MODELING OF CELLULOSE CRYSTALS
Alfred D. French
Glenn P. Johnson
USDA-ARS
New Orleans, LA
Robert J. Woods
Complex Carbohydrate Research Center
Athens, GA

Abstract
Several workers have modeled cellulose crystals. The present study concerns the minimum number of chains that can be used to create a useful model of a crystal. A number of the results appear to depend on the modeling software. One aspect that we have explored in some detail is the likelihood of proposed deviations from 2-fold screw axis symmetry in crystalline cellulose I.

Introduction
Models of crystalline cellulose have important roles in many areas of science, ranging from improved chemical treatments of cotton fabric to assisting with the biochemistry of converting crops and their residues into fuel ethanol. A recent effort by Matthews et al. (2006) showed the abilities and limitations of current models. They found that the water molecules were structured in the vicinity of the cellulose, and proposed that the structuring would slow the approach of cellulase molecules and also the escape of the carbohydrate fragments released by the enzyme. It was felt that the water did not affect the cellulose structure, with no substantial penetration of the crystallites.

During their simulation, the cellulose structure changed, with rotation of about half of the O6 atoms from tg to gg, and with small rotations and a gradual twisting of the chains of 14 glucose residues. Those movements resulted in an expansion of the crystal lattice and a conversion of the monoclinic angle to about 90°. Although Matthews et al. agreed that their model unit cell was not consistent with experiments on the usual, Iβ polymorph, they also remembered errors in previous diffraction studies of cellulose. Because molecular dynamics studies by others had found similar features, they concluded that their results were not dependent on their particular force field and left their title as "Computer simulations of microcrystalline cellulose Iβ."

We too have a long-standing interest in cellulose crystals (French et al., 1993). In our view, discrepancies between crystals and their models indicate either that the original structure is incorrect or that the combination of force field and computational method is inadequate to fully reproduce the observed experiment. The purpose of this paper is to clarify that, although Matthews et al. started with cellulose Iβ (Nishiyama et al., 2002) their model cannot describe cellulose Iβ.

We also show some of our recent work.

Methods
X-ray diffraction of a bundle of cotton fibers was carried out by Zakhia Ford on the synchrotron at the Center for Advanced Microstructures and Devices (CAMD), Baton Rouge, Louisiana. Calculations were based on AMBER modeling software and the Glycam-04 parameters designed for carbohydrates. The various AMBER molecular dynamics simulations and energy minimization studies will be described in the text. Results are taken from recent publications for comparisons with the present calculations.
Figure 1. Cotton fiber bundle from the CAMD synchrotron by Zakhia Ford.

Results and Discussion

Diffraction patterns of cellulose Iβ show three very strong reflections, indexed as labeled in Figure 1. If the γ angle were 90° then the –110 and 110 spots would merge. The enlarged unit cell of Matthews et al. has a volume of 722 Å³, while the Nishiyama et al. cell has a volume of 658 Å³. Experimental densities for cotton are expected to be lower than those for crystalline cellulose because of amorphous material in the complex cotton fiber structure. Values are about 1.55 to 1.57 g/cm³, (Davidson, 1927; Morton and Hearle, 1997) compared to 1.602 for Nishiyama et al. and 1.463 g/cm³ for Matthews et al. The lowest density for cellulose IV is 1.51 g/cm³, Aravindanath et al. 1986), so the Matthews et al. cell is also expanded compared to the more variable proposals for cellulose IV.

If half of the O6 atoms are gg, this should be reflected by the CPMAS ¹³C NMR data. Instead, there are peaks at 66.9 and 67.5 ppm, interpreted as the tg orientation (Kono et al., 2002). Besides previous x-ray work that found substantial preference for tg O6 in cellulose I, the current work (Nishiyama et al.) paid extra attention to this point by deleting O6 and then finding it with an Omit map.

The last main discrepancy with experiment is the twist of the molecules as well as a twist in the microfibril. This is more difficult because there are experiments showing twisted microfibrils, (Hanley et al., 1997) but the crystal structure is not twisted. One explanation offered by Matthews et al. was that there is a bias towards structures that do not possess a 2-fold screw axis, as shown by energy surfaces, including ours. Thus, we have some explaining to do!

We did our own simulations, using GLYCAM-04 parameters in AMBER 8, without any solvent. Mini-crystals were built with 2, 7 and 19 (Figure 2) chains having 8 glucose residues, arranged according to Nishiyama, et al. and their first hydrogen bonding system. We wanted to use the smallest reasonable model so simulations could be done for the longest time.
Figure 2. Miniature crystal or microfibril of cellulose Iβ with 19 cellooctaose chains. a) Cross section through the chains. b) View along b-axis, showing the cellooctaose molecules. The unit cell is indicated.

Figure 3 shows the RMS deviations for the different size models. The 19-chain model has a little less all-atom deviation than the 7-chain model but backbone atom deviations were similar. Note the bump at 4 ns, which validates our decision to use long simulations.
Figure 3. Trajectory for RMS values during the 10 ns simulation.

Figure 4. Minicrystal after 10 ns, final snapshot.

Figure 4 shows the model after 10 ns of simulation. As expected, there is less disorder of the chains in the center of the model than for the chains on the surfaces. The central four glucose residues of the 7 central chains are shown in Figure 5. They are quite like the initial structure.

In Figure 3, the bump at 4 ns arose from a reorientation of O6 of the residue on the reducing end of the central chain. It went to the $gg$ position but only for 0.4 ns and then returned to the $tg$ position that is found by the experimental studies. As shown in Figure 4, there is considerable deviation of the 12 external chains from the crystal structure positions. However, the O6s on the seven chains of the core in Figure 5 mostly remained where they started.

The table shows parameters of the simulated crystal, based on the positions of atoms in Figure 5. The unit cell $a$ and $b$ dimensions are slightly shrunken, while the $c$ dimension is somewhat elongated, compared to the experimental values. These discrepancies are due to parameterization. The longer model glucose rings explain most of the
difference in the \( c \) dimension. The monoclinic angle \( \gamma \) is very close to the experimental result, an important difference from the results of Matthews et al.

![Diagram](image)

Figure 5. Tetraose units from the central 7 chains from which measurements below were taken.

Table. Parameters of the simulated crystal, measured from atoms in Figure 5. Values in parentheses are from Nishiyama et al.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a )</td>
<td>7.51 ± 0.05 (7.78) Å</td>
</tr>
<tr>
<td>( b )</td>
<td>8.11 ± 0.07 (8.20) Å</td>
</tr>
<tr>
<td>( c )</td>
<td>10.58 ± 0.03 (10.38) Å</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>97.7º ± 1.3º (96.5º)</td>
</tr>
<tr>
<td>Molecular twist ((C5n-C2n-C2n+2-C5n+2))</td>
<td>2.7º ± 1.5º (0º)</td>
</tr>
<tr>
<td>Crystal twist</td>
<td>1.83º / cellobiose unit (0º)</td>
</tr>
<tr>
<td>( \phi ) origin ((O5'-C1'-O4-C4))</td>
<td>-95.9º ± 1.1º (-98.5º)</td>
</tr>
<tr>
<td>( \psi ) origin ((C1'-O4-C4-C5))</td>
<td>-142.8º ± 1.1º (-142.3º)</td>
</tr>
<tr>
<td>( \phi ) center ((O5'-C1'-O4-C4))</td>
<td>-93.3º ± 0.6º (-88.7º)</td>
</tr>
<tr>
<td>( \psi ) center ((C1'-O4-C4-C5))</td>
<td>-141.9º ± 1.4º (-147.1º)</td>
</tr>
<tr>
<td>( \tau ) origin ((C1'-O4-C4))</td>
<td>117.3º ± 0.4º (115.1º)</td>
</tr>
<tr>
<td>( \tau ) center ((C1'-O4-C4))</td>
<td>116.8º ± 0.4º (116.2º)</td>
</tr>
<tr>
<td>Monomer length O4--O4' origin</td>
<td>5.526 ± 0.01 (5.48) Å</td>
</tr>
<tr>
<td>Monomer length O4--O4' center</td>
<td>5.552 ± 0.01 (5.44) Å</td>
</tr>
</tbody>
</table>

Figure 6 shows the O6 trajectories for the eight residues of the central chain, and for the O6s on an origin chain. In Matthews et al., O6s on center chains changed to the \( gg \) position and mostly stayed there. It is likely that those changes were responsible for the enlarged unit cell in Matthews et al.

The question of a twist in the cellulose chain from the flat ribbon shape that has 2-fold screw axis symmetry is related to a long-standing question as to whether cellulose can have a 2-fold axis. Figure 7 shows a chain segment from the crystal structure in both "ball and stick" and space-filling representations. A helical thread is represented by cyan beads. This thread connects the O6 atoms, but similar threads connect each different kind of atom. By definition, successive residues are related by a rotation about the helix axis of 180º and a translation of 1/2 of the repeat distance.
Figure 6. Trajectories for O6 atoms on the central (upper) and origin (lower) chains, located in the center and to its lower right in Figure 4. All 8 O6 atoms in the two chains are monitored, with the top rows corresponding to the reducing end residues and the bottom rows corresponding to the non-reducing rings of the chain ends.

Figure 7. 2-fold screw axis symmetry and a cellulose segment. The helix axes (x, y and z) are also shown.
Figure 8. Cellobiose model having pseudo 2-fold screw axis symmetry. This structure was fully optimized with B3LYP/6-31+G(d) Density Functional quantum mechanics theory.

The reason for the debate about the 2-fold axis is that the family of structures having 2-fold screw symmetry has parallel C1'-H1' and C4-H4 bonds, placing the H1' and H4' atoms in close contact, as shown in Figure 8. It has been appealing to suggest that the short distance could be avoided by small changes in the \( \phi \) and \( \psi \) torsion angles (see Figure 8). This short distance is probably the reason that our previous energy surfaces showed elevated energies (about 1 kcal/mol) for structures falling on the diagonal line of \( \phi \) and \( \psi \) plots.

The structure in Figure 8, however, has the lowest potential energy of any that we tested with B3LYP/6-31+G(d) quantum mechanics in the region of the crystal structures (French and Johnson, 2006). In fact, two different initial structures led to this same result. They originally were at \( \phi, \psi = -80^\circ, -120^\circ \) and \( \phi, \psi = -120^\circ, -160^\circ \). This structure is very similar to the cellulose structure, with the same \( \text{tg} \) O6 position and hydrogen bonding pattern between the glucose residues. It would appear that the hydrogen bonds help hold the structure in the 2-fold conformation. This would be despite the short contact and the increased glycosidic bond angle of 119º. Based on our calculations with fluorinated cellobiose, (French et al., 2005) however, the 2-fold conformation also has very low energy at the B3LYP/6-31G(d) level, so steric factors may be responsible instead. Our HF/6-31G(d) energy map for cellobiose (French and Johnson, 2006) suggests that the 2-fold screw axis shape corresponds to a few tenths of a kcal/mol. Although this is widely accepted as a journeyman level of QM theory, the above minimization studies with B3LYP/6-31+G(d) theory are probably superior.

**Conclusions**

As have several others, we have carried out molecular dynamics studies starting with cellulose I\( \beta \). The general results are similar to those of Yui et al. (2006) who also carried out AMBER/GLYCAM-04 simulations but with water present. We found that fairly satisfactory results were obtained with a small, 19-chain microfibril, but that previously unseen transitions happened after 3.5 ns, so long simulations are needed.

The idea that the structure of cellulose I\( \beta \) is characterized by a unit cell having angles of 90º is easily ruled out by the diffraction pattern. In addition, our AMBER/GLYCAM-04 molecular dynamics studies retained the initial monoclinic angle despite some contraction of the unit cell dimensions perpendicular to the fiber axis. The increase in the fiber repeat in the model can be mostly attributed to the increases in the length of the individual glucose residues, which in turn are very dependent on bond length parameters.

Similarly, the orientation of O6 in the \( \text{tg} \) conformation was confirmed by this work for the inner glucose residues, as well as the NMR work and several diffraction studies. O6 atoms on the inner residues of our small crystal mostly inhabited their original \( \text{tg} \) locations. The same was not true for the O6 atoms on the surfaces where orientations were also \( \text{gt} \) and \( \text{gg} \), as found by others.
The question of twisting is more difficult. At this time, it appears that the twisted structures from dynamics studies are also force-field dependent. Our best QM calculation shows that 2-fold symmetry is at the energy minimum for extended structures, and other QM calculations suggest an energy of just a few tenths of a kcal. Given that the twists measure just a degree or two per glucose residue and the need to balance the many forces that govern the structure, it will be some time before there is agreement on this point.

What is needed? Better calculations. Several issues are not accounted for in GLYCAM-04, including polarization and C-H...O hydrogen bonding as proposed by the x-ray work. The hourglass contour on the GLYCAM-04 map for THP-O-THP indicates lack of parameters for the external anomeric torsional effect (Lii et al., 2005). On the other hand, the QM results are not perfect, either. Problems include hydrogen bonds that are too long to match condensed-phase results, and the need for better treatment of electron correlation.

References


Davidson, G.F. (1927) J. Textile Inst. 18, T175, T275.


