ELSEVIER

Carbohydrate Research 273 (1995) 207–223

CARBOHYDRATE RESEARCH

Crystalline cellulose Iα and Iβ studied by molecular dynamics simulation

Andreas P. Heiner *, Junji Sugiyama [1], Olle Teleman

Technical Research Centre of Finland, Biotechnology and Food Research, PO Box 1503, FIN-02044 VTT Espoo, Finland

Received 24 November 1994; accepted 14 February 1995

Abstract

A recent structure determination of native cellulose has shown that it is composed of two different crystal structures, a two-chain monoclinic phase and a single-chain triclinic phase. In this article a molecular dynamics study of the two allomorphs is presented, and a general picture of structure and energetics is provided. Consistent with experimental data, the monoclinic phase is more stable than the triclinic phase by -8.7 kJ mol^{-1} cellobiose^{-1}. In the monoclinic phase a small angle is observed between glucose residues that belong to alternate (200) planes. The glucose residues in every second plane are parallel to the (200) plane, and form more favorable intermolecular hydrogen bonds. In the triclinic phase the glucose residues are not parallel to the (200) plane. The ^{13}C NMR shifts for C-6 are fully accounted for by the distribution of the C-6 dihedral angle. The nonbonded environment is important to the splitting for C-1. The fine structure of IR spectra in the OH-stretching region can be qualitatively correlated with the number of different hydrogen bonds observed. Results indicate that chains in one set of alternating (200)-planes in the monoclinic phase resemble the triclinic phase.

Keywords: Cellulose I; Molecular dynamics; Allomorph stability; ^{13}C CP/MAS NMR-spectra; Hydrogen bonding

1. Introduction

Natural cellulose consists of long, parallel homopolymers of (1 → 4)-β-linked D-glucose monomers. Despite its importance in a number of industrial applications, relatively little is known about the detailed structure at mesoscopic or microscopic level. This is

* Corresponding author. Supplementary material available upon request.

[1] Present address: Wood Research Institute, Kyoto University, Kyoto 611, Japan.

SSDI 0008-6215(95)00103-4

due in part to the fiber structure of cellulose, with crystalline and amorphous domains, at a microscopic level, and the heterogeneity at the mesoscopic level, where the individual cellulose fibers are linked by hemicelluloses and lignins. The variation in composition of the hemicellulose is mainly responsible for the differences between hardwood and softwood [1].

The debate about the atomic structure of the crystalline part of cellulose was revived in 1984, when Atalla and vanderHart proposed two phases for the native crystalline structure based on results obtained by solid-state ^{13}C CP-MAS NMR experiments [2,3]. Further support was found by Raman spectroscopy [4] and infrared spectroscopy [5–7]. In addition to these structural data, thermodynamic data revealed that the phase most abundant in lower plants and bacteria (the Iα phase) is less stable than the Iβ phase, which preponderates in higher plants [8–10]. The exact ratio of the two phases depends critically on the origin of the cellulose. The structure of the two phases was solved by Sugiyama et al. using electron diffraction [11]. In both phases the chains were organized in a parallel-up fashion. The Iα phase was triclinic, with cell dimensions $a = 0.674$ nm, $b = 0.593$ nm, $c = 1.036$ nm (chain axis), $\alpha = 117°$, $\beta = 113°$, $\gamma = 81°$, and one cellobiose residue per unit cell. The Iβ phase was monoclinic, and very similar to the model found by Sarko and Muggli [12], with $a = 0.801$ nm, $b = 0.817$ nm, $c = 1.036$ nm (chain axis), $\alpha = \beta = 90.0°$, $\gamma = 97.3°$, and two cellobiose residues per unit cell. The unit cell volumes are 0.3395 nm^3 and 0.6725 nm^3 and the densities are 1.582 and 1.599 g cm^{-3} for the Iα and Iβ allomorphs. This two-phase model accounts for most of the observations made earlier, where some authors proposed a two-chain monoclinic model [12,13], and others an eight-chain triclinic model [14]. Mixing the triclinic and the monoclinic phase in suitable proportions yields reflections consistent with either of the two previous models. Recently, the surfaces of the allomorphs were also identified using atomic force microscopy [15].

So far, few attempts have been made to model crystalline cellulose. In most cases glucose, cellobiose, cellotetraose, or longer chains have been used as a model, both in vacuo and in solution [16–20]. In these cases crystal packing is completely ignored, which reduces the validity of the predictions considerably. In a recent study French et al. [21] calculated the relative energies of various cellulose allomorphs in two different approaches. In the first approach they took into account the crystalline environment, but used a force field with limited flexibility. This so-called rigid ring approach generally gives too narrow energy minima, and thus unreliable energy differences. In a second approach full flexibility was included using the MM3 force field in a so-called miniature crystal. The chain under consideration is embedded in a hexagonal packing consistent with crystal parameters and the miniature crystal itself is in vacuo. It is unclear whether this approach accounts for the full crystalline environment in a proper way. Only Kroon-Batenburg and Kroon have simulated cellulose Iβ in a crystalline environment using molecular dynamics [22]. Their starting structure was that of Gardner and Blackwell [23], that is, a parallel-down conformation [24], an alternative which is now generally considered invalid.

In this paper we present molecular dynamics (MD) simulations of cellulose Iα and cellulose Iβ with full atomic detail, taking into account the crystalline environment. The results will be related to available experimental data.

2. Computational procedure

Preliminary, unrefined structures of crystalline cellulose Iα and cellulose Iβ obtained by electron diffraction [11] were used as starting structures. In Fig. 1a the crystal structure of the monoclinic phase is shown with the crystal axes and planes. The chains in the (200) planes are tightly connected by intermolecular hydrogen bonds; this particular plane will also be referred to as a "sheet". The (100) plane in the triclinic phase corresponds to the (1–10) plane in the monoclinic phase, the (010/Iα) plane is equivalent to the (110/Iβ) plane and the (110/Iα) plane is equivalent to the (200/Iβ) plane. Throughout, we will use the plane nomenclature of the monoclinic phase; the two distinct (200) planes in the monoclinic phase will also be referred to as Iβ/even and Iβ/odd phase. In the monoclinic phase, the odd (200) planes are shifted $1/4c$ along the chain axis relative to the even (200) planes; in the triclinic phase each (200) plane is shifted $1/4c$ relative to the previous plane. Chains in the Iβ/odd planes are rotated by 7.4° relative to the (200) plane, in contrast to chains in the Iβ/even plane or chains in the triclinic phase, which are parallel to that plane. This indicates that chains in different (200) planes in the monoclinic phase are not equivalent. Fig. 1b is a top view of two cellulose chains in the (200) plane with the main hydrogen bonds and the nomenclature used.

The GROMOS87 force field [25] was used in conjunction with the GROMOS87 program suit [26]. The programs were extended to allow triclinic periodic boundary conditions. The force field uses a united atom approach, hence all aliphatic hydrogens were removed. Hydroxyl hydrogens (which were unresolved in the experimental structure) were added in ideal bond-length and bond-angle conformations. To allow for a reasonable non-bonded cutoff radius, unit cells were added according to their symmetry. The monoclinic phase was simulated in a periodic box of $3 \times 3 \times 3$ unit cells, yielding a total of 1512 atoms; the triclinic phase was simulated in a periodic box of $4 \times 6 \times 3$ ($a \times b \times c$) unit cells, with a total of 2016 atoms. The glucose hexamer in the simulation box was covalently connected to its periodic image to mimic an infinite chain. To prevent a bias to the crystalline conformation of the monomer, the dihedral angles of the hydroxyl hydrogens HO-2 and HO-3, and hydroxymethyl oxygen and hydrogen (O-6, HO-6) were randomized. The structures were subsequently energy-minimized. Initial velocities were taken from a Maxwellian distribution at 600 K, and the system was equilibrated at this temperature for 30 ps. Numerous dihedral transitions were observed during this period. The temperature was gradually brought down to 300 K, followed by another 20 ps of equilibration.

The system was simulated in an NVT-ensemble. Temperature was controlled using the Berendsen thermostat with a coupling constant of 0.4 ps [27]. Simulations were performed using a time step of 1.25 fs, keeping bond lengths constrained by the SHAKE-procedure with a relative tolerance of 10^{-8} [28,29]. Nonbonded interactions were calculated using the twin-range cutoff procedure with cutoff radii of 0.9 nm and 1.1 nm. The nonbonded pair list was updated every 25 fs, which appeared to be sufficient for a crystalline system. Configurations were saved every 0.1 ps. Simulations extended to 1 ns each, requiring 1 month of dedicated CPU-time on an IBM-RS6000/340 each.

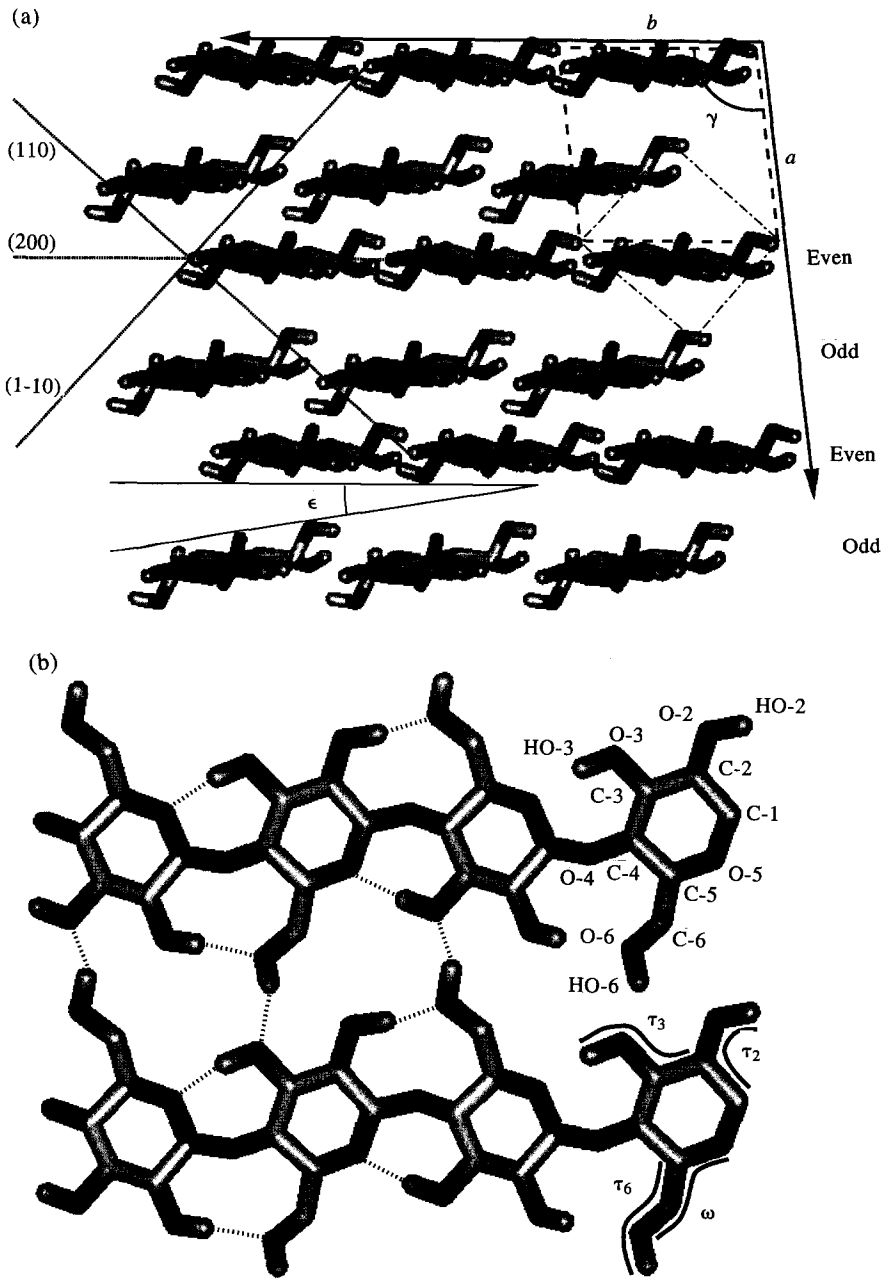

Fig. 1. (a) Relative positions prior to simulation of cellulose chains in monoclinic cellulose as viewed from the nonreducing end along the c-axis. The monoclinic a and b crystal axes are indicated by thick dashed lines, the projection of the triclinic $a-b$ plane onto the monoclinic $a-b$ plane is indicated by thin dash-dotted lines. The crystal planes (monoclinic nomenclature) are indicated by dotted lines. Chains in the odd (200) planes are rotated by an angle ϵ relative to chains in the even (200) plane. Chains in the odd (200) planes are shifted $1/4c$ relative to chains in the even (200) plane. In the triclinic phase, chains in subsequent sheets are translated by $1/4c$ relative to the previous sheet. (b) Nomenclature of the cellulose atoms and relevant dihedral angles. Intramolecular hydrogen bonds and between adjacent chains are depicted as broken lines.

3. Results and discussion

Overall structural aspects.—The root-mean-square positional difference (rmsPD) for the ring atoms and glycosidic oxygens between initial structure after energy minimization and average structure is 0.035 nm for the monoclinic phase and 0.018 nm for the triclinic phase. This is rather small, especially as the rmsPD was calculated without any rigid-body superposition. In the original Iβ structure the odd chains are off-center with an x-coordinate of 0.355 nm, while the unit cell center is at 0.397 nm. After 50 ps of equilibration the central position was attained within 0.002 nm, and the average structure over 1 ns yielded a symmetric interplanar distance of 0.397 nm. This translation constitutes the main contribution to the rmsPD of the monoclinic phase. The angle between chains in Iβ/odd and Iβ/even planes, initially 7.4°, increased to 11.5° in the average structure.

Internal coordinates.—The changes in bond angles or dihedral angles were negligible for ring atoms and the glycosidic linkage. The main difference was found for the O-5–C-5–C-6–O-6 (ω-) dihedral angle, which in the initial structure was 131.5° (Iα), 121.2° and 147.3° (Iβ/odd and Iβ/even), whereas the maximum in the distribution was at 180° for both the triclinic and the monoclinic phase. The root-mean-square positional fluctuations (rmsPF) of the ring atoms and glycosidic oxygens were rather small, 0.05 nm in both phases. Values were slightly larger for the C-6 atom and hydroxyl groups, especially the hydrogen atoms: 0.07–0.12 nm. The distributions of the dihedral angles of atoms that display large positional fluctuations are shown in Fig. 2a–d. Conformations rarely found in the Iβ/even phase are observed frequently in the triclinic phase, indicating higher flexibility in the latter phase. The distribution of the Iβ/odd phase is somewhere in between that of the Iβ/even and Iα phase, especially for the dihedrals involving hydrogens. In the Iβ/even phase the dihedral ω is almost completely restricted to the *tg* conformation (0.998). In the Iβ/odd phase other states are also populated (*tg*: 0.92, *gg*: 0.06, *gt*: 0.02), though not to the extent in the triclinic phase (*tg*: 0.81, *gg*: 0.08, *gt*: 0.11). For dihedrals involving hydrogen atoms numerous transitions were observed. The dihedrals τ_2 and τ_3 fluctuate between two states, τ_6 between three conformations, though the state around 300° is not very well defined. In the monoclinic phase a warmer region with higher dihedral transition rates was not fully equilibrated until after 500 ps. On average, the transition rate of these dihedrals during this period was roughly twice that during the second half of the simulation. The triclinic phase appears to equilibrate faster.

Transition rates of the τ_2 and ω dihedral do not differ between (200)/odd and (200)/even planes. The rates for the τ_3 and τ_6 dihedral angles, on the other hand, are approximately four times faster in the (200)/odd plane than in the (200)/even plane. This is probably due to the slight angle between chains in this plane and the plane itself. Averaging over the last 500 ps of the simulation, the transition rate in the monoclinic phase is 7.9 (τ_2), 13.0 (τ_3), 33.0 (τ_6) and 6.5 ns^{-1} (ω); for the triclinic phase the rates are 23.0, 55.0, 16.0, and 9.5 ns^{-1} (τ_2, τ_3, τ_6, and ω). The life time of the ω angle in the *gt* or *gg* states was found to be extremely short, in the order of the sampling frequency.

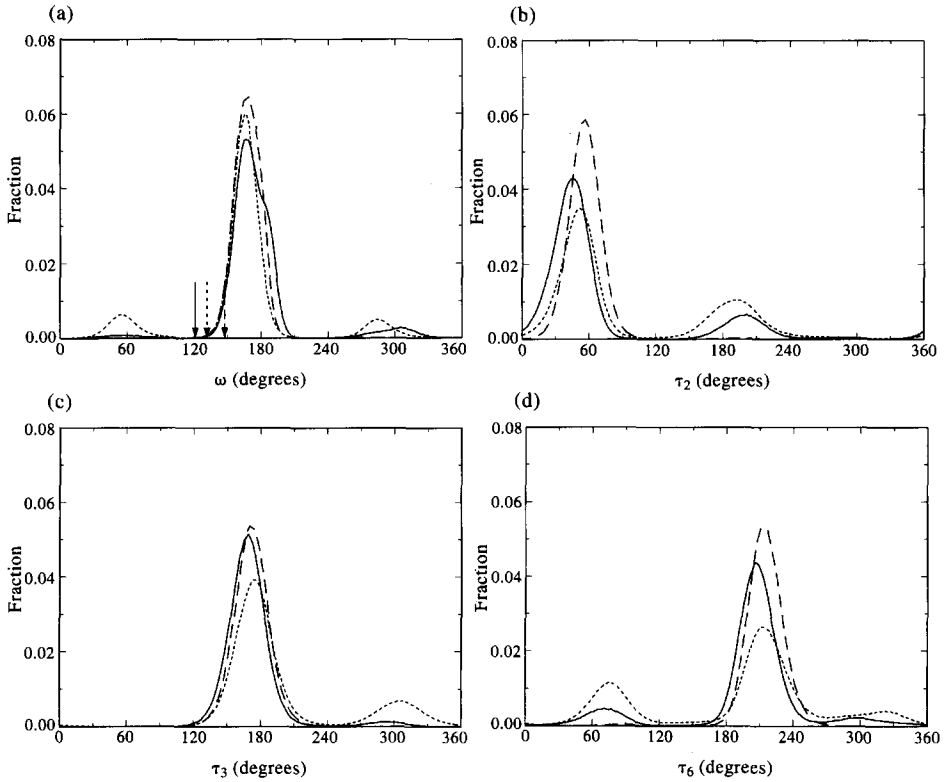

Fig. 2. Distribution of dihedral angles (**a**) ω (O-5–C-5–C-6–O-6), (**b**) τ_2 (C-1–C-2–O-2–HO-2) (**c**) τ_3 (C-2–C-3–O-3–HO-3), and (**d**) τ_6) (C-5–C-6–O-6–HO-6). Solid line: Iβ/odd phase; long dashed line: Iβ/even phase; short dashed line: Iα phase. The initial values for ω based on the unrefined diffraction coordinates are indicated by a solid arrow (121.2° Iβ/odd), a long dashed arrow (147.3°, Iβ/even), and a short dashed arrow (131.5°, Iα).

Radial distribution functions.—Similarity between structures can also be studied using the radial distribution function (RDF). It expresses the probability of finding a particular atom type in a spherical volume element dV at distance r from another atom.

The RDFs were calculated up to 0.8 nm for the C-1, C-4, and C-6 atoms (Fig. 3a–c), as these atoms are the most relevant to observed NMR spectra. As the system is crystalline and anisotropic, the radial distribution function displays pronounced structure even at 0.8 nm. The maxima in the distribution are found at almost identical distances in the three phases, especially when the target atom is not a hydrogen. The flanks of peaks and peak heights beyond the first shell (0.35 nm) show variation between the phases; this is most obvious for the hydrogen atoms. In the case of the C-6–O-2 pair the first peak in the Iβ/odd phase is split, whereas it is very wide in the triclinic phase and rather narrow in the Iβ/even phase. This has probably to do with differences in hydrogen bonding (see below); the highest occurrence of the intramolecular hydrogen bond O-2–HO-2 · · · O-6 is in the Iβ/even phase, whereas in the two other phases this

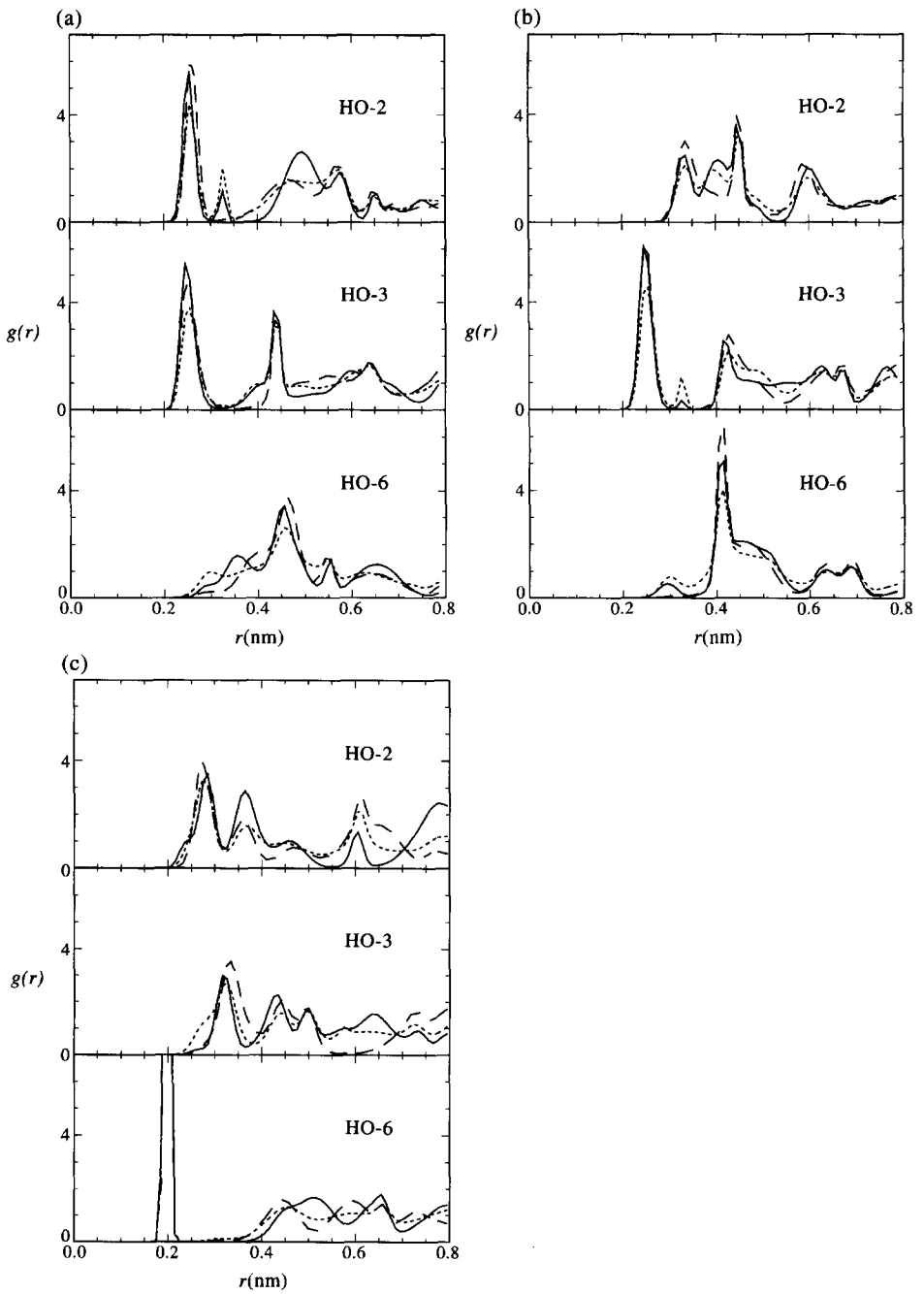

Fig. 3. Radial distribution functions of hydroxyl hydrogen atoms relative to (**a**) C-1, (**b**) C-4, and (**c**) C-6. Solid line: Iβ/odd phase; long dashed line: Iβ/even phase; short dashed line: Iα phase.

interaction is less frequent; instead, an intermolecular hydrogen bond between O-6–HO-6 · · · O-2 is formed. At larger distances the distributions diverge even further, and the RDFs for the Iβ/odd, Iβ/even, and triclinic phase are clearly distinguishable, indicating differing environments in each phase. This is most obvious when the target atom is a hydrogen, since the distributions of τ_2, τ_3, and τ_6 are different for Iα and Iβ cellulose. Assignment of the different features is not straightforward as the radial distribution function is isotropic. Hence intra- and inter-molecular contributions are highly mixed, as are intra- and inter-sheet contributions.

Differences between phases are most conveniently expressed by χ^2, the integral over space of the squared difference of the RDF in two phases for a given central atom–target atom pair. Atoms connected by a single bond were not considered in the χ^2 values. The average over all χ^2 values was largest when the Iβ/odd and Iβ/even phase were compared (typically 0.03 for C-1, C-4, or C-6), intermediate when Iβ/odd and Iα or Iβ/even and Iα were considered (0.01–0.015), and smallest for the average of Iβ/odd and Iβ/even phases compared to the triclinic phase (0.003–0.006). The order is insensitive to the precise cutoff radius used. This suggests that the environment of chains in the triclinic phase is similar to the average environment of chains in the monoclinic phase, but that the environment of chains in the two Iβ phases is very distinct. This result is not obvious from the radial distribution functions themselves.

Hydrogen bonding patterns.—Raman and IR spectra show different absorption peaks for the monoclinic and triclinic forms, especially in the OH stretching and bending regions [4–7]. From this it has been concluded that the hydrogen bonding patterns of the two allomorphs are slightly different. The hydrogen bonding characteristics of the three phases are different, the characteristics of the Iβ/odd phase being in between those of the Iβ/even phase and the Iα phase (Table 1). This could already be anticipated by the dihedral distribution of the τ_2, τ_3, and τ_6 dihedral angle.

The hydrogen bonds proposed for the static structure (intrachain O-2–HO-2 · · · O-6, O-3–HO-3 · · · O-5; intrasheet O-6–HO-6 · · · O-3) are observed most frequently, and are (as measured by r_{ha} and θ_{dha}) also the strongest. The occurrence of these hydrogen bonds is highest in the Iβ/even phase and smallest in the Iα phase, except for the intramolecular hydrogen bond O-3–HO-3 · · · O-5, which is most frequent in the Iβ/odd phase. For the intrasheet hydrogen bond O-6–HO-6 · · · O-3, the adjacent oxygen O-2 can serve as an acceptor as well, especially in the Iβ/even phase. In both the Iβ/odd and triclinic phase the intramolecular hydroxyl groups O-2 and O-3 may serve as intra-sheet donors as well. Intersheet hydrogen bond donors are observed in the Iβ/odd and triclinic phase only, but their occurrence is always smaller than 0.05, and they are energetically weak. The occurrences of intra- and inter-sheet hydrogen bonds in the three phases correlate well with the respective Coulomb energies (Table 2), emphasizing the importance of the intraplanar electrostatic interaction for the relative stability of the monoclinic phase over the triclinic phase.

Quality of the force field.—The total energy of the system during the simulation was larger than zero (Table 3). This was mainly due to electrostatic repulsion within the pyranose moiety. The GROMOS87 force field uses partial charges to represent electrostatic interactions, with exclusion of 1–2 and 1–3 interactions only. The distance between atoms at opposing sites of the pyranose ring (being 1–4 interactions) is

Table 1
Hydrogen bonds [a] observed during the simulation with occurrence > 1%

Donor–acceptor	Phase	Occurence	r_{ha} (nm)	θ_{dha} (°)
Intrachain				
O-2–HO-2···O-6 *	Even	0.951	0.190	160.6
	Odd	0.801	0.182	160.1
	Tricl	0.637	0.187	160.9
O-3–HO-3···O-5 *	Even	0.733	0.188	153.5
	Odd	0.883	0.185	157.3
	Tricl	0.634	0.189	154.3
O-6–HO-6···O-2	Odd	0.105	0.194	157.6
	Tricl	0.202	0.192	158.8
O-6–HO-6···O-3	Odd	0.011	0.215	151.5
	Tricl	0.051	0.216	153.8
Intrasheet				
O-6–HO-6···O-3 *	Even	0.874	0.184	151.8
	Odd	0.712	0.184	154.2
	Tricl	0.503	0.187	152.3
O-6–HO-6···O-2	Even	0.147	0.227	142.8
	Odd	0.087	0.224	141.6
	Tricl	0.106	0.223	143.9
O-2–HO-2···O-6	Odd	0.130	0.204	156.5
	Tricl	0.176	0.200	154.5
O-3–HO-3···O-6	Odd	0.034	0.190	154.4
	Tricl	0.122	0.197	150.9
Intersheet				
O-2–HO-2···O-3	Odd/Even	0.011	0.201	158.7
O-3–HO-3···O-6	Tricl	0.010	0.207	153.0
O-6–HO-6···O-2	Odd/Even	0.028	0.193	149.7
O-6–HO-6···O-3	Tricl	0.024	0.200	154.8
O-6–HO-6···O-4	Odd/Even	0.011	0.202	155.9
	Tricl	0.038	0.213	154.7
O-6–HO-6···O-5	Tricl	0.019	0.215	145.3
O-6–HO-6···O-6	Tricl	0.014	0.192	158.4

[a] Hydrogen bonds marked by an * are proposed for the crystalline structure. A hydrogen bond exists when the distance between hydrogen and acceptor is $r_{ha} < 0.25$ nm, and the angle between donor, hydrogen and acceptor is $\theta_{dha} > 135°$. If two phases are mentioned, the first contains the donor, the second the acceptor.

approximately 0.3 nm, causing large repulsive energies. This high energy appears to be a deficiency in the force field for saccharides, but it had no notable influence on the conformation of the pyranose rings, probably because of intraring improper dihedral angles. Apparently it has to be considered as the "zero-point" energy of the force field used.

Another measure of the quality of the force field used is the internal pressure. The pressure was approximately −0.58 GPa, and converged within 200 ps. Pressures of this magnitude are often encountered in constant volume simulations of the solid state because of the low compressibility. The pressure in the monoclinic system [2] along the

[2] For computational reasons, the pressure was monitored in the monoclinic system only.

Table 2
Nonbonded interaction energy of one cellobiose unit, divided into contributions from the two nearest neighbor chains in the same (200) plane, and contributions from the four nearest neighbor chains in adjacent (200) planes

Energy contributions (kJ mol^{-1} cellobiose^{-1})	Iβ/odd	Iβ/even	Iα	Iβ − Iα [a]
van der Waals, intraplane	−35.2 (±0.2)	−38.2 (±0.4)	−40.1 (±1.2)	3.4
van der Waals, interplane	−290.9 (±3.9)	−290.9 (±2.9)	−285.1 (±2.9)	−5.8
Coulomb, intraplane	−138.3 (±2.1)	−143.4 (±0.6)	−107.7 (±14.3)	−33.1
Coulomb, interplane	−23.0 (±4.1)	−23.0 (±4.5)	−30.2 (±7.6)	7.2

[a] Energy of a cellobiose unit in the Iβ-phase is the average of the energy of a cellobiose unit in the Iβ/odd and Iβ/even phase. Averages are taken over 9 (Iβ/odd, Iβ/even) and 24 (Iα) chains. Compared to Table 1, energies are twice as large due to double counting. Neglect of long-range interactions causes values to differ slightly.

c-axis was −0.68 GPa. The tensile strength along the chain is 135 GPa [30], requiring an adjustment of 0.5% in this direction to obtain atmospheric pressure; this is well within the experimental error of the crystal structure. The pressure along the *a*- and *b*-axis was −0.74 and −0.42 GPa, respectively. The tensile strength along these axes is not known experimentally, but Kroon-Batenburg et al. [16] estimated the tensile strength due to the intramolecular hydrogen bonds to be approximately 65 GPa. Extrapolating this to the interplanar hydrogen bonds (which are of about equal strength, see below), and correcting for the larger surface, we deduce a tensile strength of ~ 30 GPa along the *b*-axis. Comparing intra- and inter-planar energies (Table 2), the tensile strength along the *a*-axis is probably even larger. Adapting the crystal size by −2.5% and −1.4% along the *a*- and *b*-axis, respectively, would give atmospheric pressure along these axes. The latter two values are upper limits. Ramie cellulose unit cells, as determined by X-ray diffraction by three independent groups and reviewed by French et al. [24], are systematically smaller for the *a*- and *c*-axis (−1.4% and −0.22% on average), and systematically larger for the *b*-axis (+0.44% on average) compared to the unit cell used here. The γ-angle is slightly smaller, 96.55° on average vs. 97.3° reported

Table 3
Temperature and system energies during the simulations of cellulose Iα and cellulose Iβ

Energy contributions (kJ mol^{-1} cellobiose^{-1})	Iβ	Iα	E(Iβ) − E(Iα)
Total energy	305.7	314.4	−8.7
Bonded	127.5	126.9	0.5
Nonbonded	178.2	187.4	−9.2
van der Waals, intramol.	−31.4	−33.1	1.7
Coulomb, intramol.	462.7	458.3	4.3
van der Waals, intermol.	−170.4	−169.0	−1.4
Coulomb, intermol.	−82.6	−68.9	−13.7
Temperature (K)	302.3	300.3	

for the monoclinic phase studied here. Sugiyama et al. [11] suggested that the discrepancy between their unit cell dimensions and those reported was due to the phase transition induced by heating the sample to 260°C in a water-vapor saturated environment (so-called hydrothermal treatment). What effect changes of the crystal dimensions would have on pressure is unclear, as the shear elastic moduli of cellulose are not known.

Despite the deficiencies pointed out ("zero-point" energy, high pressure), the force field has been shown to be reasonably reliable in prediction of structures, hydrogen bonding patterns, and relative energies of saccharide conformations. For this reason we feel that the results presented are representative for the monoclinic and triclinic phases [17,25,30–34].

Energetics of cellulose Iα and Iβ.—The difference in energy between the triclinic and monoclinic form clearly favors the latter, consistent with hydrothermal annealing experiments (Table 3). The increased stability of the Iβ-phase is mainly due to intermolecular electrostatic interactions, which are to some extent compensated by intramolecular electrostatic interactions [$E(\mathrm{I}\beta) - E(\mathrm{I}\alpha) = 9.4$ kJ mol^{-1} cellobiose^{-1}]. The net effect of the van der Waals contributions is negligible [$E(\mathrm{I}\beta) - E(\mathrm{I}\alpha) = +0.3$ kJ mol^{-1} cellobiose^{-1}]. Since the two chains in the monoclinic phase are not equivalent, the total intermolecular energy per chain is also listed (Table 2). In three out of four cases the energy of the Iβ/even chain is lower than of the Iβ/odd chains. Considering interplanar Coulomb interactions only, the Iβ/odd chains are more stable than the Iβ/even planes. This is most likely due to the rotation of the chains in the Iβ/odd plane relative to the (200) plane. The stability of each allomorph in the direction of the (200) plane is due to electrostatic interactions, and in the direction normal to the (200) plane it is due to van der Waals interactions. The stability between (200) planes is marginally better in the Iα phase. The less favorable van der Waals interactions are fully compensated for by electrostatic contributions, probably due to more intersheet hydrogen bonds in this allomorph. The main contribution to the increased stability of the Iβ phase relative to the Iα phase is due to electrostatic interactions in the even (200) plane. The glucose residues in this sheet are better aligned to form more favorable hydrogen bonds than in the odd sheet or the triclinic phase.

Recently, Aabloo and French [35] estimated the energy difference between the Iβ and Iα phases by energy minimization using the MM3 force field. They also found cellulose Iβ to be more stable than cellulose Iα, but the energy difference was much smaller, 0.9 kJ mol^{-1} cellobiose^{-1}. This figure is too small to account for the conversion of cellulose Iα into cellulose Iβ upon hydrothermal annealing. The Boltzmann factor for this energy difference yields a ratio of ~ 1.4 at 300 K in favor of the monoclinic phase, whereas the energy difference found in our simulation gives a ratio of ~ 32. Since the free volume of the triclinic phase is lower than that of the monoclinic phase, and the distribution of the most flexible dihedrals is wider in the triclinic phase than in the monoclinic phase, entropy terms are likely to reduce the free energy difference. An earlier molecular mechanics study of cellulose in the so-called miniature crystal environment suggested that the energy difference between Iα and Iβ cellulose is due to van der Waals interactions [21]. The present model of a complete crystal instead attributes the increased stability of cellulose Iβ to intermolecular Coulomb interactions.

Correlation between structure and NMR chemical shifts.—The chemical shift of an atom in NMR spectra is a consequence of its chemical environment. The strongest influence is through covalent interactions (bonds, angles), but nonbonded interactions play a role as well.

For native *Valonia* cellulose (60–40% mixture of cellulose Iα/Iβ), three distinct NMR resonances are observed for C-1 and C-4, and two for C-6. For annealed *Valonia* cellulose and for tunicin cellulose (cellulose Iβ), two resonances are observed for all three atoms, though for C-4 a residual resonance of unknown origin is observed. The assignment order of the resonances for native *Valonia* cellulose is (increasing chemical shift) C-6:Iβ,Iα,Iβ, C-4:Iβ,Iβ,Iα, and C-1:Iβ,Iα,Iβ. The largest difference in shift is observed for C-1, the smallest for C-6. Though most authors interpret the splittings in terms of spin–spin couplings [2,3,36], comparison of the chemical shifts at 50 MHz [2,3,8,9] and 75 MHz [36] shows that shifts are independent of field strength. Hence the resonances have to be interpreted as true chemical shifts.

In an attempt to explain the shift differences for C-1, C-4, and C-6 in the different phases, we calculated the bond angles θ of the average structure. The values were very similar, and close to the energy minima. The differences between the averages for different phases were larger than the sum of the respective standard deviations, indicating that differences are significant. However, a straightforward correlation was not found; for all three atoms the order of the θ-values was Iβ,Iα,Iβ, which is consistent with the order of the NMR resonances for C-1 and C-6, but not for C-4. Also the magnitude of the $\Delta\theta$ did not correlate with the chemical shift differences for a particular atom.

Horii and co-workers observed a linear correlation between the ^{13}C chemical shift of C-1 and φ (H-1–C-1–O-4–C-4), C-4 and ψ (C-1–O-4–C-4–H-4) [3], and C-6 and χ (C-4–C-5–C-6–O-6) based on shifts and dihedral angles for various celluloses and related small sugars [36]. Given the high transition rate of the ω-dihedral, C-6 resonances are in fast exchange. Based on the distribution of states of the angle χ and the linear regression by Horii et al. [37], all calculated shifts are close to the experimental ones, especially for the triclinic phase. Basing the calculated shift on a sine function, to account for dihedral periodicity, improves the agreement with experimental shifts further. For C-6, the relative order of the calculated shifts (Table 4) is not in agreement with experiment as the chemical shifts for the monoclinic phases are overestimated by 0.2–0.4 ppm. The two low-shift *gg* and *gt* states have very low occupancy (0.002 for Iβ/even, 0.08 for Iβ/odd), which makes these states prone to sampling errors. An underestimation of the occupancy of these states of only 0.25% would cause a downshift of 0.3 ppm. Given the accuracy of the fit (0.36 ppm, no error in the distribution), agreement with experiment is surprisingly good, and the ^{13}C C-6 chemical shift can be readily accounted for with a sinusoidal relation between chemical shift and dihedral angle. It also indicates that the *gg* and *gt* states are occupied, although rarely.

[3] As a united atom force field was used, φ was taken as the average of the dihedrals O-5–C-1–O-4–C-4 and C-2–C-1–O-4–C-4; ψ was taken as the average of the dihedrals C-1–O-4–C-4–C-5 and C-1–O-4–C-4–C-3.

Table 4
^{13}C Chemical shifts [a] for selected atoms

Atom	Phase	Calcd shift (ppm)	Exptl shift (ppm)
C-6	Iα	65.65	65.7 [b]
	iβ/odd	65.86	65.5
	Iβ/even	66.21	66.0
C-4	Iβ/even	84.68	88.6
	Iβ/odd	85.35	89.4
	Iα	88.21	90.4
C-1	Iβ/odd	106.52	104.5
	Iα	106.65	105.6
	Iβ/even	106.79	106.2

[a] The calculated shifts are obtained using a linear combination of distributions and fitted lines. The assignment of experimental shifts to the two Iβ phases is based upon the calculated values. The observed shifts of the Iβ phase are those of tunicin cellulose, those of the Iα phase are for *Valonia ventricosa*.
[b] Value is interpolated assuming a 70% Iα, 30% Iβ mixture for *Valonia ventricosa*. Experimental data from Belton et al. [36].

The averages of φ and ψ are very similar in the three phases. Taking into account the distribution of φ yields the correct order for the C-1 shifts, and the most upfield calculated shift is close to the observed one (+0.6 ppm). The largest difference (2.0 ppm) is still within the uncertainty of the fit (2.2 ppm). The calculated shift difference between the two monoclinic phases (0.27 ± 0.1 ppm) is much smaller than the observed one (1.7 ppm), indicating that the dihedral effect contributes only a part of the observed shifts. Similarly calculated C-4 shifts are in correct order, but underestimated by 2.0 ppm (Iα phase) and 4.0 ppm (Iβ phase). The steep slope of the fit causes an uncertainty of approximately 3.7 ppm in the chemical shift due to ψ. The calculated shift difference between the two monoclinic phases (0.67 ppm) is close to the observed one (0.8 ppm), which indicates that the dihedral effect accounts for the chemical shift in the monoclinic phase.

Further contributions to the chemical shift arise from the nonbonded environment. The effect from packing and hydrogen environment on the ^{13}C chemical shift was estimated to be at the most 2.5 ppm [37]. The RDFs of the hydroxyl hydrogens relative to C-1 and C-4 show significant differences for the three phases at short distances (Fig. 3). For C-1–HO-2, a small, extra peak at 0.33 nm is found in the Iβ/odd and Iα phase. The height of the peak at 0.25 nm is lower for the Iα phase than for the respective Iβ phases. In addition, the height of the first peak of the C-1–HO-3 RDF varies among all three phases. The first peak of the C-4–HO-3 RDF varies in the triclinic phase and is also slightly lower than for each of the Iβ phases. This lack of density at 0.25 nm is compensated for by an increased density at 0.33 nm. The C-4–HO-3 RDFs are very alike the C-1–HO-2 RDFs. At larger distances, differences become less clear-cut, and are also of less importance to the chemical shift.

For C-1, the discrepancy between calculated and experimental shift differences is mainly due to different hydroxyl environments in the three phases. The underestimation of the calculated chemical shifts for C-4 may be due to a constant contribution of

covalent and nonbonded interactions only partly accounted for in the fit. The C-6 chemical shift is completely accounted for by the distribution of the dihedral angle χ.

The order of the C-1, C-4, and C-6 chemical shifts in the monoclinic and triclinic phases calculated from the distribution of the dihedral angles φ, ψ, and χ is consistent with experiment. It also permits an assignment of the monoclinic resonances in terms of Iβ/odd and Iβ/even phases. For all calculated shifts the difference between the Iβ/even and Iα shifts is always larger than the difference between Iβ/odd and Iα shifts.

Correlation between structure and IR and Raman spectra.—Blackwell et al. were able to correlate part of the IR and Raman spectra of cellulose to structural features, in particular the CH2 modes [38]. The main differences in IR spectra between cellulose Iα and Iβ are in the OH-stretching and bending regions [6,7]. Hence we correlate the hydrogen bonding patterns obtained with the fine structure of the IR spectrum in the OH-stretching region, considering the most important hydrogen bonds only (occurrence > 0.5, intrachain O-2–HO-2 · · · O-6, O-3–HO-3 · · · O-5; intrasheet O-6–HO-6 · · · O-3).

Formation of a hydrogen bond softens the donor-stretch [4] vibrational modes, and hardens the acceptor-stretch vibrational modes, causing a splitting of the absorbance. Modeling the hydroxyl groups by methanol at the 3-21G level in an ideal hydrogen bonding conformation ($r_{ha} = 0.180$ nm, $\theta_{dha} = 180°$), the donor stretch frequency decreases by approximately 5%, whereas the acceptor stretch frequency increases by approximately 1% relative to the gas-phase methanol OH-stretching frequency (3867.6 cm^1, uncorrected theoretical value). As the IR intensity of the acceptor-stretch modes is roughly 15–20 times smaller than that of the donor, only absorption of the donor-stretch frequencies will be observed. For pure Iβ cellulose, five different absorbances are observed experimentally; this agrees with the number of frequencies estimated from the hydrogen bonding patterns in the monoclinic phase (Table 5). The nine different hydrogen bonds observed for a mixture of Iα/Iβ cellulose generate seven different frequencies (Table 5), which correlate reasonably with the six absorbances observed experimentally (seven, if the highest frequency according to Michell [7] is different from the highest frequency by Sugiyama et al. [6]). However, the assignment of the two lowest OH-stretching frequencies in *Valonia* (3240 cm^{-1}, 3270 cm^{-1}) is Iα − Iβ [6]; the three lowest calculated frequencies are all Iβ. Also, the highest frequency (3415 cm^{-1}) in the FTIR spectrum of *Valonia* (Iβ/Iα mixture) can be assigned to the Iα phase, whereas the highest frequency in our list is assigned to the Iβ phase. Finally, the difference between the six calculated frequencies is much smaller than the difference between the six or seven experimental frequencies.

A possible source of error in the assignment of donor-stretch frequencies to particular hydrogen bonds is the constraining of bonds during the simulation. This overestimates the actual distance by up to 0.003 nm. In the case of the very stable hydrogen bond

[4] The OH-stretch frequency between an oxygen and a covalently bonded hydrogen (O-*n*–HO-*n*) will be referred to as the donor-stretch mode; the OH-stretch frequency between an oxygen and a hydrogen-bonded hydrogen atom (–HO-*m*· · · O-*n*) will be referred to as the acceptor-stretch mode.

Table 5
OH donor stretching frequencies at the 3-21G level of donors for the most frequently observed donor–acceptor pairs in cellulose. For completeness the characteristics of the hydrogen bonds are listed in the last three columns

Frequency (cm^{-1})	Donor–acceptor	Phase	Occurrence	r_{ha} (nm)	θ_{dha} (°)
3766.8	O-2–HO-2···O-6	Odd	0.801	0.182	160.1
3773.7	O-6–HO-6···O-3	Even	0.874	0.184	151.8
3773.8	O-6–HO-6···O-3	Odd	0.712	0.184	154.2
3781.4	O-3–HO-3···O-5	Odd	0.883	0.185	157.3
3786.7	O-6–HO-6···O-3	Tricl	0.503	0.187	152.3
3793.1	O-3–HO-3···O-5	Even	0.733	0.188	153.5
3793.3	O-2–HO-2···O-6	Tricl	0.637	0.187	160.9
3800.4	O-3–HO-3···O-5	Tricl	0.634	0.189	154.3
3806.9	O-2–HO-2···O-6	Even	0.951	0.190	160.6

O-2–HO-2 · · · O-6 in the Iβ/even phase, this may reduce the vibrational frequency by some 13–15 cm^{-1}. This would give better qualitative agreement for the Iα/Iβ mixture, but worse for the pure Iβ cellulose.

Several simplifications were made in the ab initio frequency calculations. Though the accuracy of the expansion is far from ideal, it is unlikely to be the main problem. All SCF ab initio Hartree–Fock methods are known to give a systematic overestimation of vibrational modes by 5–12%; a higher level approach will not solve the problem of the frequency differencies being too small, nor will it solve the problem of the order of the frequencies as inferred from the hydrogen bonding pattern. Most likely, the crucial simplification was made in modeling the hydrogen bonds as methanol–methanol and methanol–ether complexes in vacuo. Using density functional ab initio Hartree–Fock methods, Krijn and Feil [39] were able to reproduce the electron density of hydrogen bonds in crystalline α-oxalic acid dihydrate within experimental accuracy only when a large part of the electrostatic environment was included. Nonadditive effects, especially charge transfer between donor and acceptor, were found to be very important. The latter contribution is highly sensitive to the local electric field. Calculation of the electric field parallel to the hydrogen acceptor vector did not give any further clues. More accurate calculations to relate hydrogen bonding patterns to the fine structure of the IR spectra are beyond the scope of this paper.

4. Summary and conclusions

We have presented simulations of two allomorphs of native cellulose, cellulose Iα and cellulose Iβ in a crystalline environment. The GROMOS87 force field for saccharides shows some deficiencies, in particular a high potential energy due to electrostatic interactions between atoms on opposing sides of the pyranose rings. This has to be considered as a zero-point energy, and has no notable influence on the structure. The internal pressure is high, but is unlikely to affect structure because of the low compressibility of the material.

The important thermodynamic observation is that the Iβ phase is energetically favored over the Iα phase by -8.7 kJ mol^{-1} cellobiose^{-1}, consistent with an almost complete conversion of cellulose Iα into cellulose Iβ on hydrothermal treatment. Intraplanar electrostatic interactions, especially in the Iβ/even phase, account for most of the increased stability.

In the monoclinic phase the glucose residues in alternating (200) planes are not parallel, but form an angle of ca. 11.5°. Only Sarko [40] has noted this feature of the monoclinic phase before, but there the angle between the chains is only 3°. This substantially affects various properties such as energies, radial distribution functions, and hydrogen bonding patterns. The radial distribution functions for atoms in these chains differ most, whereas the RDFs for atoms in the monoclinic phase (averaged over both chains) and the triclinic phase are most similar. This suggests that, on conversion from cellulose Iα into Iβ, chains in alternate (200) planes in the Iα phase, apart from a translation, rotate about 10°. Hydroxyl hydrogens relative to C-1, C-4, or C-6 atoms in the respective phases display, especially at shorter distances, the largest differences in radial distribution function. Dihedrals involving these atoms fluctuate between various states. In conjunction with the relative rotation of the chains, this causes rather different chemical environments for the atoms with assigned NMR resonances in the different chains.

Qualitative agreement between structural features and spectroscopic data was found. In the case of IR spectra, the number of frequencies calculated on the basis of the most stable hydrogen bonds was in agreement with the number of experimentally determined frequencies, but the order of assignment was not consistent with experimental data. For a good quantitative agreement the approximations used were apparently too crude.

In the case of NMR spectra, the splitting of resonances in the monoclinic phase can be accounted for qualitatively by realizing that chains in the monoclinic phase are not equivalent from a structural point of view. The average of the dihedral angles φ, ψ, and χ does not correlate well with the observed chemical shifts in both allomorphs; one has to consider the complete dihedral distribution. With a sinusoidal relation between dihedral angle χ and chemical shift, the distribution of χ gave excellent agreement between the calculated and experimental shift of C-6 in all phases. A linear relation between dihedral angle and chemical shift gave the right order of shifts for the C-1 and C-4 resonances. The calculated C-1 shift differences based on dihedral angles are much smaller than the observed ones, but for C-1 the differences in hydroxyl environment are also much larger than for C-4 or C-6.

Dihedral distributions, hydrogen bonding patterns, energies, and calculated NMR shifts indicate that the Iβ/odd phase is more similar to the triclinic phase than is the Iβ/even phase.

Acknowledgements

The authors thank Mr A. Yliniemelä for calculating the vibrational frequencies. We thank Professor T. Drakenberg and Mr L. Kuutti for stimulating discussions.

References

[1] E. Sjöström, *Wood Chemistry Fundamentals and Applications*. Academic, New York, 1981.
[2] R.H. Atalla and D.L. vanderHart, *Science*, 223 (1984) 283–285.
[3] D.L. vanderHart and R.H. Atalla, *ACS Symp. Ser.*, 340 (1987) 88–118.
[4] J.H. Wiley and R.H. Atalla, *ACS Symp. Ser.*, 340 (1987) 151–168.
[5] A.J. Michell, *Carbohydr. Res.*, 173 (1988) 185–195.
[6] J. Sugiyama, J. Persson, and H. Chanzy, *Macromolecules*, 24 (1991) 2461–2466.
[7] A.J. Michell, *Carbohydr. Res.*, 241 (1993) 47–54.
[8] F. Horii, H. Yamamoto, R. Tyozo Kitamurz, M. Tanahashi, and T. Higushi, *Macromolecules*, 20 (1987) 2946–2949.
[9] H. Yamamoto, F. Horii, and H. Odani, *Macromolecules*, 22 (1989) 4132–4134.
[10] J. Sugiyama, T. Okana, H. Yamamoto, and F. Horii, *Macromolecules*, 23 (1990) 3198–3200.
[11] J. Sugiyama, R. Vuong, and H. Chanzy, *Macromolecules*, 24 (1991) 4168–4175.
[12] A. Sarko and R. Muggli, *Macromolecules*, 7 (1974) 486–494.
[13] K.H. Meyer and L. Misch, *Helv. Chim. Acta*, 20 (1937) 232–244.
[14] G. Honjo and M. Watanabe, *Nature (London)*, 181 (1958) 326–328.
[15] L. Kuutti, J. Peltonen, J. Pere, and O. Teleman, *J. Microsc.*, 178 (1995) 1–6.
[16] L.M.J. Kroon-Batenburg, J. Kroon, and M.G. Nordholt, *Polym. Commun.*, 27 (1986) 290–292.
[17] L.M.J. Kroon-Batenburg and J. Kroon, *Biopolymers*, 29 (1990) 1243–1248.
[18] C.J. Cramer and D.G. Thruhlar, *J. Am. Chem. Soc.*, 115 (1993) 5745–5753.
[19] B.J. Hardy and A. Sarko, *Proceedings of the Cellucon Conference*, New Orleans, 1993, pp 41–50.
[20] A.D. French and M.K. Dowd, *Proceedings of the Cellucon Conference*, New Orleans, 1993, pp 51–56.
[21] A.D. French, D.P. Miller, and A. Aabloo, *Int. J. Biol. Macromol.*, 15 (1993) 30–36.
[22] L.M.J. Kroon-Batenburg and J. Kroon, *Proceedings of the Cellucon Conference*, Lund, 1993, A8.
[23] K.H. Gardner and J. Blackwell, *Biopolymers*, 13 (1974) 1975–2001.
[24] A.F. French, W.A. Roughead, and D.P. Miller, *ACS Symp. Ser.*, 340 (1987) 15–37.
[25] J.E.H. Koehler, W. Saenger, and W.F. van Gunsteren, *Eur. Biophys. J.*, 15 (1987) 197–210.
[26] W.F. van Gunsteren and H.J.C. Berendsen, GROningen MOlecular Simulation (GROMOS) Library Manual, Biomos, Nijenborgh 4, Groningen, The Netherlands.
[27] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, and J.R. Haak, *J. Chem. Phys.*, 81 (1984) 3684–3690.
[28] J.-P. Ryckaert, G. Ciccotti, and H.J.C. Berendsen, *J. Comput. Phys.*, 23 (1977) 327–341.
[29] W.F. van Gunsteren and H.J.C. Berendsen, *Mol. Phys.*, 34 (1977) 1311–1327.
[30] I. Sakurada, Y. Nukushina, and T. Ito, *Makromol. Chem.*, 75 (1964) 1–10.
[31] J.E.H. Koehler, W. Saenger, and W.F. van Gunsteren, *Eur. Biophys. J.*, 16 (1988) 153–168.
[32] J.E.H. Koehler, W. Saenger, and W.F. van Gunsteren, *J. Biomol. Struct. Dynamics*, 6 (1988) 181–198.
[33] B.R. Leeflang, J.F.G. Vliegenthart, L.M.J. Kroon-Batenburg, B.P. van Eijck, and J. Kroon, *Carbohydr. Res.*, 230 (1992) 41–61.
[34] B.P. van Eijck, R.W.W. van Hooft, and J. Kroon, *J. Phys. Chem.*, 97 (1993) 12093–12099.
[35] A. Aabloo and A.D. French, *Macromol. Theory Sim.*, 3 (1994) 185–191.
[36] P.S. Belton, S.F. Tanner, N. Cartier, and H. Chanzy, *Macromolecules*, 22 (1989) 1615–1617.
[37] F. Horii, A. Hirai, and R. Kitamaru, *ACS Symp. Ser.*, 340 (1987) 119–134.
[38] J. Blackwell, P.D. Vasko, and J.L. Koenig, *J. Appl. Phys.*, 41 (1970) 4375–4379.
[39] M.P.C.M. Krijn and D. Feil, *J. Chem. Phys.*, 89 (1988) 4199–4208.
[40] A. Sarko, *Appl. Polym. Symp.*, 28 (1976) 729–742.