*Glycoconjugate Journal* (1997) **14:** 677–690

# EGC1
# The crystal and molecular structures of cellulose I and II

L.M.J. Kroon-Batenburg* and J. Kroon

*Department of Crystal and Structural Chemistry, Bijvoet Center for Biomolecular Research, Utrecht University Padualaan 8, 3584 CH Utrecht, The Netherlands*

**The paper describes molecular dynamics (MD) simulations on the crystal structures of the I$\beta$ and II phases of cellulose. Structural proposals for each of these were made in the 1970s on the basis of X-ray diffraction data. However, due to the limited resolution of these data some controversies remained and details on hydrogen bonding could not be directly obtained. In contrast to structure factor amplitudes in X-ray diffraction, energies, as obtained from MD simulations, are very sensitive to the positions of the hydroxyl hydrogen atoms. Therefore the latter technique is very suitable for obtaining such structural details. MD simulations of the I$\beta$ phase clearly shows preference for one of the two possible models in which the chains are packed in a parallel orientation. Only the parallel-down mode (in the definition of Gardner and Blackwell (1974) *J Biopolym* 13: 1975–2001) presents a stable structure. The hydrogen bonding consists of two intramolecular hydrogen bonds parallel to the glycosidic linkage for both chains, and two intralayer hydrogen bonds. The layers are packed hydrophobically. All hydroxymethyl group are positioned in the *tg* conformation. For the cellulose II form it was found that, in contrast to what seemed to emerge from the X-ray fibre diffraction data, both independent chains had the *gt* conformation. This idea already existed because of elastic moduli calculations and $^{13}$C-solid state NMR data. Recently, the structure of cellotetraose was determined. There appear to be a striking similarity between the structure obtained from the MD simulations and this cellotetraose structure in terms of packing of the two independent molecules, the hydrogen bonding network and the conformations of the hydroxymethyl group, which were also *gt* for both molecules. The structure forms a 3D hydrogen bonded network, and the contribution from electrostatics to the packing is more pronounced than in case of the I$\beta$ structure. In contrast to what is expected, in view of the irreversible transition of the cellulose I to II form, the energies of the I$\beta$ form is found to be lower than that of II by 1 kcal mol$^{-1}$ per cellobiose.**

***Keywords*: molecular dynamics, crystal structure, cellulose I and II**

## Introduction

Cellulose, a polymer of $\beta(1 \rightarrow 4)$ linked glucose residues, is nature's most abundant macromolecule. Plant and algal cell walls largely consist of cellulose and some bacteria produce this polymer. Cellulose has a strong tendency to form highly crystalline fibres. This makes it accessible to X-ray diffraction. It appears that at least four forms of cellulose exist. The major forms are native or cellulose I and mercerized/regenerated or cellulose II. The latter form is obtained from native material by either swelling in a sodium hydroxide solution or by derivatization, dissolution and recovery of the cellulose. Invariably the cellulose II form is then obtained. This suggests that cellulose II is the more stable modification.

Although at the end of the 1930s a rough description of both crystal structures could already be given, details of the molecular structures and hydrogen bonding were not known. In the 1970s renewed X-ray crystallographic investigations were carried out [1–6], but a final settlement of some controversies could not be given, as is evident from the still increasing number of structural studies on cellulose.

The main reason for this lack of detailed information is that X-ray diffraction on fibres does not go beyond a resolution of about 2.5 Å, which is not enough to determine the position of hydrogen atoms and not even to discriminate between a parallel or antiparallel packing. Some information on hydrogen bonding by IR and Raman spectroscopy and on the structural equivalency of C-atoms by $^{13}$C NMR spectroscopy has been obtained. In particular the latter technique has revealed that cellulose I consists of two polymorphs: I$\alpha$ and I$\beta$ [7]. The I$\alpha$ form has a triclinic space group and only one repeating cellobiose in the unit cell, thus

* To whom correspondence should be addressed.
E-mail: Loes Kroon-Batenburg/bate@chem.ruu.nl

0282–0080 

necessarily possessing a parallel structure. The ease with which the I$\alpha$ form converts to the I$\beta$ form (by hydrothermal heating) leads one to believe that the I$\beta$ form (which has two independent chains) must also be parallel. Additional evidence for the parallel cellulose I structure has already been obtained through enzymatic degradation followed by electron microscopy [8]. All X-ray studies on cellulose II indicate an antiparallel structure. Since, during the mercerization process, cellulose I also converts to cellulose II rather easily, just by swelling and therefore apparently without loss of molecular orientation, it seems a mystery how this comes about. Further diffraction experiments and molecular modelling studies may allow final conclusions.

In the present paper a study is described on the reliability of the proposed cellulose I and II structures carried out by molecular modelling studies. Molecular mechanics (MM) calculations will be shown to give evidence on hydrogen bonding and the hydroxymethyl group conformations, and molecular dynamics (MD) simulations to give details on hydrogen bonding and on the similarity of the two independent glucose units in the I$\beta$ and II structures. Recently two other attempts were made to get detailed information on the I$\alpha$ and I$\beta$ structures through MD simulations at constant volume (NVT or NVE ensembles) [9, 10]. In this paper the results of MD simulations on two models of cellulose I$\beta$ and one model of cellulose II, both at constant pressure (NPT) and at constant volume (NVT), will be discussed and compared with the earlier papers. For cellulose II the results are compared with recent structure determinations of cellotetraose, which could serve as a model for cellulose II. In a different paper [11] a more rigorous approach will be described in which a systematic search for possible packing modes is carried out.

## Computational procedures

MD simulations were carried out using the program package GROMOS [12]. The GROMOS force field uses so-called united atoms for CH and CH2 groups. The parameters that are used for carbohydrates are described by Koehler *et al.* [13]. All calculations were carried out using periodic boundary conditions. The computational box consists of $3 \times 3 \times 2$ unit cells. This gave a system of 18 cellotetraose molecules and a total of 72 glucose residues. The glucose residues are bonded across the periodic box. This initially gave problems in constant pressure calculations. In our group therefore van Eijck has adapted the GROMOS program so that the pressure is now calculated using an atom based virial, and the bonded interactions are included in the pressure calculations [14]. The program is, however, restricted to a monoclinic system. The variations of the monoclinic $\gamma$ angle was also made possible by a special adaptation [15]. Symmetry and crystallographic translations are not fixed within the computational box. The degree of conservation of the two-fold screw axis and the crystallographic translations in the average structure is an indication of the compatibility of the simulated model with experimental diffraction data.

A cut-off radius for long-range interactions of 9 Å was used throughout. A dielectric constant of 1 was used for all electrostatic calculations. The non-bonded pairs list was based on charge groups which a net charge of zero. All structural models were first subjected to energy minimization before the actual MD simulation was started. Initial velocities were taken from a Maxwellian distribution at 300 K. The time step was 0.002 ps. All bond lengths were kept fixed. The temperature was kept at 300 K by loose coupling to a temperature bath with a relaxation constant of 0.1 ps. Anisotropic pressure scaling was performed at 1 atm and a pressure relaxation constant of 2 ps and a compressibility of $1.0 \times 10^{-5}$ bar$^{-1}$ in the c-axis direction and $4.9 \times 10^{-5}$ bar$^{-1}$ for the other directions. Both constant pressure (NPT) and constant volume (NVT) simulations were performed. The first 10 ps of each simulation were used as equilibration. After that, each 0.2 ps coordinates, energies and box dimensions were stored.

The initial structure and cell dimensions of cellulose I and II were based on those published by Gardner and Blackwell [1] and Kolpak and Blackwell [5] for cellulose I and II models respectively. Since these data lack the positions of the hydroxyl hydrogen atoms, these were generated, arbitrarily, always in a transposition (HO2-O2-C2-Cl, HO3-O3-C3-C2 and HO6-O6-C6-C5 = 180 °). The kinetic energy at 300 K is expected to be enough to allow easy transitions of these hydroxyl group dihedrals to more favourable positions. This is also to some extent the case for hydroxymethyl groups. That, nevertheless, problems arise with the choice of the starting the conformations will be shown.

## Proposed structures for cellulose I$\beta$ and II

The most detailed information on the structures of cellulose I and II came from X-ray diffraction. In the 1970s several structural proposals were published. For cellulose I the structure was determined on Valonia and ramie fibres [1–4]. Some differences between these proposals remain. All studies point to a hydrophobic stacking of two independent cellulose chains in a monoclinic unit cell with only intralayer hydrogen bonds. Only French [4] suggested that the two independent chains could be packed in an anti-parallel mode. At present there appears to be consensus on the parallel packing mode in the literature. Enzymatic degradation of Valonia microcrystals through an exo-enzyme, observed with electron microscopy, has shown that they are only digested at the non-reducing end [8]. This proves that the cellulose chains all have the same orientation within one microcrystal. Also studies of the biosynthesis shows that all cellulose chains are extruded in parallel and immediately crystalize, which suggests that cellulose I must be parallel [16]. Despite the agreement there still is a slight difference in

the proposed structures. Gardner and Blackwell found the parallel-up packing to be preferred, whereas the Sarko group preferred the parallel-down packing. This difference only arises because the monoclinic angle $\gamma$ deviates from 90°, leading to somewhat different packing in these two cases. This is one of the issues that will be resolved in this paper. All these structural proposals refer to the I$\beta$ phase. They all agree on the *tg* conformation of the hydroxymethyl group. This group will be discussed in more detail. In this paper, we assume that the packing is parallel. Solid state 13C-CP/MAS NMR spectroscopy has served to prove that cellulose I consists of two different allomorphs: I$\alpha$ and I$\beta$. Atalla and van der Hart [7] showed that the various cellulose samples they investigated (ramie, cotton, bacterial and algal celluloses) were all composed of the two allomorphs, and that only the percentage of each differed. The I$\alpha$ structure is triclinic with only one independent cellulose chain, thus necessarily having a parallel packing. There is a straightforward relation between the I$\alpha$ and I$\beta$ structures. In I$\beta$ the centre chain is translated by approximately 1/4c with respect to the origin chain. The next layer of origin chains is translated by $-1/4$c with respect to the centre layer, thus leading to the monoclinic translation period. In the I$\alpha$ structure all layers are translated by 1/4c with respect to the previous ones. The monoclinic cell can be transformed approximately to the triclinic cell by (a, b, c) = (1/2a − 1/2b − 1/4c, 1/2a + 1/2b − 1/4c, c). The actual unit cell of the triclinic form has been determined by Sugiyama *et al.* [17]. In this paper we will only describe the I$\beta$ structure and use the Gardner and Blackwell unit cell: a = 8.17, b = 7.86, c = 10.38 Å, $\gamma = 97.0°$.

For cellulose II two X-ray diffraction studies were made in the 1970s on rayon fibres. Both Kolpak and Blackwell [5] and Stipanovic and Sarko [6] proposed an antiparallel orientation of the two independent cellulose chains, packed in a three-dimensional hydrogen bonded network. Also, both found the hydroxymethyl group conformation of the origin chains to be *gt* and that of the centre chains to be *tg*. In contrast Horii *et al.* [18] found, with 13C-CP/MAS NMR spectroscopy, that only the *gt* conformation could be found in cellulose II samples. From MM calculations on the elastic chain modulus of cellulose I and II models it could be shown that information on the hydroxymethyl group conformation can be obtained (see later). In this paper we will focus on this issue in detail. The unit cell we used is that of Kolpak and Blackwell: a = 8.01, b = 9.04, c = 10.36 Å, $\gamma = 117.1°$.

## Elastic moduli

An important difference between cellulose I and II fibres was found to exist in their elastic moduli. The chain modulus, i.e. the resistence to stretching the crystalline regions in the fibre direction, can be determined, for instance, by X-ray diffraction. Cellulose I (ramie) has a modulus (130–137 GN m$^{-2}$ which is about 1.5 times larger than that of cellulose II (71–90 GN m$^{-2}$). This modulus, determined at very small elongations, is purely a molecular property, because all forces acting between the molecules are to a first approximation perpendicular to the stretch direction. Our hypothesis that the difference in moduli is related to the number of intramolecular hydrogen bonds running parallel to the cellulose chains, could be confirmed [19]. A cellobiose model, with twice the *tg* hydroxymethyl group conformation and two intramolecular hydrogen bonds along the glycosidic linkage, was used for cellulose I, and a model with twice the *gt* conformation and only one such hydrogen bond was used for cellulose II. For both models the energy increase for different elongations was calculated with MM calculations. From these the force constants and elastic moduli could be calculated. Almost quantitative agreement was found between measured and calculated elastic moduli (Table 1). It confirms the effect of one additional hydrogen bond on the modulus and at the same time the all-*tg* conformation of the hydroxymethyl groups in cellulose I and the all-*gt* conformation in cellulose II.

**Table 1.** Elastic moduli for cellulose models from MM.

| | *Hydrogen bonds* | $e_c$ *calc* | $e_c$ *obs (GN m$^{-2}$)* |
|---|---|---|---|
| Cellulose I | O3-H..O5 *intra*<br>O2-H..O6 *intra* | 136 ± 6 | 130–137 |
| Cellulose II | O3-H..O5 *intra* | 89 ± 4 | 71–90 |

## MD simulations of crystal structures

### Stability of the starting structures

Initially three MD simulations were carried out. First a parallel-up model of cellulose I, which is the P1 model (Figure 1) of Gardner and Blackwell [1] and secondly their parallel-down P2 model (Figure 2), which is the same as the parallel-up model of Sarko and Muggli [2]. In both starting

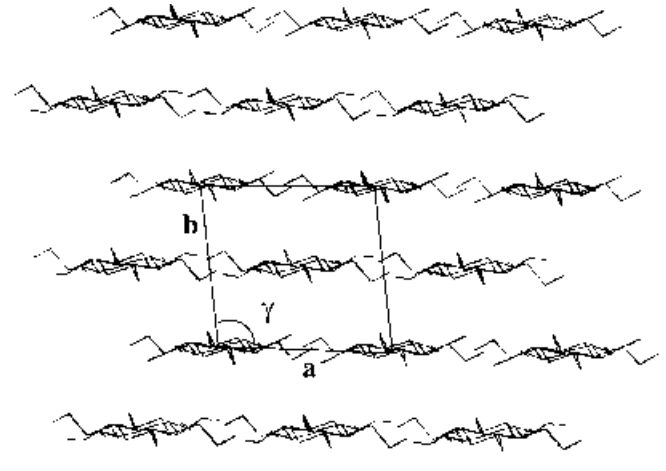


**Figure 1.** ab projection of the starting P1 model for cellulose I$\beta$.

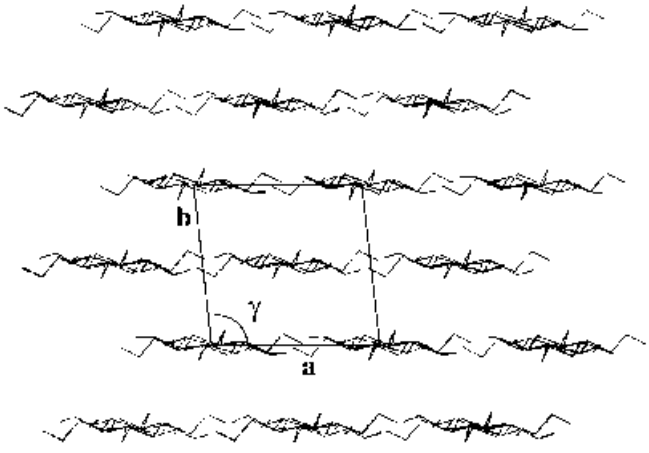


**Figure 2.** ab projection of the starting P2 model for cellulose I$\beta$.

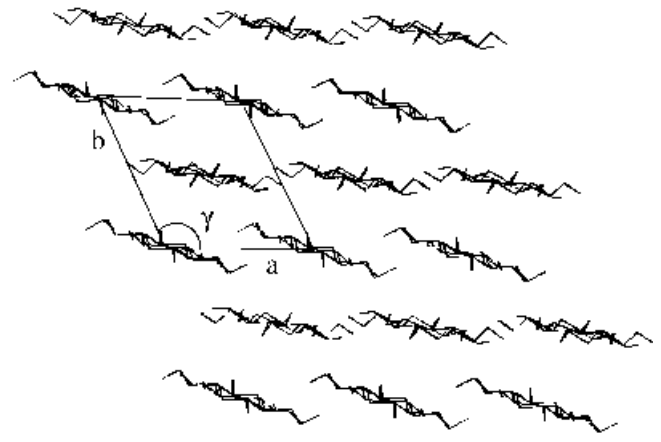


**Figure 3.** ab projection of the starting A2 model for cellulose II.

structures the hydroxymethyl group is *tg*. For cellulose II the antiparallel A2 model (Figure 3) of Kolpak and Blackwell [5] was used with the hydroxymethyl group dihedrals of the origin chains in the *gt* conformation and that of the center chains in the *tg* conformation.

From the constant pressure simulation of the IP1 structure it is clear that it does not present a stable structure. The monoclinic angle $\gamma$ (Table 2) changes from 97 to 84°, and with that is associated a shift of the layers with respect to each other. This can be seen in a display of the final structure (Figure 4) after 110 ps. In fact, this change takes place in the very first few picoseconds. Closer inspection of this structure learns that it just the IP2 structure. For instance a view from the backside (-c) shows that it has converged to a structure in the same monoclinic cell with the molecules placed in the opposite direction i.e. the P2 structure (Figure 5). Energies for the P1 structure are not listed since it does not present a stable structure. MD runs starting in the P2 structure show that it is stable. The cell dimensions are reasonably close to the experimental cell; only the b-axis shrunk by more than 0.2 Å. The energies (Table 2) of the

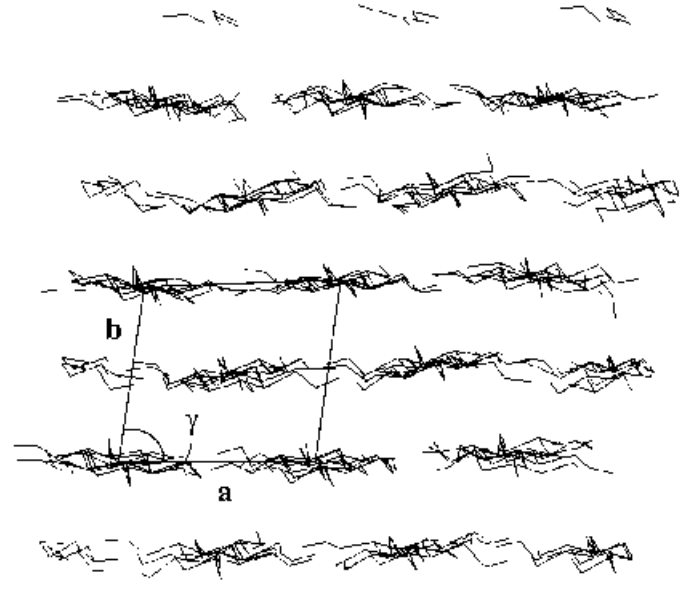


**Figure 4.** ab projection of the structure of cellulose I$\beta$ after 100 ps, that was started from the P1 model.

**Table 2.** Average potential energies and cell dimensions in MD.

| *Cellulose Iβ* | *Start* | *a* *8.17* | *b* *7.86* | *c* *10.38* (Å) | *γ* *97.0* (°) | *Final* | $E_p$ | $E_v$ (kJ mol$^{-1}$ cellobiose) | $E_{v\,-ideal}$ |
|---|---|---|---|---|---|---|---|---|---|
| P1 | tg/tg | 8.1 | 7.6 | 10.4 | 84 | tg/tg | — | | |
| P2 | tg/tg | 8.1 | 7.6 | 10.4 | 96 | tg/tg | 301.3 | 308.3 | 306.1 |
| *Cellulose II* | | *8.01* | *9.04* | *10.36* (Å) | *117.1* (°) | | $E_p$ | $E_v$ | $E_{v\,-ideal}$ |
| A2 | g/tg | 8.1 | 9.0 | 10.3 | 120 | gt/gt | — | | |
| | gt/gt | 8.1 | 8.9 | 10.3 | 119 | gt/gt | 303.1 | 310.3 | 310.0 |

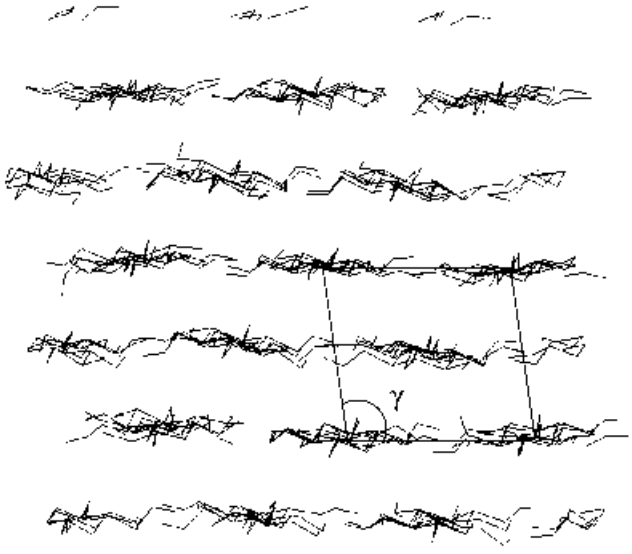


**Figure 5.** Opposite projection of Figure 4 (down -**c**).

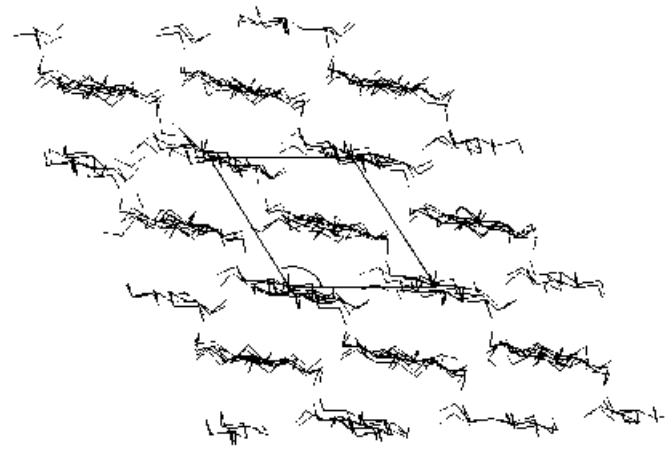

**Figure 6.** P2 structure of cellulose I$\beta$ after 210 ps.

constant pressure simulation Ep and of the constant volume simulation Ev per cellobiose unit are listed. The *tg* conformation is also well conserved, although occasionally hydroxymethyl group transitions occurred, especially in the centre layer. This will be discussed below.

The simulation of cellulose II in the IIA2 structure gave a more or less stable structure with reasonable cell dimensions compared to the experimental ones. However the hydroxymethyl group conformations of the centre chains gradually converted to *gt* instead of *tg*. Statistics of the conformational behaviour of this group will be discussed below. Clearly the proposed *gt*/*tg* combination of origin/centre layers is not preferred on an energetic basis. We repeated this calculation while starting with all hydroxymethyl groups in the *gt* conformation, where they largely stayed during the simulation. A lower energy (Table 2) was obtained which is listed. The structure after 210 ps of constant pressure simulation is depicted in Figure 6.

Before starting these calculations we thought that the hydroxyl group and also the hydroxymethyl group starting conformations were unimportant, and would convert to the preferred conformations. This turned out to be not completely true. In case of IIA2 *gt*/*tg* it appeared from the simulation that the hydroxymethyl group *gt* conformation was preferred, but to arrive at a final stable and completely ordered structure would take a very long simulation time. We typically simulated only a few hundred picoseconds. A higher temperature (simulated annealing) helps, but still it is not easy to get a completely ordered structure. Therefore we choose to restart at the more favourable position as indicated by the initial simulation. In case of IP2 the average structure turned out to be influenced by the initial conformations of the hydroxyl groups. The hydrogen bonding in the centre layer was not ordered, and it was found that those molecules that happened to have an ordered structure had a lower energy. We therefore repeated the calculation with the hydroxyl dihedrals placed in ideal positions (in order to maximize hydrogen bonding). This gave a lower energy (Table 2) for the total system (Ev-ideal). This ideal structure has of course thermal vibration and short flips to other hydroxyl group orientations occur. What has been removed by this procedure is the unacceptably long existence of high energy orientations, because the system does not reach the equilibrium structure. A similar procedure was also carried out for IIA2, but there the final average structure was similar and thus independent of the starting hydroxyl group orientations. As we will discuss below the cellulose I structure has a 2D hydrogen bonding network, whereas cellulose II has a 3D network. Transitions of hydroxyl group conformations are easier in the latter case. In a 2D network the transitions are very much cooperative, because alternative hydrogen bond donating or accepting groups cannot so easily be found. These findings demonstrate that it is not trivial to arrive at the best structure from an arbitrary starting point.

## Hydroxyl and hydroxymethyl group conformations

The data discussed here all refer to the constant volume simulations. The average torsions angles (Table 3) of hydroxymethyl and hydroxyl group and their small rms fluctuations in the IP2-ideal simulation show that all conformations are stable. A distribution of the origin and centre hydroxylmethyl group conformations (Figure 7) during a period of 100 ps in IP2-ideal is displayed. All hydroxymethyl groups stay in the *tg* conformation ($\omega = 300^\circ$). The distributions of origin hydroxyl group conformations (Figure 8) and centre hydroxyl group conformations (Figure 9) also show just one conformation.

For cellulose II the situation is somewhat different. The centre hydroxymethyl groups (Figure 10) are not exclusively in the *gt* conformation. Some 5% of *tg* conformation occurs.

**Table 3.** Average torsion angles (°) (and r.m.s. fluctuations) in cellulose I$\beta$ and II.

| | | $\phi$ | $\psi$ | $\omega$ | $\phi 2$ | $\phi 3$ | $\phi 6$ |
|---|---|---|---|---|---|---|---|
| I$\beta$P2 | Origin | − 98(7) | 101(7) | − 70(10) | 56(12) | 171(14) | − 150(14) |
| | Centre | − 97(7) | 101(7) | − 67(13) | 44(14) | 167(13) | − 154(14) |
| IIA2 | Origin | − 97(7) | 102(7) | − 169(10) | 102(13) | 162(22) | − 137(14) |
| | Centre | − 97(8) | 90(8) | − 179(23) | 138(15) | − 138(28) | 96(39) |

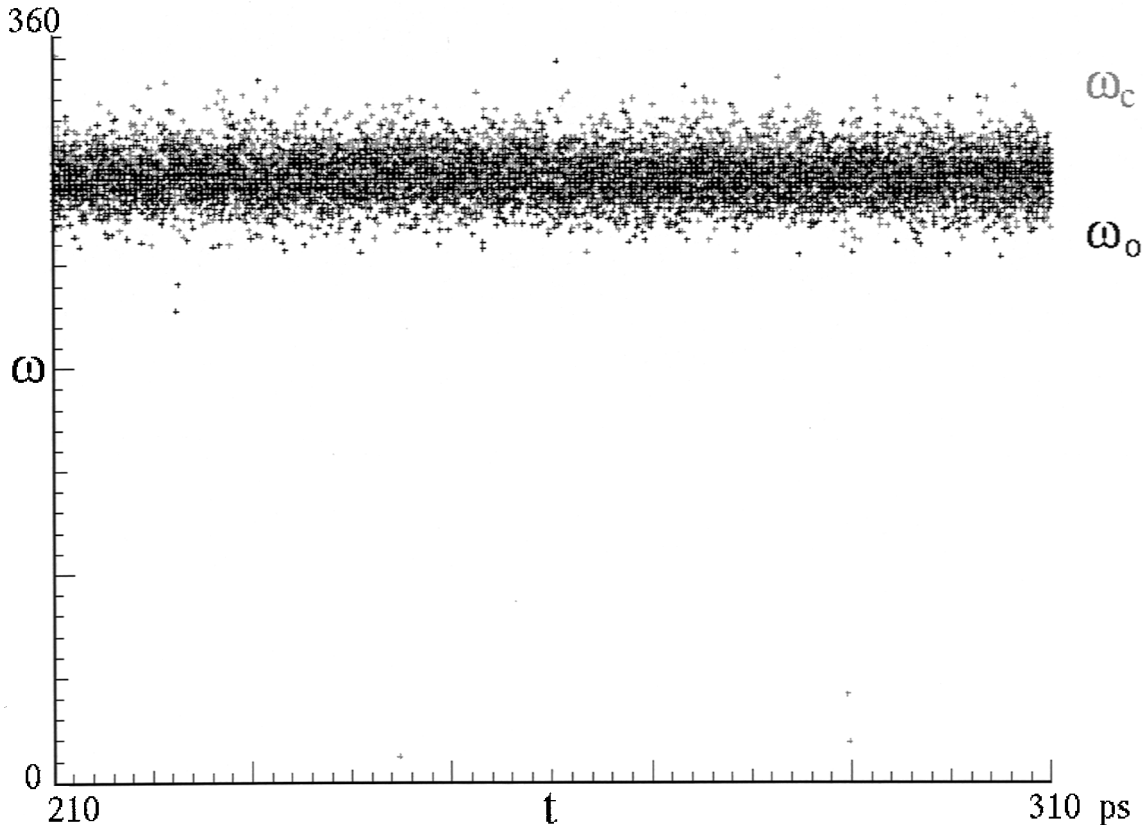


**Figure 7.** Distribution of hydroxymethyl group conformations of origin ($\omega_0$, black) and centre chains ($\omega_c$, grey) in the IP2 simulation.

If this percentage were present in the cellulose II microcrystal it would be hardly be observable. The origin hydroxyl group conformation of O3-H ($\phi_3$) (Figure 11) has a somewhat larger spread than usual for a stable conformation. In the centre layer especially the hydroxyl groups O3-H and O6-H (Figure 12) are mobile. This will be partly related to the disorder in the hydroxymethyl group conformation, but it cannot be explained by that alone. There is clearly a disorder in the hydrogen bond scheme (Table 4). This feature was not removed by restarting the simulation in the 'ideal' structure, as was the case for cellulose I.

## Hydrogen bonding and average structure

Cellulose I$\beta$ turns out to be an ordered structure consisting of hydrogen bonded layers that are packed through hydrophobic interactions. The average structure in the ab (Figure 13) and ac (Figure 14) projection can be viewed. The cellulose II structure is a 3D hydrogen bonded network. Percentages of occurrence (rounded off to 10%) of hydrogen bonds (Table 4) show that only the IA2 centre chains have a disordered hydrogen bond scheme. The O3-H hydroxyl group can either point to O5/O6 by forming an intramolecular hydrogen bond, or to O6 of an origin chain thus forming an interplane hydrogen bond. From the mobility of the centre hydroxyl groups (Figure 12) it is evident that this disorder is dynamic. The average structure in the ab (Figure 15) and ac (Figure 16) projection can be viewed. Another aspect of averaging atomic coordinates in time is that atomic displacement parameters U (temperature motions or mean square deviations from average positions) can be

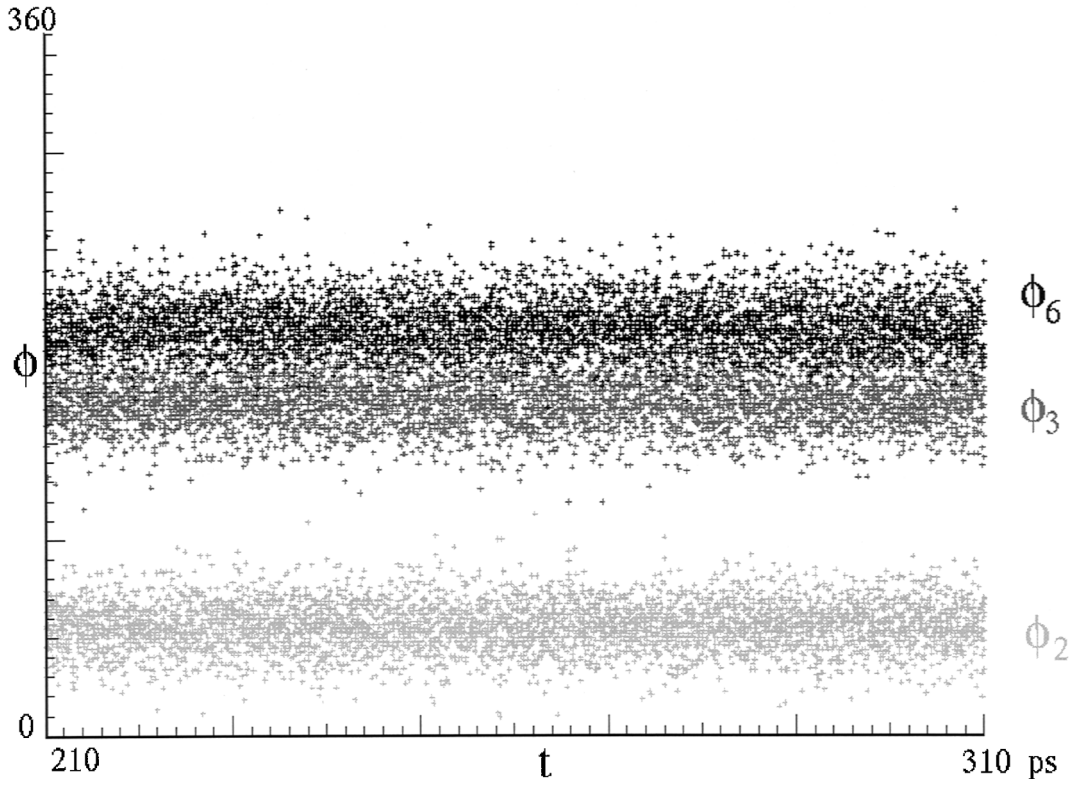


**Figure 8.** Distribution of hydroxymethyl group conformations in the origin chain of IP2 ($\phi_2$ = C1-C2-O2-H light-grey, $\phi_3$ = C2-C3-O3-H grey and $\phi_6$ = C5-C6-O6-H black).

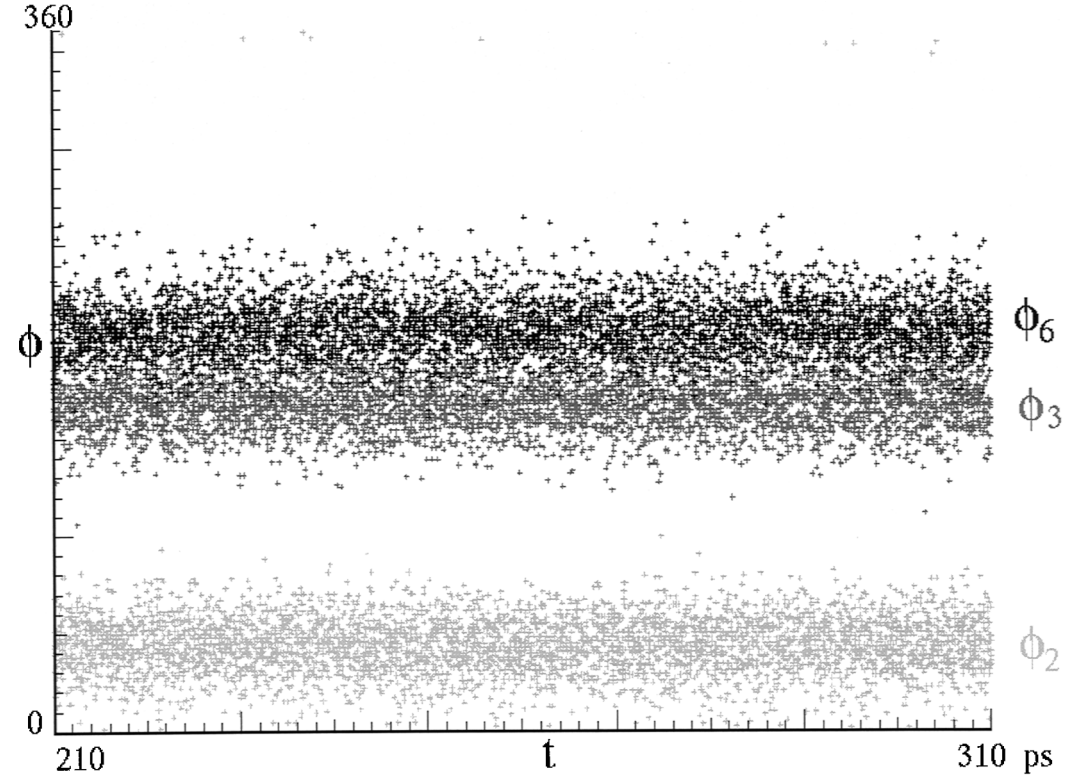


**Figure 9.** Distribution of hydroxymethyl group conformations in the centre chain of IP2 ($\phi_2$ = C1-C2-O2-H light-grey, $\phi_3$ = C2-C3-O3-H grey and $\phi_6$ = C5-C6-O6-H black).

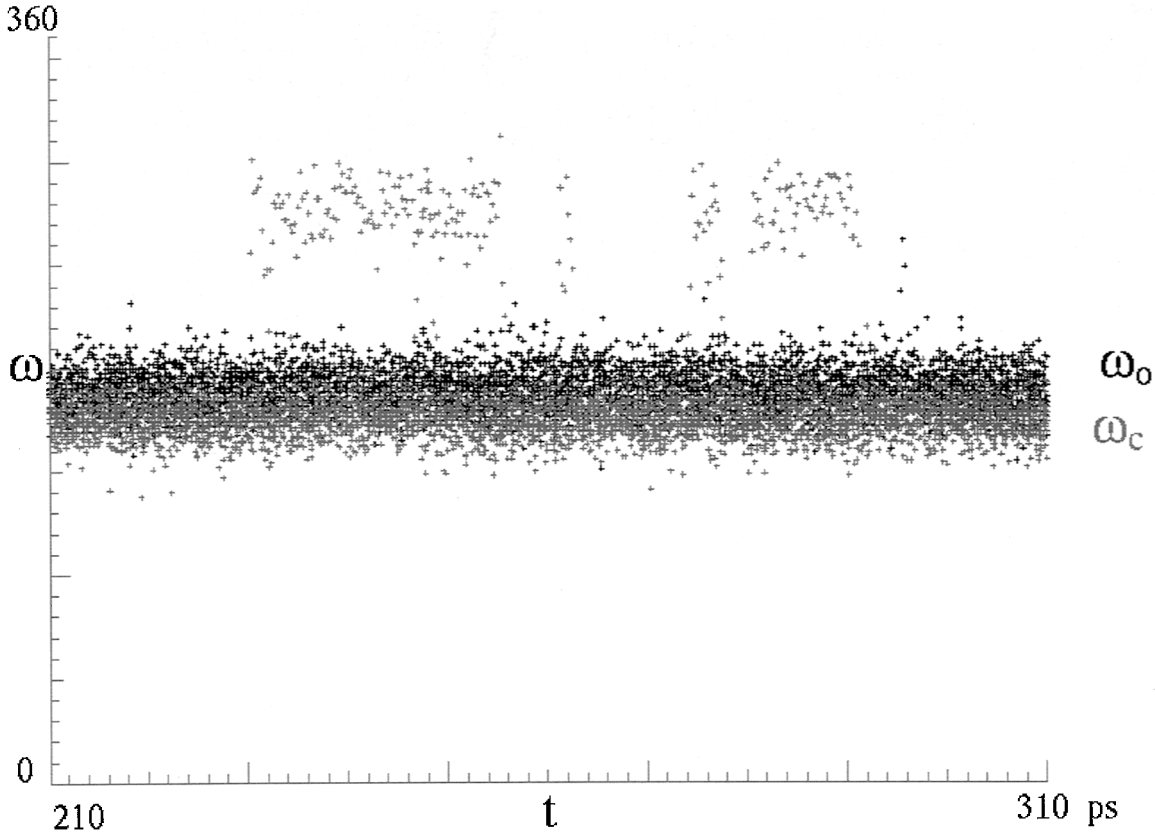


**Figure 10.** Distribution of hydroxymethyl group conformations of origin ($\omega_0$, black) and centre chains ($\omega_c$, grey) in the IIA2 simulation.

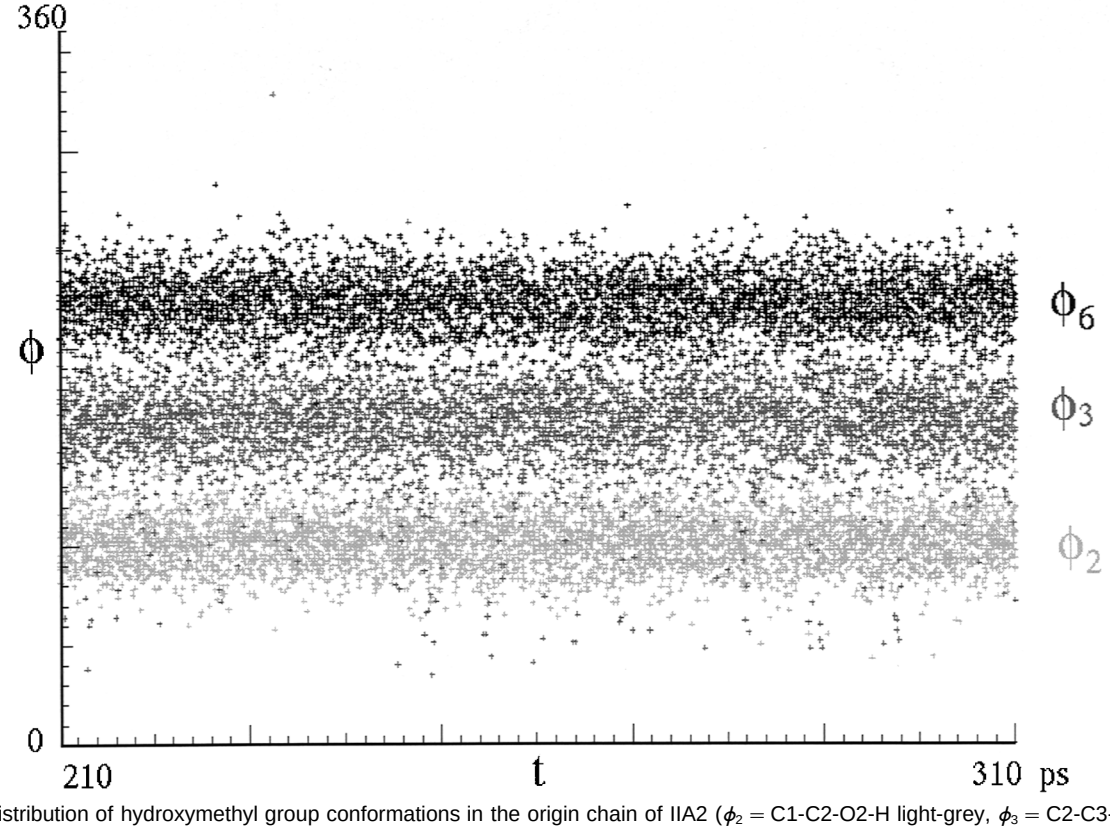


**Figure 11.** Distribution of hydroxymethyl group conformations in the origin chain of IIA2 ($\phi_2$ = C1-C2-O2-H light-grey, $\phi_3$ = C2-C3-O3-H grey and $\phi_6$ = C5-C6-O6-H black).

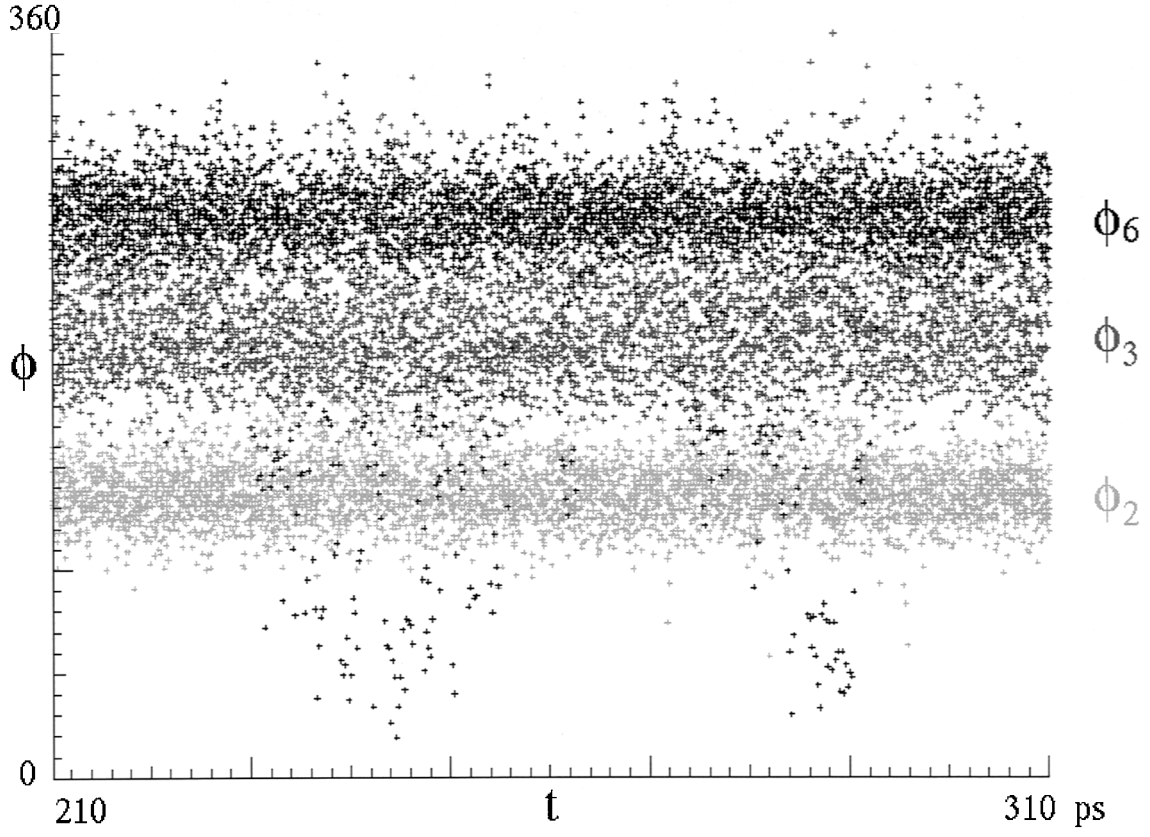


**Figure 12.** Distribution of hydroxymethyl group conformations in the centre chain of IIA2 ($\phi_2$ = C1-C2-O2-H light-grey, $\phi_3$ = C2-C3-O3-H grey and $\phi_6$ = C5-C6-O6-H black).

**Table 4.** Average percentage of existence of hydrogen bonds in cellulose I$\beta$ and II.

| | *Origin chain* | | | *Centre chain* | | |
|---|---|---|---|---|---|---|
| I$\beta$P2 | O2o-H..O6o | 100 | intra | O2c-H..O6c | 100 | intra |
| | O3o-H..O5o | 100 | intra | O3c-H..O5c | 100 | intra |
| | O6o-H..O3o | 100 | intraplane | O6c-H..O3c | 100 | intraplane |
| IIA2 | O2o-H..O2c | 100 | interplane | O2c-H..O6c | 100 | intraplane |
| | O3o-H..O5o | 100 | intra | O3c-H..O5c/O6c | 60/40 | intra |
| | | | | O3c-H..O6o | 40 | interplane |
| | O6o-H..O2o | 100 | intraplane | O6c-H..O6o | 100 | interplane |

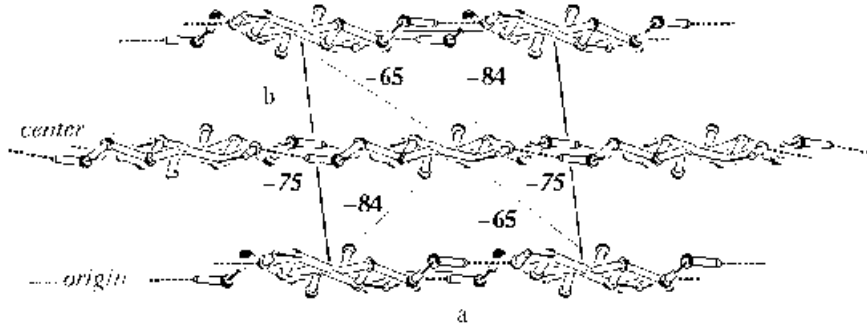


**Figure 13.** Average structure for the IP2 simulation in the ab projection. The largest interaction energies between neighbours are indicated. Coulomb interaction energies (in kJ mol$^{-1}$ per cellobiose) are given in italics and van der Waals interaction energies in bold.

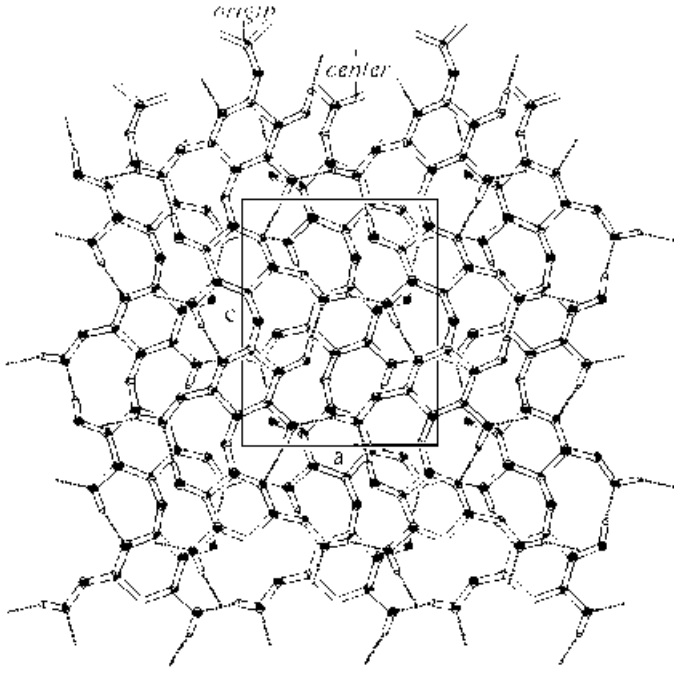


**Figure 14.** Average structure for the IP2 simulation in the ac projection.

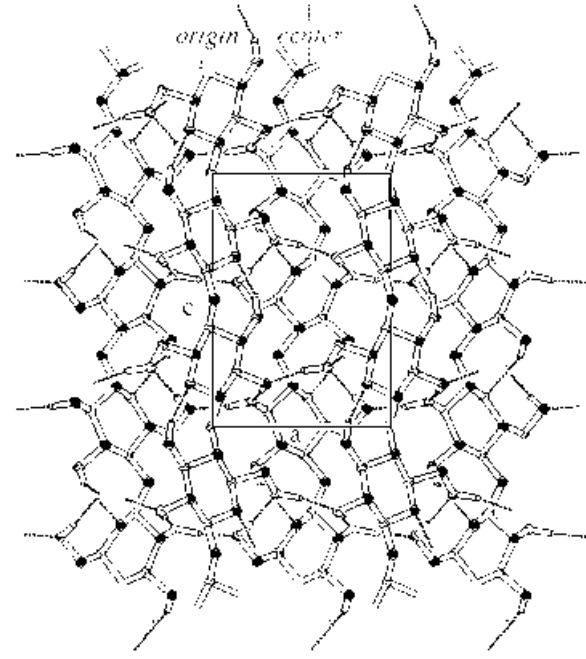


**Figure 16.** Average structure for the IIA2 simulation in the ac projection.

calculated. These are displayed as ellipsoids at a 50% probability level. For cellulose I small Us (Figure 17) are obtained in accordance with an ordered structure. For cellulose II the Us (Figure 18) of the O3-H and O6-H atoms in the centre chain are large and represent the mobility of these groups. Decomposition of the potential energies (Table 5) in the constant volume simulations gives further insight into the intra and intermolecular interactions. The potential energy Epot in cellulose II is somewhat higher than that in cellulose I. This is caused by the intramolecular energy. Obviously the molecule has a lower energy in cellulose I because of extensive intramolecular hydrogen bonding. In contrast the intermolecular interaction energy in cellulose II is larger (more negative) due to the formation of more intermolecular hydrogen bonds (Table 4). To illustrate this the largest interaction between one centre chain and its neighbours is shown for cellulose I and II. The Coulomb interactions are indicated in italics and the van der Waals interaction in bold. In cellulose I (Figure 13) the interactions with neighbours in the plane is mainly electrostatic

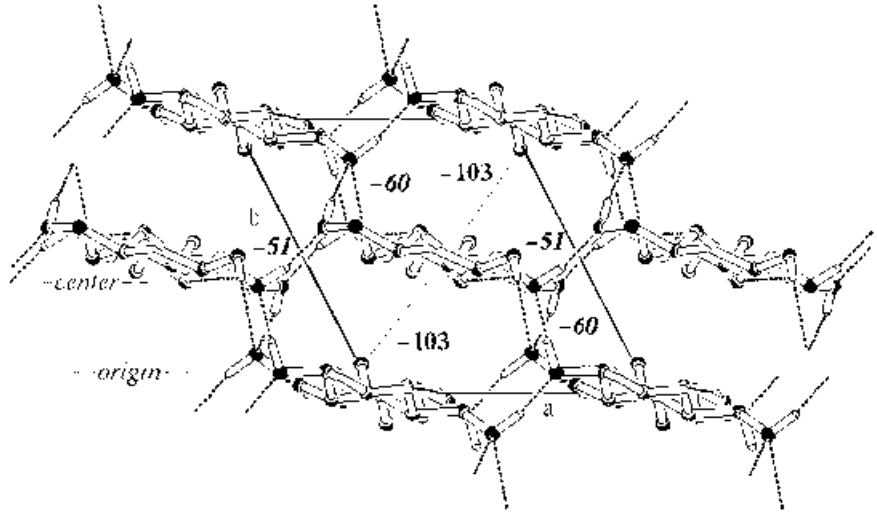


**Figure 15.** Average structure for the IIA2 simulation in the ab projection. See legend to Figure 13 for further details.

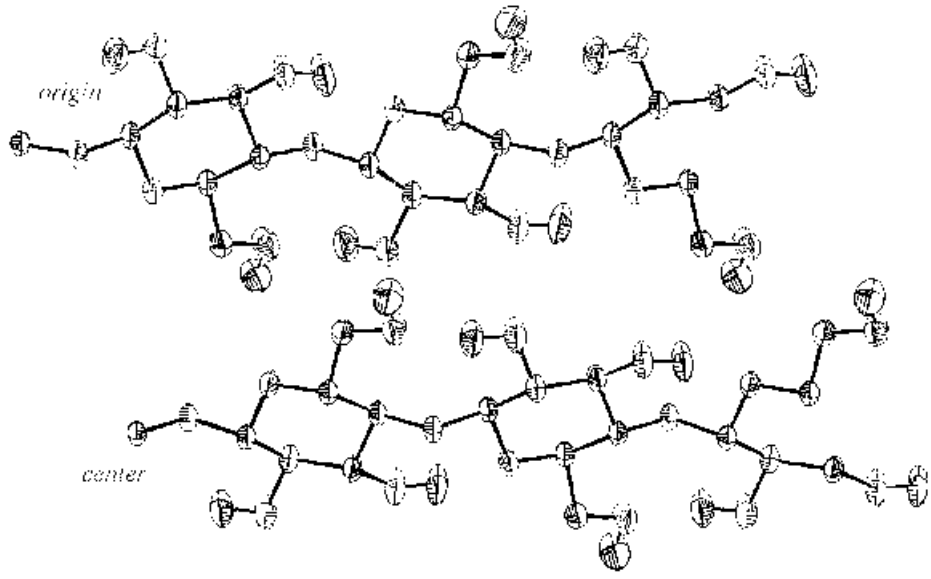


**Figure 17.** Average structure and thermal ellipsoids (drawn at the 50% probability level) of the origin and centre chain in the IP2 simulation.

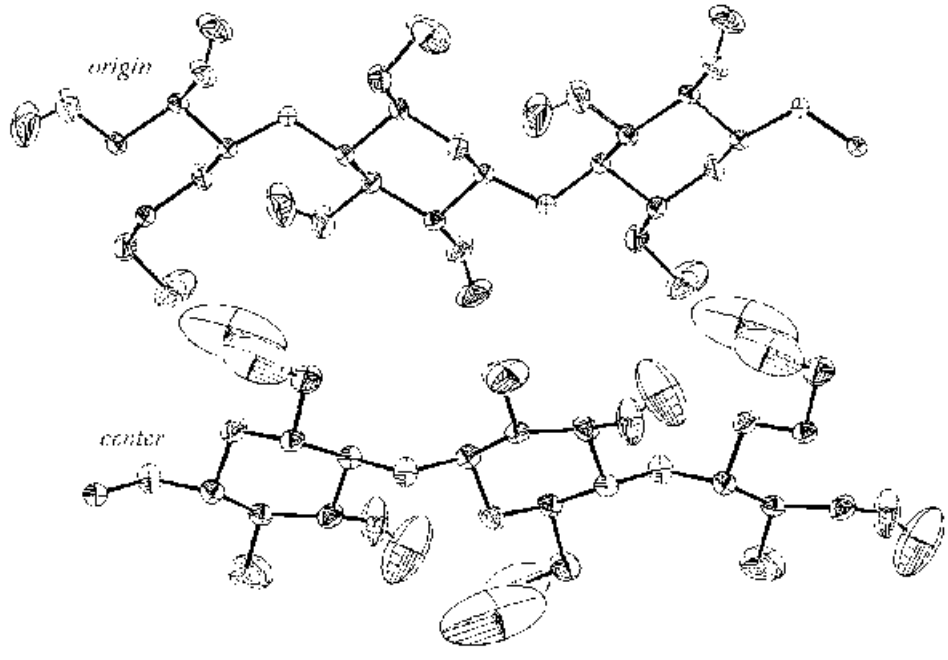


**Figure 18.** Average structure and thermal ellipsoids (drawn at the 50% probability level) of the origin and centre chain in the IIA2 simulation.

( $-75$ kJ mol$^{-1}$ per cellobiose) as expected due to the two interplane hydrogen bonds. The centre and origin chains mainly have a van der Waals interaction ( $-65$ to $-84$ kJ mol$^{-1}$ per cellobiose). Thus the interlayer interaction can be conceived as hydrophobic. In cellulose II (Figure 15) each centre chain has four neighbours with which it has mainly electrostatic interactions in line with the 3D hydrogen bond network. In the [110] direction there is a pronounced van der Waals stacking of the molecules. On the whole the contribution (Table 5) of van der Waals interactions is larger than that of electrostatic interactions, but not so evident as in case of cellulose I.

## Discussion

A similar energy analysis, using the GROMOS force field, was carried out by Heiner *et al.* [9] for cellulose I$\alpha$ and I$\beta$. Their initial structures were based on electron

**Table 5.** Decomposition of average potential energies in cellulose I$\beta$ and II in kJ/mol$^{-1}$ cellobiose.

| | | $E_{pot}$ | | $E_{elec}^{inter}$ | $E_{vdw}^{inter}$ | $E_{elec}^{intra}$ | $E_{vdw}^{intra}$ |
|---|---|---|---|---|---|---|---|
| I$\beta$P2 | tg/tg | 306.1 | Origin | − 82.1 | − 176.2 | 455.6 | − 19.7 |
| | | | Centre | − 82.6 | − 175.5 | 452.1 | − 16.8 |
| IIA2 | gt/gt | 310.3 | Origin | − 130.1 | − 161.2 | 492.6 | − 22.8 |
| | | | Centre | − 123.4 | − 155.2 | 491.2 | − 24.9 |

diffraction experiments by Sugiyama *et al.* [17]. This I$\beta$ structure is similar to our parallel-down and Sarko and Muggli's structure [2]. Small deviations between their and our I$\beta$ model occur because their unit cell was somewhat different (a = 8.17, b = 8.01, c = 10.36 Å, $\gamma = 97.3^\circ$). In particular, our b-axis is smaller, leading to a larger van der Waals interaction. Their hydrogen bond statistics is also somewhat different. This could be due to a different hydrogen bond criterion (H..O < 2.4 Å in our calculations versus H..O < 2.5 Å/O-H..O > 135° in theirs), but also to a different mobility of the hydroxyl groups. In their simulations Heiner *et al.* find frequent transitions (varying from once every 30–125 ps) of these groups in contrast to the torsional behaviour in our origin (Figure 8) and centre (Figure 9) chains, of which a 100 ps part is shown. Such transitions were also found in the centre layer of our initial, but not in the 'ideal' simulations. Since the difference in energy (Table 2) between the initial (Ev) and 'ideal' (Ev-ideal) is only 2 kJ mol$^{-1}$ per cellobiose (0.5 kcal mol$^{-1}$) it could be that in reality transitions occur. A similar reasoning applies to the hydroxymethyl groups, which also show transitions in the Heiner simulations, but not in ours. Hardy and Sarko [10] also performed MD simulations on the I$\alpha$ and I$\beta$ phase with the CHARMM force field. It appeared in their calculations as if the temperature was considerably higher than room temperature (they simulated a NVE ensemble), because considerable glycosidic torsional motions occur and frequent hydrogen bond forming and breaking takes place.

The double intramolecular hydrogen bond formation along each glycosidic linkage in cellulose I is in agreement with our earlier calculations of elastic moduli. The associated *tg* conformation of the hydroxymethyl groups was also found with 13C-CP/MAS NMR spectroscopy by Horii *et al.* [18] for all native cellulose samples. For cellulose II no earlier MD simulations were carried out. Several MM calculations are available in the literature. French *et al.* [20] used the MM3 force field for MM minimizations on miniature crystal models. They found the cellulose II structure to be lower in energy than cellulose I$\beta$ by 0.64 kcal mol$^{-1}$. This is the reverse of what we find (1 kcal mol$^{-1}$ in favour of cellulose I$\beta$). A lot of differences between their and our calculations can be pointed out. One of these is that they give the energy for a complex of 7 cellotetraose molecules. Simply dividing this by 7 does not give the proper potential energy for a cellulose crystal because the intermolecular interactions are then only partly accounted for (boundary effects). They also give the lattice energy which then turns out to be larger (more negative) in cellulose I$\beta$ than in cellulose II, again opposite to what we find. This point will be discussed more thoroughly in another paper.

The observation that cellulose I irreversibly converts to cellulose II by regeneration or mercerization would seem to imply that cellulose II is the more stable polymorph. This could not be confirmed by our relative energies. Of course one has to bear in mind that there are always model and force field errors, but in any case the energy difference between I$\beta$ and II will not be large, at least smaller than is suggested by the irreversible transformation. We anticipate therefore that energy is not the decisive factor in this transformation.

A detailed comparison of our cellulose II structure can be made with recent determinations of the crystal structure of cellotetraose from single crystal X-ray diffraction. This structure has long been considered a key structure to understanding the packing of molecules in cellulose II. Gessler *et al.* [21]) and Raymond *et al.* [22] simultaneously solved the cellotetraose structure. Two independent molecules crystallize in a triclinic space group. Apart from the *c*-axis (fibre repeating unit versus tetramer size) the unit cell can be transformed to one similar to the cellulose II cell. A view of the cellotetraose structure (Figure 19) along the tetramer axis (Raymound *et al.*) is compared with the MD cellulose II structure (Figure 15). (Do not pay attention to the orientation of the hydroxyl groups in the cellotetraose structure, because they are not well established, and not according to possible hydrogen bonds.) In both determinations the *gt* conformation for all hydroxymethyl groups was found, similar to what we persistently find and in contrast to the X-ray diffraction studies on cellulose II fibres. Just recently a renewed refinement [23] of earlier X-ray data using the cellotetraose model with all *gt* conformations, showed only a minor difference in R-value, so that this model is equally compatible with the data. Gessler *et al.* [23] report the average $\phi$ and $\psi$ dihedrals in the glycosidic linkage of both molecules. The origin chain has $(\phi, \psi) = (-95, 97)$ and the centre chain $(-94, 90)$, which is in good agreement with our dihedrals (Table 3). Even the subtle difference in $\psi$ angle between the two independent units is found in the cellotetraose model and our MD simulation. The $\omega$ angles of − 176 and 170° for the origin and centre chains differ by 7 and 11° from ours. Cremer and Polple puckering

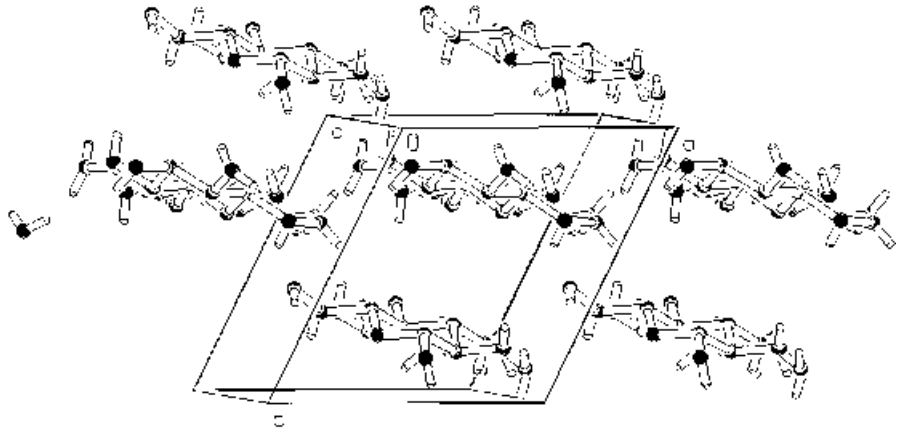

**Figure 19.** The cellotetraose structure as determined by Raymond *et al.*

**Table 6.** Puckering parameters from the average cellulose II structure

| *Residue* | *Origin* | | | *Centre* | | |
|---|---|---|---|---|---|---|
| | *Q* | *θ* | *ϕ* | *Q* | *θ* | *ϕ* |
| glc1 | 0.769 | 4.6 | 236.0 | 0.746 | 12.2 | 343.3 |
| glc2 | 0.657 | 8.1 | 55.8 | 0.662 | 23.0 | 16.7 |
| glc3 | 0.768 | 4.6 | 241.2 | 0.741 | 13.3 | 344.2 |
| glc4 | 0.672 | 6.2 | 57.0 | 0.667 | 22.1 | 16.7 |

parameters (Table 6) of the average coordinates of the IIA2 simulation, without using the two-fold screw symmetry are calculated. In both cellotetraose papers the puckering parameters for the various glucose rings are also calculated. The Q and $\theta$ parameters of the middle two glucose residues in the origin chain are 0.590/2.6 and 0.605/1.5 for Raymond and Gessler respectively, and those of the centre chain are 0.580/11.9 and 0.585/12.5 respectively. The fact that $\theta$ in the centre chain is larger than in the origin chain is similarly found in our MD simulations of the cellulose II structure. There is a notable and systematic difference in puckering of subsequent glucose residues in the MD simulation. This implies a slight deviation of the two-fold screw axis symmetry, which is not imposed on the simulation, but apparently is present in the fibre structure as judged from the X-ray data.

In the paper by Raymond *et al.* the hydroxyl hydrogen atoms have been placed rather than found from the electron density map. From the oxygen-oxygen distances one can try to locate the hydrogens, but ambiguities remain. Gessler *et al.* found several hydrogen atoms in the map: enough at least to establish a hydrogen bonds pattern. Average hydrogen bond parameters (Table 7) from MD and from those that were actually observed by Gessler *et al.* are listed. In cellotetraose exactly the same hydrogen bond scheme is found as found in the MD simulations of cellulose II. Also the geometries agree fairly well. However, no evidence was found in the crystal structure of cellotetraose for the disorder of the centre O3-H group (leading to a partial O3c-H..O6o). On the other hand only one of their O3c-H atoms

**Table 7.** Average hydrogen bond parameters in cellulose II compared with those in cellotetraose.

| | *MD* | | | *Cellotetraose (obs. by Gessler)* | | |
|---|---|---|---|---|---|---|
| | *O..O* | *H..O (Å)* | *O-H..O (°)* | *O..O* | *H..O* | *O-H..O* |
| Origin chain | | | | | | |
| O2o-H..O2c | 2.71 | 1.74 | 164 | 2.73–2.77 | 1.77–1.82 | 163–170 |
| O3o-H..O5o | 2.81 | 1.88 | 155 | 2.84–2.85 | 1.92–2.04 | 138–160 |
| O6o-H..O2o | 2.68 | 1.74 | 159 | 2.69–2.76 | 1.79–1.95 | 139–172 |
| Centre chain | | | | | | |
| O2c-H..O6c | 2.67 | 1.72 | 160 | 2.62 | 1.80 | 138 |
| O3c-H..O5c | 2.85 | 2.08 | 136 | 2.87 | 1.93 | 161 |
| O3c-H..O6c | 3.03 | 2.22 | 137 | 3.06 | 2.47 | 118 |
| O3c-H..O6o | 2.93 | 2.05 | 148 | | | |
| O6c-H..O6o | 2.69 | 1.74 | 161 | 2.74–2.86 | 1.79–1.91 | 161–163 |

could be located. Disordered hydrogen atoms are of course hardly traceable.

## Conclusion

The MD simulations indicate that cellulose I$\beta$ has a nicely ordered structure, consisting of hydrogen bonded layers, packed together through hydrophobic interactions. The orientation of the molecules in the unit cell is parallel-down and the structure is in fact the one proposed by Sarko and Mugli [2] and not that proposed by Gardner and Blackwell [1]. Cellulose II forms a tight 3D hydrogen bonding network. There is a strong indication of some disorder in the hydrogen bond pattern, due to mobility of the centre hydroxyl group O3c-H. The potential energy of cellulose II is found to be higher than that of cellulose I, in contrast to what is expected on the basis of the observed irreversible transition of cellulose I to II. The agreement between our simulated cellulose II structure and the recently determined X-ray structure of cellotetraose is to the degree that they mutually support each other as proper representations of the cellulose II structure.

## Acknowledgement

We thank B. Bouma for calculating the average structures from MD.

## References

1 Gardner KH, Blackwell J (1974) *J Biopolym* **13**: 1975–2001.
2 Sarko A, Muggli R (1974) *Macromolecules* **7**: 486–94.
3 Woodcock C, Sarko A (1980) *Macromolecules* **13**: 1183–87.
4 French AD (1978) *Carbohydr Res* **61**: 67–80.
5 Kolpak FJ, Blackwell J (1976) *Macromolecules* **9**: 273–78.
6 Stipanovic AJ, Sarko A (1976) *Macromolecules* **9**: 851–57.
7 Atalla RH, van der Hart DL (1984) *Science* **223**: 283–85.
8 Chanzy H, Henrissat B (1985) *FEBS Lett* **184**: 285–88.
9 Heiner AP, Sugiyama J, Telleman O (1995) *Carbohydr Res* **273**: 207–23.
10 Hardy BJ, Sarko A (1995) *Polymer Preprint*, ACS (Anaheim meeting) **36**: 640.
11 Kroon-Batenburg LMJ, Bouma B, Kroon J (1996) *Macromolecules* **29**: 5695–99.
12 van Gunsteren WF, Berendsen HJC, GROMOS (1987) University of Groningen, The Netherlands.
13 Koehler J, Saenger W, van Gunsteren WF (1987) *Eur Biophys J* **15**: 197–210.
14 van Eijck BP (1994) *Mol Simulation* **13**: 221–30.
15 Kouwijzer MLCE, van Eijck BP, Kroes SJ, Kroon J (1993) *J Comp Chem* **14**: 1281–89.
16 Haigler CH, Brown RM Jr, Benziman M (1980) *Science* **210**: 903–06.
17 Sugyiama J, Vuong R, Chanzy H (1991) *Macromolecules* **24**: 4168.
18 Horii F, Hirai F, Kitamaru R (1983) *Polym Bull* **10**: 357–61.
19 Kroon-Batenburg LMJ, Kroon J, Northolt MG (1986) *Polym Comm* **27**: 290–92.
20 French AD, Miller DP, Aabloo A (1993) *Int J Biol Macromol* **15**: 30–36.
21 Gessler K, Krauss N, Steiner T, Betzel C, Sandmann C, Saenger W (1994) *Science* **266**: 1027–29.
22 Raymond S, Heyraud A, Tran Qui D, Kvick A, Chanzy H, (1995) *Macromolecules* **28**: 2096–100.
23 Gessler N, Krauss N, Steiner T, Betzel C, Sarko A, Saenger W (1995) *J Am Chem Soc* **117**: 11397–406.