UPDATE ON MODELS OF CELLULOSE CRYSTALS

Alfred D. French
Glenn P. Johnson
USDA-ARS
New Orleans, LA

Abstract

This paper describes progress in the computational modeling of cellulose crystals since our 2007 report. These crystal models are needed to better understand the interactions of cotton cellulose with water, enzymes and chemical finishing agents. Previous models resulted from molecular dynamics simulation, but many of the differences between model and experiment could be revealed with much less computer time by the energy minimization technique. The deviations of our models, and those of other workers, could be described in terms of twists of both the individual molecules and the entire crystal. Other deviations described last year have been accepted by the original authors as artifacts of the modeling techniques. The present work describes the twisting of the molecules in terms of deviations from a two-fold screw-axis, and presents results from a periodic boundary calculation with quantum mechanics.

Introduction

Our previous model cellulose crystals (French and Johnson, 2007) were similar to those of other workers who also assembled short cellulose chains, e.g., eight glucose residues long, in arrays based on the published crystal structure of cellulose Iβ (Nishiyama, Langan and Chanzy, 2002). A common problem is that those cellulose chains tended to twist somewhat (Figure 1) in the modeling studies, and it was understood that this also caused the crystal itself to twist, based on the forces on the atoms from the empirical force fields and room-temperature molecular dynamics (MD). This twisting from the untwisted experimental crystal structure was found for both the CSFF force field used in the CHARMM program by Matthews et al. (2006), and for the GLYCAM-04 force field used in the AMBER program. The latter system was used by both Dr. Yui's group (Yui et al., 2006) and ourselves. Additional problems of large changes in the unit cell dimensions that arose from rotation of the hydroxymethyl groups on half of the model cellulose chains, as well as the rotations themselves, have subsequently been accepted by the original authors as artifacts of the CSFF force field (Brady, 2007). Since carrying out our work reported last year, we have acquired increased computer capacity and speed, so we took the classic approach to the problem of reducing the assumptions and trying more sophisticated modeling systems to learn if more of the results are likely to be due to artifacts of the modeling method.

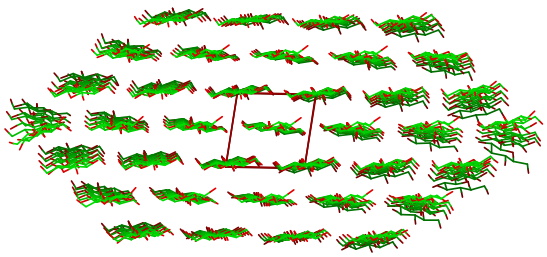

Figure 1. Cellulose model with 37 cellooctaose chains after minimization by AMBER 9 with the GLYCAM-04 parameters. The view is down the axes of the chains, and the *a*- and *b*- axes of the unit cell are shown. The hydrogen atoms are not shown.

Our hypothesis was that the twisting of the model chains was due to an inadequate force field, the collection of equations that describe the various interatomic relationships in molecular mechanics programs such as AMBER. In the current crystal structure of cellulose Iβ, the molecules conform to two-fold screw-axis symmetry. The twists after application of the modeling force field result in deviations from that symmetry. To some extent, they disrupt

the entire crystal, but there is more disruption on the surfaces of the crystal than in the interior. We think that such twisted structures would not account for the experimentally observed diffraction pattern (we hope to test that soon).

Many workers think that two-fold screw conformations are inherently higher in potential energy than structures that deviate from that symmetry. Therefore, cellulose chains would not possess true symmetry, even in the crystal. For example, nearly 20 years ago (French, 1989), we considered that idea, and a variation was published more recently (Viëtor et al., 2000). As we showed last year, the screw axis places the hydrogen atoms on the C1' and C4 atoms in short contact. This causes high energy when calculated by some methods. However, our structure of cellobiose used to illustrate that conformation was the result of energy minimization at a good level of quantum mechanics. In other words, a very good modeling method found the two-fold structure to be at a minimum in the energy. Therefore, it is by no means certain that close contacts of these hydrogen atoms will result in higher energy than calculated for other conformations. We are planning another paper that will discuss the calculated and high-accuracy experimental structures that provide further evidence for the viability of the two-fold screw axis. The present work documents some of our efforts to demonstrate that the likely molecular and crystal shapes are dependent on the particular choice of modeling method.

Approaches

We have observed over the years that strong hydrogen bonding forces in many modeling systems can interfere with the prediction of the full range of observed structures. This is especially the case for prediction of molecular shapes in the condensed phase when the model is studied in isolation, i.e., the gas phase (Johnson, Stevens and French, 2007). (Unfortunately, it is not yet feasible to predict molecular shapes in crystalline solids using models that include neighboring molecules.) Therefore, we tried several different force fields and modeling software combinations for modeling the miniature crystals, such as shown in Figure 1, as well as for modeling the isolated cellooctaose chains and just cellobiose. In these calculations, we used different values of the dielectric constant (ε) as a simple method to adjust the balance between the forces from hydrogen bonding and from other force field components. Raising the dielectric constant decreases the forces arising between atoms that have different charges because of the equation $F = q_a * q_b / \varepsilon$, where the values of q are the charges on atoms a and b. Thus, if a particular, hydrogen-bonded pair of atoms is stabilized by 5 kcal/mol at $\varepsilon = 1$, that interaction will only have a hydrogen bonding energy of 1 kcal/mol at $\varepsilon = 5$. This procedure is not widely accepted in the modeling community because it brutally overrides the careful parameterization of the force field. In the case of the prediction of condensed phase structures from isolated molecules, however, the reduced hydrogen bonding strength may compensate for the absence of neighboring molecules in the calculation. In the condensed phase, the importance of intramolecular hydrogen bonding may be reduced because of the possibility of forming intermolecular hydrogen bonds that often have superior geometries, compared to the intramolecular arrangements. In the case of crystal structures, arguments could be made for reduced hydrogen bonding strength because of their elevated dielectric constants, compared to vacuum.

Unlike the previous report, we have used energy minimization in the present work instead of molecular dynamics (MD) simulation. Based on previous work (Yui, 2006) that suggested that the end result was similar, we temporarily abandoned MD. Ultimately, MD will be needed to take into account the water that will surround and interact with the crystal models, but for the present, we feel that minimization can find flaws in a particular method without the added computational expense of MD. In the case of the AMBER program, we used the LBFGS method of the XMIN minimizer. The XMIN minimizer is part of the Low-Mode methods package (Kolossváry, 1996).

We nearly doubled the size of our model crystals from the work in our 2007 report, increasing the number of chains from 19 to 37, keeping a model crystal shape that keeps the surface molecules in close contact with as many interior molecules as possible. Another advance from last year was the addition of periodic boundary calculations with quantum mechanics (QM) calculations. This technology permits the modeling of just the repeated cellobiose moiety of the unit cell so that there are no end effects. It also reduces the number of atoms in the calculation so that QM methods, which take at least 10^5 times as long as the empirical methods or more, can be successfully applied. Even so, only two-dimensional calculations were feasible with our current resources.

Besides improving the model crystals, we have also done supporting calculations. In particular, we relied on all 181 arrangements of hydroxyl group orientations that we studied with QM (French and Johnson, 2006) to make energy surfaces for cellobiose with AMBER, CHARMM, MM3 and MM4. In other calculations, we minimized isolated

cellooctaose chains started in the cellulose Iβ crystal structure conformation to learn whether the molecule would twist, independent of its neighbors. The point in both cases was to see how variations in the force field, including the dielectric constant, affected the results. There is certainly considerable variation but most of the detailed results are not presented.

Results and Discussion

Figure 2 shows the problem with high energy for 2-fold screw-axis structures, based on cellobiose conformations. The AMBER/FF99 map with ε = 1.0 has a local minimum in the energy some 60° to the left of the 2-fold screw-symmetry line. That minimum is also somewhat distant from the observed crystal structures. The lowest energy encountered by the line representing 2-fold screw-axis structures is over 6 kcal/mol. The map based on AMBER9/GLYCAM-04 at ε = 8.0 seems much more predictive of the crystal structures, and the 2-fold line encounters a much more reasonable minimum energy of about 0.4 kcal/mol. Perhaps the most remarkable feature of the ε = 8.0, map is its similarity in the central region to the HF/6-31G* QM map that we made with the same starting geometries. One might have expected the AMBER9 map made with ε = 1 to better resemble the QM map. These results certainly indicate that an elevated ε might be useful in studies of cellulose crystals. Unlike our QM map, the central minimum with the crystal structures has the lowest energy on our ε = 8 map, another advantage of an elevated dielectric constant. With QM and with the AMBER9 map with ε = 1, the lowest energy is found near the centers of the side edges.

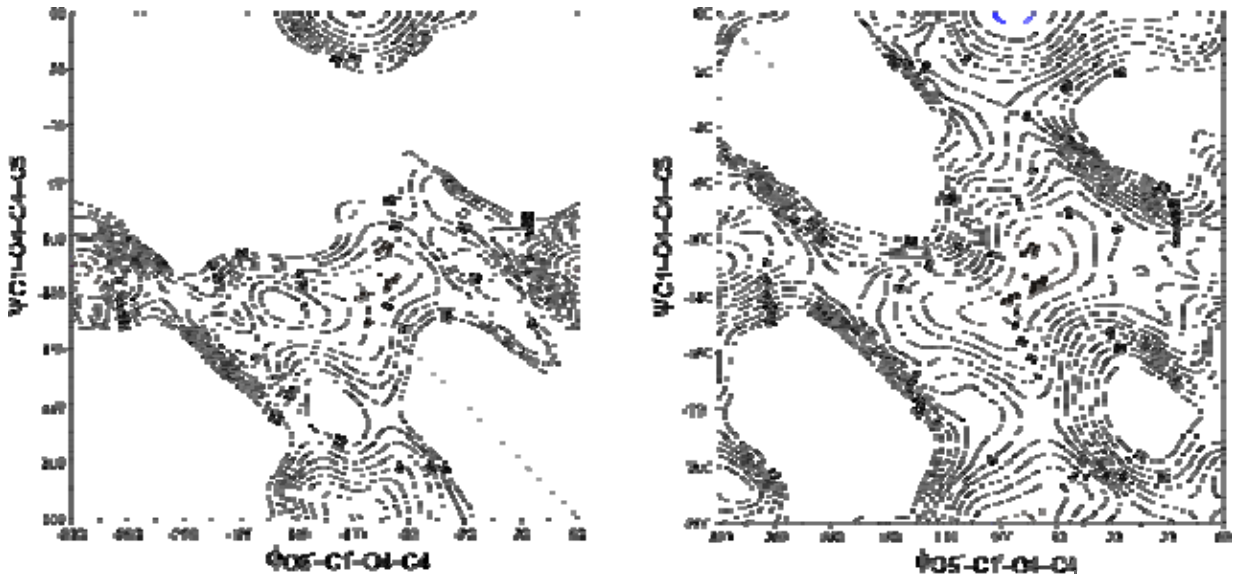

Figure 2. Phi (φ), Psi (ψ) (or Ramachandran) energy surfaces for β-cellobiose, based on the same 181 variations of the hydroxyl and hydroxymethyl group orientations used in our QM study. The dashed diagonal lines represent structures having 2-fold screw pseudosymmetry. The dots represent the experimentally observed values of φ and ψ in crystal structures of small molecules such as cellobiose and the related lactose molecule. The contour lines are marked in kcal/mol of relative energy, and the contour line inside the 1 kcal/mol line is at 0.25 kcal/mol. Contours above 12 kcal/mol are not shown. The map on the left was created using the AMBER9 program and the FF99 force field (ε = 1.0), while the map on the right was created with the same program but the GLYCAM-04 force field specifically intended for carbohydrate molecules and with ε = 8.0.

Figure 3 shows isolated cellooctaose chains, projected down the helix axis, optimized with AMBER9/GLYCAM-04 and six different values of ε. It is interesting that the models that incorporate the three highest ε values are closest to 2-fold screw conformations as evidenced by the near overlap of the atom positions. Although this is not our viewpoint, many workers associate the 2-fold screw conformation with strong intramolecular hydrogen bonding between O3 and O5' and between O6 and O2' on adjacent glucose residues. In any case, the results again show that a 2-fold screw conformation is strained in the AMBER9/GLYCAM force field when ε = 1.0, the usual value for an isolated molecule in that system. Table 1 shows that the unit cell is somewhat collapsed in the ***a*** and ***b*** directions with AMBER/GLYCAM-04 and low dielectric constants, regardless of using MD or minimization.

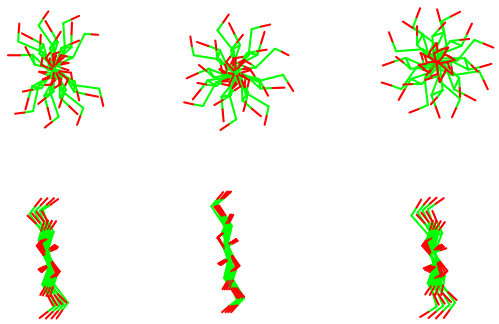

Figure 3. Cellooctaose chains minimized with AMBER/GLYCAM-04 and ε = 1, 2, 4, (top left to top right), and 6, 8 and 10 (bottom left to bottom right). The view is down the chain axis, and hydrogen atoms are not shown. The numbers of residues per turn are: –2.19, –2.28, –2.69, –2.03, 2.01 and 2.04, respectively, where the negative sign indicates a left-handed helix; positive values are for right-handed structures.

Table 1. Unit cell dimension discrepancies (Å and °) for cellulose Iβ models

Method	Cell axis lengths and angles	***a***	***b***	***c***	α	β	γ
MD	Matthews* CHARMM/CSFF	0.69	-0.09	0.13	0.0	0.0	-6.5
MD	AMBER/Glycam MD (ε = 1.0)	-0.27	-0.09	0.20	0.0	0.0	1.2
Minimize	AMBER/Glycam (ε = 1.0)	-0.25	-0.04	0.19	0.1	-0.2	0.8
Minimize	AMBER/Glycam (ε = 2.0)	-0.15	0.12	0.20	0.0	-0.2	1.6
Minimize	AMBER/Glycam (ε = 6.0)	0.12	0.17	0.19	-1.3	1.2	6.4
Minimize	HF/6-31G* (2-dimensions)	na	0.07	0.03	0.1	na	na
Minimize	PBEPBE/6-31+G** (2-d)	na	0.02	0.08	0.1	na	na

In our quest to make a model that is satisfactory from all standpoints, we also considered QM. We have tried both Self-Consistent Field theory (HF/6-31G*) and Density Functional Theory (PBEPBE/6-31+G**). Instead of the fairly large but finite minicrystal models above, we instead used a relatively new capability in the Gaussian-03 program that allows periodic structures to be studied. In this work the model consisted of cellobiose that was periodically reproduced in both the fiber axis and 010 sheet directions, with no control in the *a*-axis direction. Only the center chain in the unit cell was studied; these calculations are very expensive. Full 3-dimensional calculations are impractical at the moment but eagerly anticipated.

Of the two quantum methods, we consider the PBEPBE results shown in Figure 4 to be somewhat better than the HF results because of the better hydrogen bond lengths in the minimized model. Some of these hydrogen bonds are parallel to the b-axis of the unit cell, and the unit cell *b*-dimension in Table 1 shows the expansion. On the other hand, the *c*-dimension increased more with PBEPBE, so neither model is entirely satisfactory. Compared to the reported experimental values for the hydrogen bonding, the O···O distances increased when using PBEPBE, but the H···O distances diminished. There was a small amount of chain rotation, most easily seen in the "Before" and "After" views that show the projections down the cellulose chain axis.

Conclusions

Cellulose models are under continual development. People are hard at work on the empirical force fields used in AMBER and CHARMM, and as mentioned, MM4 is also undergoing revisions to improve hydrogen bonding. The above work shows that the energy penalty for adherence to a 2-fold screw axis is very dependent on the particular force field (a change in dielectric constant is effectively a change in the force field.) Besides the work with empirical force fields, we are now able to apply quantum mechanics to cellulose models, although it will be some time before many of the issues with cellulose structure can be reliably resolved with QM.

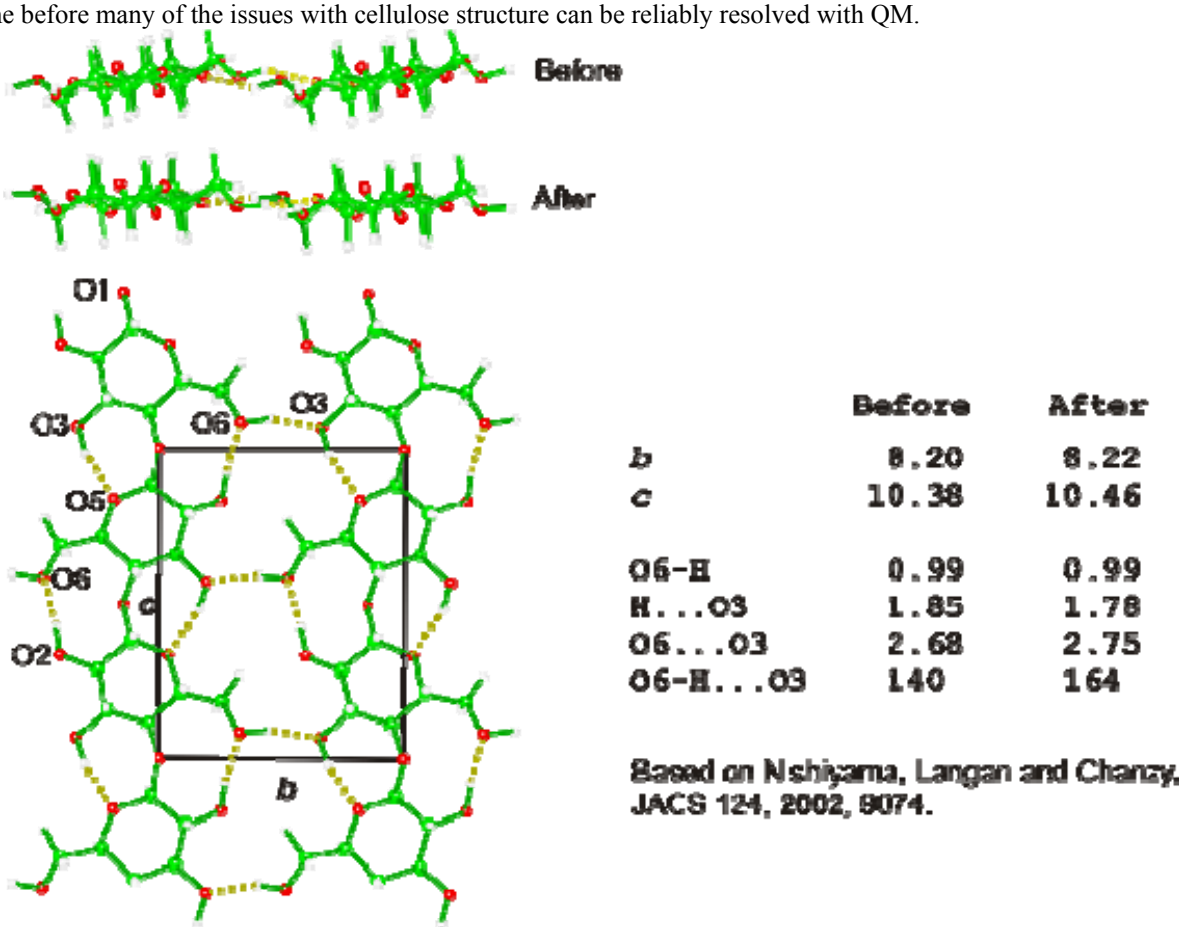

Figure 4. Two cellulose chain segments after energy minimization with PBEPBE/6-31+G**, and some geometric changes for the unit cell *b*- and *c*-dimensions. The indicated unit cell is only for measurement purposes; the *c*-axis should be shifted slightly to the left coincide with the 2-fold screw axes.

References

Brady, J. W. (2007) Lecture at the International Cellulose Conference, October 23, Tokyo, and following discussion.

French, A. D. (1989) Cellulose and Wood - Chemistry and Technology, Scheurch, C., ed., Wiley and Sons. pp. 103–118.

French, A. D., Johnson, G.P. (2006) *Can. J. Chem.* 84, 603–612.

French, A. D., Johnson, G. P. (2007) Proceedings of the Beltwide Cotton Conferences, 1261–1268.

Johnson, G. P., Stevens, E. D., French, A. D. 2007. *Carbohydr. Res.* **342** 1210–1222.

Kolossváry, I., Guida, W. C. (1996) *J. Am. Chem. Soc.* **118** 5011–5019.

Matthews, J. F., Skopec, C. E., Mason, P. E., Zuccato, P., Torget, R. W., Sugiyama, J., Himmel M. E., Brady, J. W. (2006) *Carbohydr. Res.* **341** 138–152.

Nishiyama, Y., Langan, P., Chanzy, H. (2002) *J. Am. Chem. Soc.* **124**, 9074–9082.
Viëtor, R. J., Mazeau, K., Lakin, M., Pérez, S. (2000) *Biopolymers* **54** 342–354.

Yui, T., Nishimura, S., Akiba, S., Hayashi, S. (2006) *Carbohydr. Res.* **341***,* 2521–2530