Atomic and Electronic Structures of Molecular Crystalline Cellulose Iβ: A First-Principles Investigation

Xianghong Qian,*,† Shi-You Ding,‡ Mark R. Nimlos,‡ David K. Johnson,‡ and Michael E. Himmel‡

Rx-Innovation, Inc., Fort Collins, Colorado 80525, and National Bioenergy Center, National Renewable Energy Laboratory, Golden, Colorado 80401

Received July 29, 2005; Revised Manuscript Received October 12, 2005

ABSTRACT: A theoretical model based on the competition between hydrogen-bonding energy and strain energy was constructed to explain the size of native cellulose Iβ. The cellodextrins in native crystalline cellulose Iα and Iβ are unusually stable compared to other polysaccharides, not easily prone to hydrolysis even though they are only nanometers in diameter. The stability of crystalline cellulose Iβ is most likely due to its greatly enhanced hydrogen-bonding (HB) network. We carried out ab initio calculations to determine the native crystalline cellulose Iβ atomic and conformational structures. For crystalline cellulose, we found that every hydroxyl group in the cellulose structure is hydrogen bonded as both a donor and an acceptor. This agrees well with published X-ray and neutron diffraction data. We also determined the electronic structures and the energetics for one cellodextrin chain, one to four sheets of cellodextrins in cellulose, and the bulk cellulose Iβ.

I. Introduction

Native crystalline cellulose consists of two phases, Iα and Iβ. Both are frequently found to coexist in cell wall structures together with amorphous cellulose.[1−10] Bacterial and algal celluloses are predominantly of the Iα type, whereas higher plants and tunicate celluloses are mostly of the Iβ type. Both cellulose Iα and Iβ are metastable and can only be synthesized by the living organisms.[11] Thermodynamically, the cellulose Iβ format is found to be more stable than the Iα.[12−14] Cellulose Iα was converted to Iβ by annealing at around 200 °C in a number of different solvent media.[12−15] In addition, there exist several mainly synthetic crystalline cellulose allomorphs II, III, and IV, which differ vastly from native cellulose in their atomic conformational structures.[7,16−20] Even though it is well-known that the hydrogen-bonding interaction is the main binding force for maintenance of these molecular crystals, the detailed hydrogen-bonding networks in native crystalline cellulose remained elusive until recently. This situation was due primarily to the coexistence of both cellulose Iα and Iβ in most plant cell walls and the small size of the microfibrils with which they are associated, typically only several nanometers in width.[4,21−25] As a result of recent synchrotron X-ray and neutron diffraction analyses of native cellulose Iα and Iβ by Nishiyama and coworkers,[7,8] the basic atomic structures and hydrogen-bonding networks for these cellulose forms are now known experimentally.

Cellulose Iα was found to have a triclinic P1 structure with one cellobiose chain in each unit cell (a = 6.717 Å, b = 5.9962 Å, c = 10.400 Å, α = 118.08°, β = 114.80°, and γ = 80.37°), whereas cellulose Iβ is found to have a monoclinic $P2_1$ structure with two cellobiose chains in each unit cell (a = 7.784 Å, b = 8.201 Å, c = 10.380 Å, $\alpha = \beta$ = 90°, γ = 96.5°). Both crystalline cellulose Iα and Iβ are formed by stacking planar cellulose sheets, but in different ways. The cellulose sheets are in turn composed of linear cellulose chains bounded by the interchain hydrogen-bonding interactions between the O6H (donor) in one chain and O3 (acceptor) in the neighboring chain. Besides interchain hydrogen-bonding interactions, there are intrachain hydrogen bonds between O3H and the ring O5 and between O2H and O6. For cellulose Iβ, the lattice a direction is the cellulose sheet stacking direction and the b direction is perpendicular to the chain direction in the sheet plane. The c direction is the chain direction with ∼10.4 Å cellobiose repeating unit length in both Iα and Iβ. The two chains as shown in Figure 1 in cellulose Iβ unit cell lying on two neighboring sheets, designated as the origin and center chains respectively, are conformationally different. The center chain is shifted in the chain direction by $^1/_4c$. In cellulose Iα, the chains lying on the two neighboring sheets are also shifted with respect to the c direction by $^1/_4c$. However, these two chains are conformationally identical to each other. The stacking order of cellulose sheets in Iβ is *ABAB*..., whereas in cellulose Iα is *ABCABC*.... The binding forces between the sheets were thought to be mainly of van der Waals interaction for both Iα and Iβ; however, substantial intersheet C−H···O hydrogen-bonding interactions were also believed to play a role in the cohesion of these cellulose sheets.[7] The differences between Iα and Iβ are very subtle, and both have almost the same density. Cellulose Iβ has a slight higher density than that of Iα. It is perhaps important to note that Iβ has a slightly compressed unit length in the c (chain) direction than Iα (10.38 Å in Iβ vs 10.40 Å in Iα).[7,8] This indicates that one of the native cellulose allomorphs is more stressed than the other and that could affect the stability of their corresponding structures. A recent investigation by Ding and Himmel[26] suggested that the cellulose microfibril in plant cell walls is composed of 36 cellulose chains with a total of six sheets only. Of these 36 chains, only the inner chains are crystalline in nature, whereas the outer ones are noncrystalline. The crystalline part of the inner cellulose of the microfibril is dominated by Iβ

† Rx-Innovation, Inc.
‡ National Renewable Energy Laboratory.
* Corresponding author.

10.1021/ma051683b CCC: $30.25
Published on Web 11/12/2005

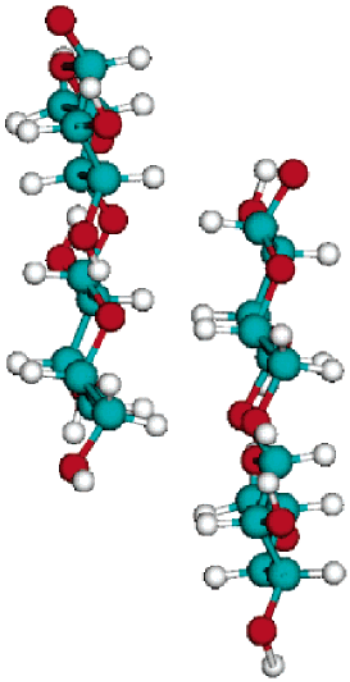

Figure 1. Unit cell containing two chain cellulose Iβ with a monoclinic $P2_1$ symmetry.

structure. In addition, it is known from small-angle neutron scattering studies[27] that there can exist alternating crystalline and noncrystalline regions on cellulose crystals along the long axis. The length of the crystalline and noncrystalline regions are only about 300 glucose residues and 4−5 residues, respectively.[27] Native crystalline cellulose Iα and Iβ can be converted to other more stable crystalline cellulose allomorphs II, III, and IV through heating or chemical treatment.[7,16−20] The difference between native and synthetic crystalline cellulose fibers lies mainly in their hydrogen-bonding networks. Strong O−H···O hydrogen bonds were also found between the neighboring sheets in synthetic crystalline cellulose fibers.[7,16−20,28] Since cellulose Iβ is the main crystalline allomorph in higher plants, our initial theoretical modeling study will focus mainly on cellulose Iβ.

Despite the experimental elucidation of native crystalline cellulose Iα and Iβ structures, there remain a number of important questions concerning the unusual stability and resistance to hydrolysis of cellulose in plant cell walls. Examples are as follows: (1) What is the nature of the interaction between the neighboring cellulose sheets? (2) Why is it so difficult to obtain pure allomorphs of cellulose Iα and Iβ rather than a mixture of the two forms with various proportions? (3) Why is native cellulose so recalcitrant to hydrolysis even though crystals larger than 10−20 nm in diameter are rarely observed? (4) What is the role of disordered amorphous cellulose in cell wall structures? (5) Finally, why is there a biphasic behavior for enzymatic hydrolysis of cellulose, where a slower conversion is followed by generally an initial fast digestion? Does this phenomenon arise from the differences between different regions of the substrate cellulose structure? To shed some light on these questions, theoretical understanding and molecular modeling studies become indispensable. The structures of experimentally determined native crystalline cellulose Iβ[7] became the basis for our current theoretical and modeling investigation.

II. Computational Procedures

The atomic and electronic bulk structures of native cellulose Iβ were investigated on the basis of experimentally determined atomic structures using the ab initio Car−Parrinello[29] molecular dynamics simulation code CPMD.[30] CPMD is based on density-functional theory, which treats the semicore and valence electrons quantum mechanically. Core electrons are frozen during the course of simulation because they contribute very little to the interatomic potential. The nuclei are treated using classic mechanics. CPMD is capable of both static structural optimization and dynamics calculations. A plane-wave basis set was used to construct the electronic wave functions. CPMD was adopted in an earlier study to investigate the crystalline glucose atomic and hydrogen-bonding structures.[31] A good agreement was found between the theoretical results and experimental data. In our study, the unit cell contains two cellobiose chains with periodic boundary conditions of $P2_1$ symmetry, as shown in Figure 1. For the bulk cellulose Iβ, the initial electron wave functions for the experimentally derived structure (a = 7.784 Å, b = 8.201 Å, c = 10.380 Å, $\alpha = \beta = 90°$, $\gamma = 96.5°$) were optimized. This structure was subsequently annealed at 300 K for 250 steps (~32 fs) with confined symmetry. It was then quenched at 0 K and the atomic structure optimized. The energy and force convergence criteria of 1.0×10^{-5} Ha and 1.0×10^{-5} N/m were adopted. On the basis of the optimized structure, the atomic structure and the hydrogen-bonding network were compared with the experimental data. In addition, the energies of electron orbitals including highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) were examined. Moreover, the lattice dependence of the energy was also investigated. The HOMO and LUMO orbital energy variations were examined as a function of the lattice spacing.

Since the actual native crystalline cellulose Iβ is only nanometers in diameter, systems with one chain, one sheet, two sheets, and three and four sheets were also investigated. The atomic structures, hydrogen-bonding networks, and the HOMO and LUMO orbital energies were compared with each other and with the bulk crystalline cellulose Iβ. The same procedure was applied to that of bulk cellulose. To study the binding energy between the cellulose chains in one cellulose sheet, and the binding energy between the sheets, the distances between the chains and between the sheets were varied and the energies of their corresponding systems were determined and examined as a function of the interchain and intersheet distances.

An electron mass of 800 au and a time step of 0.125 fs were used in these calculations. The Becke, Lee, Yang, and Parr (BLYP)[32,33] gradient corrected functional was used. Only the energy at the Γ-point was calculated. The plane-wave basis set cutoff is 70 Ry, which was shown[34−36] to be sufficient for biomolecular simulations.

III. Hydrogen-Bonding Structures in Cellulose Iβ

Figures 2−5 exhibit the theoretically optimized hydrogen-bonding structures of bulk crystalline cellulose Iβ and the corresponding results from the X-ray and neutron diffraction study[7,8] for both the origin and center chains, respectively. The hydrogen-bonding (O−H···O) cutoff criteria are 3.0 Å for the bond length and 160° for the bond angle for the theoretically determined

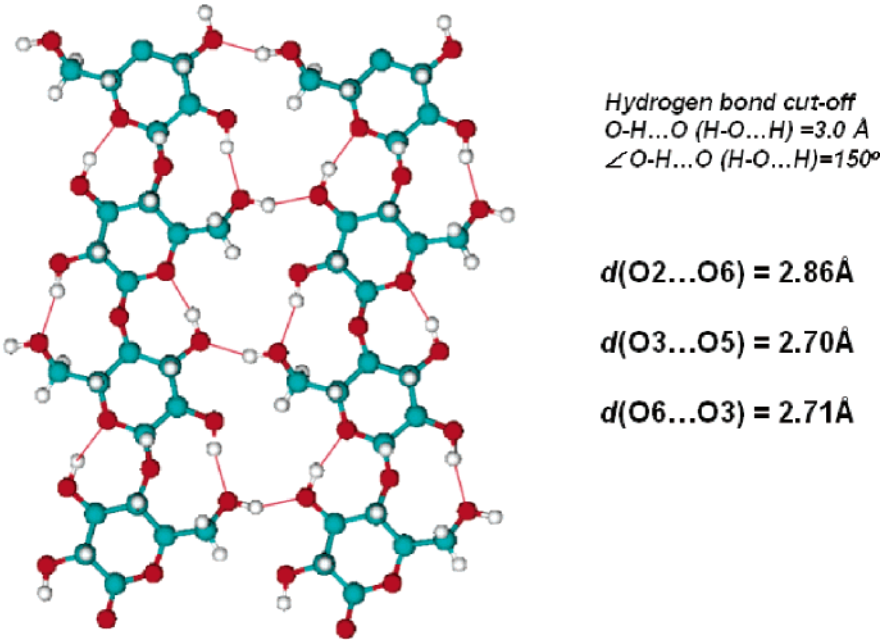

Figure 2. Experimentally determined hydrogen-bonding structure (O−H···O) of the center chain in native crystalline cellulose Iβ. The cutoffs for the hydrogen-bonding interaction are 3.0 Å for the bond length and 150° for the bond angle.

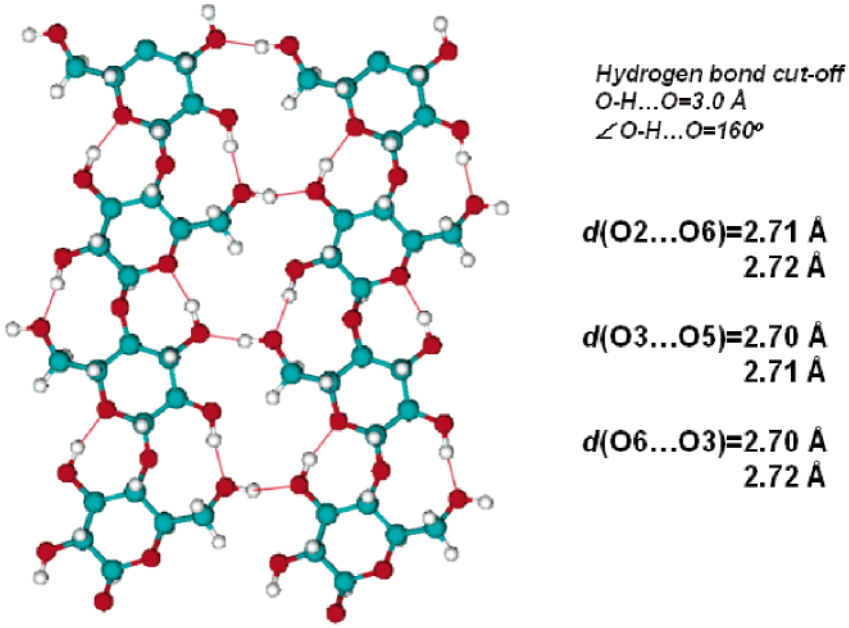

Figure 3. Calculated hydrogen-bonding structure (O−H···O) of the center chain in native crystalline cellulose Iβ. The cutoffs for the hydrogen-bonding interaction are 3.0 Å for the bond length and 160° for the bond angle.

structures, which means that only hydrogen bonds shorter than 3.0 Å and ∠O−H···O ≥ 160° are shown. The hydrogen bond length cutoff for the experimentally resolved structures is the same as the calculated data at 3.0 Å, while the hydrogen bond angles are 150° and 130° for the center and origin chains, respectively, in order to show all the intrachain and interchain hydrogen bonds.

For the center sheet, the intrachain experimental hydrogen bond length O2−H···O6 and O3H···O5 are 2.86 and 2.70 Å, respectively. The interchain hydrogen bond length O6−H···O3 is 2.71 Å. The theoretical intrachain hydrogen bond lengths are 2.71 and 2.72 Å for O2−H···O6 and 2.70 and 2.71 Å for O3−H···O5. The interchain values are 2.70 and 2.72 Å for O6−H···O3. Since cellulose chains are polymers of the cellobiose unit, there exist two slightly different sets of hydrogen bonds for all of the hydrogen-bonding interactions. Experimentally, only one set of hydrogen-bonding lengths were identified due to the very similar values for each type of hydrogen bond lengths. It can be seen that the theoretical intrachain O3−H···O5 and interchain O6−H···O3 hydrogen bond lengths agree very well with those measured experimentally. The theoretical intrachain O2−H···O6 hydrogen bond length, however, is much smaller than the experimental value by almost 0.15 Å. This means that the intrachain O2−H···O6 hydrogen bonds in the experimental sample are much weaker than their corresponding theoretical ones. All bulk theoretical hydrogen-bonding angles as shown in Table 1 are larger than their corresponding experimental values. These data show that there may be consider-

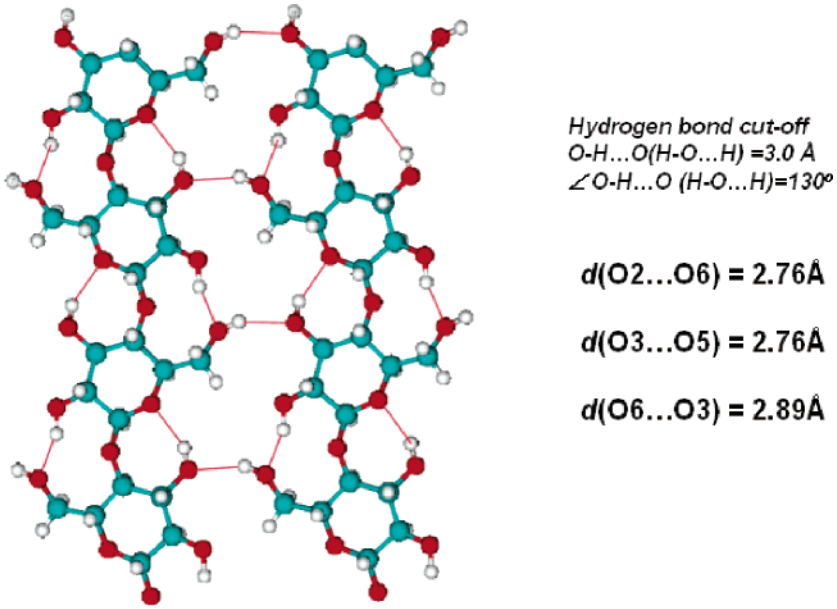

Figure 4. Experimentally determined hydrogen-bonding structure (O−H···O) of the center chain in native crystalline cellulose Iβ. The cutoffs for the hydrogen-bonding interaction are 3.0 Å for the bond length and 130° for the bond angle.

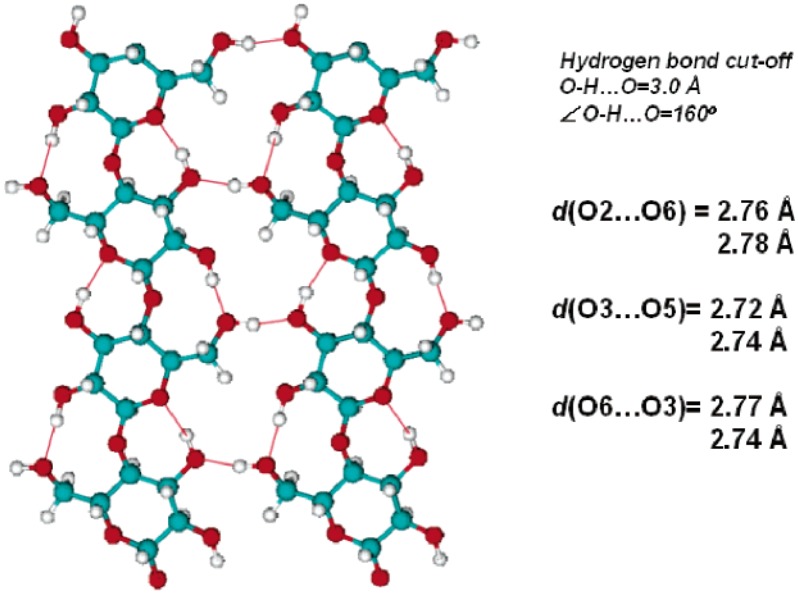

Figure 5. Calculated hydrogen-bonding structure (O−H···O) of the origin chain in native crystalline cellulose Iβ. The cutoffs for the hydrogen-bonding interaction are 3.0 Å for the bond length and 160° for the bond angle.

Table 1. Comparison of Hydrogen Bond Lengths and Angles between Experiment and Theory for a Single Chain, One Isolated Sheet, the Origin Sheet (OS), and the Center Sheet (CS) in the Bulk Crystalline Cellulose Iβ

	OS (exp) (Å)	CS (exp) (Å)	1 chain (Å)	1 sheet (Å)	OS (theor) (Å)	CS (theor) (Å)
$d_{O2-H\cdots O6}$	2.76	2.86	2.96, 3.00	2.77, 2.79	2.76, 2.78	2.71, 2.72
$d_{O3-H\cdots O5}$	2.76	2.7	2.79, 2.87	2.82, 2.84	2.72, 2.74	2.70, 2.71
$d_{O6-H\cdots O3}$	2.89	2.71		2.77, 2.74	2.68, 2.69	2.70, 2.72
$d_{O6-H\cdots O2}$		3.21		3.27		3.31, 3.34
$\angle_{O2-H\cdots O6}$	150, 159	152, 165	171, 175	173, 170	171, 173	167, 170
$\angle_{O3-H\cdots O5}$	137	162	149, 158	158, 163	162, 164	164, 164
$\angle_{O6-H\cdots O3}$	144	157		162, 164	168, 168	166, 167

able disorder in the experimental crystal samples used in the X-ray and neutron scattering measurement. There are several possible reasons to explain this difference between experiment and theory. The first one is that the crystal samples used in experiments are only several nanometers in size. We know that the surface and interface effect become very dominant in nanometer scale materials. Since the O6H hydroxyl group is conformationally most free to rotate, the surface cellulose chains minimize their energies by changing their conformational structures of the O6H and other hydroxyl groups to a lesser degree. As a result, a relatively longer O2−H···O6 hydrogen bond is observed experimentally. Indeed, all hydrogen-bonding angles in the experiments were found to be smaller than their corresponding theoretical ones. The second reason is that theoretical structures were optimized at 0 K due to the temperature limitations of the current density-func-

tional theory (DFT). The experimental measurements were carried out at room temperature. It is known that materials at higher temperature generally have more disorder than those at lower temperature (the Derby−Waller factor), thereby accounting for the hydrogen-bonding structure difference between theory and experiment. The third reason is the entropy effect, which is related to the second reason. At a certain finite temperature, there is an advantage to increase the disorder of the system, i.e., entropy, because the term $T\Delta S$ in free energy $\Delta G = \Delta H - T\Delta S$ will be able to more than compensate for the increase of enthalpy ΔH as a result of more disorder, equivalent here to weaker hydrogen-bonding interactions in cellulose.

For the origin chain, the hydrogen-bonding network is different from that of the center chain. The hydrogen bond lengths are 2.76 and 2.78 Å for the theoretical intrachain O2−H···O6 bond and 2.72 and 2.74 Å for the intrachain O3−H···O5 bond. The theoretical hydrogen bond lengths are 2.68 and 2.69 Å for the interchain O6−H···O3 bond. Experimentally, the intrachain O2−H···O6 and O3−H···O5 hydrogen bond lengths are the same at 2.76 Å. The interchain hydrogen bond length is much longer at 2.89 Å. For the center chain, the agreement is very good for the intrachain hydrogen bond lengths between experiment and theory whereas there is more disorder in the interchain O6−H···O3 interaction. Furthermore, all experimental hydrogen-bonding angles are smaller than their corresponding theoretical ones. Once again the disorder in the origin chain can be explained by the same rationale as that for the center chain. However, the reason that the disorder occurs for the intrachain O2−H···O6 in the center chain and for the interchain O6−H···O3 in the origin chain is not clear. Perhaps it is due to the stacking of the cellulose sheets in Iβ structure.

Consistent with the greater disorder associated with O6H in both cellulose center and origin chains in actual structures compared to those in calculated structures, it can be seen from Table 1 that the origin and the center chains are different in their conformational hydrogen-bonding networks for both the experimentally determined and theoretically calculated structures. Earlier studies[20] also show that the one chain is conformationally more strained compared to the other. This has important implications in the electronic structure and stability of crystalline cellulose Iβ, which will be discussed in more detail in the electronic structure section. These theoretical data also demonstrate that the hydrogen bonds in one chain are slightly stronger than the other, which is in agreement with experimental observations.

Listed in Table 1 are also the hydrogen bond lengths and angles in one chain and one sheet of cellulose. For a single chain, the hydrogen bond lengths are 2.96 and 3.00 Å for the intrachain O2−H···O6 bonds and 2.79 and 2.87 Å for the intrachain O3−H···O5 bonds. These hydrogen bond lengths are considerably longer than the corresponding ones in the bulk. The bond angles, however, are comparable to the theoretical bulk values. For a single cellulose sheet composed of hydrogen-bonded cellulose chains, the hydrogen-bonding network is again seen quite different. The intrachain hydrogen bond lengths are 2.77 and 2.79 Å for the O2−H···O6 bonds and 2.82 and 2.84 Å for the O3−H···O5 bonds. The interchain O6−H···O3 hydrogen bond lengths are 2.77 and 2.74 Å. Again, it is interesting to see that the overall hydrogen-bonding interactions become stronger in bulk crystalline cellulose Iβ compared to those in one chain and one sheet, and the cellulose single sheet has a stronger hydrogen-bonding interaction than the cellulose single chain. This implies that as the chains are packed into sheets, and sheets into layered crystals, the hydrogen-bonding interaction is actually enhanced during the process.

One possible explanation for this finding is that when the hydroxyl groups become more confined as more chains are packed together, they gradually lose their conformational freedom. The "system" thus compensates for the increase of free energy due to this entropy decrease by increasing the overall number of hydrogen bonding interactions. As a result, the system minimizes its free energy. It is therefore puzzling why there appears to be no native crystalline cellulose Iβ larger than $\sim 10^2$ nm in diameter observed in nature. This is perhaps due to the fact that cellulose chains in the crystalline cellulose Iβ are conformationally constrained in order to maximize their hydrogen-bonding interactions which causes the electronic energy of the system to increase. The competition between hydrogen-bonding energy and electronic energy is likely to be one of the major factors that limits the size of native crystalline cellulose Iβ.

Besides strong inter- and intrachain O−H···O hydrogen-bonding interactions, there exists weak intersheet C−H···O hydrogen-bonding interaction in addition to the van der Waals interaction.[7] These hydrogen bonds are much weaker, not only due to the fact that C is a much weaker hydrogen bond donor than O but also that the bond length is around 3.5 Å, which is significantly longer than those of O−H···O type (around 2.8 Å). Nevertheless, this weak hydrogen-bonding interaction together with van der Waals interaction plays an important in the cohesive binding between the cellulose sheets and is one of the reasons for the recalcitrance of biomass to hydrolysis.

IV. Binding Energies in Cellulose Iβ

To understand the recalcitrant nature of crystalline cellulose, the binding energies between two neighboring chains in a single sheet were calculated by varying the interchain unit cell length, which determines the interchain distance. A single sheet of cellulose is used so that the intersheet interaction is excluded from the interaction energy. Figure 6 plots the relative energies of one cellulose sheet as a function of the interchain unit length. The relative energy increases monotonically as the interchain distance increases. The curve in Figure 6 appears to follow an $-1/r$ function. This result is reasonable because the interchain interaction is dominated by the hydrogen bonding between the O6H hydroxyl group in one chain and O3H in the neighboring chain. As is well-known, more than 90% of the hydrogen-bonding interaction is electrostatic in origin and is due fundamentally to Coulomb interaction. The binding energy between these two chains can be estimated by taking the difference between the maximum energy (when the chains are far apart) and the minimum energy, which is close to the experimental interchain unit length at 8.5 Å. This energy is about 0.6 eV (58 kJ/mol) per unit cell. When the chains come close to form the sheet, a total of four hydrogen bonds are formed between the two chains. As a result, two unit cells share four hydrogen bonds. Therefore, the energy

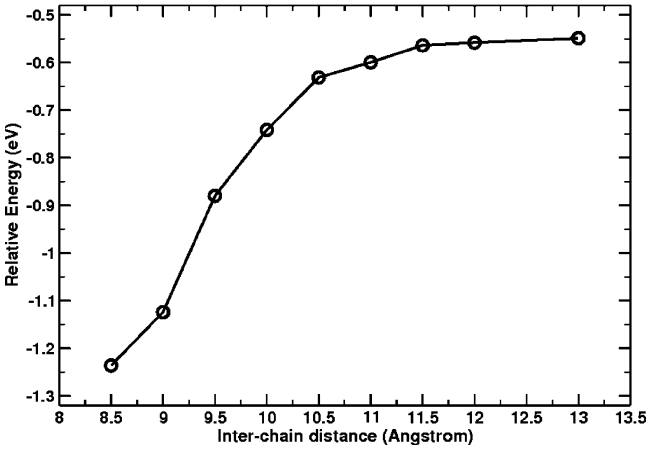

Figure 6. Relative interchain binding energy between the two neighboring cellulose chains in an isolated sheet with a Iβ structure as a function of interchain unit cell distance.

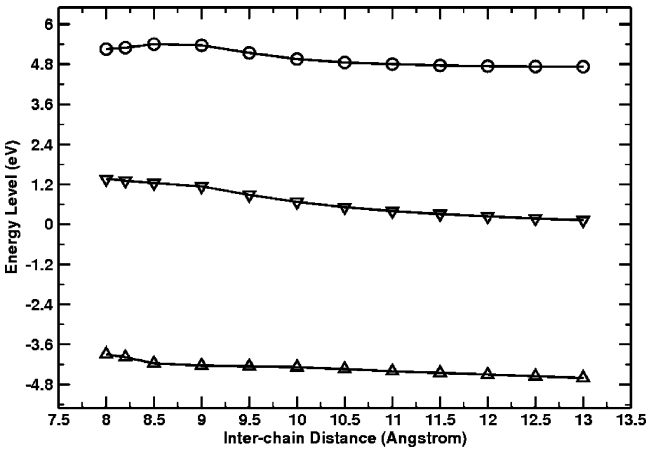

Figure 7. Relative LUMO and HOMO orbital energies and the corresponding Γ-point band gaps for a single sheet as a function of the interchain lattice parameter. The symbols △, ▽, and ○ represent HOMO, LUMO, and band-gap energies, respectively.

for hydrogen bonding is about 29 kJ/mol. This hydrogen-bonding energy is very strong compared to conventional hydrogen-bonding energy ~20 kJ/mol. However, the contribution from the van der Waals interaction between the two chains was also included in the interaction. This strong network of interaction energy may account partially for resistance of cellulose to hydrolysis.

Figure 7 shows the HOMO and LOMO orbital energies in electronvolts and the band gaps at Γ-point as a function of interchain unit cell length. The variations of the HOMO and LOMO orbital energies are very small. The change that occurs does so close to the experimental interchain distance, i.e., between 8 and 10 Å. This is expected, as the interaction energy decreases rapidly as the separation distance between the two chains varies from 8.5 to 10 Å.

In addition to the interchain interaction energy, the intersheet interaction energy was also estimated as a function of intersheet distance. Figure 8 plots the relative energies of the two interacting sheets. The unit cell consists of two chains belonging to the two neighboring sheets. The unit cell has the same symmetry and unit cell parameters as the experimental bulk crystalline cellulose Iβ, except for the intersheet unit length parameter. It can be seen from Figure 8 that intersheet interaction energy is more complex as a function of

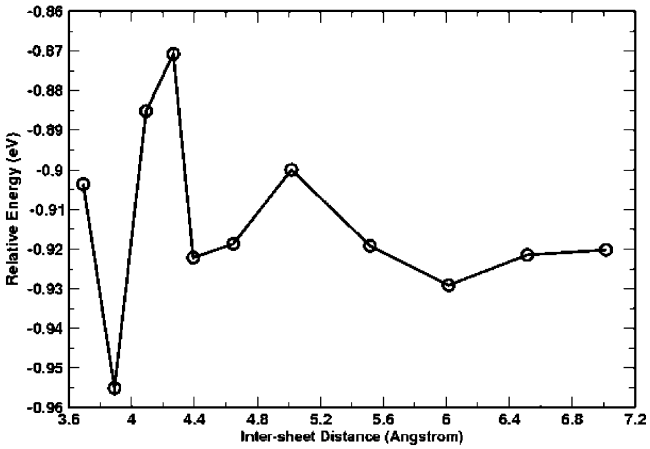

Figure 8. Relative intersheet binding energy as a function of intersheet distance for the two neighboring cellulose Iβ sheets.

intersheet distance. The minimum energy was found at the experimental intersheet distance of 3.892 Å. Any small increase or decrease of the intersheet distance increases the interaction energy dramatically. However, the magnitude of the variation is around 0.08 eV, which is almost 8 times smaller than that of the interchain interaction energy. This is perhaps due to the nature of intersheet interaction. It was thought earlier that the intersheet interaction is dominated by van der Waals interaction. Later experimental X-ray and neutron scattering results[7] confirmed that there is extensive C−H···O hydrogen-bonding interactions between the neighboring sheets. Since C−H···O hydrogen-bonding interaction is considerably weaker than the O−H···O interaction energy, a sharp decrease in intersheet binding energy can be expected. Furthermore, from the complex energy variation with intersheet distance shown in Figure 8, it could be inferred that both van der Waals and intersheet C−H···O hydrogen-bonding interactions contribute to the binding energy between the two neighboring sheets. The complex change is most likely due to the superposition of the two interactions, perhaps of more or less equal magnitude. The experimental intersheet distance is situated at a small, but sharp, minimum. This perhaps also contributes to the resistance for hydrolysis.

V. Size Dependence of Electronic Structures and Energetics of Cellulose Iβ

In addition to the binding energy study described above, the electronic HOMO and LUMO orbital energies of the cellulose 1 chain, 1 sheet, 2−4 sheets together with the bulk were calculated. For the cellulose 1 chain, the structure is infinite in the cellulose chain direction with repeated cellobiose polymeric unit. For cellulose 1−4 sheets, the sheets are translationally invariant in the sheet plane with the same symmetry as that of the bulk. For all of these calculations, the initial structures are taken from the experimental data,[7] and then each system was annealed at 300 K and quenched to 0 K. Structural optimization was carried out until the force and energy were converged according to the convergence criteria. Figure 9 exhibits the energy levels as a function of cellulose size. It can be seen from this figure that both the HOMO and LUMO orbital energies increases as the cellulose size increases. Their difference, i.e., band gap at Γ-point, changes only very slightly. For cellulose 1 chain, 1−3 sheets, the LUMO orbitals lie below vacuum

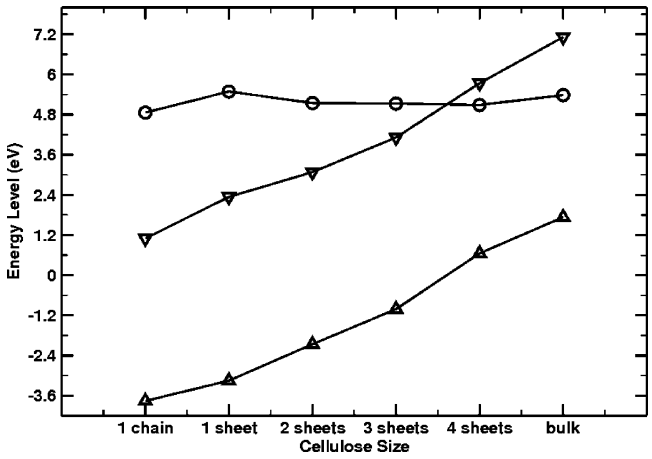

Figure 9. LUMO and HOMO orbital energies and the corresponding band gaps for 1-chain, 1−4 sheets of cellulose, and bulk crystalline cellulose Iβ. The symbols △, ▽, and ○ represent HOMO, LUMO, and band-gap energies, respectively.

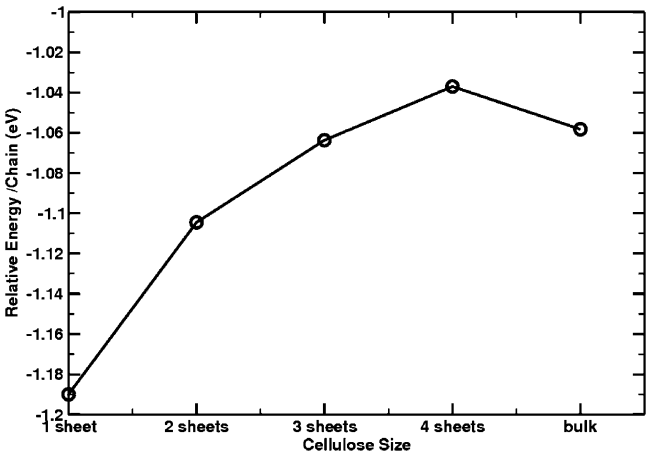

Figure 10. Calculated relative energies for the 1−4 sheets of celluloses and bulk crystal with Iβ allomorph structure.

level, which means these systems are stable. However, for cellulose 4 sheets and the ideal bulk structure, the LUMO orbitals lie above vacuum level, which indicates that perhaps crystalline cellulose Iβ with sizes larger than 4 sheets structure is unstable electronically. All the HOMO orbitals lie above vacuum level but follow the same trend as the LUMO orbital energy, resulting in little change in their differences.

It is not too difficult to understand the LUMO and HOMO orbital energy variations as a function of cellulose size. As mentioned earlier, one of the two chains in cellulose Iβ unit cell is strained with respect to the other in order to form a more energetically favorable hydrogen-bonding network. This strain undoubtedly will increase the electronic energy of the system since not all atoms are at their electronically favorable locations. However, at relatively small sizes, the hydrogen-bonding energy overcomes this negative electronic effect caused by the strain. The overall energy of the system is lowered as a result of stronger hydrogen bonds at the price of higher strain energy. As the size of the system increases, the hydrogen-bonding energy is no longer able to compensate for the increase of strain energy. At a certain critical size, the system will release this increased strain energy by causing disorder in the system. That is perhaps why there are periodic ordered and disordered cellulose structures along its long axis. The periodic disorder, however, breaks down the hydrogen-bonding network but releases strain energy. This phenomenon is not unique among solid-state materials, most notably of magnetic domains in ferromagnetic materials. Another benefit of periodic disorder is to increase the entropy of the system, which at finite temperature will decrease the free energy of the system.

Figure 10 compares the relative energies per cellulose chain in 1 sheet, 2−4 sheets, and the bulk. From Figure 6, it has already been shown that cellulose 1 sheet with interchain hydrogen bonding has a much lower energy per chain than that of an isolated single chain. Figure 10 shows that the energy of the system increases as the number of cellulose sheets increases from 1 to 4. The difference in energy per chain between 4 cellulose sheets and 1 cell sheet is only around 14.5 kJ/mol, which is of the magnitude of hydrogen-bonding energy. This increase in energy is due to the increased strain energy as a result of forming cellulose Iβ structure as cellulose thickness increases. For example, in cellulose 2 sheets, one sheet is more strained compared to the other, resulting a higher total energy than a single sheet only. As cellulose Iβ thickness increases, this strain energy accumulates (see Figure 10). When the strain energy exceeds the gain from forming more favorable hydrogen-bonding energy, the cellulose Iβ structure will collapse, resulting in a region of disorder as mentioned earlier. This disorder will be mostly demonstrated by the disorder of the hydrogen-bonding network[7,8] as the hydroxyl groups are the conformationally most flexible. From Figure 10, the bulk has a slightly lower energy than the 4-sheet cellulose. The reason is perhaps due to the presence of surface energy in the finite systems. If the surface energy can be approximated as the difference in energies between the bulk and the 4-sheet celluloses, then the hypothetical bulk energy adding surface energy would be higher than those of the finite systems. Since the hydrogen-bonding interaction is stronger in the bulk than those in finite-sized sheets, the increase in energy has to be due to the increase in strain energy arising from conformational constraint. This agrees with the HOMO and LUMO energy trend observed as a function of system size.

The increase in HOMO and LUMO orbital energies and relative energy per chain can be explained using the analogy of the pressure effect. To maximize its hydrogen-bonding interaction between the chains on the same sheet and the weak C−H···O hydrogen bonding between the sheets, the system is slightly "pressurized". Because of this increased pressure, the orbital energies are pushed up and the total electronic energy also becomes larger. This is partly compensated by the lowering of hydrogen-bonding energy. As a result, there exists a finite size of the cellulose Iβ crystal at which the balance between the hydrogen-bonding energy and electronic energy is reached. At this point, hydrogen-bonding energy would no longer compensate the increased electronic energy, thus reaching the size limit of the crystalline cellulose Iβ. This could perhaps explain why the native crystalline cellulose Iβ is of limited size in the range of nanometers in diameter. Our theoretical results show that the maximal possible thickness of crystalline cellulose Iβ is only about 4 layers thick, which is approximately 2−3 nm only. Recent high-resolution atomic force microscopy (AFM)[26] shows that the cellulose fibers in plant cell walls are indeed less than 5 nm in diameter. Furthermore, NMR data[24] show that these cellulose fibers have a crystalline core surrounded by the amorphous cellulose. The actual crystal-

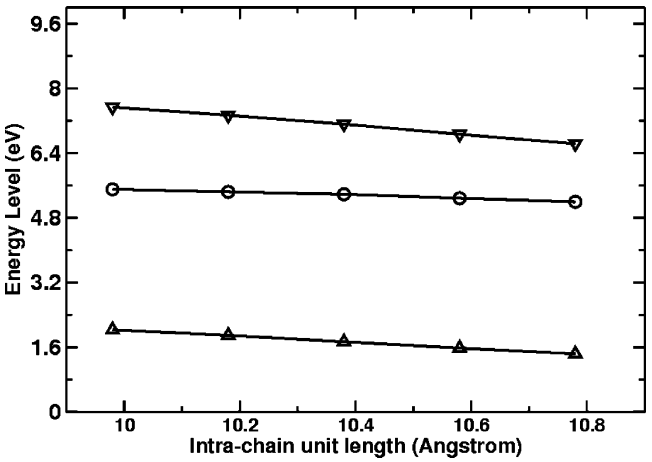

Figure 11. LUMO and HOMO orbital energies and the corresponding Γ-point band gaps for the bulk crystalline cellulose Iβ as a function of intrachain lattice parameter. The symbols △, ▽, and ○ represent HOMO, LUMO, and band-gap energies, respectively.

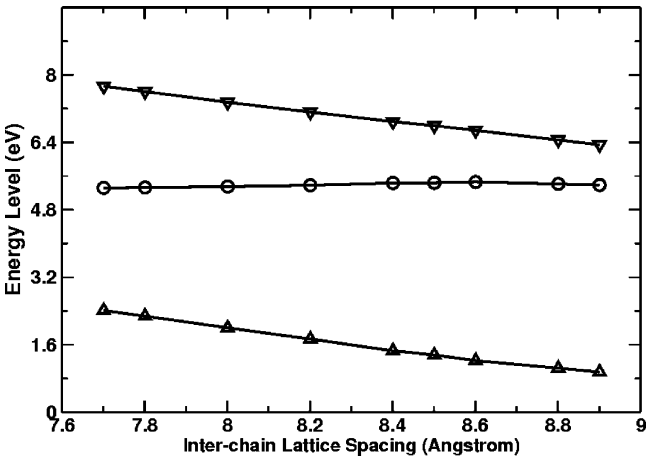

Figure 12. LUMO and HOMO orbital energies and the corresponding Γ-point band gaps for the bulk crystalline cellulose Iβ as a function of interchain lattice parameter. The symbols △, ▽, and ○ represent HOMO, LUMO, and band-gap energies, respectively.

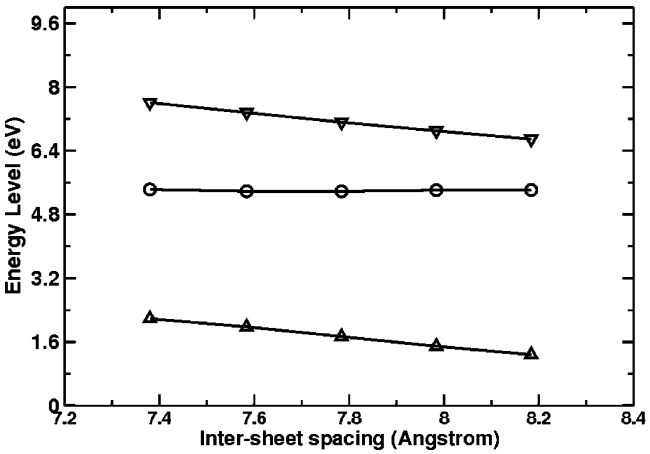

Figure 13. LUMO and HOMO orbital energies and the corresponding Γ-point band gaps for the bulk crystalline cellulose Iβ as a function of intersheet lattice parameter. The symbols △, ▽, and ○ represent HOMO, LUMO, and band-gap energies, respectively.

line cellulose Iβ is thus only 2−3 nm in diameter. Moreover, the cellulose fiber in the long axis is composed of both crystalline and amorphous regions. These amorphous wrappings and the intermittent amorphous regions in the long axis may serve the function of releasing some of the conformational pressure in native crystalline celluloses.

The electronic structure calculations carried out here are only valid for a single-crystal cellulose Iβ. The results for polycrystalline cellulose Iβ and for cellulose Iα could be quite different. In a polycrystalline cellulose Iβ, the hydrogen-bonding interaction along the two single crystalline interface will not be as strong as within a single crystalline cellulose Iβ. As a result, there is minimum strain energy built up between the single-crystalline interfaces. For crystalline cellulose Iα, the situation is also quite different. Since cellulose Iα has a different symmetry and hydrogen-bonding network, the size limit of cellulose Iα, if it exists, could be very different than that of Iβ.

VI. Lattice Dependence of the Orbital Energies and the Energetics of Bulk Cellulose Iβ

To further understand the properties of the native crystalline cellulose Iβ, its lattice dependence of the electronic structure and the relative energy were investigated. Even though the actual crystalline cellulose Iβ is of limited size, it is still illuminating to study the relatively large bulk structures to understand why they cannot be synthesized biologically nor artificially. The structures were optimized at the corresponding lattice parameters with the same symmetry as determined by earlier experiments.[7] From the converged structures, the electronic structure and energy were determined. Figures 11−13 plot the HOMO and LUMO orbital energies and the band gaps at Γ-point as a function of the intrachain, interchain, and intersheet unit lattice parameters. From these figures, it can be seen that both the HOMO and LUMO orbital energies lie above the vacuum level, thus rendering the bulk structure unstable. Furthermore, it can be seen that as the lattice parameter increases, both the HOMO and LUMO orbital energies become lower, whereas the band gap remains more or less constant. This observation supports our earlier argument that the system is strained in order to form stronger hydrogen bonds. When the strain in the system was released by systematically increasing the lattice parameters, a decrease in orbital energies was observed.

Figures 14−16 show the relative energy of the bulk crystalline Iβ as a function of lattice parameter. These lattice-dependent energy figures show that the variation of the energy is relatively small at lattice parameters larger than the experimentally determined lattice constants at a = 7.784 Å (intersheet), b = 8.201 Å (interchain), and c = 10.380 Å (intrachain). The change in energy is more dramatic, as the lattice parameters become smaller than the experimental values. The theoretical equilibrium lattice constants, where the energy minima are located, are close to experimental data at a = 7.78 Å, b = 8.4 Å, and c = 10.38 Å. In most metallic or inorganic materials, a parabolic energy dependence of the lattice parameter is generally found.[37,38] Here a very different function of energy vs lattice was observed, due most likely to the nature of bonding interactions existing in the crystalline cellulose Iβ structures. Since the binding force in crystalline cellulose is dominated by the inter- and intrachain O−H···O and intersheet C−H···O hydrogen bonding

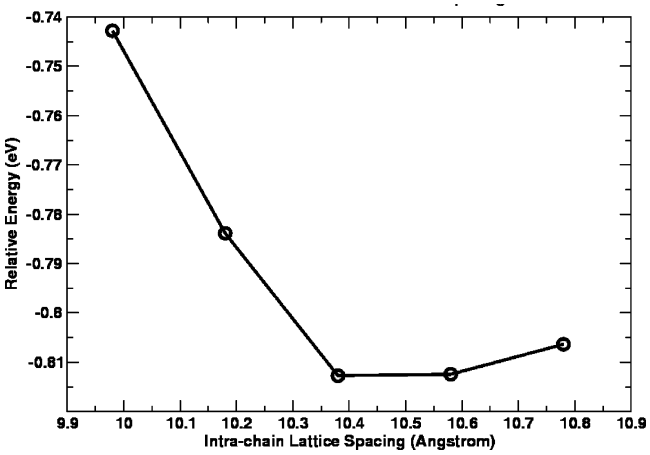

Figure 14. Relative energy of the bulk crystalline cellulose Iβ as a function of intrachain lattice spacing.

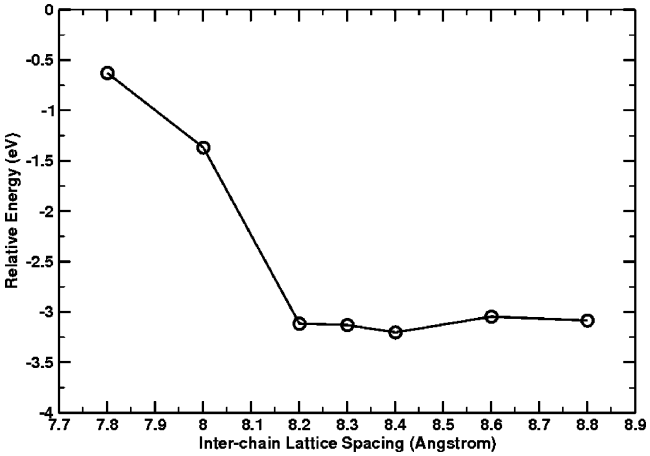

Figure 15. Relative energy of the bulk crystalline cellulose Iβ as a function of interchain lattice spacing.

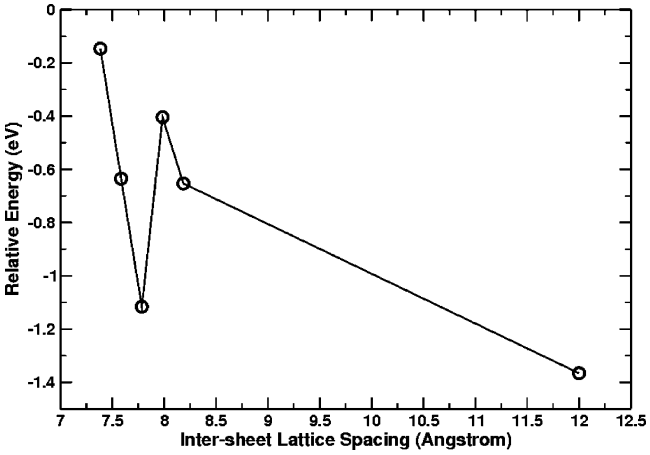

Figure 16. Relative energy of the bulk crystalline cellulose Iβ as a function of intersheet lattice spacing.

interactions, these interactions are electrostatic in origin and thus of long-range nature. As a result, at large lattice parameters the change of energy is relatively small in contrast to the large variations at small lattice parameters, due to strong short-range, high-order repulsive interaction.

VII. Summary

The atomic and electronic structures of native crystalline cellulose Iβ were investigated using the ab initio pseudopotential method. The hydrogen-bonding interaction in crystalline cellulose is the binding force forming these highly recalcitrant molecular crystals. A model based on the competition between the hydrogen-bonding energy and electronic strain energy was established to explain the size of native crystalline cellulose. The binding energies between the chains in one-cellulose sheet and between the neighboring sheets were determined. It was found that the hydrogen-bonding interaction dominates the interchain interaction, whereas both hydrogen-bonding and van der Waals interaction are of significance for intersheet interaction. The intersheet binding energy was found to be about 8 times smaller than that of interchain binding energy.

Acknowledgment. The authors acknowledge helpful discussions with Dr. Mark Davis at the National Renewable Energy Laboratory (NREL) in Golden, CO. This work was carried out at the San Diego Supercomputing Center and the Computational Science Center at NREL. Funding for this work was from the Office of the Biomass Program, Department of Energy, via a subcontract from the National Renewable Energy Laboratory #ACO-4-33101-01.

References and Notes

(1) Atalla, R. H.; VanderHart, D. L. *Science* **1984**, *223*, 283−285.
(2) Sugiyama, J.; Persson, J.; Chanzy, H. *Macromolecules* **1991**, *24*, 2461−2466.
(3) Imai, T.; Sugiyama, J. *Macromolecules* **1998**, *31*, 6275−6279.
(4) Šturcová, A.; His, I.; Apperley, D. C.; Sugiyama, J.; Jarvis, M. C. *Biomacromolecules* **2004**, *5*, 1333−1339.
(5) Nishiyama, Y.; Isogai, A.; Okano, T.; Muller, M.; Chanzy, H. *Macromolecules* **1999**, *32*, 2078−2081.
(6) Sugiyama, J. J.; Vuong, R.; Chanzy, H. *Macromolecules* **1991**, *24*, 4168−4175.
(7) Nishiyama, Y.; Langan, P.; Chanzy, H. *J. Am. Chem. Soc.* **2002**, *124*, 9074−9082.
(8) Nishiyama, Y.; Sugiyama, J.; Chanzy, H.; Langan, P. *J. Am. Chem. Soc.* **2003**, *125*, 14300−14306.
(9) Kono, H.; Yunoki, S.; Shikano, T.; Fujiwara, M.; Erata, T.; Takai, M. *J. Am. Chem. Soc.* **2002**, *124*, 7506−7511.
(10) Kono, H.; Erata, T.; Takai, M. *Macromolecules* **2003**, *36*, 3589−3592.
(11) Saxena, I. M.; Brown, R. M. *Ann. Bot.* **2005**, *96*, 9−21.
(12) Debzi, E. M.; Chanzy, H.; Sugiyama, J.; Tekely, P.; Excoffier, G. *Macromolecules* **1991**, *24*, 6816−6822.
(13) Yamamoto, H.; Horii, F. *Macromolecules* **1993**, *26*, 1313−1317.
(14) Yamamoto, H.; Horii, F.; Odani, H. *Macromolecules* **1989**, *22*, 4130−4132.
(15) Horii, F.; Yamamoto, H.; Kitamaru, R.; Tanahashi, M.; Higuchi, T. *Macromolecules* **1987**, *20*, 2946−2949.
(16) Sjostrom, E. *Wood Chemistry Fundamentals and Applications*; Academic Press: New York, 1981.
(17) Wada, M.; Heux, L.; Isogai, A.; Nishiyama, Y.; Chanzy, H.; Sugiyama, J. *Macromolecules* **2001**, *34*, 1237−1243.
(18) Wada, M.; Heux, L.; Sugiyama, J. *Biomacromolecules* **2004**, *5*, 1385−1391.
(19) Wada, M.; Chanzy, H.; Nishiyama, Y.; Langan, P. *Macromolecules* **2004**, *37*, 8548−8555.
(20) Langan, P.; Nishiyama, Y.; Chanzy, H. *Biomacromolecules* **2001**, *2*, 410−416.
(21) Jakob, H. F.; Fengel, D.; Tschegg, S. E.; Fratzl, P. *Macromolecules* **1995**, *28*, 8782−8787.
(22) Ha, M. A.; Apperley, D. C.; Evans, B. W.; Huxham, M.; Jardine, W. G.; Vietor, R. J.; Reis, D.; Vian, B.; Jarvis, M. C. *Plant J.* **1998**, *16*, 183−190.
(23) Jakob, H. F.; Fratzl, P.; Tschegg, S. E. *J. Struct. Biol.* **1994**, *113*, 13−22.
(24) Newman, R. H. *Solid State Nucl. Magn. Reson.* **1999**, *15*, 21−29.
(25) Mueller, M.; Hori, R.; TItoh, T.; Sugiyama, J. *Biomacromolecules* **2002**, *3*, 182−186.
(26) Ding, S.-Y.; Himmel, M. E., submitted to *Biomacromolecules.*

(27) Nishiyama, Y.; Kim, U. J.; Kim, D. Y.; Katsumata, K. S.; May, R. P.; Langan, P. *Biomacromolecules* **2003**, *4*, 1013−1017.
(28) Sarko, A.; Muggli, R. *Macromolecules* **1974**, *7*, 486-494.
(29) Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471−2474.
(30) CPMD. Version 3.7 ed., 2004; pp copyrighted jointly by IBM Corp, and by Max-Planck Institute, Stuttgart.
(31) Molteni, C.; Parrinello, M. *Chem. Phys. Lett.* **1997**, *275*, 409−413.
(32) Becke, A. D. *Phys. Rev. A* **1988**, 3098−3100.
(33) Lee, C.; Yang, W.; Parr, R. C. *Phys. Rev. B* **1988**, *37*, 785−789.
(34) Qian, X.; Nimlos, M. R.; Davis, M.; Johnson, D. K.; Himmel, M. E. *Carbohydr. Res.* **2005**, *340*, 2319−2327.
(35) Andreoni, W.; Curioni, A. *Parallel Comput.* **2000**, *26*, 819−842.
(36) Qian, X.; Nimlos, M. R.; Johnson, D. K.; Himmel, M. E. *Appl. Biochem. Biotechnol.* **2005**, *121−124*, 989−997.
(37) Marcus, P. M.; Qian, X.; Huebner, W. *Phys. Rev. B* **1999**, *60*, 16088−16093.
(38) Qian, X.; Wickramasinghe, S. R. *J. Phys.: Condens. Matter* **2005**, *17*, 581−586.

MA051683B