ELSEVIER

Carbohydrate Research 306 (1998) 205–220

CARBOHYDRATE RESEARCH

Comparison of the interface between water and four surfaces of native crystalline cellulose by molecular dynamics simulations

Andreas P. Heiner [a,*], Lauri Kuutti [b], Olle Teleman [a]

[a] *VTT Biotechnology and Food Research, POB 1500, FIN-02044 Espoo, Finland*
[b] *VTT Chemical Technology, POB 1400, FIN-02044 Espoo, Finland*

Received 14 July 1997; accepted 11 October 1997

Abstract

Molecular Dynamic (MD) simulations were performed for the four surfaces of native crystalline cellulose. In all cases, only the topmost surface layer of the crystalline cellulose is structurally affected by the water outside the surface. Except for the glucose orientation repeat symmetry, the monoclinic 110 surface and the triclinic 010 surfaces are very similar. Likewise, the monoclinic 1-10 surface is very similar to the triclinic 100 surface. The two latter surfaces are denser and were found to be more hydrophilic than the two former. All surface layer molecules are equivalent in the monoclinic 110 and triclinic 010 surfaces, i.e., the odd/even duplicity breaks down for the monoclinic 110 surface. On the other hand, alternate molecules have different geometric and energetic properties in the monoclinic 1-10 and triclinic 100 surfaces, such that solvation of the triclinic 100 surface induces a translational asymmetry reminiscent of the monoclinic form. The results are discussed with respect to electron microscopy, scanning force microscopy, solid state NMR and protein binding data.

Keywords: Cellulose; Surface structure; Molecular Dynamics; Solvent-induced changes; Hydrophilicity

1. Introduction

Cellulose research has undergone a major revival in recent years. Improved experimental procedures and theoretical methods have shown natural crystalline cellulose to consist of two allomorphs, the triclinic cellulose Iα containing one cellobiose moiety per unit cell, and the monoclinic cellulose Iβ containing two non-equivalent cellobiose moieties per unit cell [1–3]. The triclinic form dominates in lower organisms such as bacteria and algae, whereas the monoclinic form, which is the more thermostable one, dominates in higher organisms such as plants [4–7]. Thus the monoclinic form is the industrially most relevant form, especially in the textile and pulp and paper industries. The increasing demand for biodegradable products from renewable resources has brought renewed attention to the cellulose research. This includes modern biotechnology approaches such as enzyme-aided fibre processing and bleaching, or

* Corresponding author. Tel.: +358-9-456-51-05; fax: +358-9-455-21-03; e-mail: andrepeter.heiner@vtt.fi.

0008-6215/98/$19.00

chimerae which immobilize an enzymic activity onto cellulose as a carrier material.

All biologically and technically relevant processes involving cellulose occur at the interface between cellulose and water. Only a few experimental studies have dealt with crystalline cellulose surfaces. Kuutti et al. were able to identify the monoclinic 1-10 surface from Scanning Force micrographs [8]. Larsson et al. tentatively assigned two surplus resonances of equal intensity near the C-4 resonance in CP-MAS ^{13}C NMR spectra to the two different monoclinic surfaces [7]. Which resonance belonged to what surface or from how many surface layers it originated could not be ascertained. Another question raised by these experiments is the structure of the so-called paracrystalline cellulose, which shows up as a broad resonance superimposed on the pure crystal resonances.

Solvent affects the conformation of small oligosaccharides. In vacuum the hydroxymethyl group in β-D-glucose invariably adopts the tg-conformation (t relative to O-5–C-5–C-6–O-6, g relative to C-4–C-5–C-6–O-6), whereas this conformation is absent in polar liquids [9]. In methyl β-cellobioside a strong intramolecular hydrogen bond O-3–H-3 $\cdots$ O-5 is found both in vacuum and in non-polar solvents, but not in polar liquids like pure water or methanol–water mixtures [10]. The tg-conformation for the hydroxymethyl group and the O-2–H-2 $\cdots$ O-6 hydrogen bond are characteristic of the cellulose I structure.

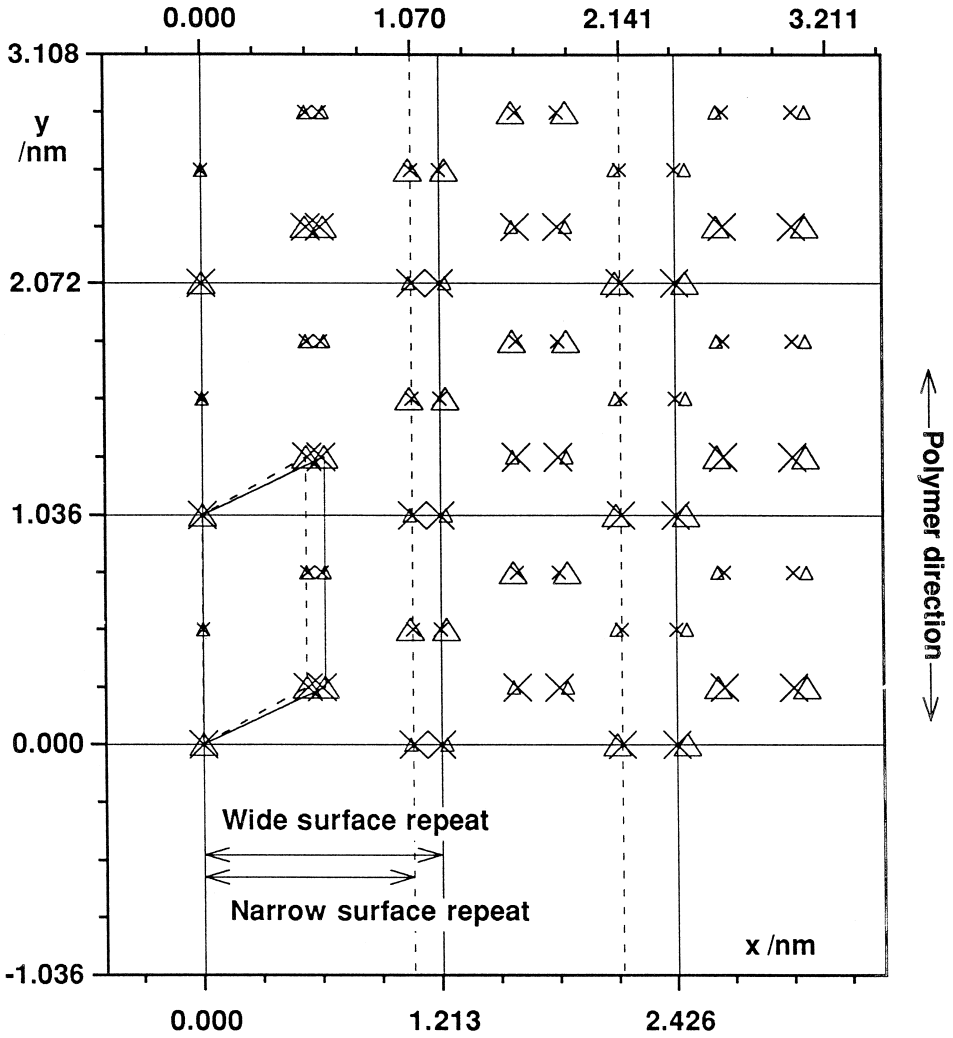

Fig. 1. Schematic representation of the surfaces en face. Crosses denote the two monoclinic surfaces' glucose positions and triangles the two triclinic surfaces' glucose positions. A larger mark denotes a C-6-outward glucose and a smaller a C-6-inward glucose. The cellulose polymers run from top to bottom in the picture, which also shows the narrow (1.070 nm) and the wide (1.213 nm) repeat distance. Two out of four polymers are out phase regarding C-6-outward or -inward between a monoclinic and triclinic surface, otherwise the similarity is striking.

Using molecular dynamics simulations, it is also possible to distinguish between cellulose Iα and Iβ, and between the two chains in the monoclinic cell, based on their distribution and population [11]. This is also true for the O-3–H-3 · · · O-5 hydrogen bond.

The orientation of hydroxyl groups is also important for the stability of carbohydrate complexes. Upon association, hydroxyl groups of either substrate or ligand may reorient, even when they are not directly involved in the recognition of the complex. The reorientation is thought to be due to solvation/desolvation processes, and therefore entropy-driven [12].

So far most theoretical work on cellulose has concerned itself with the structure of the bulk [11,13,14]. Theoretical studies of molecular surfaces have concentrated on model systems to assess hydrophobic–hydrophilic aspects [15], or self-assembled monolayers and membranes [16]. Only recently, we published the first paper on the interface between crystalline cellulose and water [17].

The cellulose surface was found to be very similar to the bulk except for the orientation of exposed hydroxymethyl and hydroxyl groups, which are less restrained or even freely rotating. The cellulose surface can thus be seen as a rigid surface with periodically placed flexible hydrogen-bonding groups. We expect that the extent of the hydrogen bonding, and therefore the hydrophilicity of the surface, depends on the precise distribution of the hydroxyl groups, and their ability to adjust to the solvent to maximise hydrogen bonding.

In this paper we present Molecular Dynamics simulations of the four surfaces of native crystalline cellulose (monoclinic 110 and 1-10 and triclinic 100 and 010) exposed to water. The analysis will concentrate on a qualitative description of the surface structure, with emphasis on the hydrophilic/hydrophobic nature of the surface and with solvent-induced structural changes.

2. Computational procedure

Although the two crystal forms are different, the glucose ring spacings are quite similar on the four surfaces (Fig. 1). A molecular graphics representation is found in Ref. [17]. Two surfaces, monoclinic 110 and triclinic 010, have a wider spacing between adjacent molecules. These two surfaces differ only in repeat of up vs. down orientation of the glucose moieties (C-6 outward or inward). The other two surfaces, monoclinic 1-10 and triclinic 100 have a narrower spacing between adjacent molecules but also differ in the repeat of the hydroxymethyl group orientation. We refer to these surfaces as Mwide, Twide, Mnarrow and Tnarrow.

The monoclinic 110 surface system (Mwide) was constructed as in Ref. [17]. Six cellulose layers, each consisting of six chains of three cellobiose units, were placed in the center of a monoclinic periodic box with dimensions $L_x = 3.639$ nm, $L_y = 3.108$ nm, $L_z = 3.211$ nm, with the normal of the monoclinic 110 surface parallel to the *z*-axis; the *c*-axis was chosen parallel to the *y*-axis (Fig. 2). The angle

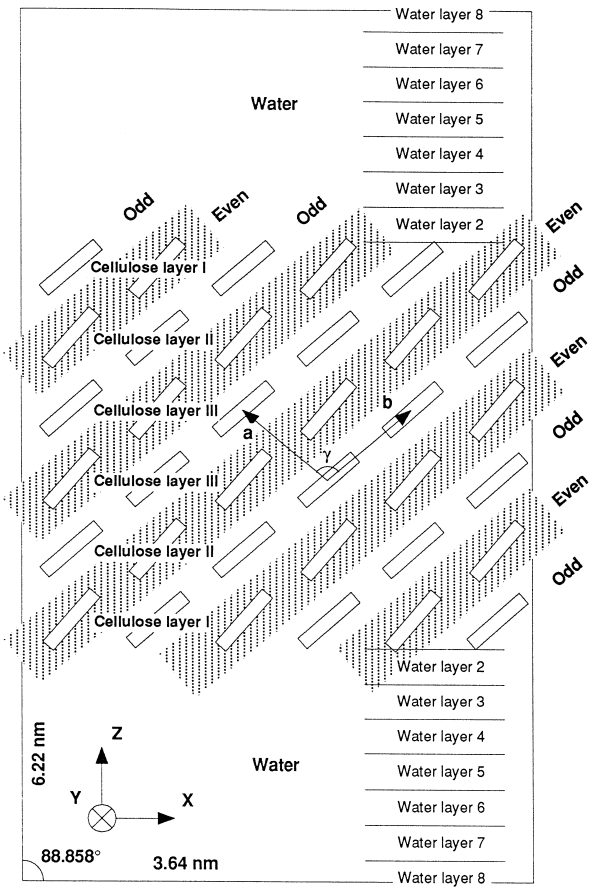

Fig. 2. Schematic placement of monoclinic crystalline cellulose in the periodic simulation box. Each cellulose layer has been numbered. All properties have been averaged over both layers, i.e., a surface layer property has been averaged over both layers I. In the triclinic systems, the upper surface (100 or 010) is the mirror image of the lower (-100 or 0-10), but this does not affect any geometric or energetic analysis, since water is not a chiral molecule. For analysis purposes, the water has been divided into slabs of 0.25 nm thickness. During the initial equilibration the box was shorter and only contained the cellulose. The numerical values in the figure are for the Mwide system.

between the L_x and L_z sides is then equal to the angle between the crystal 110 and 1-10 vectors, i.e., $\alpha_y = 88.858°$. The initial cellulose conformation was obtained by applying crystal symmetry operations on the average structure over 500 ps of a previous cellulose Iß simulation [11], based on electron diffraction data by Sugiyama et al. [3]. To prevent a bias towards the initial average crystal structure, side-group dihedrals of the glucose moieties were randomised followed by a short simulation using full periodic boundary conditions. L_z was then increased to 6.211 nm (cf. Fig. 2), and the additional space was filled with 1132 SPC/E water molecules [18], equivalent to a density of 0.998 g cm^{-3}. The complete system consists of 3024 cellulose (united) atoms and 3396 water atoms.

The monoclinic 1-10 (Mnarrow), triclinic 100 (Twide) and triclinic 010 (Tnarrow) surface systems were constructed in a similar fashion but in each case producing a slightly different total volume. In each case the amount of water was chosen to make the water slab at least 3 nm thick. All four systems contained 216 glucose rings, and in all systems the crystal *c*-axis was chosen parallel to the *y*-axis. The system properties have been summarised in Table 1.

The system was energy minimised and subsequently thermalized for 60 ps at 300 K while restraining the cellulose chains to their initial position using a force constant of 3000 kJ nm^{-2} during the first 10 ps. After releasing the position restraints the cellulose contracted by approximately 3% along the *z*-axis. The water density away from the interface was reduced slightly more due to the formation of hydration layers. Water molecules were added, and the *c*-axis was reduced linearly during the first 60 ps, to maintain an approximate water density of 0.97 g cm^{-3}. After the start-up procedure all systems were equilibrated for another 60 ps.

All systems were simulated in an NVT ensemble using the Gromos87 force field [19] using a locally adapted GROMOS program suit to allow for triclinic periodic boundary conditions. The temperature of the system was controlled using the Berendsen thermostat with coupling constants of 0.4 ps for cellulose and solvent separately [20]. A time step of 1.25 fs was employed, and bond lengths were constrained using the SHAKE algorithm with a relative tolerance of 10^{-8}. Non-bonded interactions were calculated using the twin-range procedure with cutoff radii of 0.9 nm and 1.1 nm. The charge group-based pairlist was updated every 10 steps. All simulations extended for 1.0 ns, and conformations were saved every 0.2 ps. The simulations were run on an IBM RS6000/390 and a SGI Indigo2 (R10000) station in 64 bit precision; each simulation took approximately 4 weeks of dedicated CPU time.

3. Results and discussion

The following terms will be used to simplify the discussion (Fig. 2). The cellulose layers II and III (the four central layers) will be referred to as bulk, while the simulations performed in the absence of solvent [11] will be referred to as 'crystal' simula-

Table 1
Summary of the simulated systems

System	Mwide	Mnarrow	Twide	Tnarrow
Lattice type	Monoclinic	Monoclinic	Triclinic	Triclinic
Crystal surface	110	1-10	010	100
Box dimensions				
L_x (nm)	3.639	3.211	4.044	3.558
L_y (nm)	3.108	3.108	3.108	3.108
L_z (nm)	6.110	6.54	6.825	7.14
α_x (nm)	90.0°	90.0°	63.0°	67.0°
α_y (nm)	88.86°	91.14°	99.0°	81.0°
α_z (nm)	90.0°	90.0°	113.0°	63.0°
Surface area (nm^2)	11.312	9.980	11.568	9.852
Number of water molecules	1032	999	1065	986
Water density (g cm^{-3})	0.968	0.964	0.971	0.975
Cellulose temperature (K)	302.3	302.8	302.4	302.4
Water temperature (K)	312.3	312.1	312.0	312.0
Total temperature (K)	307.6	307.5	307.6	307.2
Water potential energy (kJ mol^{-1})	−43.00	−42.95	−43.04	−43.04

tions. 'Odd' and 'Even' indicate alternate (200) planes in the monoclinic crystal, as defined in Ref. [17]. At the surface, glucose moieties have the hydroxymethyl group either exposed to the solvent (C-6-outward) or buried (C-6-inward). This also holds for the O-2- and O-3-hydroxy groups, which are exposed when C-6 is pointing inward, and vice versa.

Except where otherwise stated, energy values will be reported as kJ mol^{-1} meaning mol of water or mol of cellobiose as pertinent.

To ensure a valid comparison of the four surfaces it is necessary to establish that the four trajectories represent similar thermodynamic conditions. Bulk water temperatures, densities and energies are very similar in the four simulations ($T = 312.1 \pm 0.1$ K, $\rho = 0.970 \pm 0.004$ g cm^{-3}, $E = -43.01 \pm 0.04$ kJ mol^{-1}). Also the energy difference between the cores of the Mwide and Mnarrow systems (-1.7 kJ mol^{-1}) and the Twide and Tnarrow systems (0.92 kJ mol^{-1}) are relatively small. Bulk dihedral distributions, bulk pucker parameters and bulk hydrogen bonding patterns are similar in the two monoclinic systems and in the two triclinic systems. All these properties also agree very closely with the data obtained from pure crystal simulations [11]. In addition, the potential energy difference between the monoclinic and triclinic bulk phase (layers II and III) in the present simulations (-7.3 kJ mol^{-1}) agrees well with those reported for pure crystal simulations (-8.7 kJ mol^{-1} Ref. [11]). For these reasons we conclude that the systems have been simulated under essentially the same thermodynamic conditions, and that a valid comparison of the surfaces can be made. Only the present simulations and those of Ref. [11] reproduce the strong preference for the monoclinic crystal form observed experimentally.

The thickness of the cellulose slab decreased in all cases by 3.5% to 4.5%. This resulted in small differences between the bulk simulated and experimental crystal parameters (Table 2). The main decrease was observed normal to the 200 layers (the 110 layer in the triclinic systems). This is to be expected, since the chains are kept together by strong hydrogen bonds within these layers, whereas the packing normal to the 200 (triclinic 110) layers is mainly due to van der Waals interactions. A similar contraction has been observed for other saccharides [21–23] but did not cause any significant change in energetics, dynamics or hydrogen bonding. Furthermore, cellulose unit cell parameters vary slightly with the species of origin (Dr. J. Sugiyama, unpublished experiments).

Density profile.—The density profiles normal to the surfaces are shown in Fig. 3. In each cellulose layer the total cellulose density is largest in the middle of the layer. In cellulose layers in narrow surfaces two additional maxima on either side flank the central peak. In contrast, wide surfaces have only one additional maximum on either side. The profiles of the Mwide and Twide surfaces are very similar. The peak heights of the Mnarrow surface are larger than for the Tnarrow surface. For Mnarrow surfaces, the density peak closest to the central peak is lower than the one further away. For Tnarrow surfaces, this order is precisely reversed. The small differences between monoclinic and triclinic surfaces are to be expected, since the main difference between the two crystal forms is a translation of the chain normal to the z-direction.

On going from the bulk cellulose (layer III) to the interfacial cellulose (layer I), the heights of the (central) peak reduce. This effect is most pronounced for the Twide surface, and least for the Mnarrow surface. At the wide surfaces one can discriminate between 'inward' and 'outward' glycosidic oxygens, given the two peaks for this atom in each cellulose layer ($z = 1.17$ and $z = 1.38$ nm in layer I). Narrow surfaces have the glycosidic oxygen density at the same level ($z = 1.47$ nm in layer I). The hydroxyl group density at the narrow surfaces is more localised given the higher intensity (1.07 g cm^{-3} vs. 0.81 g cm^{-3}). Since hydroxyl groups are the most solvent-exposed groups, the cellulose density increases more slowly to its first peak for the wide surfaces than for the narrow surfaces. For other surfaces such as the dipalmitoyl phosphatidylcholine or decane bilayer system differences in $\delta\rho/\delta z$ have been associated with the hydrophilicity/hydrophobicity of the surface, but this is unlikely to apply to cellulose where the hydroxyl groups are bound to an almost rigid surface.

In all systems the solvent density quickly reaches (within one water diameter) its first maximum, which in all cases is rather small. Although such a profile is characteristic for hydrophobic surfaces [15], three-dimensional density plots (Fig. 4) reveal pronounced structure in the x and y directions with strong hydration maxima. The surfaces appear, in consequence, to be very well hydrated.

Solvent accessibility and radial distribution functions.—The number of water molecules in the first hydration layer (up to 0.36 nm from the cellulose) is approximately the same for all surfaces (12.8 nm^{-2}). Closer to the surface the number of water molecules is larger for the wide surface, i.e., these surface have

Table 2
Crystal parameters in the cellulose bulk

Crystal parameters		Monoclinic experiment	Mwide simulation	Mnarrow simulation	Triclinic experiment	Twide simulation	Tnarrow simulation
Vectors	a (nm)	0.801	0.773	0.764	0.674	0.674	0.657
	b (nm)	0.817	0.820	0.818	0.593	0.569	0.584
	c (nm)	1.036	1.036	1.036	1.036	1.036	1.036
Angles	α	90.00°	90.02°	90.01°	117.00°	116.60°	115.23°
	β	90.00°	90.01°	89.99°	113.00°	113.00°	114.71°
	γ	97.30°	99.09°	94.96°	81.00°	82.21°	83.13°
Spacings	d(110 or 100)[a]	0.535	0.516	0.534	0.528	0.509	0.527
	d(1-10 or 010)[a]	0.607	0.605	0.582	0.621	0.620	0.595
	d(200 or 110)[a]	0.820	0.820	0.818	0.822	0.822	0.825

[a]The two Miller indices refer to the monoclinic and triclinic unit cells.

a more open structure. Given the different interchain spacings this is to be expected.

A Connolly surface [24] was calculated to assess the solvent accessibility more accurately. The probe radius was 0.14 nm to mimic a water molecule. The ratio of the accessible surface A_{Conn} and the cross section A_{cross} gives an idea of the structure of a particular surface. The two wide surfaces are slightly more rough (A_{Conn}/A_{cross}), 1.24, than the narrow surfaces, 1.18. The small difference is not surprising, given the similar amount of water molecules in the first hydration shell. Accessibilities are very similar for the two wide surfaces, and for the two narrow. The hydroxymethyl group is always highly accessible. Larger differences in solvent exposure between the wide and narrow surfaces are observed for C-2, C-3 and C-5.

The radial distribution function of water molecules around the exposed hydroxyl groups is shown in Fig. 5. Due to the surface only half of the space is accessible to solvent, and the radial distribution function approaches 0.5 for large distances. For the monoclinic surfaces, the first hydration peak is very similar for all but the Mnarrow/O-3 distributions of the odd chain, where it is considerably lower. Differences in these functions between odd or even chains correlate with the probability of hydrogen bond formation of this group (see Section 3.6). The O-6 distribution is less affected since the O-6 protrudes into the water. At the Tnarrow surface moderate differences apply to the first hydration peak around the O-6–H and O-2–H hydroxyl groups from alternate molecules.

Energetics.—Cellulose molecules at the microfibril surface have a higher energy than in the bulk. For chains at monoclinic surfaces the energy increase is approximately 25 kJ mol^{-1} $cellobiose^{-1}$, for chains at the triclinic surface 19 kJ mol^{-1} $cellobiose^{-1}$. In all cases this is due to a loss of crystalline van der Waals interactions, not compensated by interactions with the solvent. In contrast to expectation alternate chains at the Tnarrow surface have an energy difference of 6.5 kJ mol^{-1} $cellobiose^{-1}$. The surface potential energies, defined as the difference between

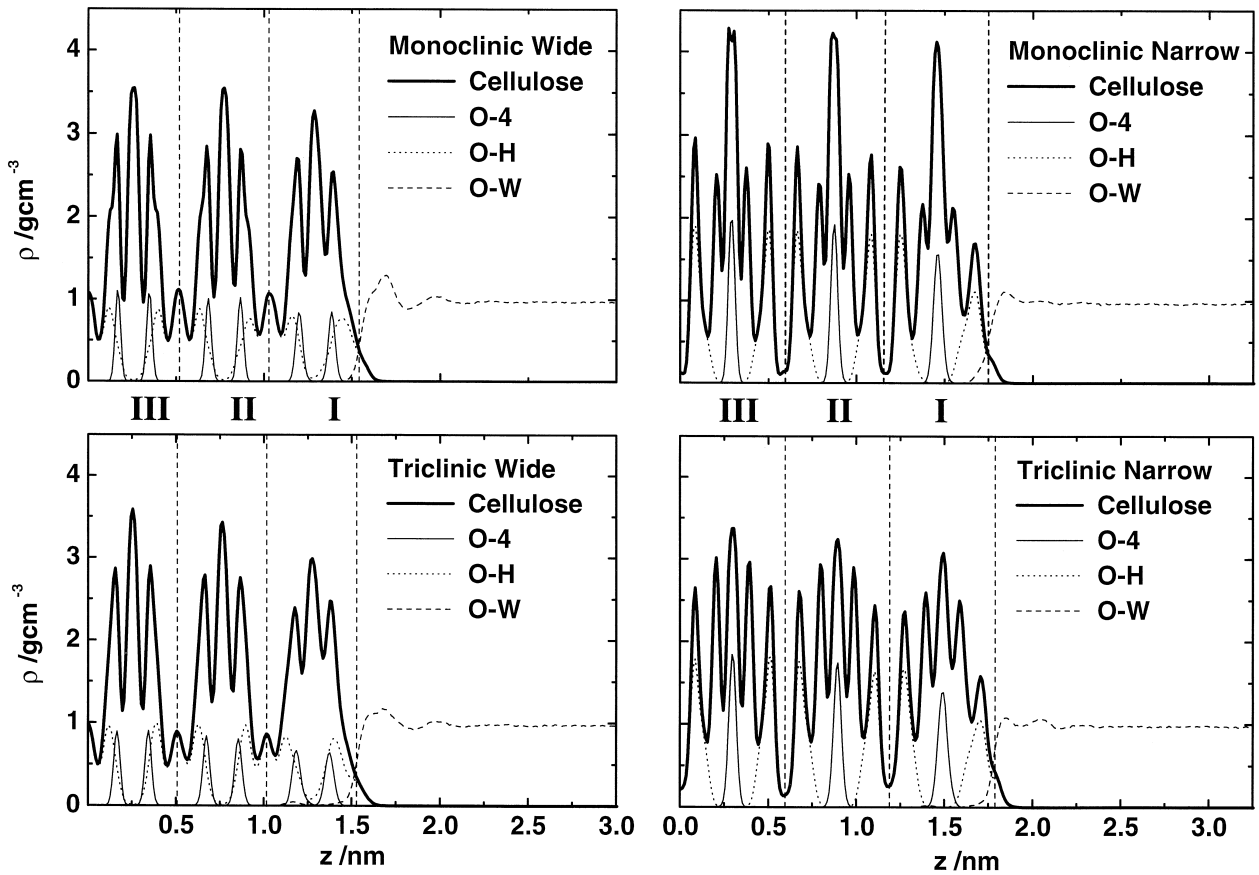

Fig. 3. Density profiles in g cm^{-3} as a function of z, i.e., normal to plane of the interface. O-4 is the glycosidic oxygen, OH refers to the O-2–H and O-3–H hydroxyl groups and OW to the water density. Cellulose is the entire cellulose density.

potential energy at the interface and the bulk per unit area, are given in Table 3.

The interaction energy between cellulose and solvent may give a first indication of which surface is more hydrophilic. The average interaction of a solvent molecule in the first hydration layer of the surface with cellulose is more favourable by 0.9 kJ mol^{-1} for the narrow surfaces than for the wide surfaces. The net cellulose–water interaction energy as well as the total surface potential energy is given in Table 3. The cellulose–water interaction energy favours the narrow surfaces, while the total surface energies seem to favour triclinic surfaces over monoclinic.

Molecular structure of surface chains.—Backbone torsion angles at the surface are within 5° of their bulk values, but distributions are generally wider (Table 4). Differences in φ are largest for odd chains in the narrow surfaces. Smaller differences are found in ψ values for odd chains in monoclinic surfaces and in the Tnarrow surface. For all torsions the average of C-6-outward and C-6-inward is similar to the bulk value. We note, that due to the twofold screw axis changes in φ/ψ have to compensate each other.

The distribution of pucker parameters (Fig. 6) is characteristic for the monoclinic odd/even and the triclinic cellulose chains [25,26]. The Θ distribution is slightly wider for solvent-exposed cellulose molecules. The distinction between odd and even chains remains at the Mnarrow surface. At the Mwide surface the even and odd chain conformations converge and both resemble the triclinic bulk conformation [17]. The Θ distributions of both triclinic sur-

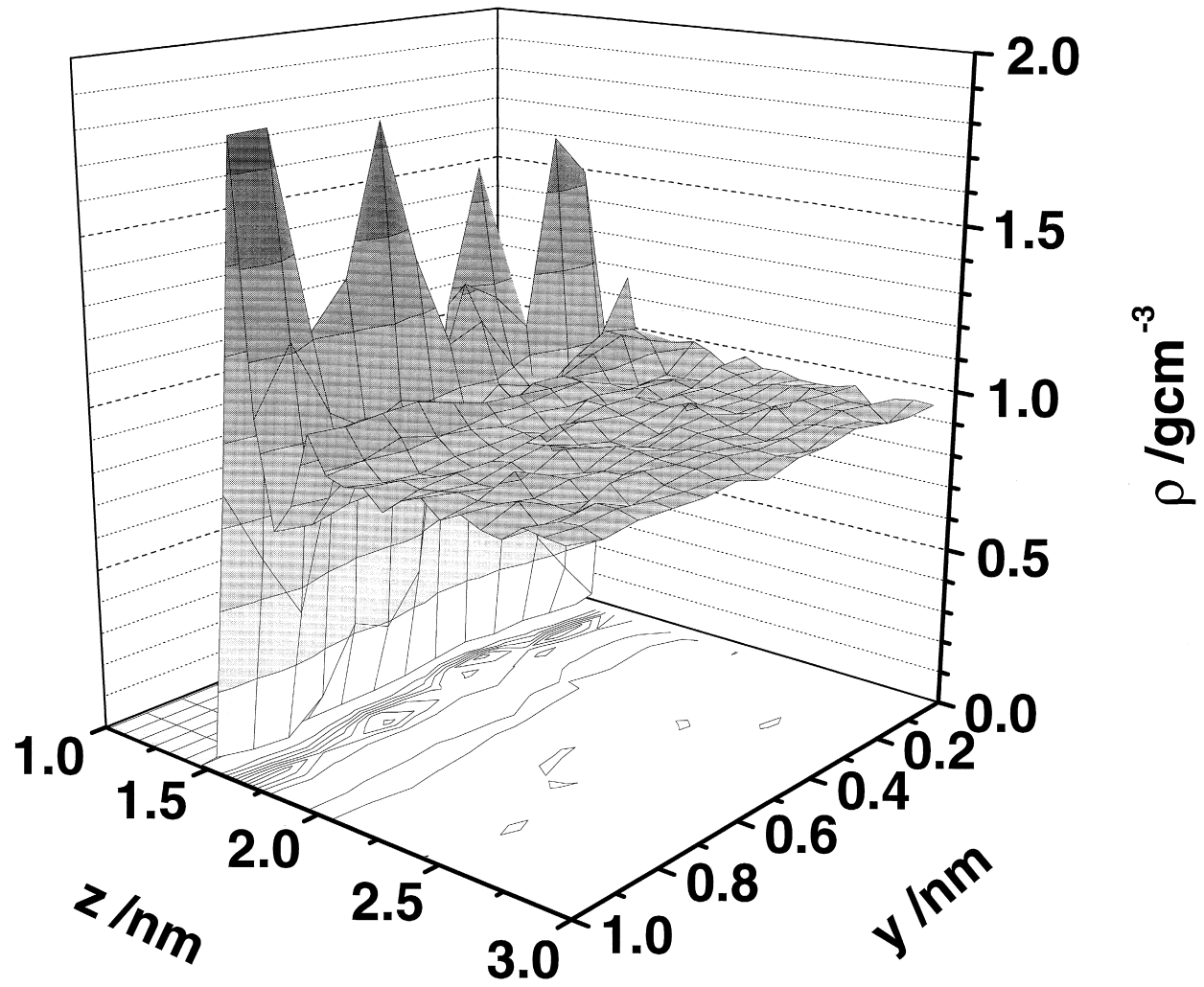

Fig. 4. Water density outside the Mwide surface as a function of z and y. The range in y corresponds to about one cellobiose moiety along the polymer (1.036 nm). Four pronounced hydration peaks ($y = 0.15$, 0.4, 0.7 and 0.9 nm) are resolved in the y direction.

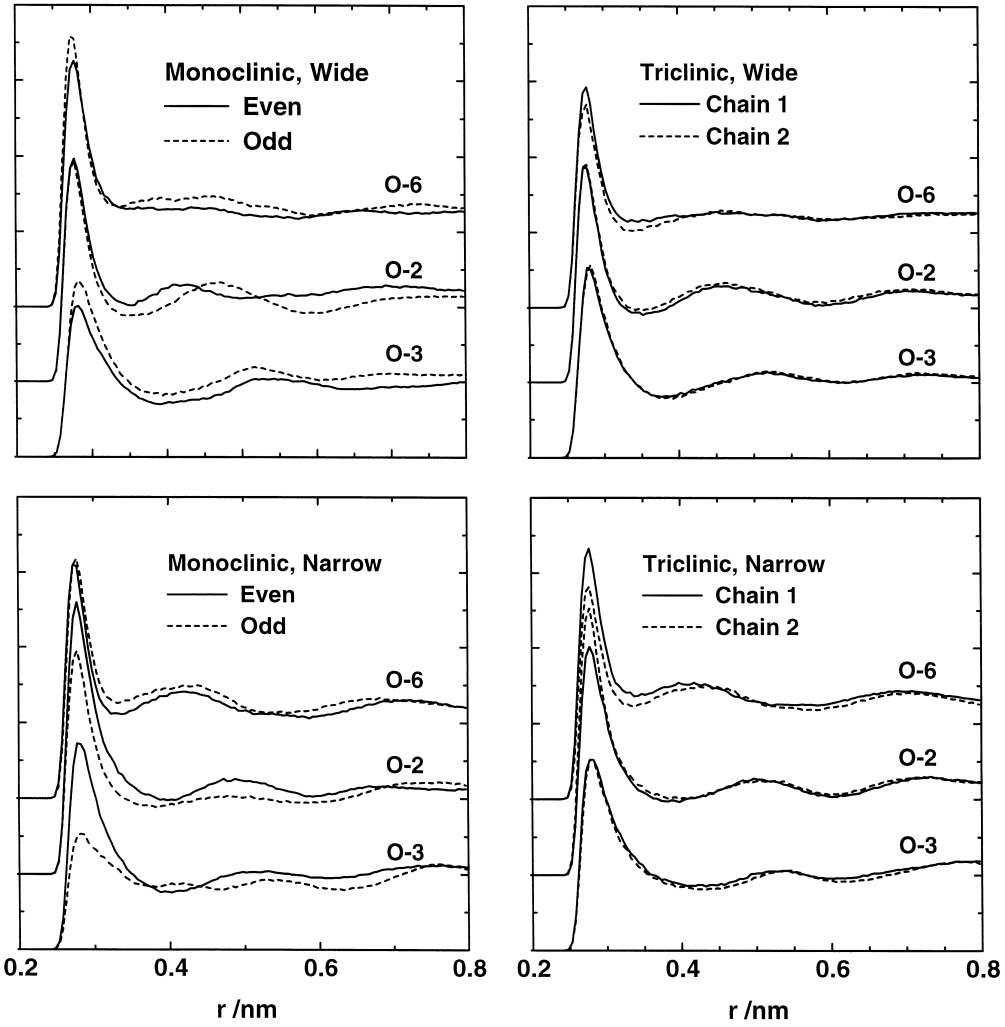

Fig. 5. Water radial distribution functions around exposed cellulose hydroxyl oxygens.

faces converge to the same distribution and is similar to the monoclinic odd distribution. The pucker amplitude distribution (Q_{ampl}) is wider at the surface, but the average values do not change significantly. The Φ distribution at the Mwide surface deviates considerably from that in the bulk, with the C-6-outward even and odd distributions converging towards a form resembling a triclinic Φ distribution. At the Mnarrow surface the even chain Φ distribution remains typical of the even subphase, the odd chain Φ distribution similarly remains typical of the odd subphase. The Φ distributions at triclinic surfaces are less sensitive to solvent exposure.

Conformation of exposed hydroxyl groups.—The distributions of the χ dihedral (C-4–C-5–C-6–O-6) and the three hydroxyl group torsion angles (τ_2: C-1–C-2–O-2–H-2; τ_3: C-2–C-3–O-3–H-3; τ_6: C-5–C-6–O-6–H-6) are shown in Fig. 7. In the crystal, the hydroxymethyl group is invariably in the tg conformation ($-60°$), but this preference disappears at the surfaces. The tg conformation often coincides with an intramolecular hydrogen bond. This also occurs here, where the population of the tg conformation correlates well with the occupancy of the intrachain hydrogen bond O-2–H-2 · · · O-6 or O-6–H-6 · · · O-2 at the surface.

Table 3
Excess surface energy[a]

	Mwide	Mnarrow	Twide	Tnarrow
Cellulose[b]	39.76	49.96	27.47	32.52
Solvent[b]	−33.72	−34.93	−36.96	−35.28
Total	6.04	15.03	−9.49	−2.77
Cellulose–water interaction	−117.26	−127.61	−124.26	−134.03

[a]Numbers listed are energy differences in kJ nm^{-2} between interface and bulk for cellulose and solvent.
[b]Includes the cellulose–water interaction, which has been attributed with 50% to each interagent.

The population maximum in the distribution near $\tau_2 = 60°$ disappears, and the trans-conformation becomes dominant. For the Mnarrow surface a third maximum near $-60°$ appears, occupied at the narrow surfaces but not for the wide. For hydroxyl groups in the Mnarrow even chains the distribution near the gauche + conformation is so wide that it no longer is a separate conformer.

The τ_3 trans-conformation occupancy decreases on solvent exposure but remains dominant in all cases. For all surfaces, the population of this rotamer correlates well with the fraction of O-3–H-3 · · · O-5 hydrogen bonds. Gauche + is somewhat populated at wide surfaces, but there is almost always more gauche − torsion angles.

The strong preference for the τ_6 trans-conformation in the crystal disappears completely on solvent exposure. All conformations become widely populated with a mild preference for gauche + at narrow surfaces. High population of intermediate rotamers show that rotation barriers are low.

Clear differences between even and odd torsion angle distributions exist for all four torsions at the Mnarrow surface but only for τ_6 and χ at the Mwide surface. Somewhat surprisingly alternate molecules in the Tnarrow simulation showed significantly different torsion angle distributions, particularly for τ_2 and τ_3.

Hydrogen bonds.—Table 5 gives the effect of solvent exposure on the three hydrogen bonds observed in the crystal. The strength of a formed hydrogen bond, as reflected by hydrogen-acceptor distance and donor–hydrogen-acceptor angle, is typically invariant. The amount of cellulose–solvent hydrogen bonding differs considerably between alternate molecules at the Tnarrow surface, but not at the Twide surface. The occupancy of inward oriented hydrogen bonds is similar to the occupancy in the bulk, except for the inward-oriented O-3–H-3 · · · O-5 hydrogen bond at the Twide surface which is less populated. When solvent exposed, the intra-chain hydrogen bond O-2–H-2 · · · O-6 is better preserved at narrow surfaces than at the wide surfaces. The opposite is true for the intra-chain O-3–H-3 · · · O-5 hydrogen bond, which is generally better preserved at wide than at narrow surfaces.

Hydrogen bonds between cellulose and water are listed in Table 6. The differences between even and odd molecules are large at the Mnarrow surface (e.g., O-3–H and O-5), but considerably smaller at the Mwide surface. Alternate molecules have the same hydrogen bonding at the Twide surface, but substantial variations arise at the Tnarrow surface (e.g., for O-2–H and O-6–H). Discrepancies typically coincide with discrepancies in torsion angle distribution but not always. The narrow surfaces show more hydrogen bonding for O-2–H and O-5, while O-4 hydration is better at the wide surfaces.

The number of water–cellulose hydrogen bonds is related to the surface hydrophilicity. The Mnarrow and Tnarrow surfaces have more hydrogen bonds with the solvent (7.18 nm^{-2} and 6.92 nm^{-2}) than the wide surfaces (Mwide 5.73 nm^{-2} and Twide 5.69 nm^{-2}). Dividing by the number of water molecules in the first hydration shell gives the probability for a water molecule to form solvent–solute hydrogen bonds. This probability is also larger for the narrow surfaces (Mnarrow and Tnarrow both 0.55) than at the wide (both 0.44). Estimated in this way the wide surfaces are some 25% more hydrophobic than the narrow.

Differences in hydration of small monosaccharides have been explained in terms of the relative hydroxyl positions on the ring [27]. The better the carbohydrate donor and acceptor groups fit into the water hydration structure, the more hydrated the compound is. Detailed computer simulations support this [28]. The

Table 4
Backbone torsion angles[a]

		Mwide		Mnarrow		Twide	Tnarrow
		Even	Odd	Even	Odd		
φ	C-6-outward	23.4° (8.2)	22.1° (8.8)	21.0° (8.1)	26.5° (7.8)	20.1° (10.1)	24.2° (9.9)
	C-6-inward	19.8° (8.1)	24.5° (8.0)	19.6° (7.9)	18.6° (8.3)	20.9° (11.6)	15.4° (10.7)
	Core	21.2° (6.5)	23.2° (6.7)	20.5° (6.5)	23.8° (7.0)	20.5° (8.7)	20.6° (8.3)
ψ	C-6-outward	−18.4° (7.6)	−15.2° (7.4)	−17.5° (7.3)	−23.4° (6.7)	−16.6° (10.0)	−21.2° (8.3)
	C-6-inward	−18.2° (7.2)	−22.5° (7.2)	−18.1° (7.2)	−16.8° (7.4)	−19.6° (9.0)	−15.3° (9.0)
	Core	−18.0° (6.2)	−19.2° (4.4)	−18.0° (6.1)	−19.5° (4.5)	−18.3° (8.1)	−18.1° (7.9)

[a]The dihedral φ is named C-6-outward if the C-1–H-1 atoms belong to a pyranosyl ring that has the hydroxymethyl group pointing into the solvent. Similarly, ψ is C-6-outward when the C-4–H-4 belongs to a glucose moiety that has the C-6 pointing into the solvent. Numbers in parentheses are widths of the distributions.

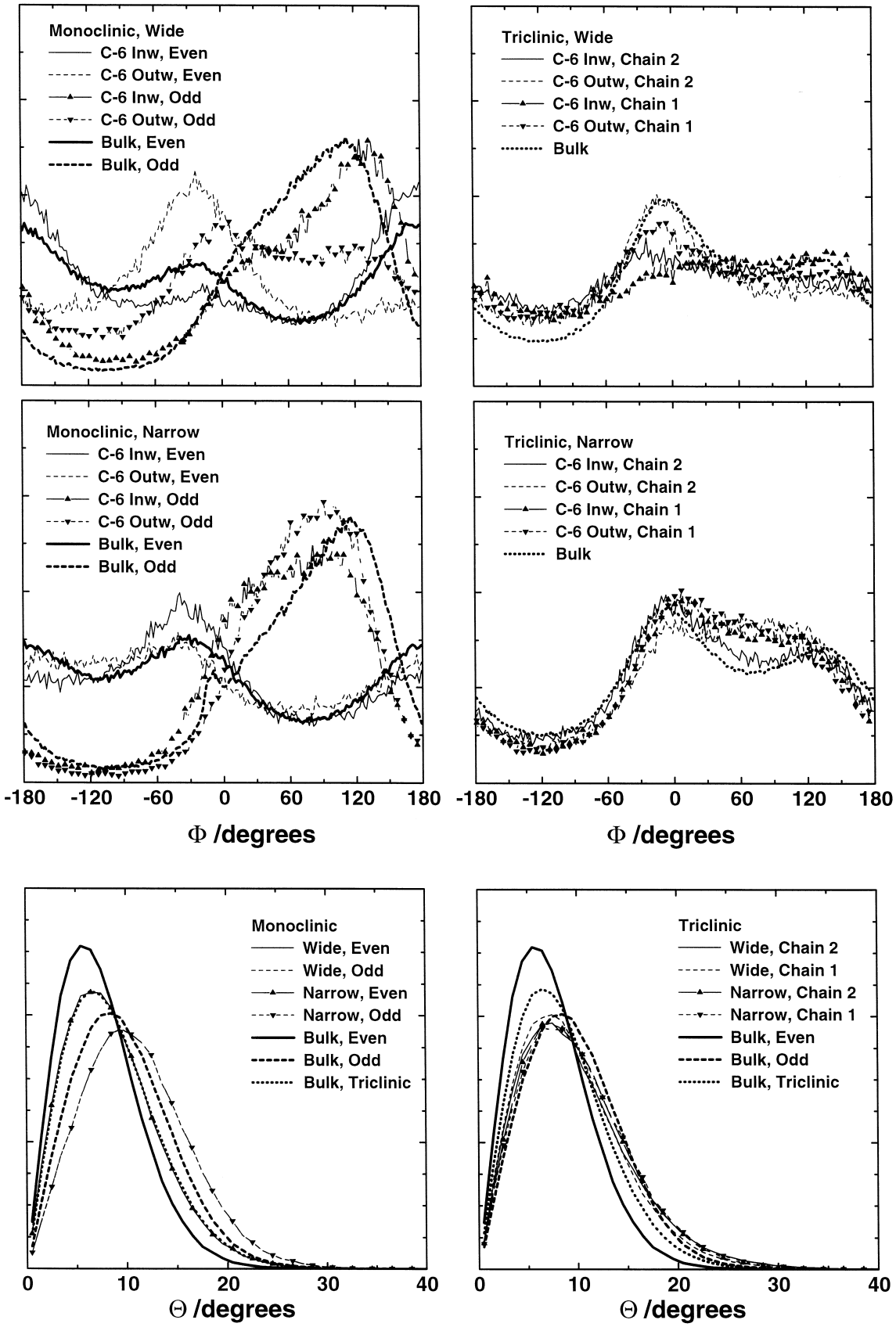

Fig. 6. Φ (top four diagrams) and Θ pucker parameters at the surfaces.

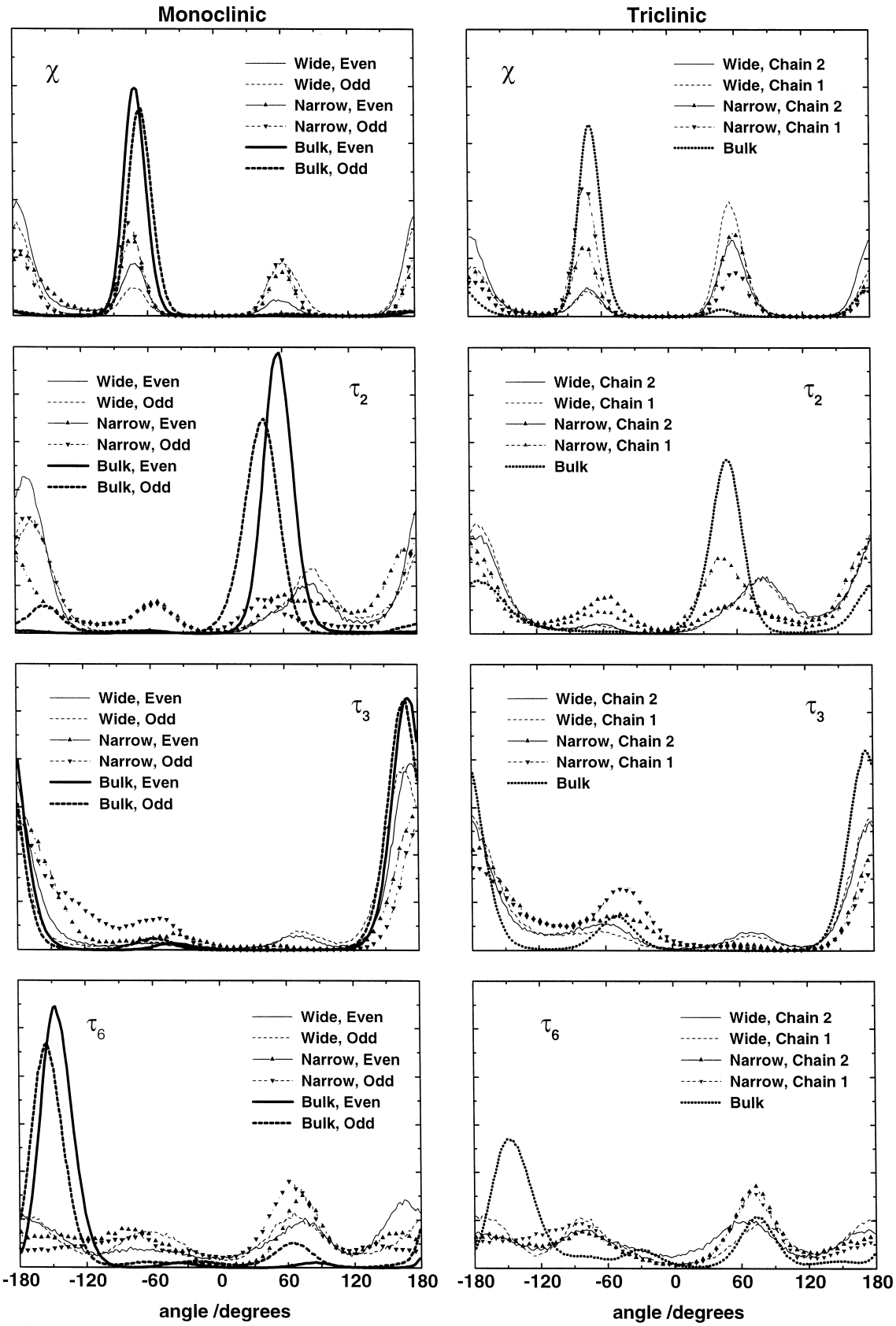

Fig. 7. Torsion angle distributions.

Table 5
Cellulose hydrogen bonds[a]

		Hydrogen bond occupancy					
		O-2–O-6*		O-3–O-5*		O-6–O-3*	
		Mwide	Mnarrow	Mwide	Mnarrow	Mwide	Mnarrow
Layer I, Even	solvent exposed	0.050	0.120	0.510	0.419	0	0
	inward oriented	0.919	0.913	0.709	0.772	0.818	0.819
Layer I, Odd	solvent exposed	0.055	0.139	0.697	0.435	0	0
	inward oriented	0.841	0.677	0.739	0.865	0.715	0.587
Bulk, Even		0.878	0.931	0.692	0.755	0.797	0.862
Bulk, Odd		0.868	0.833	0.888	0.880	0.712	0.757
		Twide	Tnarrow	Twide	Tnarrow	Twide	Tnarrow
Layer I, Chain 1	solvent exposed	0.078	0.293	0.365	0.296	0	0
	inward oriented	0.627	0.598	0.437	0.612	0.598	0.525
Layer I, Chain 2	solvent exposed	0.060	0.104	0.347	0.242	0	0
	inward oriented	0.588	0.610	0.440	0.669	0.520	0.532
Bulk		0.667	0.626	0.677	0.680	0.539	0.544

[a]The number listed is the occupancy, i.e., the probability of finding a hydrogen bond at any given time. Only hydrogen bonds with an individual occupancy over 0.05 are listed. The donor is given first. For layer I, solvent exposed and inward oriented hydrogen bonds are listed separately. Bulk occupancies are calculated over layers II and III. A hydrogen bond exists when the distance between hydrogen and acceptor is less than 0.24 nm, and the donor–hydrogen-acceptor angle is larger than 135°. Alternate chains at the triclinic surface are named chain 1 or chain 2.

preference of the β-anomer of xylose in water (0.67 vs. 0.33) over the α-anomer concurs with an increased hydrogen bonding at the C-1-position (estimates vary between 0.3 and 0.8 more hydrogen bonds). Also the orientation of the C-2 hydroxyl group plays a role. The reversed anomeric ratio ($\alpha:\beta \approx 65:35$) for D-mannose in water is probably a consequence of the axial C-2 hydroxyl group.

The main difference between the four surfaces is the relative position of the hydroxy groups (Fig. 1). A good fit of the solvent with the lateral hydroxy group distribution should be reflected in the lateral solvent

Table 6
Number of hydrogen bonds between cellulose and solvent for the glucose moieties in which the relevant oxygen atom is solvent-exposed

Cellulose atom	Hydrogen bond occupancy							
	Mwide		Mnarrow		Twide		Tnarrow	
	Even	Odd	Even	Odd	Chain 1	Chain 2	Chain 1	Chain 2
O-2, as donor	0.78	0.76	0.74	0.69	0.74	0.73	0.57	0.75
O-3, as donor	0.09	0.14	0.42	0.41	0.20	0.25	0.59	0.54
O-6, as donor	0.58	0.69	0.69	0.56	0.55	0.58	0.57	0.65
O-2, as acceptor	0.43	0.48	0.78	0.53	0.43	0.42	0.58	0.64
O-3, as acceptor	0.60	0.66	0.60	0.18	0.75	0.70	0.41	0.45
O-4, as acceptor	0.26	0.16	0.00	0.21	0.25	0.26	0.16	0.10
O-5, as acceptor	0.03	0.00	0.08	0.33	0.09	0.07	0.32	0.25
O-6, as acceptor	0.77	0.79	0.65	0.79	0.59	0.64	0.58	0.72
O-2, donor + acceptor	1.21	1.23	1.51	1.22	1.18	1.16	1.15	1.39
O-3, donor + acceptor	0.70	0.80	1.02	0.59	0.95	0.95	1.00	0.98
O-4, donor + acceptor	0.26	0.16	0.00	0.21	0.25	0.26	0.16	0.10
O-5, donor + acceptor	0.03	0.00	0.08	0.33	0.09	0.07	0.32	0.25
O-6, donor + acceptor	1.35	1.48	1.34	1.35	1.14	1.22	1.15	1.38
Total per cellobiose	3.02	3.68	3.97	3.70	3.60	3.64	3.78	4.09
Total per cellobiose	3.36		3.84		3.62		3.94	

The existence of a hydrogen bond has been defined in Table 5.

distribution. To test this hypothesis the (normalised) solvent–solvent radial distribution function in a layer covering the first peak in the normal distribution function was calculated (Fig. 8). The hydration peaks for the narrow surfaces are more pronounced than for the wide surfaces, and coincide better with the periodicity of the hydroxymethyl groups at the surface. This is most obvious for the third hydration peak near 1.0 nm (equivalent to one cellobiose repeat unit), which is well defined for the narrow surfaces, but absent for the wide surfaces. The differences in the τ_2, τ_3 and χ dihedral distributions for alternate molecules at the Tnarrow surface may thus be a consequence of optimising water hydrogen-bonding.

The relative hydrophilicity.—The surface excess potential energy (Table 3) favours triclinic surfaces over monoclinic. This is unlikely to possess biological significance, since the choice of crystal form is determined by the bulk energetics, which favours the monoclinic form, in combination with other factors such as the cellulose biosynthesis framework.

Several of the analysed surface properties but notably the hydrogen bonding and the cellulose–water interaction energies suggest that the narrow surfaces are more hydrophilic than the wide. We estimated that the total surface free energy is positive for the Mwide surface, such that nature would tend to minimise the relative amount of surface, i.e., maximise microfibril cross-sections [17]. With the narrow surfaces being less hydrophobic, the solvation penalty would be smaller. This is in fact consistent with estimates of microfibril cross-sections [29,30], by which there is more narrow surface than wide (40×15 nm). Again, the biosynthesis will influence the microfibril cross-section but rearrangement may be prevented by the interference of other compounds, notably hemicellulose.

It is further noteworthy that once the microfibril dimension is given, its solvation is spontaneous ($\Delta G_{\text{solvation}} < 0$). Previously we estimated the free energy of solvation for the Mwide surface to be -12.7 kJ nm^{-2} and we expect that the free energy of solvation will be negative for all four cellulose surfaces. This is supported by the fact that it is very difficult to remove all water from a cellulose surface and that cellulose samples are considered 'dry' at a residual water content of 5–10% [7].

The present simulation is longer than our previous one [17] and produced a lower value for the total surface energy. Although the change is moderate, it illustrates that even several hundred ps of simulation may not produce full equilibration.

The narrow surface dichotomy.—The odd/even duplicity is almost absent at the monoclinic wide surface [17], but persists at the monoclinic narrow surface. More surprisingly, alternate molecules at the triclinic narrow surface feature significantly different geometries and hydrogen bonding although the triclinic bulk only has one cellobiose moiety in its asymmetric unit. It thus appears that all molecules are essentially equivalent in the two wide surfaces, but that alternate molecules segregate in the two narrow surfaces. With our arbitrarily chosen nomenclature the Tnarrow surface molecules of the Chain 1 set are somewhat even-like, and correspondingly the Chain 2 molecules somewhat odd-like. It appears that the narrow geometry at the surface produces an alternation in terms of molecular properties. This again

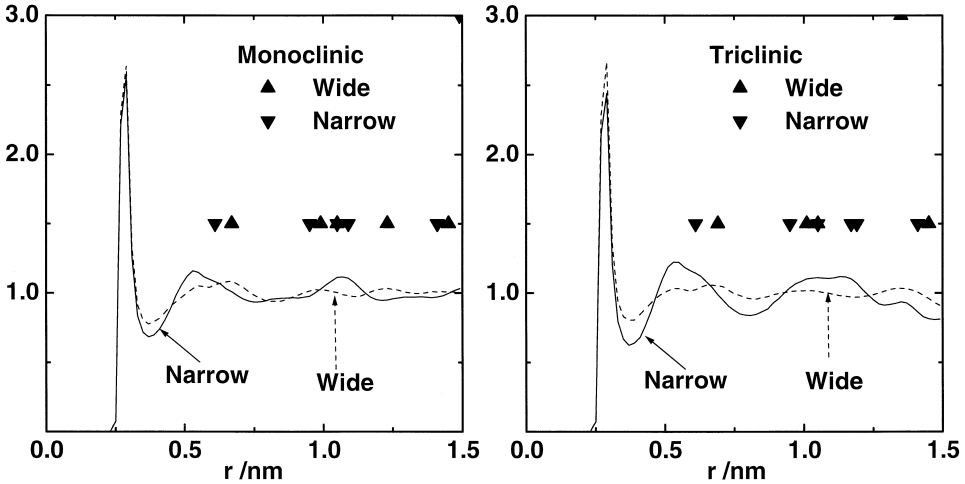

Fig. 8. Normalised solvent–solvent radial distribution function calculated in a layer of 0.15 nm around the first hydration peak. The triangles denote repeat distances in the cellulose surface.

stresses that the two wide surfaces are very similar as are the two narrow surfaces, with the orientation repeat symmetry as the only difference.

AFM observation of cellulose surfaces.—Atomic force microscopy has been used to identify individual crystalline faces from *Valonia macrophysa* cellulose microfibrils [8]. Images with a resolution of 0.4 nm were obtained for the monoclinic narrow face. Moderately good images have been obtained for the monoclinic wide face but so far it has been impossible to obtain high quality images of triclinic crystal faces. One of the suggested reasons for this has been that the triclinic faces may be softer than the monoclinic and therefore so far deformed by the AFM tip that image resolution is completely lost. The fluctuation, hydrogen bonding and energy data presented here suggest that the triclinic surfaces are about as hard as the monoclinic. Thus the reason for their non-observation must be sought in other factors. The monoclinic 110 surface appears to be slightly softer and its hydration layer less structured than at the 1-10 surface which may explain the worse image quality obtained for the 110 surface.

CP-MAS ^{13}C NMR experiments.—The conformational basis for ^{13}C chemical shifts in oligosaccharides is poorly understood. For the hydroxymethyl group, chemical shift differences can be predicted from the χ dihedral distribution [25] but attempts to relate C-1 and C-4 shifts to backbone torsion angles have failed [4,25]. Jarvis has suggested [31] that the C-4 chemical shift is influenced by through-space delocalisation of axially oriented lone-pairs on O-3, O-4 and O-5. This would link the C-4 chemical shift to the presence of an O-3–H-3 $\cdots$ O-5 hydrogen bond.

On the NMR time scale all hydrogen bonds will be in fast exchange, resulting in a downfield shift proportional to the occupancy of the hydrogen bond. The chemical shifts for cotton cellulose [7] can be correlated to our simulated O-3–H-3 $\cdots$ O-5 hydrogen bond occupancies as follows. The resonances at δ 88.7 and δ 87.9 ppm are tentatively assigned to odd and even bulk molecules. The former resonance has been assigned to both Iα and Iβ, but the similar intensity of the two resonances suggests that the former is mostly Iβ. In the same fashion the surface resonances at δ 84.1 and δ 83.2 ppm would arise from the wide and narrow surface respectively. There appears to be a qualitative correlation with the O-3–H-3 $\cdots$ O-5 hydrogen bond occupancy (data not shown), but it does not account for the entire C-4 chemical shift differences.

The interaction with cellulose binding domains.—The enzymic degradation of cellulose is initiated by the cellulase anchoring on a crystalline cellulose surface by means of a cellulose binding domain (CBD) [32]. Electron microscopy data led Chanzy et al. [33] to conclude that *Trichoderma reesei* cellobiohydrolase I preferentially adsorbs either to a wide surface or to its edge. We later found that the binding capacity corresponds to a monomolecular layer on the cellulose surface [34], such that the enzyme primarily binds to the surface. Site-directed mutagenesis has demonstrated the importance of certain aromatic residues on the CBD binding face [35] but a hydrophobic effect also contributes to the affinity of the CBD to the cellulose [34]. The analysis presented here suggests that the wide surfaces are less hydrophilic than the narrow, which is consistent with a hydrophobic effect and the observed preferential binding to a wide surface.

4. Corrigendum

We regret that our paper ‘Interface between monoclinic crystalline cellulose and water’ [Langmuir, 13 (1997) 511–518] contains the following printing error. In Fig. 2, p. 514, top of left column, bold and thin curves represent the odd and even subphases, respectively, rather than vice versa as stated.

Acknowledgements

We are grateful to Drs. Tommy Iversen and Tomas Larsson for valuable discussion and to Dr. Mike Jarvis for pointing out the possible relation between the O-3–H-3 $\cdots$ O-5 hydrogen bond and the C-4 chemical shift. This work was financially supported by the European Commission under the Human Capital and Mobility (Grant No. ERBCHICT941552) and FAIR (grant FAIR-CT-961624) programs.

References

[1] R.H. Atalla and D.L. VanderHart, *Science*, 223 (1984) 283–285.
[2] D.L. VanderHart and R.H. Atalla, *ACS Symp. Ser.*, 340 (1987) 88–118.
[3] J. Sugiyama, R. Vuong, and H. Chanzy, *Macromolecules*, 24 (1991) 4168–4175.
[4] F. Horii, H. Yamamoto, R. Kitamaru, M. Tanahashi, and T. Higushi, *Macromolecules*, 20 (1987) 2946–2949.

[5] M. Wada, J. Sugiyama, and T. Okano, *Mokuzai Gakkaishi*, 40 (1994) 50–56.
[6] P.T. Larsson, U. Westermark, and T. Iversen, *Carbohydr. Res.*, 278 (1995) 339–343.
[7] P.T. Larsson, K. Wickholm and T. Iversen, *Carbohydr. Res.*, 302 (1997) 19–25.
[8] L. Kuutti, J. Peltonen, J. Pere, and O. Teleman, *Microsc.*, 178 (1995) 1–6.
[9] L.M.J. Kroon-Batenburg and J. Kroon, *Biopolymers*, 29 (1990) 1243–1248.
[10] 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.
[11] A.P. Heiner, J. Sugiyama, and O. Teleman, *Carbohydr. Res.*, 273 (1995) 207–223.
[12] J. Jimenez-Barbero, E. Junquera, M. Martin-Pastor, S. Sharma, C. Vincent, and S. Panedas, *J. Am. Chem. Soc.*, 117 (1995) 11198–11204.
[13] S. Reiling and J. Brickmann, *Macromol. Theory Simul.*, 4 (1995) 725–743.
[14] L.M.J. Kroon-Batenburg, B. Bouma, and J. Kroon, *Macromolecules*, 29 (1996) 5695–5699.
[15] J.R. Grigera and S.G. Kalko, *Langmuir*, 12 (1996) 154–158.
[16] D.B. Adolf, D.J. Tildesley, M.R.S. Pinches, J.B. Kingdon, T. Madden, and A. Clark, *Langmuir*, 11 (1995) 237–246; J.I. Siepmann and I.R. McDonald, *Langmuir*, 9 (1993) 2351–2355; A.R. van Buuren, J. de Vlieg, and H.J.C. Berendsen, *Langmuir*, 11 (1995) 2957–2965.
[17] A.P. Heiner and O. Teleman, *Langmuir*, 13 (1997) 511–518.
[18] H.J.C. Berendsen, J.R. Grigera, and J.P. Straatsma, *J. Phys. Chem.*, 91 (1987) 6269.
[19] J.E.H. Koehler, W. Saenger, and W.F. van Gunsteren, *Eur. Biophys. J.*, 15 (1987) 197–210; W.F. van Gunsteren and H.J.C. Berendsen, *Groningen Molecular Simulation (GROMOS) Library Manual*, Biomos, Nijenborgh 4, Groningen, Netherlands, 1987.
[20] 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.
[21] M.C.L.E. Kouwijzer, B.P. van Eijck, S.J. Kroes, and J. Kroon, *J. Comput. Chem.*, 14 (1993) 1281–1289; L.M.J. Kroon-Batenburg, B. Bouma, and J. Kroon, *Macromolecules*, 29 (1996) 5695–5699.
[22] D.A. French, W.A. Roughead, and D.P. Miller, *ACS Symp. Ser.*, 340 (1987) 15–37.
[23] M. Wada, T. Okano, J. Sugiyama, and F. Horii, *Cellulose*, 5 (1995) 223–233.
[24] M.L. Connolly, *J. Appl. Cryst.*, 18 (1985) 499.
[25] D. Cremer and J.A. Pople, *J. Am. Chem. Soc.*, 97 (1975) 1354–1358.
[26] A.P. Heiner and O. Teleman, *Pure Appl. Chem.*, 68 (1996) 2187–2192.
[27] S.A. Galema, E. Howard, J.B.F.N. Engberts, and J.R. Grigera, *Carbohydr. Res.*, 265 (1994) 215–225; S.A. Galema, *The Effect of Stereochemistry on Carbohydrate Hydration in Aqueous Solution, Groningen University*, Netherlands, 1992.
[28] R.K. Schmidt, M. Karplus, and J.W. Brady, *J. Am. Chem. Soc.*, 118 (1996) 541–546.
[29] S. Kuga and R.M. Brown Jr., *Electron Microsc. Technol.*, 6 (1987) 349–356.
[30] A.R. White and R.M. Brown Jr., *Proc. Natl. Acad. Sci. U.S.A.*, 78 (1981) 1047–1051.
[31] M. Jarvis, *Carbohydr. Res.*, 259 (1994) 311–318.
[32] P. Tomme, R.A.J. Warren, and N.R. Gilkes, *Adv. Microb. Physiol.*, 37 (1995) 1–81.
[33] H. Chanzy, B. Henrissat, and R. Vuong, *FEBS Lett.*, 172 (1984) 193–197.
[34] T. Reinikainen, O. Teleman, and T.T. Teeri, *Proteins Struct. Funct. Genet.*, 22 (1995) 392–403.
[35] M. Linder, M.-L. Mattinen, M. Kontteli, G. Lindeberg, J. Ståhlberg, T. Drakenberg, T. Reinikainen, G. Pettersson, and A. Annila, *Protein Sci.*, 4 (1995) 1056–1064.