Role of van der Waals, electrostatic, and hydrogen-bond interactions for the relative stability of cellulose Iβ and II crystals

Richard Kullmann, Martina Delbianco, Christian Roth, and Thomas R. Weikl

Naturally occuring cellulose Iβ with its characteristic parallel orientation of cellulose chains is less stable than cellulose II, in which neighbouring pairs of chains are oriented antiparallel to each other. While the distinct hydrogen-bond patterns of these two cellulose crystal forms are well established, the energetic role of the hydrogen bonds for crystal stability, in comparison to the van der Waals and overall electrostatic interactions in the crystals, is a matter of current debate. In this article, we investigate the relative stability of cellulose Iβ and II in molecular dynamics simulations and energy minimizations. We find that the larger stability of cellulose II results from clearly stronger electrostatic interchain energies that are only partially compensated by stronger van der Waals interchain energies in cellulose Iβ. A decomposition of the electrostatic interchain energies into interaction energies of neutral subgroups of atom leads to a consistent multipole description of hydrogen bonds and to interchain hydrogen-bond energies that account for roughly 80% of the interchain electrostatics in cellulose II.

Introduction

Cellulose is the most abundant biopolymer and a sustainable source for a large variety of materials[1–3]. Naturally occurring cellulose biopolymers are assembled in crystalline arrays, termed cellulose I, in which the polymer chains are oriented in parallel to each other[4–6]. Cellulose I, however, is not the most stable crystalline assembly of cellulose chains. Dissolving and recrystallizing cellulose I leads to cellulose II[7], in which neighboring chains are oriented antiparallel to each other[8]. For synthetic[9] or enzymatically generated[10] cellulose oligosaccharides, only cellulose II is observed in crystalline assemblies. Cellulose I and II have characteristic, distinct hydrogen-bond patterns established decades ago[5,8], but the energetic role of these hydrogen bonds for crystal stability, compared to van der Waals, hydrophobic, and the overall electrostatic interactions, is still a matter of current debate[11–15].

Molecular modelling and simulations have been used extensively to explore the hydrogen-bond networks[16,17] and unit cell parameters[18] of cellulose crystals, the twist of cellulose I fibrils[19–21], the elastic[22–26] and thermal response[27–29] of cellulose, and the assembly and interactions of few cellulose chains[30–32]. The electrostatic and van der Waals (or London dispersion) intrachain and interchain energies in cellulose crystals have been recently calculated with density functional theory (DFT) methods in conjunction with three popular generations of dispersion correction, which lead to differences in dispersion energies of up to about 50%[13]. The dispersion corrections are necessary to empirically include the long-range dispersion interactions in the approximative quantum-mechanical DFT approach[33,34]. In classical atomistic force fields used in molecular dynamics (MD) simulations, long-range van der Waals interactions are included in the Lennard-Jones pair interaction of atoms (see Methods). The mathematical form and numerous atom-type-specific parameters of force fields have been optimised over decades[35], in particular for proteins, resulting in rather accurate descriptions of the structure and dynamics of proteins[36,37]. Current standard carbohydrate force fields tend to overestimate attractive carbohydrate-carbohydrate interactions in carbohydrate solutions, which has led to recalibrations of the Lennard-Jones potentials for the van der Waals interactions[38–41].

In this article, we investigate the relative stability of cellulose Iβ, the dominant form of cellulose I, and cellulose II in molecular dynamics (MD) simulations and energy minimizations with the popular standard force field GLYCAM06[42] and the recalibrated force field $\text{GLYCAM06}^{\text{TIP5P}}_{\text{OSMOr14}}$[39]. In simulations with $\text{GLYCAM06}^{\text{TIP5P}}_{\text{OSMOr14}}$, cellulose Iβ and II nanocrystals composed of 52 hexameric chains are unstable (see Fig. 1), with a faster decrease of interchain contacts in cellulose Iβ, which appears to indicate that the lower stability of cellulose Iβ is reproduced by the force field. In simulations with the standard force field GLYCAM06, both cellulose Iβ and II nanocrystals are stable on the microsecond simulation timescale. Our simulation results thus indicate different absolute stabilities of cellulose crystals in the two force fields, which is plausible due to the rescaled, weaker van der Waals interactions in $\text{GLYCAM06}^{\text{TIP5P}}_{\text{OSMOr14}}$[39].

Energy minimization of cellulose Iβ and II crystals, however, lead to rather similar relative stabilities of the two crystal forms in both force fields. From interpolations of minimization results for different crystal sizes to eliminate surface effects, we obtain a bulk energy difference of about 3 kcal/mol per glucose ring in favour of cellulose II. This bulk energy difference arises from differences in the electrostatic and van der Waals interchain energies of cellulose Iβ and II, i.e. from clearly stronger electrostatic interchain energies in cellulose II that are only partially compensated by stronger van der Waals interchain energies in cellulose Iβ. The electrostatic interchain energies per glucose monomer in the two force fields are about -12 to -13 kcal/mol for cellulose Iβ and -19 to -20 kcal/mol for cellulose II, while the van der Waals interchain energies per glucose monomer are -13 to -16 kcal/mol for cellulose Iβ and -10 to -12 kcal/mol for cellulose II.

To determine the energetic contributions of the three OH groups of the glucose monomers, and of the hydrogen bonds formed by these hydroxyl groups, we decompose the overall electrostatic and van der Waals interchain energies into interaction energies of

Max Planck Institute of Colloids and Interfaces, Department of Biomolecular Systems, Am Mühlenberg 1, 14476 Potsdam, Germany

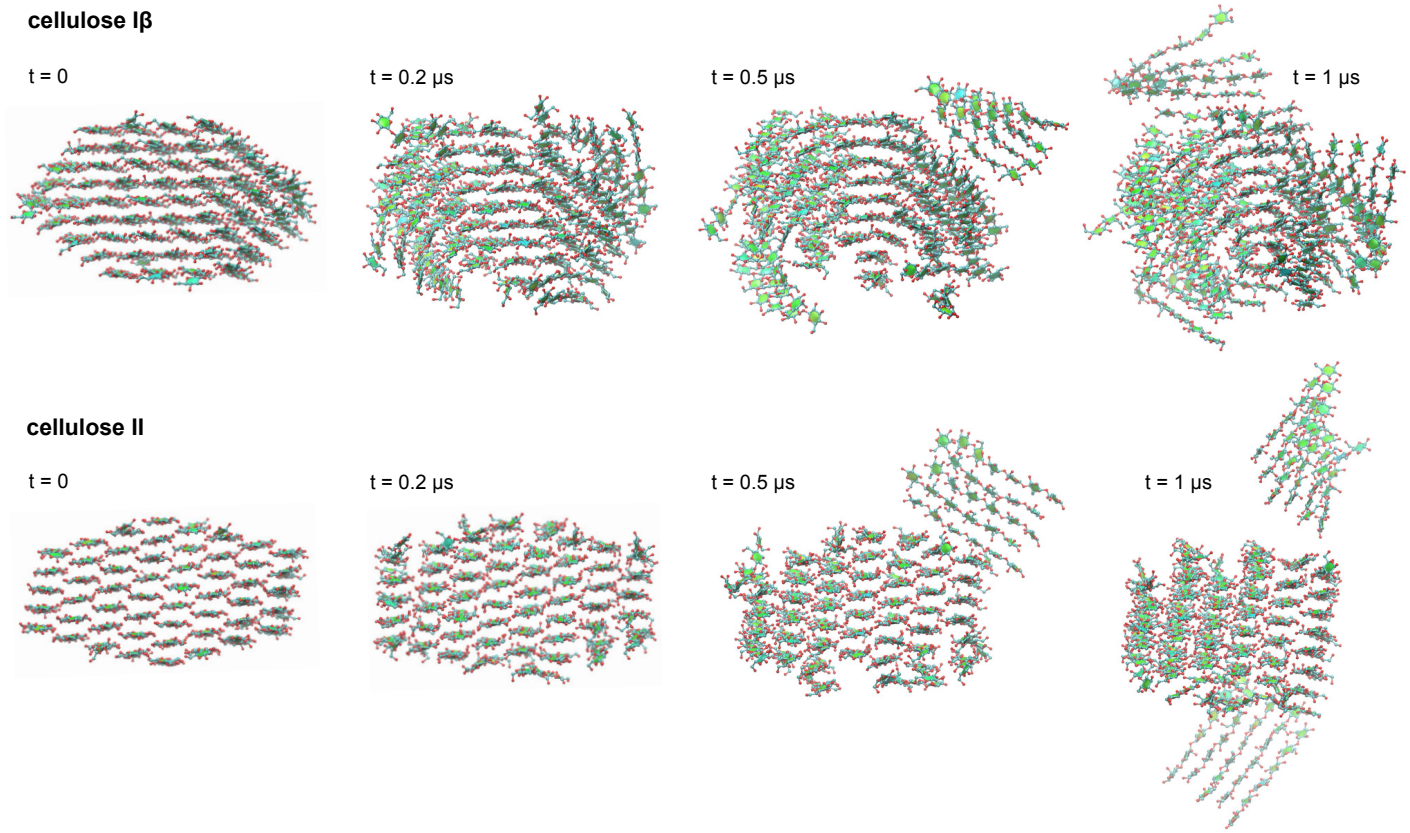

Fig. 1 Snapshots from molecular dynamics simulation trajectories of cellulose Iβ and II nanocrystals composes of 52 hexameric chains with the force field GLYCAM06$^{\mathrm{TIP5P}}_{\mathrm{OSMOr14}}$ in which both nanocrystals decay within microseconds. The nanocrystals are shown in top view along the directions of the cellulose chains.

(nearly) neutral subgroups of atoms. For the OH groups, these neutral subgroups include the C atom to which the groups are bound as third atom, because the O atoms of the hydroxyl groups "draw" their negative partial charge both from the bound H and C atoms. The energetic decomposition of the interchain energies of the glucose monomers leads to a consistent multipole description of the hydrogen bonds and to hydrogen-bond energies between about -4 and -7 kcal/mol in good agreement with estimates based on infrared band shifts for cellulose Iβ [12]. In cellulose Iβ, the single interchain hydrogen bond per glucose monomer contributes about 60% of the interchain electrostatic interactions. In cellulose II, the electrostatic interactions of the two interchain hydrogen bond per glucose monomer sum up to about 80% of the interchain electrostatics. The interchain electrostatic energies in cellulose Iβ and II thus can be seen to be dominated by, but do no result entirely from hydrogen bonding.

Methods

Molecular dynamics simulations

We generated initial structures of cellulose Iβ crystals [5] and cellulose II crystals [43] composed of 52 cellulose 6-mers with the software cellulose-builder [44] and solvated the crystals in octahedral simulation boxes with periodic boundaries using the software Amber 20 [45]. The simulation box sizes were chosen so that the distances between the crystal surfaces and the box edges were at least 15 Å. To equilibrate the solvated crystal structures, we first performed two rounds of energy minimization. In the first round, which consisted of 5000 steps of the steepest-descent method followed by 15000 steps of the conjugate-gradient minimization method, the cellulose atoms were harmonically restrained with a force constant of 25 kcal mol^{-1}Å^{-2}. In the second round, the minimization was repeated with reduced force constant 1 kcal mol^{-1}Å^{-2}. We next heated the simulation systems in three stages from 0 to a temperature of 300 K. In each simulation stage in the NVT ensemble, the temperature was first linearly increased by 100 K within 45000 simulation time steps and then maintained in 5000 subsequent simulation time steps. In these simulations and all other simulations, hydrogen mass repartitioning was employed to reach a simulation time step of 4 fs [46], bonds involving hydrogen atoms were constrained with the SHAKE algorithm [47], and the temperature was controlled by a Langevin thermostat with a collision frequency of 1 ps^{-1}. A cutoff length of 10 Å was used for non-bonded interactions, and long-range electrostatic interactions were treated with the Particle Mesh Ewald (PME) method [48,49]. After heating, we slowly released the restraints on cellulose atoms in 10 equilibration steps with a length of 1 ns by lowering the force constant to 0.7, 0.49, 0.34, 0.23, 0.16, 0.11, 0.07, 0.04, 0.02, and 0.01 kcal mol^{-1}Å^{-2} in the steps. We finally performed production simulations in the NPT ensemble with a trajectory lengths of 1 μs with Amber GPU using a Monte Carlo barostat for pressure control.

Minimum-energy calculations

Our energy calculations are based on energy-minimized structures of cellulose Iβ and II nanocystals composed of 52 cellulose 6-mers, 8-mers, 10-mers, and 12-mers. These nanocrystal differ in their volume-to-surface ratio, which we use to extract bulk (volume) energies of cellulose Iβ and II (see Results). We generated initial crystal structures with the software cellulose-builder[44], solvated the structures, and performed a first minimization round with harmonic restraints on cellulose atoms as described above in the setup for our MD simulations. In second minimization rounds, we fully removed the constraints on cellulose atoms, and generated six energy-minimized structures per crystal by varying the number of the initial minimization steps with the steepest-descent method from 2000 to 7000 in steps of 1000. The minimizations were completed with conjugate-gradient steps to reach a total number of 20000 minimization steps. Our energy calculations include averages over the energy-minimized structures per crystal.

From the partial charges q of the atoms in units of the elementary charge, the electrostatic interaction of two atoms i and j with distance r in units of kcal/mol is calculated in the GLYCAM force fields considered here as the Coulomb interaction

$$V_{i,j}^{\mathrm{elec}}(r) = \frac{q_i q_j}{4\pi\varepsilon_0 r} = \frac{Q_i \cdot Q_j}{r} \tag{1}$$

with $Q = 18.2223\,q$. The partial charges q of the atoms in a central glucose ring of the cellulose chains are listed in Table 1. The sum of these partial charges is 0 because the central glucose rings of the cellulose chains are neutral, which leads to overall electrostatic interactions between two such cellulose rings that are short-ranged compared to the Coulomb interactions of atom pairs. The overall electrostatic interactions between two neutral glucose rings are composed of shorter-ranged interactions of charge dipoles and higher charge multipoles. The charged cellulose atoms in Table 1 consist of four groups of atoms that are nearly neutral. To avoid artefacts in the calculation of bulk energies from long-range Coulomb interactions of the charged terminal glucose rings of the cellulose chains, which would be neutralized by the surrounding solvent not considered in our electrostatic calculations, we adjusted the partial charges of the H atoms at the chain termini in these calculations so that also the terminal glucose rings are neutral.

The van der Waals interaction is calculated from the Lennard-Jones potential

$$V_{i,j}^{\mathrm{LJ}} = \varepsilon\left((R_{\min}/r)^{12} - 2\,(R_{\min}/r)^{6}\right) \tag{2}$$

with $R_{\min} = (R_i + R_j)/2$ and $\varepsilon = \sqrt{\varepsilon_i \varepsilon_j}$ for atom-specific van der Waals radii R_i and R_j and ε parameters ε_i and ε_j. In the force field GLYCAM06$^{\mathrm{TIP5P}}_{\mathrm{OSMOr14}}$, most ε parameters of the original force field GLYCAM06 have been slightly rescaled by 0.94 to reproduce experimentally measured osmotic pressures of carbohydrate solutions, which reflect carbohydrate-carbohydrate interactions, in simulations with the TIP5P water model[39]. The GLYCAM06$^{\mathrm{TIP5P}}_{\mathrm{OSMOr14}}$ force field employs the TIP5P water model because this water model leads to more reliable carbohydrate-carbohydrate interactions in GLYCAM06, compared to the standard TIP3P water model[38,50].

Table 1 Partial charges of cellulose atoms in GLYCAM force fields in units of the elementary charge e and atoms groups used in decomposing interactions

group	atom	charge	group charge
	C2	0.246	
gO2	O2	−0.713	−0.03
	H_{O2}	0.437	
	C3	0.286	
gO3	O3	−0.699	0.014
	H_{O3}	0.427	
	C6	0.276	
gO6	O6	−0.682	0.012
	H_{O6}	0.418	
	O4	−0.468	
	C1	0.509	
gO5	O5	−0.574	0.004
	C5	0.283	
	C4	0.254	

Results

Stability of cellulose Iβ and II in molecular dynamics simulations

To explore the stability of cellulose Iβ and II in MD simulations, we have generated micro-second long simulation trajectories starting from equilibrated crystal structures composed of 52 cellulose 6-mers. In simulations with the force field GLYCAM06$^{\mathrm{TIP5P}}_{\mathrm{OSMOr14}}$, cellulose Iβ and II nanocrystals are unstable and start to decay within the micro-second long simulations (see Fig. 1). To quantify this decay, we have run five independent simulations for each crystal type and have calculated the number of interchain contacts of non-hydrogen atoms along the simulation trajectories. The initial numbers of interchain contacts differ between the two crystal forms, with a total number of about 13050 non-hydrogen atom contacts

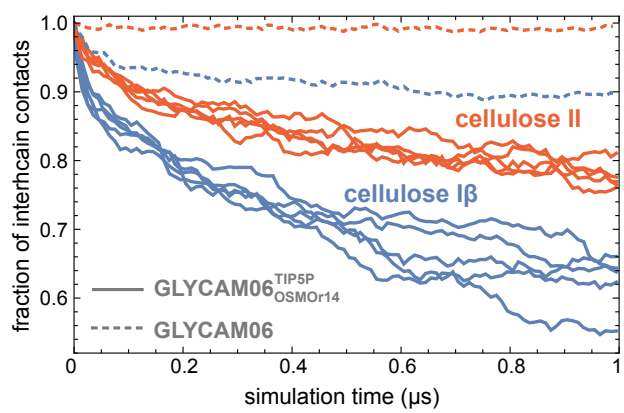

Fig. 2 Fraction of interchain contacts of non-hydrogen atoms, relative to the initial numbers of contacts in the crystals, along five microsecond-long simulation trajectories with the force field GLYCAM06$^{\mathrm{TIP5P}}_{\mathrm{OSMOr14}}$ (full lines), starting from equilibrated crystal structures of cellulose Iβ and II crystals (full lines), and along single simulation trajectories with the standard force field GLYCAM06 (dashed lines). Two non-hydrogen atoms are taken to be in contact for distances smaller than 4.5 Å.

between the cellulose chains in the equilibrated, initial structure of cellulose Iβ, and a total number of about 11280 interchain contacts in the initial structure of cellulose II. Because of the decay of the crystals, the total number of interchain contacts decreases over time in the simulations with GLYCAM06$^{\text{TIP5P}}_{\text{OSMOr14}}$. Fig. 2 shows the fraction of interchain contacts, i.e. the number of interchain contacts at a given timepoint divided by the initial number of interchain contacts, along the trajectories. For cellulose Iβ, the fraction of contacts decreases faster over time along all five simulation trajectories in the force field GLYCAM06$^{\text{TIP5P}}_{\text{OSMOr14}}$, compared to the five simulation trajectories for cellulose II, which appears to indicate that the stability of the cellulose Iβ crystal is lower than the stability of cellulose II crystal in this force field. In simulations with the standard force field GLYCAM06, the cellulose Iβ and II crystals appear stable on the microsecond simulation timescale, with a noticeable decrease of the number of interchain contacts occuring only for cellulose Iβ.

Bulk energy of cellulose Iβ and II from minimization

To further investigate the relative stability of cellulose Iβ and II in the two force fields, we have analyzed structures of cellulose Iβ and II obtained from energy minimizations with the force fields GLYCAM06$^{\text{TIP5P}}_{\text{OSMOr14}}$ and GLYCAM06. The overall energy of a crystal is the sum of its bulk and surface energy. We focus on the bulk energy of cellulose Iβ and II crystals, because the recrystallization of cellulose II from dissolved cellulose Iβ does not seem to be affected by crystal size and therefore likely results from a lower bulk energy of cellulose II compared to cellulose Iβ, and because the surface energies of the crystals include contributions from water interactions that are not directly accessible with energy minimization. To determine the bulk energies of the crystals, we have performed energy minimizations of cellulose Iβ and II crystals composed of 52 cellulose chains with varying numbers of glucose rings per chain. The data points in Fig. 3 represent the interchain electrostatic and van der Waals energy per chain obtained for energy-minimized crystals composed of 52 cellulose 6-mers, 8-mers, 10-mers, and 12-mers. To reduce surface effects from the outer chains in the crystal, the energies in Fig. 3 are averaged over the interchain energies of the 30 central chains in the crystals indicated in Fig. 4. The electrostatic and van der Waals interchain interaction energies of a central chain is calculated as the sum of pairwise energies between the atoms in this chain and the atoms in all other chains of the crystals, divided by two to avoid a double-counting of atom pairs in the averaging over the central chains. The data points in Fig. 3 fall on lines with slopes that reflect energy changes per glucose ring from chain elongation. These energy changes from elongation by glucose rings are equivalent to bulk energies of the cellulose Iβ and II crystals per glucose ring. Table 2 summarizes the interchain electrostatic and van der Waals bulk energies per glucose ring obtained from the linear fits of Fig. 3 with errors estimated as standard errors of the linear fits. The two values per energy term in Table 2 are the energies obtained in the two force fields GLYCAM06$^{\text{TIP5P}}_{\text{OSMOr14}}$ (upper value) and GLYCAM06 (lower value). In addition, Table 2 includes the intrachain bulk energy of Iβ and II crystals from linear fitting of the intrachain

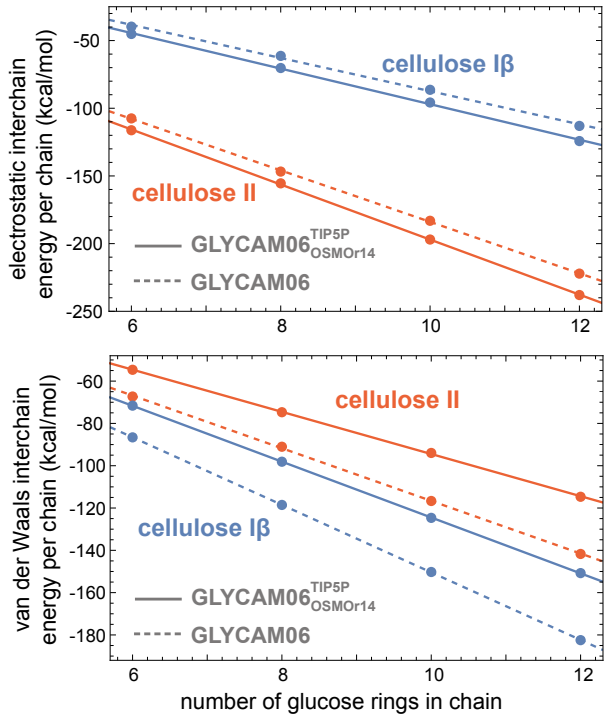

Fig. 3 Interpolation of electrostatic and van der Waals interchain energies obtained for the 30 central chains of energy-minimized Iβ and II crystals composed of 52 cellulose 6-mers, 8-mers, 10-mers, and 12 mers (data points). The slope of the fit lines is the bulk interchain energy per glucose ring, i.e. the energy change per glucose ring from chain elongation.

energies averaged over the 30 central chains of the crystals of 6-mers, 8-mers, 10-mers, and 12-mers akin to Figure 3. The overall bulk energy of cellulose Iβ and II per glucose monomer is the sum of the electrostatic and van der Waals interchain energies and the total intrachain energies of Table 2.

For both force fields, we obtain an overall bulk energy per glucose ring for cellulose II that is about 3 kcal/mol lower than the overall bulk energy for cellulose Iβ (see Table 2). This bulk energy difference arises from differences in the electrostatic and van der Waals interchain energies of cellulose Iβ and II, i.e. from clearly stronger electrostatic interchain energies in cellulose II that are only partially compensated by stronger van der Waals interchain

Table 2 Bulk energies per glucose ring in the force fields GLYCAM06$^{\text{TIP5P}}_{\text{OSMOr14}}$ (upper value) and GLYCAM06 (lower value) in kcal/mol

	Iβ	II	Iβ − II
electrostat. interchain	-13.1 ± 0.3	-20.3 ± 0.2	7.2 ± 0.3
	-12.3 ± 0.4	-19.0 ± 0.2	6.7 ± 0.5
vdW interchain	-13.2 ± 0.1	-10.0 ± 0.1	-3.2 ± 0.1
	-16.0 ± 0.1	-12.4 ± 0.1	-3.6 ± 0.1
total intrachain	119.5 ± 0.2	120.6 ± 1.6	-1.1 ± 1.6
	119.2 ± 0.1	119.2 ± 0.6	0.0 ± 0.6
overall energy	93.2 ± 0.5	90.3 ± 1.7	2.9 ± 1.8
	90.9 ± 0.4	87.8 ± 0.5	3.1 ± 0.7

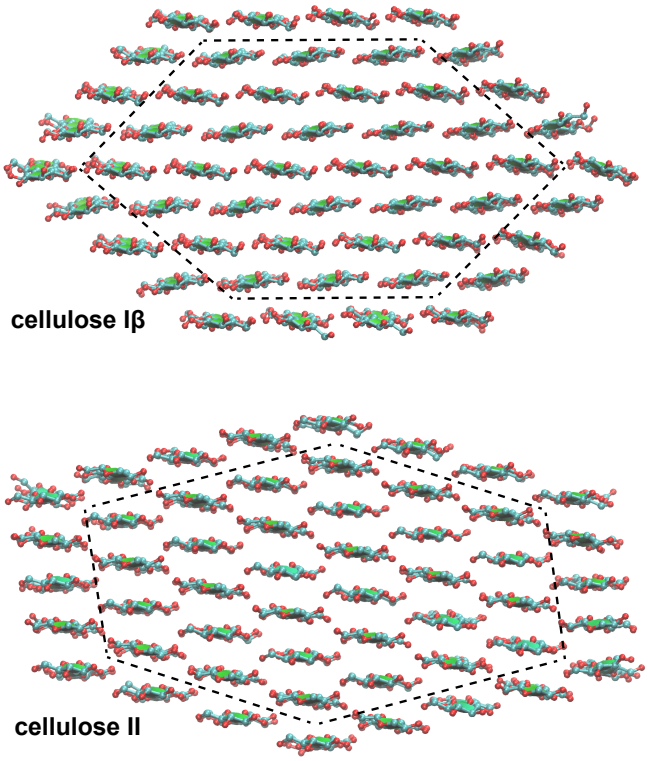

Fig. 4 Energy minimized structures of cellulose Iβ and II nanocystals composed of 52 cellulose 6-mers. For clarity, only the four central glucose rings of the 6-mers are shown in the top-view representations of the crystal structures. The dashed lines indicate the 30 central chains of the crystals used in the energy calculations.

energies in cellulose Iβ. The total intrachain energy in cellulose Iβ and II, in contrast, is rather similar for both force fields, despite the different conformations of the cellulose monomers in both crystals, in particular of the atom O6 of the hydroxymethyl group (see Fig. 5). In cellulose Iβ, O6 is oriented in trans (distal) to O5 and in gauche (proximal) to C4, which is denominated as trans-gauche (tg) conformation of the dihedral angles O6-C6-C5-O5 and O6-C6-C5-C4 of the hydroxymethyl group. In cellulose II, in contrast O6 is oriented in gauche to O5, and in trans to C4, which is denominated as gauche-trans (gt) conformation.

We find that the rather similar total intrachain energies in cellulose Iβ and II result from a near cancellation of differences in the intrachain energy terms of the force fields (see Table 3). These energy terms are the bond, angle, and dihedral energies, which penalize deviations from ideal bond lengths, bond angles, and dihedral angles in force fields, and the “non-bonded" electrostatic and van der Waals interactions between atom pairs that are separated by at least three bonds. The electrostatic and van der Waals interactions are further split up into so-called 1-4 interactions between atom pairs that are separated by three bonds and the remaining electrostatic and van der Waals interactions between atom pairs separated by more than three bonds. The overall 1-4 electrostatic interactions are strongly positive due to repulsive interactions between atom pairs separated by three bonds in the glucose rings (see Table 3), which leads to positive total intrachain energies. If we focus on the conformation-dependent differences of the intrachain energy terms in the last column of Table 3, we see that the 1-4 electrostatic interactions are about 6.3 to 6.5 kcal/mol per glucose ring lower in the cellulose Iβ conformation of the chains, while the angle, dihedral, other electrostatic, and van der Waals energies are lower in the cellulose II conformation, leading to equal total intrachain energies for both conformations within error bounds.

The differences in the 1-4 electrostatic interactions in cellulose Iβ and II can be traced back to the different orientations of O6 and of the H_{O2} hydrogen atom of O2. Table 4 lists the average 1-4 electrostatic interaction energies of the 10 atom pairs (out of a total of 28 1-4 pairs in cellulose) for which the absolute values of the energy difference between the cellulose Iβ and II conformations are larger than 0.5 kcal/mol in the force field $\text{GLYCAM06}^{\text{TIP5P}}_{\text{OSMOr14}}$. The energies in Table 4 have been averaged over all 1-4 atom pairs in the central chains of the energy-minimized crystal composed of 12-mers for which at least one atom of the pairs is located within the 8 central glucose rings of the 12-mers. The numbers in brackets indicate the standard deviation of the energies in the last digit(s) to illustrate variations within the crystal. The 10 energy differences in the last column of Table 4 add up to a total value of -6.6 kcal/mol, which is close the value -6.49 ± 0.06 kcal/mol per glucose ring in Table 3 obtained from linear interpolation of the overall 1-4 energies for crystals composed of 6-mers, 8-mers, 10-mers, and 12-mers with $\text{GLYCAM06}^{\text{TIP5P}}_{\text{OSMOr14}}$. The 1-4 electrostatic interaction energies of O6 and the associated hydrogen atom H_{O6} with atoms O5, C4, and C5 are clearly lower in the tg conformation of cellulose Iβ, compared to the gt conformation of cellulose II. In contrast, the 1-4 electrostatic interaction energy of H_{O2} with C1 is clearly lower for the orientation of the H_{O2} hydrogen atom in cellulose II. Also the differences in dihedral energies in cellulose Iβ and II can be traced back to the different conformations of the

Table 3 Intrachain energy components per ring for the force fields $\text{GLYCAM06}^{\text{TIP5P}}_{\text{OSMOr14}}$ (upper values) and GLYCAM06 (lower values) in kcal/mol

	Iβ	II	Iβ − II
bond	2.77 ± 0.03	2.98 ± 0.08	-0.21 ± 0.09
	2.70 ± 0.02	2.94 ± 0.05	-0.24 ± 0.05
angle	5.26 ± 0.02	3.31 ± 0.08	1.95 ± 0.08
	5.29 ± 0.06	3.40 ± 0.06	1.89 ± 0.08
dihedral	7.30 ± 0.05	5.95 ± 0.39	1.35 ± 0.39
	7.22 ± 0.02	5.63 ± 0.18	1.59 ± 0.19
1-4 electrost.	215.52 ± 0.04	222.01 ± 0.04	-6.49 ± 0.06
	215.70 ± 0.18	221.96 ± 0.28	-6.26 ± 0.34
electrostatic	-114.64 ± 0.22	-115.8 ± 1.3	1.2 ± 1.3
	-114.13 ± 0.19	-116.37 ± 0.75	2.24 ± 0.78
1-4 vdW	4.62 ± 0.03	4.95 ± 0.06	-0.33 ± 0.07
	4.36 ± 0.03	4.84 ± 0.07	-0.48 ± 0.08
vdW	-1.36 ± 0.06	-2.78 ± 0.28	1.42 ± 0.29
	-1.92 ± 0.06	-3.21 ± 0.13	1.29 ± 0.14
total	119.5 ± 0.2	120.6 ± 1.6	-1.1 ± 1.6
	119.2 ± 0.1	119.2 ± 0.6	0.0 ± 0.6

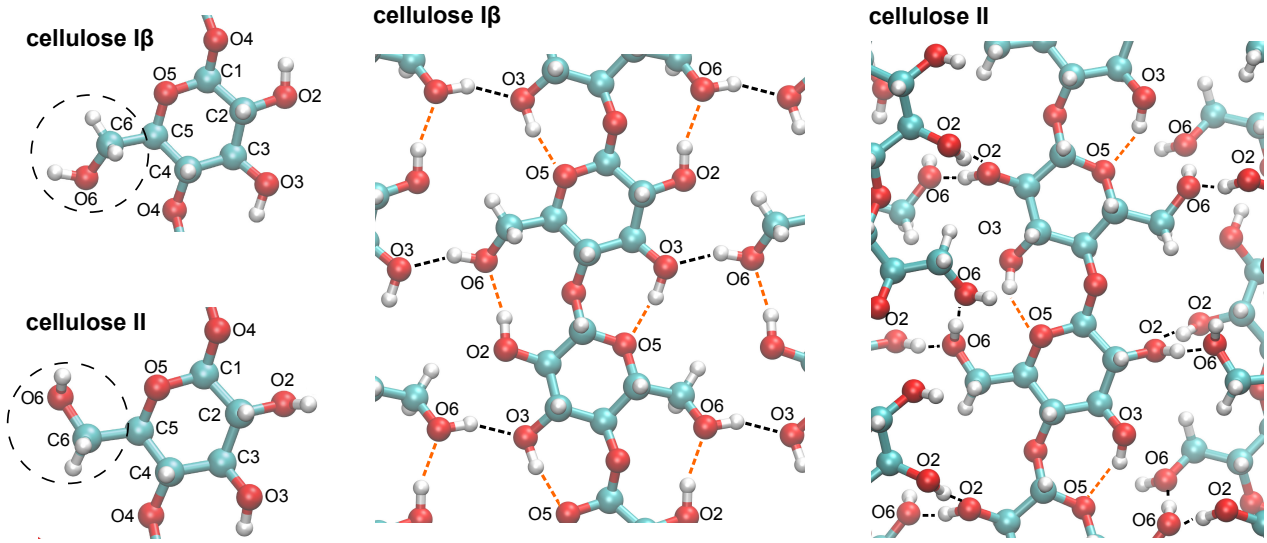

Fig. 5 Molecular conformations and hydrogen bonds in cellulose Iβ and II.

hydroxymethyl group, in particular to the O6-C6-C5-O5 dihedral angle with an average energy that is about 1.7 kcal/mol lower for the gauche orientation in cellulose II, compared to the trans orientation in cellulose Iβ. This energy difference of 1.7 kcal/mol is close to the total dihedral energy difference of 1.35 ± 0.39 kcal/mol in Table 3 for GLYCAM06$^{\mathrm{TIP5P}}_{\mathrm{OSMOr14}}$. The energy for the O6-C6-C5-C4 dihedral angle of the hydroxymethyl group, in contrast, is essentially equal in the gauche conformation of cellulose Iβ and trans conformation of cellulose II.

Table 4 1-4 electrostatic interactions of selected atom pairs in GLYCAM06$^{\mathrm{TIP5P}}_{\mathrm{OSMOr14}}$ in kcal/mol

		Iβ	II	Iβ−II
O5	O6	34.7(1)	45.0(11)	−10.2(11)
C4	O6	−18.5(1)	−15.1(1)	−3.4(1)
C5	H_{O6}	12.0(2)	13.2(9)	−1.2(9)
C3	H_{O2}	12.5(1)	13.5(6)	−1.0(6)
C6	O4	−14.0(2)	−14.6(1)	0.6(2)
O2	O3	56.1(3)	55.4(8)	0.7(8)
C4	H_{O3}	14.5(1)	13.7(5)	0.8(5)
O2	O4	38.1(2)	37.2(6)	0.9(6)
O3	O4	37.5(2)	36.2(4)	1.3(4)
C1	H_{O2}	29.0(6)	24.1(9)	4.9(11)

The table includes 10 pairs out of a total of 28 1-4 pairs with absolute energy differences larger than 0.5 kcal/mol between cellulose Iβ and II in the last column. The averages have been calculated for 1-4 pairs with at least one atom in the 8 central rings of the chains with 12 sugar rings. Numbers in brackets indicate standard deviations for the last digit(s).

Electrostatic and van der Waals interchain energies between atom groups

To explore the role of the three OH groups for the interchain energies of cellulose Iβ and II, we first aim at a decomposition of the overall electrostatic and van der Waals interchain energies per glucose ring in Table 2. The charged atoms of a central glucose ring listed in Table 1 consist of four groups of atoms that are nearly neutral and, thus, constitute a natural starting point for decomposing the overall electrostatic interchain interactions of the glucose rings into electrostatic interactions that are of similarly short range as the interactions of the neutral rings. The first group, which we call gO2, includes O2 and the two atoms H_{O2} and C2 bound to O2 from which O2 “draws" its partial charge. The second and third atom group, gO3 and gO6, are similar to this first group in containing an OH group and the bound C atom, while the fourth atom group, termed gO5, includes the remaining five cellulose atoms with nonzero partial charge. Because the remaining, small group charges indicated in Table 1 still lead to divergences in the sum of electrostatic interactions between groups, we “neutralize" the groups by shifting a charge of -0.03 e from C2 to C1, a charge of 0.014 e from C3 to C4, and a charge of 0.012 e from C6 to C5, and use the slightly shifted partial charges of these “neutralized" groups in calculating the electrostatic interchain group interactions listed in Table 5. The electrostatic interchain and intrachain energies per ring in Tables 2 and 3, in contrast, have been calculated for the original partial charges of the force fields, which have been used also in all energy minimizations and simulations.

The interchain energies between pairs of atom groups in Table 5 are calculated akin to the overall interchain electrostatic and van der Waals energies of Table 2. To determine, e.g., the gO2–gO3 electrostatic interchain interaction, we sum up all interchain pair interactions of gO2 and gO3 groups in the crystals of 6-mers, 8-mers, 10-mers, and 12-mers, for which at least one group is located within the central 30 chains of the crystal, and divide this sum by the number 30 of central chains to obtain energies per chain, and further by a factor of two to avoid a double counting of group pairs in the calculation. To avoid complications from end groups in the terminal rings of the cellulose chains, we consider only the interactions between groups in interior cellullose rings,

Table 5 Interchain energies of atom groups per glucose ring in GLYCAM06$^{\mathrm{TIP5P}}_{\mathrm{OSMOr14}}$ (upper values) and GLYCAM06 (lower values) in kcal/mol

	electrost. Iβ	electrost. II	vdW Iβ	vdW II
gO2 – gO2	-0.50 ± 0.01	$\mathbf{-4.24 \pm 0.07}$	-0.40 ± 0.01	0.57 ± 0.07
	-0.52 ± 0.01	$\mathbf{-3.75 \pm 0.04}$	-0.48 ± 0.01	0.43 ± 0.07
gO2 – gO3	-2.05 ± 0.02	-0.56 ± 0.06	-0.97 ± 0.01	-0.76 ± 0.02
	-2.01 ± 0.01	-0.64 ± 0.06	-1.17 ± 0.01	-0.92 ± 0.02
gO2 – gO5	-0.19 ± 0.09	1.58 ± 0.07	-2.17 ± 0.02	-2.14 ± 0.01
	0.04 ± 0.01	1.62 ± 0.10	-2.55 ± 0.01	-2.52 ± 0.01
gO2 – gO6	-1.48 ± 0.04	$\mathbf{-11.75 \pm 0.22}$	-0.58 ± 0.01	0.86 ± 0.07
	-1.47 ± 0.03	$\mathbf{-11.04 \pm 0.10}$	-0.93 ± 0.01	0.55 ± 0.07
gO3 – gO3	-1.10 ± 0.01	0.36 ± 0.08	-0.39 ± 0.01	-0.39 ± 0.01
	-1.07 ± 0.01	0.34 ± 0.06	-0.47 ± 0.01	-0.49 ± 0.01
gO3 – gO5	-0.76 ± 0.01	-0.70 ± 0.07	-2.04 ± 0.01	-2.00 ± 0.01
	-0.63 ± 0.04	-0.55 ± 0.10	-2.34 ± 0.01	-2.30 ± 0.01
gO3 – gO6	$\mathbf{-7.35 \pm 0.02}$	-0.40 ± 0.13	-0.37 ± 0.01	-1.08 ± 0.05
	$\mathbf{-6.68 \pm 0.04}$	-0.34 ± 0.02	-0.72 ± 0.03	-1.23 ± 0.01
gO5 – gO5	0.30 ± 0.10	-1.56 ± 0.07	-2.26 ± 0.01	-2.38 ± 0.03
	0.20 ± 0.17	-1.62 ± 0.05	-2.61 ± 0.01	-2.77 ± 0.02
gO5 – gO6	0.58 ± 0.01	1.28 ± 0.07	-3.47 ± 0.01	-2.79 ± 0.03
	0.42 ± 0.10	1.12 ± 0.07	-3.97 ± 0.01	-3.24 ± 0.01
gO6 – gO6	-0.36 ± 0.05	$\mathbf{-4.53 \pm 0.11}$	-0.63 ± 0.01	0.22 ± 0.03
	-0.32 ± 0.02	$\mathbf{-4.21 \pm 0.02}$	-0.74 ± 0.01	0.05 ± 0.03
total	-12.9 ± 0.2	-20.5 ± 0.4	-13.3 ± 0.1	-9.9 ± 0.1
	-12.0 ± 0.2	-19.1 ± 0.2	-16.0 ± 0.1	-12.4 ± 0.1

i.e. in 4 interior out 6 rings of the 6-mer chains, 6 interior out of 8 rings of the 8-mer chains, etc., in this summation. The interaction energies per glucose ring in Table 5 then are determined again from linear interpolation of the interaction energies per chain obtained for each crystal as in Fig. 3. In calculating the interchain van der Waals energies between the groups, we also include the noncharged hydrogen atoms of each group, on top of the charged atoms listed in Table 1. An important consistency check of the calculations is that the total of all group interchain interactions in Table 5 agrees with the overall interchain interaction energies in Table 2 within the error bounds obtained from linear interpolations and error propagation, both for cellulose Iβ and II, for the electrostatic and van der Waals interchain interactions, and for calculations with GLYCAM06$^{\mathrm{TIP5P}}_{\mathrm{OSMOr14}}$ and GLYCAM06.

Electrostatic interchain interactions of atom groups that are involved in intrachain hydrogen bonds in cellulose Iβ and II are highlighted in bold font in Table 5. The electrostatic interactions of these atom groups are indeed significantly larger than the electrostatic interactions of other groups, even though the interactions are calculated from averaging over all pairs of, e.g., gO3 and gO6 groups in cellulose Iβ, including distant groups in the crystals, not only for the small subset of hydrogen-bonded gO3 and gO6 groups. In cellulose Iβ, intrachain hydrogen bonds are formed between O6 as donor and O3 as acceptor, with one interchain hydrogen bond per glucose ring (see Fig. 5). In cellulose II, there are two interchain hydrogen bonds per glucose ring, formed by the O2 and O6 atoms in different combinations, in which O2 and O6 atoms can be donor or acceptor, and in which, e.g., an O2 atom can be bound to another O2 atom, or to an O6 atom (see Fig. 5). The electrostatic interchain group interaction gO2 – gO6 in cellulose II, which includes two different types of hydrogen bonds in which either O2 or O6 is donor, is largest. The electrostatic group interactions gO2 – gO2 and gO6 – gO6 include only one type of hydrogen bond and are smaller than the gO2 – gO6 interactions, but still significantly larger than the interactions between groups that do not form hydrogen bonds. Interestingly, the van der Waals interactions between groups involved in hydrogen bonds in cellulose II are positive, from short distances of O atoms in the hydrogen bonds at which the van der Waals radii of these O atoms are already overlapping, leading the repulsive, positive energies in the Lennard-Jones interactions of Eq. 2 (see next section). Apart from these positive interactions, the van der Waals interactions between groups in cellulose Iβ and II tend to be more uniform than the electrostatic interactions, with larger interaction energies for group interactions involving gO5, which can be understood from the larger number of atoms in this group, compared to the other three groups.

Energies of hydrogen bonds

To further assess the role of hydrogen bonds in the interchain interactions, we now focus on the energies and geometry of the hydrogen bonds formed by the three OH groups in cellulose Iβ and II. A simple electrostatic view of hydrogen bonds depicts the OH group of the donor oxygen atom as a dipole with oppositely equal charges $-\delta$ and $+\delta$ on the O and H atom, respectively. An electrostatic attraction between the donor OH group and the acceptor O atom with negative partial charge then directly results from

Table 6 Hydrogen-bond geometry and energetics (in kcal/mol) for cellulose Iβ in the force fields GLYCAM06$^{\mathrm{TIP5P}}_{\mathrm{OSMOr14}}$ (upper values) and GLYCAM06 (lower values)

	geometry			electrost.		vdW	
	d_{HO} (Å)	d_{OO} (Å)	angle (°)	gO – O	gO – gO	gO – O	gO – gO
O3oH...O5o intra	1.74(1)	2.69(1)	163(1)	−12.4(2)	−4.7(1)	1.0(1)	0.9(1)
	1.78(1)	2.73(1)	163(1)	−11.8(1)	-4.3(1)	1.1(1)	0.9(1)
O3cH...O5c intra	1.71(1)	2.68(1)	169(1)	−12.9(2)	−4.7(1)	1.1(1)	0.9(1)
	1.75(1)	2.72(1)	169(1)	−12.4(1)	-4.3(1)	1.2(1)	0.9(1)
O2oH...O6o intra	1.77(1)	2.74(1)	172(1)	−14.6(2)	-8.0(1)	0.9(1)	0.7(1)
	1.83(1)	2.81(1)	172(1)	−13.6(2)	-7.3(1)	0.8(1)	0.6(1)
O2cH...O6c intra	1.71(1)	2.68(1)	166(1)	−14.9(2)	-8.6(1)	1.7(1)	1.5(1)
	1.76(1)	2.72(1)	167(1)	−14.2(1)	-8.0(1)	1.6(1)	1.4(1)
O6oH...O3o inter	1.72(2)	2.68(1)	162(2)	−14.0(5)	−8.3(3)	1.6(2)	1.4(2)
	1.79(2)	2.74(1)	161(1)	−12.9(3)	−7.5(2)	1.4(2)	1.2(2)
O6cH...O3c inter	1.86(3)	2.83(3)	171(2)	−13.4(5)	−6.7(3)	0.4(1)	0.1(1)
	1.93(3)	2.89(2)	168(1)	−12.4(4)	−6.1(3)	0.3(1)	0.1(1)

the fact that the acceptor O atom is closer to the H atom than to the O atom of the donor group in the hydrogen bond, leading to an attractive Coulomb interaction between H and acceptor O that dominates over the repulsive Coulomb interaction of the two Os. In force fields, however, the situation is more complex, with an absolute value of the partial charge on the O atom of an OH group that is significantly larger than the partial charge of the H atom. In the GLYCAM force fields considered here, the negative charge on the O atom of three cellulose OH groups is nearly balanced by the positive charge of the bound H and C atom in the atom groups gO2, gO3, and gO6 of Table 1. For the hydrogen bond geometries obtained in our energy-minimized cellulose crystals, the repulsive Coulomb interaction of the two O atoms in a hydrogen bond dominates over the attractive Coulomb interaction between the H atom of the hydrogen bond and the acceptor O atom, leading to an overall positive, repulsive electrostatic interaction between the OH group and the acceptor O. A realistic depiction of the hydrogen bond energetics therefore needs to include the C atoms of the gO2, gO3, and gO6 groups and the negative, attractive interaction energy between these C atoms and the acceptor O atoms in hydrogen bonds in which these groups are donors.

Tables 6 and 7 list results for the hydrogen-bond geometry and the electrostatic and van der Waals interaction energies between the gO2, gO3, and gO6 group as donor and the acceptor O atom of the hydrogen bonds as well as the interaction energies between these donor groups and the whole group gO2, gO3, gO5, or gO6 of the acceptor O atom. Because the crystal cells of cellulose Iβ and II include two chains with slightly different conformations, an origin (o) chain and a center (c) chain[5,8,12], we specify and distinguish the hydrogen bonds based on these chain types using a standard distance- and angle-based geometric criterion for identifying hydrogen bonds. For cellulose II, we obtain the hydrogen-bond pattern B first described by Chen et al.[16] The results in Tables 6 and 7 are averages obtained for the hydrogen bonds in the energy-minimized crystals composed of cellulose 12-mers in which the acceptor group is located in the central cellulose chains of the crystals and in the 8 central glucose rings of the 12-mer chains and, thus, in the crystal interior. As in Table 4, numbers in brackets in Tables 6 and 7 indicate standard deviations for the last digit(s) to illustrate variations within the crystal. In cellulose Iβ, there are two intrachain and one interchain hydrogen bond per glucose ring. The electrostatic energy between the donor gO6 group and the acceptor O3 atom of the interchain hydrogen bonds O6oH...O3o and O6c...O3c is −14.0 and −13.4 kcal/mol for GLYCAM06$^{\mathrm{TIP5P}}_{\mathrm{OSMOr14}}$ and −12.9 and −12.4 kcal/mol for GLYCAM06 and, thus, exceeds the overall electrostatic interchain interactions per glucose ring in cellulose Iβ in the two force fields (see Table 2). The electrostatic energy between the donor gO6 group and the gO3 group of the acceptor O3 atom is −8.3 and −6.7 kcal/mol for GLYCAM06$^{\mathrm{TIP5P}}_{\mathrm{OSMOr14}}$, and −7.5 and −6.1 kcal/mol for GLYCAM06, with averages of −7.5 kcal/mol for GLYCAM06$^{\mathrm{TIP5P}}_{\mathrm{OSMOr14}}$ and −6.8 kcal/mol for GLYCAM06 that agree with the interchain energies of the groups in Table 5. It is important to recall that the interchain energies of the groups gO3 and gO6 in Table 5 involve summations of the interaction of a gO3 group with all other gO6 groups in the crystals, whereas Table 6 focuses on the interaction of a gO3 group with its hydrogen-bonded gO6 group. The agreement of average interchain group interaction energies for gO3 and gO6 in Tables 5 and 6 thus indicates that the sum over group pairs in Table 5 is dominated by the hydrogen-bonded groups in the summation. For consistency with Table 5, the electrostatic interactions in Tables 6 and 7 are calculated for neutralized groups with slightly shifted charges on C atoms (see previous section). For the original force field charges in Table 1, we obtain the slightly larger average values of −14.5 kcal/mol in GLYCAM06$^{\mathrm{TIP5P}}_{\mathrm{OSMOr14}}$ and −13.5 kcal/mol in GLYCAM06 for the electrostatic interaction between the donor gO6 group and the acceptor O3 atom of the O6oH...O3o and O6c...O3c bonds, but the same average values of −7.5 kcal/mol and −6.8 kcal/mol in the two force fields for the electrostatic interactions between the donor gO6 group and the gO3 group of the acceptor O3 atom.

In cellulose II, the on average two interchain hydrogen bonds per glucose ring occur in different combinations of the gO2 and gO6 groups involved in these hydrogen bonds (see Fig. 5 and Table 7). For the interchain hydrogen bonds O2oH...O2c and

Table 7 Hydrogen-bond geometry and energetics (in kcal/mol) for cellulose II in the force fields GLYCAM06$^{\mathrm{TIP5P}}_{\mathrm{OSMOr14}}$ (upper values) and GLYCAM06 (lower values)

	geometry			electrost.		vdW	
	d_{HO} (Å)	d_{OO} (Å)	angle (°)	gO – O	gO – gO	gO – O	gO – gO
O3oH...O5o intra	1.79(3)	2.68(2)	150(3)	−11.6(5)	−5.3(2)	1.1(2)	1.1(2)
	1.88(4)	2.76(2)	148(3)	−10.5(6)	-4.6(2)	0.9(2)	0.8(2)
O3cH...O5c intra	1.67(1)	2.62(1)	164(2)	−13.4(2)	−5.7(2)	1.7(2)	1.6(2)
	1.71(1)	2.66(2)	164(1)	−13.0(2)	-5.4(1)	1.9(2)	1.7(2)
O2cH...O6c inter	1.66(1)	2.64(2)	169(2)	−17.1(4)	-9.0(3)	2.2(2)	2.0(2)
	1.71(1)	2.69(1)	169(1)	−16.2(4)	-8.3(2)	2.2(2)	1.9(2)
O2oH...O2c inter	1.70(1)	2.68(1)	175(2)	−17.4(3)	-7.8(3)	1.7(2)	1.5(2)
	1.75(2)	2.73(2)	176(1)	−16.5(3)	-7.0(3)	1.6(2)	1.4(2)
O6cH...O6o inter	1.71(1)	2.68(1)	172(2)	−16.1(3)	-8.3(2)	1.7(2)	1.4(2)
	1.76(2)	2.73(2)	173(2)	−15.2(4)	-7.5(3)	1.6(2)	1.3(2)
O6oH...O2o inter	1.69(1)	2.65(2)	167(2)	−16.9(4)	-8.5(2)	2.0(3)	1.8(3)
	1.74(2)	2.70(2)	166(2)	−16.1(4)	-7.9(2)	2.0(2)	1.8(2)

O6cH...O6o, the group interaction energies gO2 – gO2 and gO6 – gO6 in Table 7 are roughly twice the corresponding group interactions per ring in Table 5, which can be understood from the fact that these hydrogen bonds occur on average for every second glucose ring in the crystal. For the interchain hydrogen bonds O2cH...O6c and O6oH...O2o, the group interaction energies in Table 7 are noticably smaller than the gO2 – gO6 group interaction in Table 5, which appears to indicate that also other, more distant pairs of gO2 and gO6 groups in the crystal contribute to the overall electrostatic interchain interaction of the groups. The distances d_{HO} between the donor H atom and the acceptor O atom and the distances d_{OO} between the two O atoms of the hydrogen bonds are smaller for the force field GLYCAM06$^{\mathrm{TIP5P}}_{\mathrm{OSMOr14}}$, because the van der Waals interactions have been slightly reduced by rescaling of ε parameters in GLYCAM06$^{\mathrm{TIP5P}}_{\mathrm{OSMOr14}}$ (see Computational methods), which leads to a weaker repulsion of the O atoms, closer distances, and stronger electrostatic interactions in the hydrogen bonds.

Discussions and Conclusions

The overall consistency of the electrostatic group interactions in Table 5 with the gO – gO electrostatic hydrogen-bond interactions of Tables 6 and 7 indicates that the hydrogen-bond energetics in atomistic force fields of cellulose can be quantified as a multipole interaction between groups of atoms. The gO – O electrostatic interactions of the hydrogen bonds between a donor group and an acceptor O atom, in contrast, appear excessively large. The average gO – O electrostatic interaction energy of the four interchain hydrogen bonds of cellulose II in Table 7 , for example, is −33.8 kcal/mol per glucose ring in GLYCAM06$^{\mathrm{TIP5P}}_{\mathrm{OSMOr14}}$ and −32.9 kcal/mol in GLYCAM06 and, thus, much larger than the overall electrostatic interchain interaction per glucose ring in Table 2. The average gO – gO electrostatic interaction energy of the four interchain hydrogen bonds of cellulose II, in contrast, is −16.8 kcal/mol per glucose ring in GLYCAM06$^{\mathrm{TIP5P}}_{\mathrm{OSMOr14}}$ and −15.4 kcal/mol in GLYCAM06, which is slightly more than 80% of the overall electrostatic interchain energies per glucose ring in Table 2. The average gO – gO electrostatic interaction energy of the interchain hydrogen bonds O6oH...O3o and O6c...O3c in cellulose Iβ is −7.5 kcal/mol per glucose ring in GLYCAM06$^{\mathrm{TIP5P}}_{\mathrm{OSMOr14}}$ and −6.8 kcal/mol in GLYCAM06, which constitutes somewhat less than 60% of the overall electrostatic interchain interaction per glucose ring in Table 2. The quantification of electrostatic hydrogen-bond energies as multipole interactions between (nearly) neutral groups of atoms is reminiscent of the classical approach of Kabsch and Sander[51] to determine the energy of hydrogen bonds in protein secondary structures as electrostatic dipole-dipole interactions between the backbone CO group with oppositely equal partial charges $\pm q_1$ of the C and O atom and the backbone NH group with oppositely equal partial charges $\pm q_2$ of H and N. To quantify hydrogen-bond energies in cellulose, we have used multipoles of three or more atoms because of the absence of dipoles with oppositely equal charges in the charge distributions of the force fields (see Table 1).

The total energy of the hydrogen bonds is the sum of the negative, attractive electrostatic interactions and the positive, repulsive van der Waals interactions. The van der Waals interactions are repulsive because the distances d_{OO} of the donor and acceptor oxygen atoms that result from the minimization of the overall energy of the crystals are clearly smaller than the van der Waals radii of 3.442 Å for the oxygen atoms in the force fields, which leads to positive, repulsive values of the Lennard-Jones potential in Eq. (2). The total energy of the hydrogen bonds in cellulose Iβ obtained as the sum of the gO – gO electrostatic and van der Waals interactions in Table 6 ranges from −3.8 to −7.3 kcal/mol for the force field GLYCAM06$^{\mathrm{TIP5P}}_{\mathrm{OSMOr14}}$, and from −3.8 to −6.6 kcal/mol for GLYCAM06. These ranges of hydrogen bond energies are in agreement with the range from −4.0 to −7.0 kcal/mol estimated based on infrared band shifts for cellulose Iβ[12].

The ranges of the overall electrostatic and van der Waals interchain energies per glucose monomer in Table 2 are comparable to ranges recently obtained from density functional theory (DFT) calculations for cellulose crystals[13]. Depending on the generations of dispersion correction approaches in the DFT calculations, Li et

al.[13] obtained values in the range from −11.7 to −14.8 kcal/mol for the van der Waals interchain energy per glucose monomer in cellulose Iβ, and −12.2 to −15.3 kcal/mol for the van der Waals interchain energy in cellulose II. From energy minimization in the two forced fields considered here, we obtain the range −13 to −16 kcal/mol for the van der Waals interchain energy in Iβ, and −10 to −12 kcal/mol for cellulose II. For the electrostatic interchain energies per glucose monomer, Li et al.[13] obtain the range −11.2 to −12.4 kcal/mol for cellulose Iβ, and −16.7 to −17.9 kcal/mol for cellulose II. The ranges of electrostatic interchain energies obtained from our force field minimizations are −12 to −13 kcal/mol for cellulose Iβ and −19 to −20 kcal/mol for cellulose II.

An important aspect in electrostatic calculations is the long range $\sim 1/r$ of the Coulomb interaction (1) between partially charged atoms. This long range leads to divergences $\sim r_c^2$ in the summation ("volume integration") of the Coulomb interactions of a given partially charged atom of type A with all partially charged atoms of type B in the crystal, where r_c is a length reflecting the crystal size. In contrast, the shorter range $\sim 1/r^4$ of the effective electrostatic interaction of a charge dipole of type A with other charge dipoles of type B in the crystal leads to convergence $\sim 1/r_c$ in the summation of interactions. For this reason, the decomposition of the electrostatic interchain interactions per ring into the group interactions of Table 5 requires to define overall neutral groups of atoms. A central consistency check of this decomposition is that the total sum of all group interchain interactions in Table 5, calculated with "neutralized groups", agrees with the overall interchain interaction energies per ring in Table 2, which were obtained with the original force field charges, within error bounds. Divergences with increasing crystal size can also occur for Coulomb interactions with the charged terminal glucose rings of the cellulose chains, in particular for cellulose Iβ crystals, in which these terminal charges lead to charged surface layers because of the parallel orientation of the chains. In MD simulations, these surface charges are balanced by opposing charges in the surrounding water. In our electrostatic calculations of energy-minimized crystals, we are only interested in bulk energies obtained from linear fits of results for different crystal sizes as in Fig. 3. We therefore safely eliminate any long-range electrostatic interactions with the crystal surfaces by simply neutralizing the terminal glucose ring, which we achieve by adjusting the charges of the terminal H atoms for the electrostatic calculations.

In this article, we have explored the absolute stability of cellulose Iβ and II crystals composed of 52 hexameric chains in MD simulations, and have determined the relative stability of large cellulose Iβ and II crystals based on bulk crystal energies deduced from energy minimizations. Our MD simulations with $\text{GLYCAM06}^{\text{TIP5P}}_{\text{OSMOr14}}$ indicate that the simulated crystals are instable, or in other words, that the free energy of the dissolved state of the 52 chains is lower than the free energy of the intact crystal in this force fields, because the crystals quickly decay in the simulations (see Figs. 1 and 2). In simulations with the standard force field GLYCAM06, in contrast, the crystals are at least metastable on the microsecond simulation timescale. For determining the relative stability of cellulose Iβ and II crystals, it is central to note that the dissolved states of the two crystals are identical, and, thus, also the free energies of these states. Stability differences of cellulose Iβ and II crystals therefore need to result from free energy differences of the crystals, and the bulk energies determined from our minimization approach correspond to such free energies in the limit of zero temperature and large crystal size. In principle, stability differences may also result from kinetic rather than thermodynamic free-energy differences, e.g. from different kinetic, or entropic, bottlenecks in the formation or dissolution of two structures. However, at least for cellulose chains composed of rather few glucose monomers, a larger kinetic barrier for forming cellulose Iβ versus II appears implausible. We have found that the relative stability, i.e. the bulk energy difference, of cellulose Iβ and II crystals is rather similar in the two force fields considered here, despite rather clear differences in absolute stabilities of cellulose crystals observed in MD simulations with these force fields.

In summary, we have determined the interchain and intrachain bulk energies in cellulose crystals from linear modeling of force-field-based minimization results for differently sized crystals. Our calculations allow to quantify the role of electrostatic and van der Waals energies in cellulose crystals and provide new insights on the energetics of hydrogen bonds in the crystals. While the dynamics of hydrogen bonds has been well explored in atomistic simulations of aqueous systems[52,53], standard approaches focusing on donor OH groups and acceptor O atoms do not lead to realistic descriptions of the hydrogen-bond energetics in cellulose crystals[12]. We have shown that including the C atoms to which the OH groups are attached in the calculation of hydrogen-bond energies, for both donor and acceptor atom groups, leads to consistent results for hydrogen-bond energies that agree with estimates based on infrared band shifts for cellulose Iβ[12].

Acknowledgements

This work was funded by the International Max Planck Research School (IMPRS) on Multiscale Bio-Systems and the Max Planck Society.

Availability of data

The MD simulation trajectories and the structures and force-field terms of the energy-minimized cellulose crystals of this work are available in the Edmond data repository at https://doi.org/10.17617/3.1FPW2C[54].

Notes and references

1 S. J. Eichhorn, A. Etale, J. Wang, L. A. Berglund, Y. Li, Y. Cai, C. Chen, E. D. Cranston, M. A. Johns, Z. Fang, G. Li, L. Hu, M. Khandelwal, K.-Y. Lee, K. Oksman, S. Pinitsoontorn, F. Quero, A. Sebastian, M. M. Titirici, Z. Xu, S. Vignolini and B. Frka-Petesic, *J. Mater. Sci.*, 2022, **57**, 5697–5767.

2 R. J. Moon, A. Martini, J. Nairn, J. Simonsen and J. Youngblood, *Chem. Soc. Rev.*, 2011, **40**, 3941.

3 D. Klemm, B. Heublein, H.-P. Fink and A. Bohn, *Angew Chem Int Ed Engl*, 2005, **44**, 3358–93.

4 R. H. Atalla and D. L. VanderHart, *Science*, 1984, **223**, 283–285.

5 Y. Nishiyama, P. Langan and H. Chanzy, *J. Am. Chem. Soc.*, 2002, **124**, 9074–9082.

6 S. Pérez and D. Samain, *Adv. Carbohydr. Chem. Biochem.*, 2010, **64**, 25–116.

7 B. Medronho and B. Lindman, *Curr. Opin. Colloid Interface Sci.*, 2014, **19**, 32–40.

8 P. Langan, Y. Nishiyama and H. Chanzy, *J. Am. Chem. Soc.*, 1999, **121**, 9940–9946.

9 G. Fittolani, D. Vargová, P. H. Seeberger, Y. Ogawa and M. Delbianco, *J. Am. Chem. Soc.*, 2022, **144**, 12469–12475.

10 T. Serizawa, Y. Fukaya and T. Sawada, *Polym. J.*, 2018, **50**, 799–804.

11 M. Wohlert, T. Benselfelt, L. Wågberg, I. Furó, L. A. Berglund and J. Wohlert, *Cellulose*, 2022, **29**, 1–23.

12 Y. Nishiyama, *Philos Trans A Math Phys Eng Sci*, 2018, **376**, year.

13 Y. Li, C. Yan, Y. Chen, X. Han, Z. Shao, H. Qi, X. Li, Y. Nishiyama, T. Hu and P. Chen, *Cellulose*, 2023, **30**, 8127–8138.

14 M. Norgren, C. Costa, L. Alves, A. Eivazi, C. Dahlström, I. Svanedal, H. Edlund and B. Medronho, *Molecules*, 2023, **28**, 4216.

15 M. C. Jarvis, *Cellulose*, 2023, **30**, 667–687.

16 S. Chen, T. Desai, J. A. McNew, P. Gerard and P. J. Novick, *Proc. Natl. Acad. Sci. USA*, 2015, **112**, 418–423.

17 S. P. S. Chundawat, G. Bellesia, N. Uppugundla, L. da Costa Sousa, D. Gao, A. M. Cheh, U. P. Agarwal, C. M. Bianchetti, G. N. Phillips, Jr, P. Langan, V. Balan, S. Gnanakaran and B. E. Dale, *J. Am. Chem. Soc.*, 2011, **133**, 11163–74.

18 P. Chen, Y. Nishiyama and K. Mazeau, *Cellulose*, 2014, **21**, 2207–2217.

19 S. Paavilainen, T. Róg and I. Vattulainen, *J. Phys. Chem. B*, 2011, **115**, 3747–3755.

20 J. A. Hadden, A. D. French and R. J. Woods, *Biopolymers*, 2013, **99**, 746–56.

21 L. Bu, M. E. Himmel and M. F. Crowley, *Carbohydr. Polym.*, 2015, **125**, 146–152.

22 S. Eichhorn and G. Davies, *Cellulose*, 2006, **13**, 291–307.

23 J. Wohlert, M. Bergenstrahle-Wohlert and L. A. Berglund, *Cellulose*, 2012, **19**, 1821–1836.

24 X. Wu, R. J. Moon and A. Martini, *Cellulose*, 2014, **21**, 2233–2245.

25 C. Zhang, S. Keten, D. Derome and J. Carmeliet, *Carbohydr. Polym.*, 2021, **258**, 117682.

26 A. Gupta, A. Khodayari, A. C. T. van Duin, U. Hirn, A. W. Van Vuure and D. Seveno, *Biomacromolecules*, 2022, **23**, 2243–2254.

27 M. Bergenstrahle, L. A. Berglund and K. Mazeau, *J. Phys. Chem. B*, 2007, **111**, 9138–9145.

28 Q. Zhang, V. Bulone, H. Ågren and Y. Tu, *Cellulose*, 2011, **18**, 207–221.

29 K. Kulasinski, S. Keten, S. V. Churakov, D. Derome and J. Carmeliet, *Cellulose*, 2014, **21**, 1103–1116.

30 H. Miyamoto, M. Umemura, T. Aoyagi, C. Yamane, K. Ueda and K. Takahashi, *Carbohydr. Res.*, 2009, **344**, 1085–1094.

31 M. Bergenstrahle, J. Wohlert, M. E. Himmel and J. W. Brady, *Carbohydr. Res.*, 2010, **345**, 2060–6.

32 P. Heasman, A. Y. Mehandzhiyski, S. Ghosh and I. Zozoulenko, *Carbohydr. Polym.*, 2023, **311**, 120768.

33 L. A. Burns, A. Vázquez-Mayagoitia, B. G. Sumpter and C. D. Sherrill, *J. Chem. Phys.*, 2011, **134**, 084107.

34 M. J. Gillan, D. Alfe and A. Michaelides, *J. Chem. Phys.*, 2016, **144**, 130901.

35 P. Dauber-Osguthorpe and A. T. Hagler, *J. Comput. Aided Mol. Des.*, 2019, **33**, 133–203.

36 K. Lindorff-Larsen, P. Maragakis, S. Piana, M. P. Eastwood, R. O. Dror and D. E. Shaw, *PLoS One*, 2012, **7**, e32131.

37 P. Robustelli, S. Piana and D. E. Shaw, *Proc. Natl. Acad. Sci. USA*, 2018, **115**, E4758–E4766.

38 J. Sauter and A. Grafmüller, *J. Chem. Theory Comput.*, 2015, **11**, 1765–1774.

39 J. Sauter and A. Grafmüller, *J. Chem. Theory Comput.*, 2016, **12**, 4375–4384.

40 W. K. Lay, M. S. Miller and A. H. Elcock, *J. Chem. Theory Comput.*, 2016, **12**, 1401–1407.

41 W. K. Lay, M. S. Miller and A. H. Elcock, *J Chem Theory Comput*, 2017, **13**, 1874–1882.

42 K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. Gonzalez-Outeirino, C. R. Daniels, B. L. Foley and R. J. Woods, *J. Comput. Chem.*, 2008, **29**, 622–655.

43 P. Langan, Y. Nishiyama and H. Chanzy, *Biomacromolecules*, 2001, **2**, 410–416.

44 T. C. F. Gomes and M. S. Skaf, *J Comput Chem*, 2012, **33**, 1338–46.

45 D. A. Case, K. Belfon, I. Ben-Shalom, S. Brozell, D. Cerutti, T. Cheatham III, V. Cruzeiro, T. Darden, R. Duke, G. Giambasu, M. Gilson, H. Gohlke, A. Goetz, R. Harris, S. Izadi, S. Izmailov, K. Kasavajhala, A. Kovalenko, R. Krasny, T. Kurtzman, T. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, V. Man, K. Merz, Y. Miao, O. Mikhailovskii, G. Monard, N. Nguyen, A. Onufriev, F. Pan, S. Pantano, R. Qi, D. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. Simmerling, N. Skrynnikov, J. Smith, J. Swails, R. Walker, J. Wang, L. Wilson, R. Wolf, X. Wu, Y. Xiong, Y. Xue, D. York and P. A. Kollman, *AMBER 2020*, University of California, San Francisco, 2020.

46 C. W. Hopkins, S. Le Grand, R. C. Walker and A. E. Roitberg, *J. Chem. Theory Comput.*, 2015, **11**, 1864–1874.

47 J.-P. Ryckaert, G. Ciccotti and H. J. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327 – 341.

48 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.

49 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.

50 R. J. Woods, *Chem. Rev.*, 2018, **118**, 8005–8024.

51 W. Kabsch and C. Sander, *Biopolymers*, 1983, **22**, 2577–2637.

52 A. Luzar and D. Chandler, *Nature*, 1996, **379**, 55–57.

53 R. Kumar, J. R. Schmidt and J. L. Skinner, *J. Chem. Phys.*, 2007, **126**, 204107.

54 R. Kullmann and T. Weikl, *MD simulations and energy minimizations of cellulose crystals. Edmond, V1.* *https://doi.org/10.17617/3.1FPW2C*, 2023.