, properties and function. Some fungal cellulases have enhanced activity on substrates of specific cellulose I α /I β composition. These developments indicate that the crystal and molecular structures of cellulose I have to be revised in light of this dimorphism. Electron diffraction studies have recently shown the I α and I β forms to correspond to one-chain triclinic and two-chain monoclinic unit cells, respectively [15]. A combination of biochemical techniques and electron microscopy have allowed the polarity of the chains packed in these cells to be determined [16]. A priority now is to obtain new X-ray and neutron diffraction data from pure I α and I β fibres, in order to determine precise atomic coordinates for these allomorphs.

There has also been controversy over the unit cell of cellulose II. The X-ray structure, determined from diffraction studies on regenerated fibres, has defined the crystals of this polymorph as consisting of two antiparallel and crystallographically independent chains [17, 18]. The case of mercerised cellulose is less clear. Some authors believe that in the cellulose II crystals obtained by mercerisation the chains are antiparallel [19]. This is contradicted by other authors who propose a parallel-chain system, because of the topographical problems involved in converting from a parallel-chain cellulose I to an antiparallel-chain cellulose II [20, 21]. The antiparallel X-ray structure has a monoclinic unit cell where the chains are aligned on the two-fold screw axis of the cell. Both chains have equivalent backbone and sugar conformations but differ in the conformation of their primary alcohol (commonly called hydroxymethyl groups); *gt* for the chain located at the cell origin and *tg* for the centre chain [22]. This model has been challenged by observations resulting from a number of ^{13}C NMR studies. In the cellulose II spectra, the C6 resonance occurs as a singlet near 64ppm and not as the expected doublet with resonances near 64 and 66ppm if both *gt* [23-27] and *tg* conformations are coexistent in the crystalline structure [17-19].

The model of cellulose II has been further challenged by the recent determination of the crystalline structures of two cellulose oligomers, β -cellotetraose [28, 29] and methyl β -cellotrioside [30] that are known to crystallise in the same type of lattice as cellulose II. Their molecular configurations are similar to that of the cellulose II model except in two main respects; all hydroxymethyl groups are in the *gt* conformation and the sugar and backbone conformations are slightly different for the two chains. On the basis of these observations it is clear that the structure of cellulose II should also be re-examined.

Another important feature of all the crystalline cellulose allomorphs that needs to be re-examined is that of the hydrogen bonding system. There are significant differences in the hydrogen bonding schemes proposed for β -D cellotetraose by Gessler *et al.* [28] and Raymond *et al.* [29]. Recent MD simulations [31] would appear to support the scheme proposed by Gessler *et al.* Both these hydrogen bonding schemes differ from any potential hydrogen bonding network in the structure of cellulose II determined in the earlier fibre diffraction studies

[17,18]. A resolution of the hydrogen bonding scheme in cellulose II and the determination of hydrogen bonding schemes in the other cellulose allomorphs 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.

The power of neutron fibre diffraction for locating hydrogen atoms [32] and investigating hydrogen bonding [33, 34] 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 ($-0.37\times 10^{-12}\text{cm}$ for H, $0.667\times 10^{-12}\text{cm}$ for D, $0.665\times 10^{-12}\text{cm}$ for C and $0.58\times 10^{-12}\text{cm}$ for O) [35]. At the resolution of most fibre diffraction studies, where individual atoms cannot be resolved, the scattering length of hydrocarbon and hydroxyl groups is small ($0.291\times 10^{-12}\text{cm}$ for CH, $-0.083\times 10^{-12}\text{cm}$ for CH_2 and $0.206\times 10^{-12}\text{cm}$ for OH), but of deuteroyl groups is large ($1.245\times 10^{-12}\text{cm}$). It has already been shown that, in cellulose-II, a partial replacement of the OH moieties by OD can lead to meaningful neutron fibre diffraction patterns where a substantial contrast can be observed when comparing data from the deuterated and hydrogenated forms [36]. Unfortunately, in this earlier work, the substitution of OH by OD was only partial and the fibres poorly oriented so that the exact positions of these moieties within the lattice could not be determined.

We are involved in a long term study using neutron and synchrotron X-ray techniques to determine the precise structures of the various cellulose polymorphs. Here we report on the first results from our neutron diffraction studies. We have developed methods for replacing H atoms involved in hydrogen bonding in cellulose fibres with D, without any loss in crystalline perfection [37,38]. The deuterated fibres give high resolution neutron diffraction patterns with intensities that are substantially different from the intensities observed in neutron fibre diffraction patterns obtained from hydrogenated fibres. So far we have collected neutron diffraction data from the deuterated and hydrogenated forms of cellulose I β , cellulose I α /I β and cellulose II. Measured intensities from these diffraction patterns are being used to compute Fourier difference maps

leading to full descriptions of the hydrogen bonding systems. In this report we present the results of our first full analysis: the hydrogen bonding system in cellulose II.

Neutron Fibre Diffraction Patterns

Neutron diffraction data were collected on diffractometer D19 at the Institut Laue Langevin, Grenoble, using generic data collection strategies that have been described elsewhere [38]. Cellulose I β samples were prepared from the cellulosic mantles of tunicates (*Halocynthia roretzi*) and cellulose I β +I α samples from the walls of green algae (*Cladophora sp.*). The proportion of I α and I β in *Cladophora* is approximately 3:1. Cellulose II samples were prepared from flax fibres. Preparing highly crystalline, hydrogenated and deuterated, samples large enough for neutron diffraction required a number of innovative steps that have been described elsewhere [37, 39]. Neutron diffraction patterns are shown in Figures 1(a)-(f). Figures 1(a), (c) and (e) correspond to diffraction from hydrogenated samples and Figures 1(b), (d) and (f) to diffraction from deuterated samples.

Figures 1(a) and 1(b) correspond to cellulose I β (*Halocynthia*). Diffraction features from both the hydrogenated and deuterated samples extend well beyond atomic resolution (0.9Å). Several hundred diffraction spots can be measured, far exceeding the number of data available from reported X-ray fibre diffraction studies of cellulose I. A comparison of hydrogenated and deuterated patterns (Figures 1(a) and 1(b)) reveals substantial differences in the

relative distribution of intensity, in particular on the first, third, fourth, seventh and eighth layerlines. Figures 1(c) and 1(d) correspond to cellulose I α +I β (*Cladophora*). The resolution of the data in these diffraction patterns is similar to that from I β . The patterns from *Cladophora* and *Halocynthia* show a great deal of similarity. Differences can nevertheless be seen in the third and fifth layer lines in the patterns from deuterated samples (Figures 1(b) and 1(d)). Figures 1(e) and 1(f) correspond to cellulose II. The difference between the patterns are striking particularly along the fibre axis where the deuterated pattern presents a strong 002 meridional which is totally absent from the hydrogenated pattern. There are also large differences on the first, second and fifth layer lines. Diffraction features extend to a resolution of $\sim 1.2\text{\AA}$.

The neutron diffraction patterns we have recorded extend to a much higher resolution than any X-ray diffraction patterns from cellulose published so far. In fact, the resolution of the patterns from *Halocynthia* and *Cladophora* matches closely that of electron diffraction diagrams obtained on 1 micron of carefully selected specimens [40]. With electron diffraction, however, it is not yet possible to establish with certainty the relation between the intensity of the diffraction spots and the structure factors. The intensities measured from these neutron diffraction patterns are being used to determine the crystal structure and hydrogen bonding in cellulose. We have just completed a full analysis of the cellulose II data [41], the results of which are represented below.

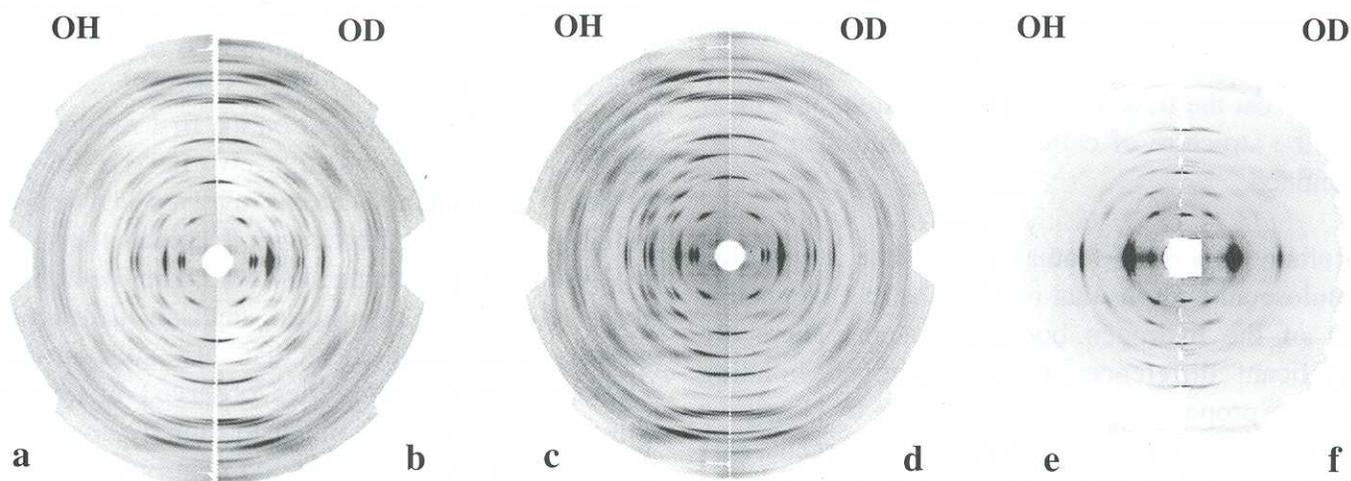


Figure 1: Series of neutron fibre diffraction patterns for cellulose I and cellulose II with vertical fibre axes and printed at the same camera length; (a) from a reconstituted sample of tunicin (cellulose I β) microcrystals; (b) as in (a), but after substituting all OHs by ODs; (c) from a reconstituted sample of *Cladophora* cellulose (cellulose I α +I β); (d) as in (c) but after substitution of OHs by ODs; (e) from mercerized flax in standard NaOH/H₂O; (f) as in (e), but mercerized in NaOD/D₂O.

A Revised Structure and Hydrogen Bonding System for Cellulose II [41]

Measured intensities from neutron diffraction data collected from cellulose II have been combined with phases calculated from an X-ray model in order to compute Fourier difference maps leading to a full description of the hydrogen bonding system. In fact, since there are two competing models in the literature we decided to re-refine the structure of cellulose II against the X-ray data. Both models agree on several points, namely that the structure of cellulose II is based on a two chain unit cell where the chains are antiparallel and that the chains are located on the 2_1 axes of the monoclinic cell. In model A the chains have different conformations for their hydroxymethyl groups [17, 18]. In model B, derived from the crystal and molecular structure of cellulose oligomers, the hydroxymethyl groups are in the same conformation for both chains [28, 29]. The carbon and oxygen atom positions of both models were refined against the X-ray fibre diffraction data allowing the sugar and backbone geometries of the independent chains to change. We could not differentiate between the resulting two models on the basis of their agreement with the X-ray data alone. The H/D atom positions of both models identified in the neutron Fourier difference maps were then refined against our neutron diffraction data. Model B was in significantly better agreement with the data than model A and we were able to reject model A. A final 2Fd-Fc map is shown in Figure 2. Model B has the same basic conformational features as molecules in crystals of β -D cellotetraose. In particular the two chains have different backbone and sugar conformations. The sugar of the central chain is strained and the chains are displaced relative to each other by $\sim 0.24c$. In both the Fd-Fh and 2Fd-Fh Fourier maps there is a difference density feature that cannot be assigned to a deuterium atom. This peak, identified by an arrow in Figure 2(a), is in a position that would be occupied by a hydroxymethyl group near the *tg* position. Allowing the hydroxymethyl group of the central chain to be shared between *tg* and *gt* positions significantly improved the agreement with data and indicated an occupancy of 30% and 70% respectively. It is interesting to note that in MD simulations the hydroxymethyl groups are not exclusively in the *gt* conformation [31]. The hydrogen bonding system is shown schematically in Figure 3. A systematic three-centre intrachain hydrogen bond [42] is observed in both chains. This bond has a major component between O3 and O5,

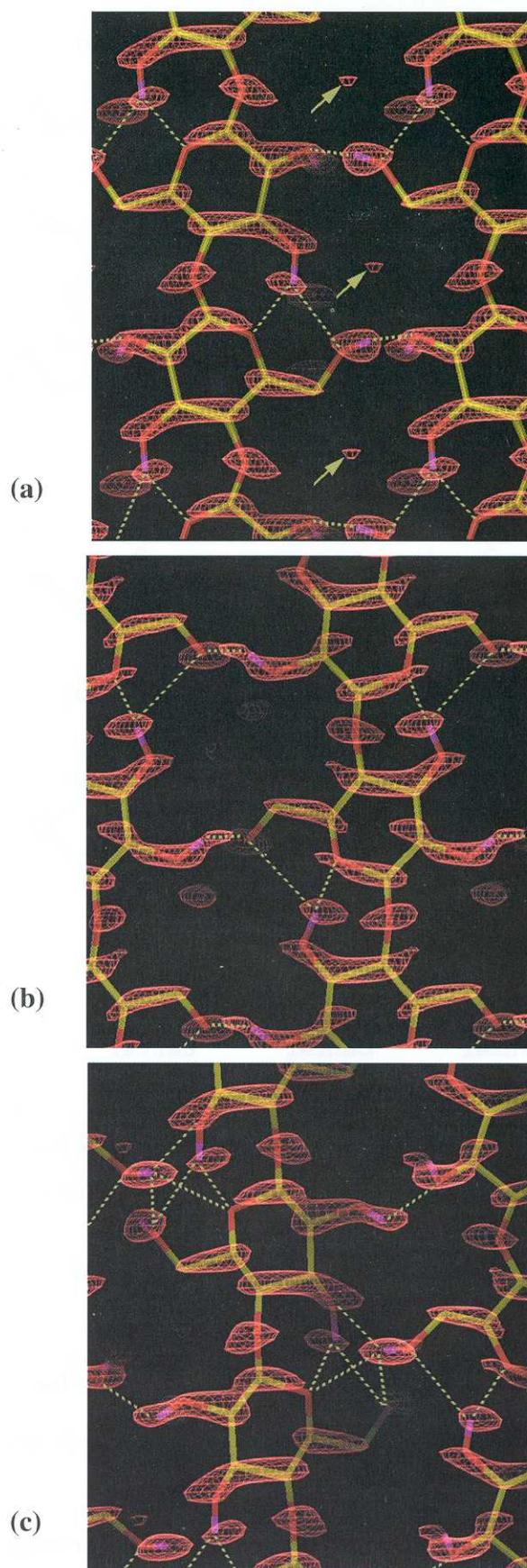


Figure 2: The final 2Fd-Fc map (red density) for model B', showing views of the planes containing a) the centre chains b) the origin chains and c) origin and centre chains. Cellulose chains are represented by a skeletal model in which carbon atoms are yellow, oxygen atoms are red and labile hydrogen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

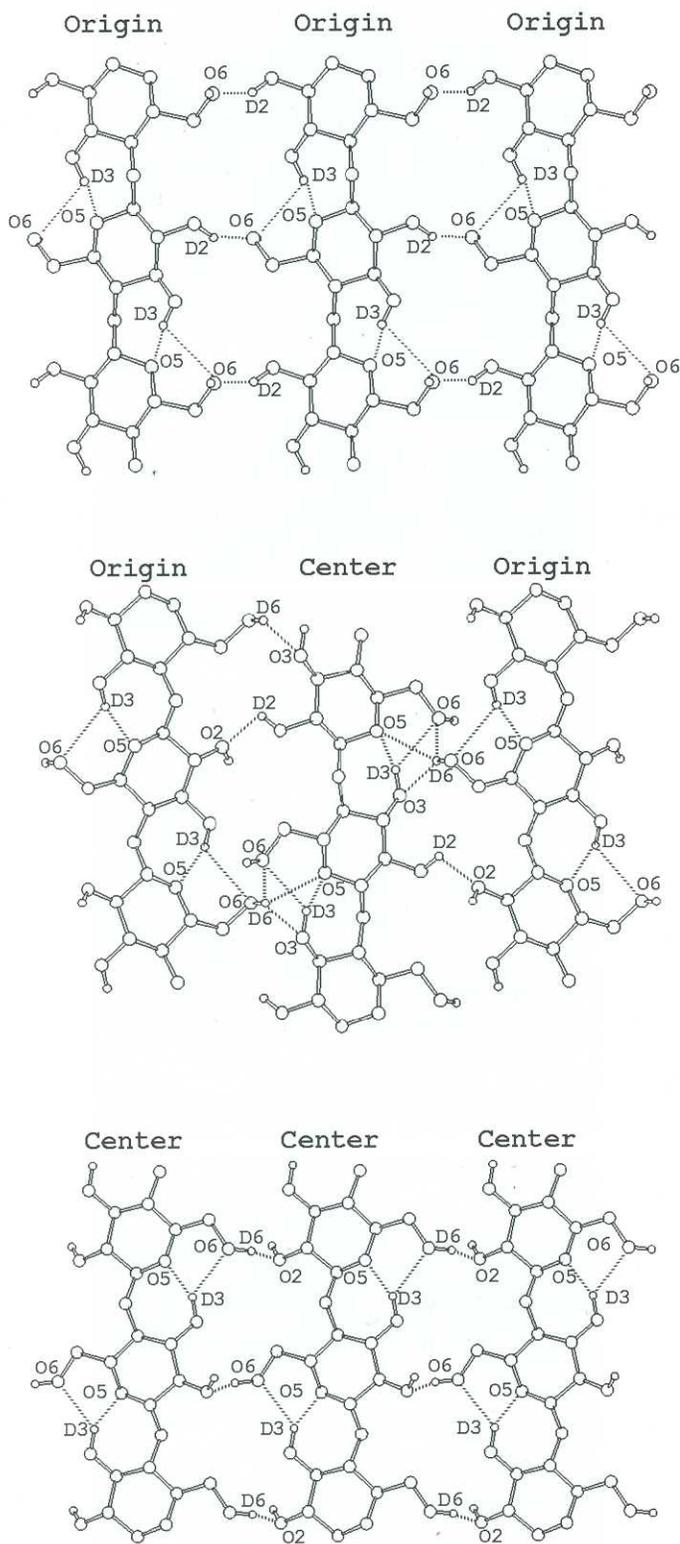


Figure 3: A schematic representation of the hydrogen bonds in cellulose II. Only atoms involved in hydrogen bonding are labeled. Hydrogen bonds are represented by dotted lines. Intermolecular hydrogen bonds are O2-D...O6 in sheets containing only origin molecules and O6-D...O2 in sheets containing only centre molecules. In the sheet containing both centre 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.

with O3 as donor. A similar three-centre hydrogen bond interaction is observed in the β -D cellotetraose structures [28, 29]. The intermolecular hydrogen bonding differs substantially from that observed in β -D-cellotetraose. One consequence of this difference is that O6 of the origin chain can donate a hydrogen bond to three possible acceptors, the major component being to O6 of the centre chain. These three acceptors already interact with each other through a three-centre hydrogen bond. It is unclear to what extent disorder of the O6 group of the centre chain is responsible for this intricate hydrogen bonding arrangement.

Conclusion

Our study on cellulose II 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. A similar analysis of the neutron diffraction data collected from cellulose I β (*Halocynthia*) and cellulose I α +I β (*Cladophora*) is under way. We are also in the process of collecting neutron diffraction data from the newly discovered cellulose from *Glaucocystis* which has been shown to be essentially cellulose I α [12, 13] and also cellulose III. It has been said that in the history of polymer science, cellulose has most often been a trailblazer, advancing many analytical methods such as crystallography and microscopy [43, 44]. These neutron diffraction studies are providing the first three dimensional descriptions of hydrogen bonding systems in fibrous polysaccharides. They are also driving the development of new instrumentation such as the Neutron Diffraction Structural Biology Station being built at Los Alamos, New Mexico [45].

Acknowledgements

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.

References

- [1] Hermans, P.H., *Physics and Chemistry of Cellulose Fibers* (Elsevier, New York, 1949).
- [2] Nishikawa, S. and Ono, S., *Proc. Tokyo. Math. Phys. Soc.* (1913) **7**, 131.
- [3] Roelofsen, P.A., *The Plant Cell-Wall* (Gebruder Borntraeger, Berlin, 1959).
- [4] Davis, W.E., Barry, A.J., Peterson, F.C. and

- King, A.J., *J. Am. Chem. Soc.* (1943) **65**, 1294.
- [5] Krassig, H.A., *Cellulose, Structure, Accessibility and Reactivity* (Gordon and Breach Pub., Yverdon, Switzerland, 1993).
- [6] Hess, K. and Kiessig, H., *Z. Phys. Chem. Abt. B* (1941) **49**, 235.
- [7] Meyer, K.H. and Misch, L., *Helv. Chim. Acta* (1937) **20**, 232.
- [8] Honjo, G. and Watanabe M., *Nature* (1958) **181**, 326.
- [9] Atalla, R.H. and VanderHart, D.L., *Science* (1984) **223**, 283.
- [10] VanderHart, D.L. and Atalla, R.H., *Macromolecules* (1984) **17**, 1465.
- [11] Belton, P.S., Tanner, S.F., Cartier, N. and Chanzy, H., *Macromolecules* (1989) **22**, 1615.
- [12] Sugiyama, J., Persson, J. and Chanzy, H., *Macromolecules* (1991) **24**, 2461.
- [13] Imai, T., Sugiyama, J., Itoh, T. and Horii, F., *J. Struct. Biol.* (in press).
- [14] Atalla, R.H., in *Comprehensive Natural Products Chemistry* (Ed. B.H. Piuto), Volume 3 (Elsevier, Cambridge 1999).
- [15] Sugiyama J, Vuong R, and Chanzy H., *Macromolecules* (1991) **24**, 4168.
- [16] Koyama, M., Helbert, W, Imai T., Sugiyama J. and Henrissat, B., *Proc. Natl. Acad. Sci.* (1997) **94**, 9091.
- [17] Kolpak, K.J. and Blackwell, J., *Macromolecules* (1976) **9**, 273.
- [18] Stipanovic, A. and Sarko, A., *Macromolecules* (1976) **9**, 851.
- [19] Kolpak, K.J., Weih, M. and Blackwell, J., *Polymer* (1978) **19**, 123-131.
- [20] Fengel, D., *Das Papier* (1993) **12**, 695.
- [21] Kroon-Batenburg, L.M.J., Bouma B. and Kroon, J., *Macromolecules* (1996) **29**, 5695.
- [22] The conformation of the hydroxymethyl group is defined by two letters, the first referring to the torsion angle c (O5-C5-C6-O6) and the second to the torsion angle c' (C4-C5-C6-O6). Thus, an ideal *gt* conformation would be defined as the set of two angles : 60° , 180° .
- [23] Dudley, R.L., Fyfe, C.A., Stephenson, P.J., Deslandes, Y., Hamer, G.K. and Marchessault, R. H., *J. Am. Chem. Soc.* (1983) **105**, 2469.
- [24] Fyfe, C.A., Stephenson, P.J., Veregin, R.P., Hamer, G. and Marchessault, R.H., *Carbohydr. Chem.* (1984) **3**, 663.
- [25] Isogai, A., Usuda, M., Kato, T., Uryu, T. and Atalla, R.H., *Macromolecules* (1989) **22**, 3168.
- [26] Horii, F., Hirai, A., Kitamaru, R. and Sakurada, I., *Cellulose Chem. Technol.* (1985), **19**, 513.
- [27] Horii, F., Hirai, A., Kitamaru, R. and Sakurada, I., *Polym. Bull.* (1983) **10**, 3168.
- [28] Gessler, K., Krauss, N., Steiner, T., Betzel, C., Sandman, C. and Saenger, W., *Science* (1994) **266**, 1027.
- [29] Raymond, S., Heyraud, A., Tran Qui, D., Kvik, Å. and Chanzy, H., *Macromolecules* (1995) **28**, 2096.
- [30] Raymond, S., Henrissat, B., Tran Qui, D., Kvik, Å. and Chanzy, H., *Carbohydr. Res.* (1995) **277**, 209.
- [31] Kroon-Batenburg, L.M.J., Kroon, J., *Glycoconjugate J.* (1997) **14**, 677.
- [32] Shotton, M.W., Pope, L.H., Forsyth, V.T., Langan, P., Denny, R.C., Giesen, U., Dauvergne, M-T. and Fuller, W., *Biophys. Chem.* (1997) **69**, 85.
- [33] Mahendrasingam, A., Al-Hayalee, Y., Forsyth, V.T., Langan, P., Fuller, W., Oldman, R.J., Blundell, D.J. and Mason, S. A., *Physica B* (1992) **180&181**, 528.
- [34] Langan, P., Forsyth, V.T., Mahendrasingam, A., Pigram, W.J., Mason, S.A. and Fuller, W., *J. Biomol. Struct. Dyn.* (1992) **10**, 489.
- [35] Koester, L. and Rauch, H., *Summary of Neutron Scattering Lengths*, IAEA contract 2517/Rb (International Atomic Energy Agency, Vienna, 1981).
- [36] Langan, P., Denny, R.C., Mahendrasingam, A., Mason, S.A., Jaber, A., *J. Appl. Cryst.* (1996) **29**, 383.
- [37] Nishiyama, Y., Isogai, A., Okano, T., Müller, M. and Chanzy, H., *Macromolecules* (1999) **32**, 2078.
- [38] Nishiyama, Y., Okano, T., Langan, P. and Chanzy, H., *Int. J. Biol. Macromol.* (1999) **26**, 279.
- [39] Nishiyama, Y., Kuga, S., Wada, M. and Okano, T., *Macromolecules* (1997) **30**, 6395.
- [40] Favier, V., Chanzy, H. and Cavailé, J-Y., *Macromolecules* (1995) **28**, 6365.
- [41] Langan, P., Nishiyama, Y. and Chanzy, H., *J. Am. Chem. Soc.* (1999) **121**.
- [42] Jeffrey, G.A. and Saenger, W., *Hydrogen Bonding in Biological Structures* (Springer Verlag, Berlin, 1991).
- [43] Marchessault, R.H. and Sundararajan, P.R., in *The Polysaccharides* (Ed. G. Aspinall), Volume 2 of Molecular Biology (Academic Press, New York, 1983).
- [44] O'Sullivan, A., *Cellulose* (1997) **4**, 173.
- [45] See <http://lsdiv.lanl.gov/nsb> and Schoenborn, B.P. and Langan, P., *Biophys. J.* (1999) **76**, A392.