THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Ab Initio Study of Molecular Interactions in Cellulose Iα

Ajitha Devarajan,[†,‡] Sergiy Markutsya,[†,§] Monica H. Lamm,[†,§] Xiaolin Cheng,[∥] Jeremy C. Smith,[∥,⊥] John Y. Baluyut,[†,‡] Yana Kholod,[†,‡] Mark S. Gordon,[†,‡] and Theresa L. Windus*[,†,‡]

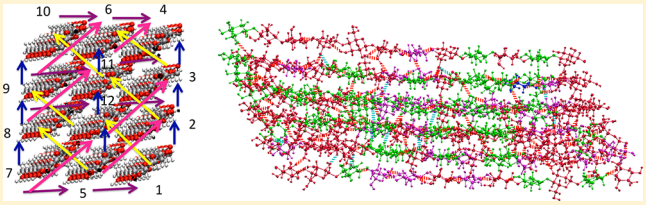
[†]Ames Laboratory, [‡]Department of Chemistry, [§]Chemical and Biological Engineering Department, Iowa State University, Ames, Iowa 50011, United States

[∥]UT/ORNL Center for Molecular Biophysics, Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, Tennessee 37831-6309, United States

[⊥]Department of Biochemistry and Cellular and Molecular Biology, University of Tennessee, M407 Walters Life Sciences, 1414 Cumberland Avenue, Knoxville, Tennessee 37996, United States

Supporting Information

ABSTRACT: Biomass recalcitrance, the resistance of cellulosic biomass to degradation, is due in part to the stability of the hydrogen bond network and stacking forces between the polysaccharide chains in cellulose microfibers. The fragment molecular orbital (FMO) method at the correlated Møller–Plesset second order perturbation level of theory was used on a model of the crystalline cellulose Iα core with a total of 144 glucose units. These computations show that the intersheet chain interactions are stronger than the intrasheet chain interactions for the crystalline structure, while they are more similar to each other for a relaxed structure. An FMO chain pair interaction energy decomposition analysis for both the crystal and relaxed structures reveals an intricate interplay between electrostatic, dispersion, charge transfer, and exchange repulsion effects. The role of the primary alcohol groups in stabilizing the interchain hydrogen bond network in the inner sheet of the crystal and relaxed structures of cellulose Iα, where edge effects are absent, was analyzed. The maximum attractive intrasheet interaction is observed for the GT-TG residue pair with one intrasheet hydrogen bond, suggesting that the relative orientation of the residues is as important as the hydrogen bond network in strengthening the interaction between the residues.



I. INTRODUCTION

Lignocellulose, comprising cellulose, hemicellulose, and lignin, is abundantly present in wood, grasses, and nonedible parts of other plants and forms a significant component of the global carbon cycle.[1−3] Because of its abundance, lignocellulosic biomass is recognized as a potential sustainable source for the production of biofuels. However, biofuel production requires chemical pretreatment of plant biomass to make the polysaccharides easily amenable to enzymatic attack and the release of fermentable monosaccharides. Hemicellulose hydrolysis requires only mild chemical pretreatments, while cellulose hydrolysis requires much harsher pretreatments.[4] The resistance of lignocellulosic biomass to degradation by chemical, enzymatic, and microbial methods is termed biomass recalcitrance and is largely responsible for the high cost of biofuel production.[5−8] This recalcitrance is attributed in part to the extensive hydrogen bond network and the nonbonded interactions between the polysaccharide chains in cellulose fibers that make cellulose a very stable structure.[7,8] An electronic and atomic level understanding of the factors that bind the cellulosic chains together is likely to be of use in designing cost-effective pretreatment processes.

Cellulose is an unbranched homopolysaccharide with linear chains of D-glucose residues linked via β(1→4) glycosidic bonds. Nishiyama et al. determined the crystal structure of native cellulose Iα using atomic resolution synchrotron X-ray and neutron fiber diffraction experiments.[9] Cellulose Iα crystallizes in the space group *P*1 where the crystallographic unit cell is triclinic and contains one cellobiose (disaccharide) unit. While the hydrogen bond network depends on the stacking of the chains and is different for different polymorphs, both the cellulose Iα and Iβ crystal structures exhibit primary alcohol groups exclusively in the trans−gauche (TG) conformation.

Recently, Crowley and co-workers[10] conducted a systematic molecular dynamics simulation of finite cellulose Iβ microfiber structures using the CHARMM35,[11] GLYCAM06,[12] and Gromos45a4[13] force fields. Simulations with the three force fields predict structures that diverge from the Iβ structure, and it is unclear what interactions need refining in the examined force fields. Hence, thorough experimental and theoretical work is required to understand the structure and nonbonded interactions of small diameter cellulose microfibrils.

To understand the nonbonded interactions and hydrogen bond network in cellulose, an ab initio quantum mechanics (QM) study of the cellulose structure is important. Only QM methods that correctly incorporate electron correlation are able

Received: June 25, 2013
Revised: August 5, 2013
Published: August 12, 2013

to account for all of the potential contributions to nonbonded interactions. However, due to the high computational cost, ab initio QM studies of cellulose have been limited to oligomers of glucose.[14] Fortunately, fragmentation based approaches, in which the system is fragmented into subsystems, have been shown to be effective in studying large molecular systems.[15−17] In the present work, the ab initio fragment molecular orbital (FMO)[18−20] theory is used to model cellulosic interactions. Ab initio computations were carried out to understand the interaction behavior of a finite cellulose Iα structure at the chain and residue scales. The FMO method at the Møller−Plesset second order perturbation (MP2) level of theory was used to study the chain pair interactions between chains on the same and neighboring sheets. These sheets are formed from cellulose chains that have the same mean plane as that of the glucose rings. Residue scale interactions in the cellulose structure between residues that exhibit different primary alcohol conformations and that have one or more interchain hydrogen bonds were analyzed via the fragment pair interaction energy (PIE) analysis and the fragment pair interaction energy decomposition analysis (PIEDA).

The paper is organized as follows: Brief details of the cellulose structure, the all atom classical molecular dynamics (MD) simulations used to obtain a relaxed structure, the FMO computations, and the PIEDA analysis are given in section II. In section III, the results of the simulations and the analysis of the interaction energies and the primary alcohol conformations are presented. Concluding remarks are presented in section IV.

II. MODELS, METHODS, AND COMPUTATIONAL DETAILS

A. Crystalline Cellulose Iα. A finite cellulose Iα structure consisting of 12 chains and 144 glucose residues was constructed using the triclinic unit cell reported by Nishiyama et al.[9] Since one of the goals of this work is to examine the chain interactions within a cellulose structure, this system was chosen because it contains an inner chain that is surrounded on all sides by neighboring chains and it is of a size that is computationally feasible. Oxygen terminal residues were capped with hydrogen atoms and carbon terminal residues were capped with OH groups, using the GROMACS[21,22] program, to obtain a finite cellulose structure. This crystalline structure was used for energy calculations using the FMO/MP2 method (described below).

B. Relaxed Cellulose Structure. A relaxed structure was generated by a molecular dynamics simulation, using the crystalline cellulose Iα structure as the initial configuration. The simulation was performed with GROMACS, and the polysaccharides were modeled with the CHARMM36[23] all-atom force field. An energy minimization on the crystal structure was performed before starting the molecular dynamics simulation. After energy minimization, the resulting 7.1 nm × 2.1 nm × 2.3 nm cellulose structure was placed in a 10.4 nm × 9.1 nm × 8.9 nm simulation box and an MD simulation in implicit solvent was carried out. While a simulation with solvent would be more representative of the experimental system, this study will allow for faster sampling of relaxed cellulose structures while treating the solvent effect in a simplified but reasonable manner. In future studies, solvent will be included to elucidate solvent effects on an extended cellulose structure.

A cutoff distance of 1.6 nm was used for long-range van der Waals interactions. The reaction field method[24,25] with a cutoff of 1.2 nm was used for the electrostatic interactions. The so-called "Reaction-field-zero" option was used to obtain energy conservation. In this approach, the potential is set to zero and an infinite dielectric constant is used outside of the cutoff radius, simulating the solvent implicitly. The reaction field method has been shown to yield cellulose structures of comparable accuracy to the particle mesh Ewald method while performing much more efficiently on high performance computers.[26] Additional analysis of the MD simulation is provided in the Supporting Information.

The system was equilibrated for 8 ns using 0.5 fs time steps at 300 K in the NVT ensemble. A multichain Nosé-Hoover thermostat[27] was used to maintain the temperature at 300 K. Production runs were continued for an additional 12 ns with the same conditions. The cellulose structure quickly developed a twist and remained stable as a twisted structure for the rest of the simulation. The existence of a twisted cellulose microfibril was reported by Hanley et al. using atomic force microscopy and transmission electron microscopy[28] and also by Zhao et al. with MD simulations.[29] Matthews et al. reported that when MD simulations are run for a much longer time, the twisted cellulose microfbril untwists itself.[30] However, the focus of the present study is not to debate the existence of twisted vs untwisted microfibril, but rather to study the fundamental molecular interactions in the twisted cellulose structure, the existence of which has been supported by many microscopic studies.[31−41] From the production run trajectory, an MD equilibrated structure was chosen for further study using the FMO method; this structure is referred to as the "relaxed structure" in this paper.

C. Residue, Atom, and Hydrogen Bond Labeling Conventions. Based on the primary alcohol group conformations, dihedral angles were assigned as gauche (G) when the dihedral is less than 90° and trans (T) for those that are greater than or equal to 90°. A two-letter code was assigned to each of the glucose residues as GT, TG, and GG conformations in which the first letter corresponds to the relative O6−O5 orientation and the second letter corresponds to the relative O6−C4 orientation. These conformations are shown in Figure 1 where the first row in the figure shows the atom labeling for these conformations, and the second row shows the conformations color coded as GG (red), GT (green), and TG (purple).

The hydrogen bond networks in the cellulose Iα structures were analyzed using the Visual Molecular Dynamics (VMD) program.[42] For these purposes, a hydrogen bond is defined as

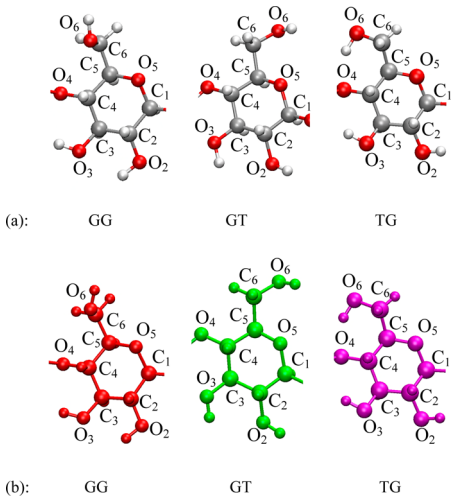


Figure 1. (a) Atom labeling for different glucose primary alcohol conformations. (b) Color coded glucose primary alcohol conformations. The same color coding is used in all of the figures where relevant.

dx.doi.org/10.1021/jp406266u | J. Phys. Chem. B 2013, 117, 10430−10443

having an O donor to O acceptor distance of less than 3.2 Å and an O donor−H donor−O acceptor angle of 150−180°.[42]

D. Fragment Molecular Orbital (FMO). High performance, multilevel parallel,[20] FMO computations were carried out for both the crystalline and the relaxed structures. Each glucose residue of the cellulose Iα structure was considered to be a monomer fragment and up to dimer interactions were considered in the FMO2 method. In the FMO2 method[18] the total energy of the system is expressed as

$$E = \sum_i E_i + \sum_i \sum_{j<i} [E_{ij} - E_i - E_j] \quad (1)$$

where E_i is the energy of a monomer fragment i in the Coulomb field of other monomer fragments, and E_{ij} is the energy of a dimer fragment formed from monomers i and j in the field of the other monomer fragments.

First principles based quantum chemical calculations generally provide only the total energy and properties of interest such as charges, dipoles, and frequencies. To interpret the physical and chemical aspects of binding, an energy decomposition analysis (EDA) approach was proposed by Morokuma,[43] and later by Morokuma and Kitaura,[44] in which the binding energy is divided into familiar contributions, that is, electrostatic, charge transfer, dispersion, and exchange repulsion interactions. The original method was developed for the analysis of binding between two units (e.g., molecules); Chen and Gordon extended the method to the general case of an arbitrary number of units.[45]

Fedorov and Kitaura extended the EDA to the pair interaction energy decomposition analysis for the FMO method[46] in order to treat the binding between both standalone and covalently bound subsystems (fragments). This PIEDA extension is used to obtain total pair interaction energies (PIE) and the components of the interaction energies. For the purpose of the pair interaction analysis, the FMO2 energy expression (1) can be rewritten as

$$E = \sum_i E_i' + \sum_i \sum_{j<i} \Delta E_{ij}^{\text{int}} \quad (2)$$

$$E_i' = E_i - \text{Tr}(\mathbf{D}^i\mathbf{V}^i) \quad (3)$$

In eqs 2 and 3, $\mathbf{D}^i$ is the electron density of the ith monomer, $\mathbf{V}^i$ is the potential on the ith monomer from all other monomers, and $\Delta E_{ij}^{\text{int}}$ is the pair interaction energy between the monomer fragments i and j. The latter quantity can, in turn, be decomposed into contributing components as

$$\Delta E_{ij}^{\text{int}} = \Delta E_{ij}^{\text{ES}} + \Delta E_{ij}^{\text{EX}} + \Delta E_{ij}^{\text{CT}} + \Delta E_{ij}^{\text{DISP}}$$

where the superscripts ES, EX, CT, and DISP designate the electrostatic, exchange repulsion, charge transfer, and dispersion contributions, respectively. The CT term also includes other contributions that are not easily characterized and that are generally small.[45,46] These quantities are computed by the GAMESS software package[47] according to equations given by Fedorov and Kitaura.[46] The PIEDA module can be used with any level of theory that is available with the FMO method.

In the present work, the chain−chain interaction energy between two given chains is defined as a sum of the interaction energies of the fragments in these chains:

$$\Delta E_{IJ}^{\text{chain},x} = \sum_{i\in I} \sum_{j\in J} \Delta E_{ij}^{x};\ x \in \{\text{ES, EX, CT, DISP}\} \quad (4)$$

where I, J enumerate chains; i, j enumerate monomers belonging to chains I, J, respectively; the superscript x can take a value of ES, EX, CT, or DISP.

Multilevel parallel FMO2/MP2 computations using the 6-31G basis set were carried out using the BlueGene/P computer at Argonne National Laboratory. A total of 4096 processor cores were employed for 8 h to complete a single point energy computation. Restricted Hartree−Fock computations were carried out with 48 groups of processors and MP2 computations used 192 groups of processors. Utility programs were written to compute the two-dimensional (2D) map of the total chain PIE, the 2D map of components of the chain PIEDA, the fragment PIE, and the components of the fragment PIEDA. VMD plugins were developed in the Tcl scripting language to analyze dihedral angles in the cellulose structure, to assign a two-letter code to each residue and to color the structure accordingly.

III. RESULTS AND DISCUSSION

A. FMO/MP2 Chain Interaction Analyses in Crystalline Cellulose Iα. Chains in the cellulose structure are held together by a hydrogen bond network and by stacking forces.[14] To characterize the interactions between chains on the same sheet and between chains on neighboring sheets, an analysis of chain PIE was carried out. FMO chain PIEs for all of the chains are given in the Supporting Information, and the most significant chain PIEs for the crystal structure are included in Table 1. To characterize the nature of the electronic interactions binding the chains together chain PIEDA analysis was also carried out.

Table 1. Cellulose Iα Intrasheet and Intersheet Chain Pair Interaction Energy (kcal/mol)[a]

chain type	crystal structure	relaxed structure	ΔE
Intersheet (1−2)	−131.6	−125.7	−5.9
intersheet (9−10)	−129.6	−168.3	38.7
intersheet (1−5)	−87.6	−125.3	37.7
intersheet (6−10)	−86.4	−96.4	10.0
intersheet (7−8)	−130.1	−121.0	−9.1
intersheet (3−4)	−130.5	−159.9	29.4
intersheet (11−12)	−128.0	−109.8	−18.2
intersheet (8−9)	−130.3	−99.8	−30.5
intersheet (6−11)	−127.3	−146.5	19.2
intersheet (5−12)	−128.0	−89.9	−38.1
intersheet (2−3)	−131.4	−119.2	−12.2
intrasheet (2−5)	−79.6	−94.9	15.3
intrasheet (6−9)	−78.9	−127.6	48.7
intrasheet (7−12)	−80.4	−132.1	51.7
intrasheet (4−11)	−81.9	−131.8	49.9
intrasheet (3−12)	−81.2	−108.1	26.9
intrasheet (8−11)	−80.4	−150.5	70.1

[a]The second and third columns provide the chain pair interaction energies for the crystal and relaxed structures. The final column gives the interaction energy for the crystal structure minus the interaction energy for the relaxed structure (a negative sign means the crystal structure has a greater attractive interaction energy), ΔE.

Two-dimensional (2D) maps that illustrate the chain PIE and chain PIEDA values are shown in Figure 2. A 2D map of the total chain PIE is shown in Figure 2a. Chain indices and the location of each chain in the crystalline structure are given in Figure 2f. The sheets are arranged in Figure 2f diagonally from lower left to upper right (for example, chains 3, 12, and 7 form one of the sheets). The intrasheet plane (plane that is parallel to the

dx.doi.org/10.1021/jp406266u | J. Phys. Chem. B 2013, 117, 10430−10443

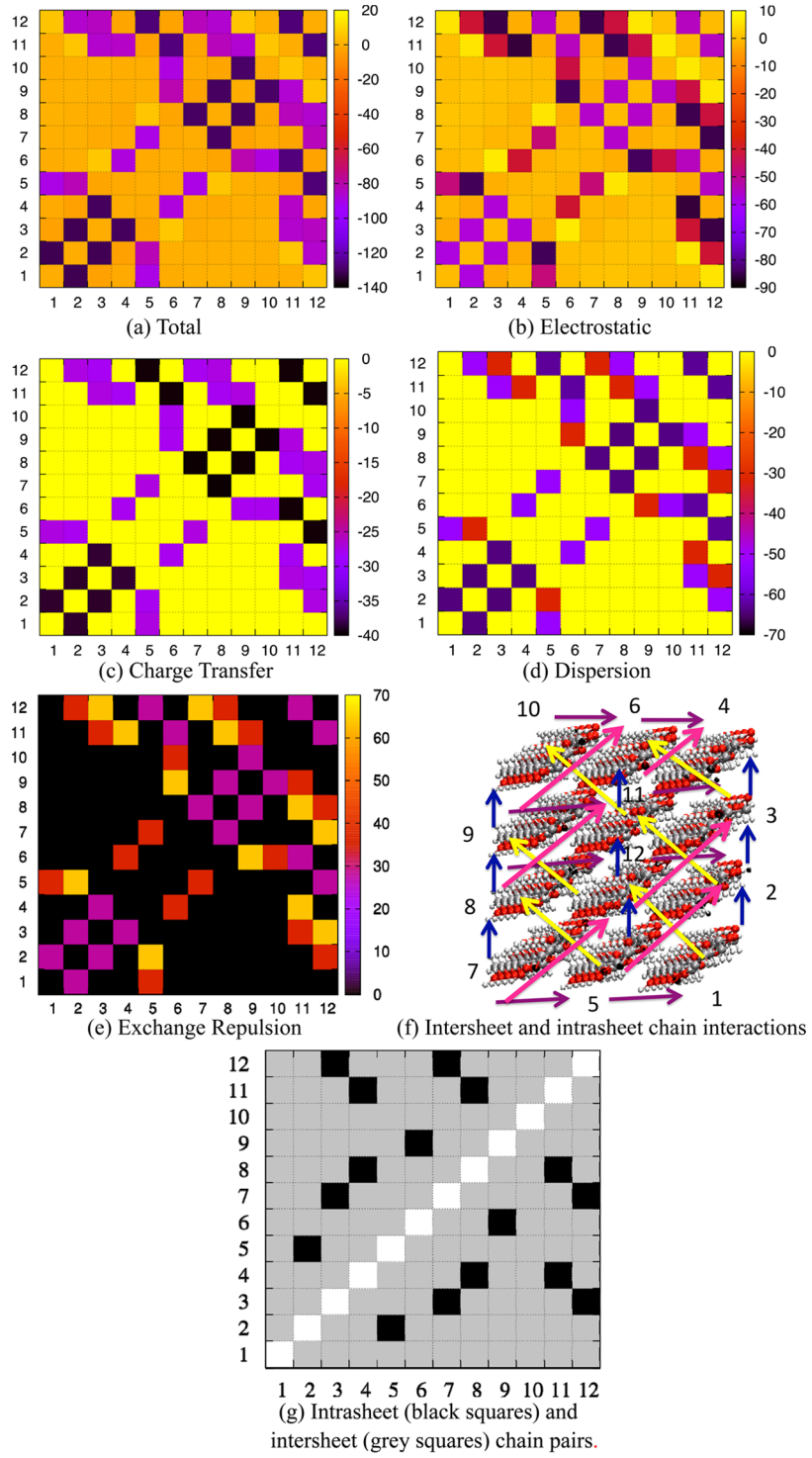


Figure 2. (a−e) Color-coded 2D maps of the chain PIE values and components of the chain PIE values computed for the crystalline structure of cellulose Iα: (a) total, (b) electrostatic, (c) charge transfer, (d) dispersion, and (e) exchange repulsion. In the 2D maps, the *x*- and *y*-axes represent the chain numbers. The rectangular boxes on the right side of each figure show how the color scheme relates to the energy values (in kcal/mol) in the 2D map. (f) Chain indices and location of chains in the crystalline structure of cellulose Iα. Color coded arrows in (f) represent the strength of the chain PIE as shown in (a) with the darker arrows representing the stronger attractive interaction. The intrasheet chain interaction is shown as pink arrows. The blue and purple arrows show pair interactions with chains on neighboring sheets in the vertical (stacking) and horizontal (staggered stacking) directions respectively. Yellow arrows represent pair interactions between chains on next neighboring sheets. (For clarity, not all possible yellow arrows have been included.) (g) Color coded 2D map of intrasheet chain pairs (black squares) and intersheet chain pairs (gray). The white squares on the diagonal represent the interaction of the chain with itself. The same legend schemes are used in Figure 3.

glucopyranose ring) is parallel to the (110) crystalline cellulose face. Figure 2g characterizes the chain pair indices as intrasheet chains (black squares) and chains (gray squares). The most attractive chain pair interactions (darkest arrows in Figure 2f, darkest squares in Figure 2a and data in Table 1) occur between chain pairs 1−2, 2−3, 3−4, 5−12, 6−11, 7−8, 8−9, 9−10, and

dx.doi.org/10.1021/jp406266u | *J. Phys. Chem. B* 2013, 117, 10430−10443

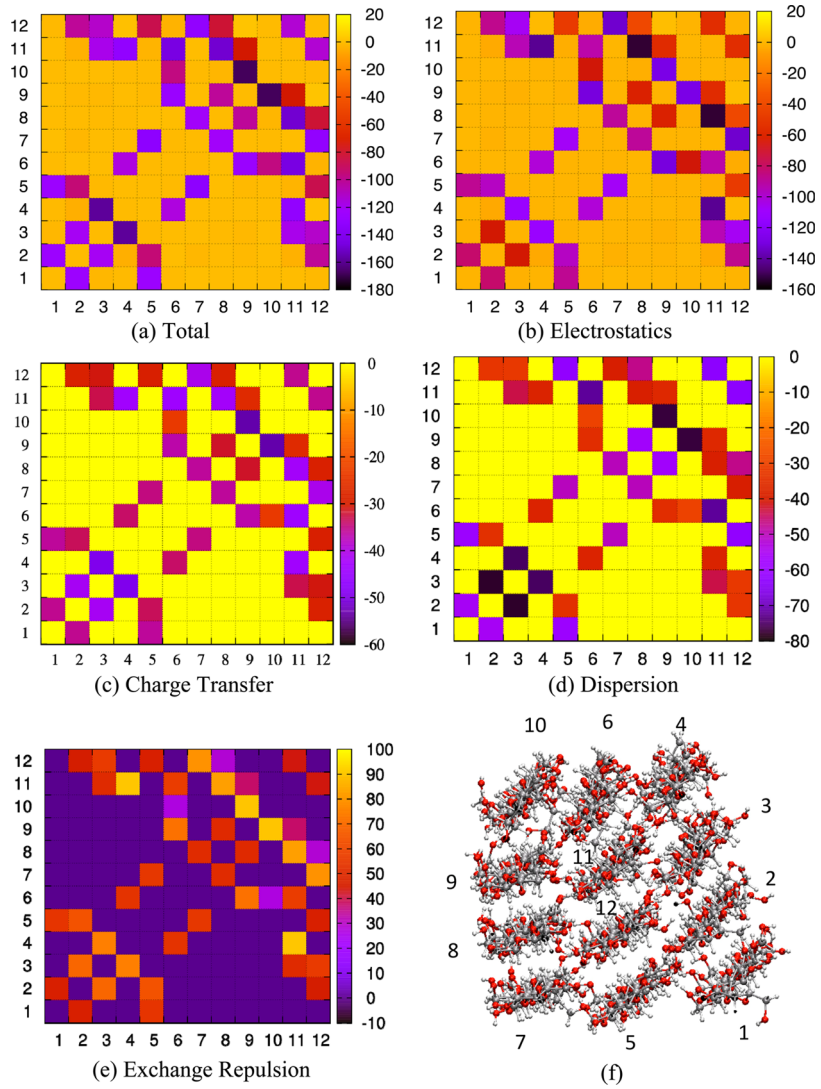


Figure 3. (a–e) Color-coded two-dimensional maps of chain PIE values and components of chain PIE values computed for the relaxed structure of cellulose Iα (in kcal/mol): (a) total, (b) electrostatics, (c) charge transfer, (d) dispersion, and (e) exchange repulsion. (f) Chain indices and location of the chains in the relaxed structure.

11–12; these chain pairs are intersheet chain pairs seen as gray squares in Figure 2g. These chain pairs are located on neighboring sheets in the vertical direction in Figure 2f. The next most attractive chain pair interactions are between chains 1–5, 5–7, 2–12, 8–12, 3–11, 9–11, 4–6, and 6–10. These are also intersheet chain pairs located on neighboring sheets (in the horizontal direction) in Figure 2f. The intrasheet chain pair interactions between chains 2–5, 3–12, 7–12, 4–11, 8–11, and 6–9 (shown as pink arrows in Figure 2f and as black squares in 2g) are only slightly less attractive than the second set of interactions (see Table 1 for numerical data).

A chain PIEDA was carried out and the resulting contributions from electrostatics, charge transfer, dispersion, and exchange repulsion are also shown as 2D maps in Figure 2b–e. Symmetry that is related to the crystalline structure is apparent in the PIE and PIEDA maps for the chains with similar neighbors. There are significant differences between the relative strengths of the electrostatic contributions to the chain–chain interactions (Figure 2b) and the total interaction energies (Figure 2a). This suggests that electrostatic interactions alone are not sufficient to reproduce the overall trends in the chain–chain total interaction energies. For example, the electrostatic contribution to the intrasheet pair interactions between chains 2–5, 3–12, 4–11, 6–9, 7–12, and 8–11 are the strongest. On the other hand, the total interaction energy is similar to those for the dispersion and charge transfer PIEDA maps. The exchange repulsion interactions are all positive (repulsive), but do follow the total PIE trend. It is clear that dispersion and charge transfer interactions, often incompletely modeled in simulation studies, are necessary to predict the chain pair interactions correctly. It is important to mention here that MP2 is the minimum first principles method that correctly accounts for the dispersion contributions without the addition of parametrized, ad hoc terms.

Cellulose binding is usually attributed to hydrogen bonds and stacking forces. The strengths of hydrogen bonds (intrasheet and intersheet) in cellulose are usually assumed to arise from electrostatic and charge transfer interactions, while the stacking interactions are usually ascribed to van der Waals interactions (including London dispersion). The chain PIEDA shows that electrostatic interactions are often the largest for both intrasheet *and* intersheet interactions. The electrostatic interactions contribute a much greater percentage of the intrasheet

dx.doi.org/10.1021/jp406266u | J. Phys. Chem. B 2013, 117, 10430–10443

interactions than they contribute to the intersheet interactions. This is in keeping with the notion that hydrogen bonding is the most important contribution to the intrasheet interactions since only the intrasheet interactions have hydrogen bonds in the crystal structure. Dispersion interactions (usually associated with stacking interactions) contributes more to the strength of the intersheet interactions in the vertical (stacking) direction than do the charge transfer interactions and are often about as strong as the electrostatic interactions. This suggests that the traditional "stacking interactions" are really a complex mixture of the four types of interactions described here. It is true that dispersion plays a much larger percentage role, but it is certainly not the only important contribution.

The computed results presented here for the intrasheet and intersheet interactions are only qualitatively consistent with recent density functional theory (DFT) results using the M06-2X functional for the stacked cellobiose model of Iβ crystalline cellulose.[14] The DFT model caps each cellobiose unit with −H or −OH as appropriate, whereas in the current work each polysaccharide chain (consisting of 12 glucose units) is capped, that is, each cellobiose unit is generally uncapped. This difference allows the model in the current work to better represent the inner core of the cellulose structure. Of course, sheets in cellulose Iα are staggered in a different way than are the sheets in cellulose Iβ. The DFT calculations on stacked cellobiose Iβ models suggest that the strengths of individual hydrogen bonds and stacking interactions are comparable.[14] The results in the present work for the stacked cellulose chains in cellulose Iα show that the total chain PIEs are stronger for the intersheet interactions than the intrasheet interactions. However, since there are more intersheet interactions than intrasheet interactions, it might be reasonable to assume that these two types of interactions are similar in strength overall, in accord with ref 14. More details on a per residue basis are explored in the rest of the paper.

The results presented here are also consistent with recent MD studies by Gross and Chu of both Iα and Iβ crystalline cellulose fibers in water that have demonstrated that intersheet interaction energies per residue are larger than interchain (intrasheet) interactions per residue.[48] Gross and Chu have suggested that the strength of the intersheet interactions does not display a noticeable dependence on solvent exposure.[48] Earlier studies by French et al. have suggested that van der Waals (stacking) forces are more important than hydrogen bonding for the intersheet interaction.[49] Their observation is consistent with the predicted importance in this work of the nonelectrostatic interactions.

B. FMO/MP2 Chain Interaction Analyses in a Relaxed Cellulose Structure. To characterize the energetics of chain scale and residue scale interactions in a relaxed cellulose structure at 300 K, an equilibrated structure from a molecular dynamics simulation was chosen for FMO/MP2 calculations. Chain PIE and chain PIEDA for the relaxed structure were carried out and the results are shown as 2D maps in Figure 3a−e. As can be seen in Figure 3f, the relaxed structure has a bend in the sheets that significantly distorts the structure from the crystalline configuration.

The total chain PIE (Figure 3a, numerical data in Table 1) shows that the intersheet interaction between the corner chain (chain 10) and the neighboring sheet chain (chain 9) is the most attractive interaction. The second most stabilizing (attractive) interaction is the intersheet chain interaction between chains 3 and 4. Thus, the most attractive interactions are intersheet interactions, as observed with the crystalline structure. The 2D map of the electrostatic contributions between the chains is shown in Figure 3b. As observed with the crystalline structure, the electrostatic interactions in the relaxed structure do not reflect the total PIE, even though the electrostatic interactions do usually make the largest attractive contributions. For example, the intrasheet electrostatic interactions between chains 11 and 4 and chains 11 and 8 are the most attractive, suggesting (incorrectly) that intrasheet chain pair interactions are the strongest. As seen with the crystal structure, none of the interaction types alone can replicate the total PIE pattern.

Analysis of the different 2D chain PIEDA maps of dispersion and charge transfer contributions for the relaxed structure suggest that the dispersion contribution is greater than the charge transfer contribution. For the relaxed structure, dispersion interactions (stacking interactions) add more strength to the intersheet interactions in the vertical stacking direction than the charge transfer interactions make to the hydrogen bond network. Electrostatic, exchange repulsion, charge transfer, and dispersion are all essential to account for the trend observed in the total chain PIE.

The change in the interaction energies obtained by subtracting the relaxed structure chain pair interaction energies from those of the crystal structure, ΔE, are given in Table 1. A positive ΔE means that the attractive interaction energy in the relaxed structure is greater than that in the crystal structure, while a negative ΔE means the opposite. In general, most of the changes show that the relaxed structure interactions are more attractive than those in the crystal structure.

The average cellobiose−cellobiose pair interaction energies, referred to as the normalized cellobiose interactions, can be calculated by determining the average of the chain pair interaction energies divided by 12 since there are 6 cellobiose units in each of the two chains in the interacting pair of chains. For the intrasheet interactions, the normalized cellobiose interaction energies are −6.7 and −10.4 kcal/mol for the crystal structure and the relaxed structure, respectively. The intersheet interaction energies for the vertical (horizontal) pairs are −10.8 (−7.1) and −10.6 (−8.9) kcal/mol for the crystal structure and the relaxed structure, respectively. To the best of the authors' knowledge, a direct comparison is not available in other work. However, using the M06-2X density functional method for a Iβ cellobiose model, the normalized cellobiose intersheet interaction for cellulose Iβ is predicted by Parthhasarathi and coworkers to be approximately −12 kcal/mol.[14] Beckham and coworkers found normalized intersheet cellobiose free energies for Iα decrystallization simulations ranging from −4 to −6 kcal/mol using the CHARMM36 potential for the cellulose and TIP3P for water.[50] The FMO pair interaction energies presented in this study are consistent with those from the Beckham study but systematically predict a larger attractive interaction.

C. FMO/MP2 Fragment Conformation and Interaction Analysis in Crystalline and Relaxed Cellulose Structures. Hydroxymethyl (primary alcohol) groups play a key role in the interactions within the cellulose fibril as well as with solvent molecules. It has been demonstrated that the formation of hydrogen bonds, for example, O3−H···O5 (atoms numbered as in Figure 1) and O2−H···O6 between adjacent glucose residues, critically depends on the conformation of the hydroxymethyl group.[5,7,14,51,52] Disruption of these two hydrogen bonds may lead to an easier disassembly of the cellulose fiber.

In the following subsections, each of the intrasheet and intersheet interactions for the crystalline and relaxed structures

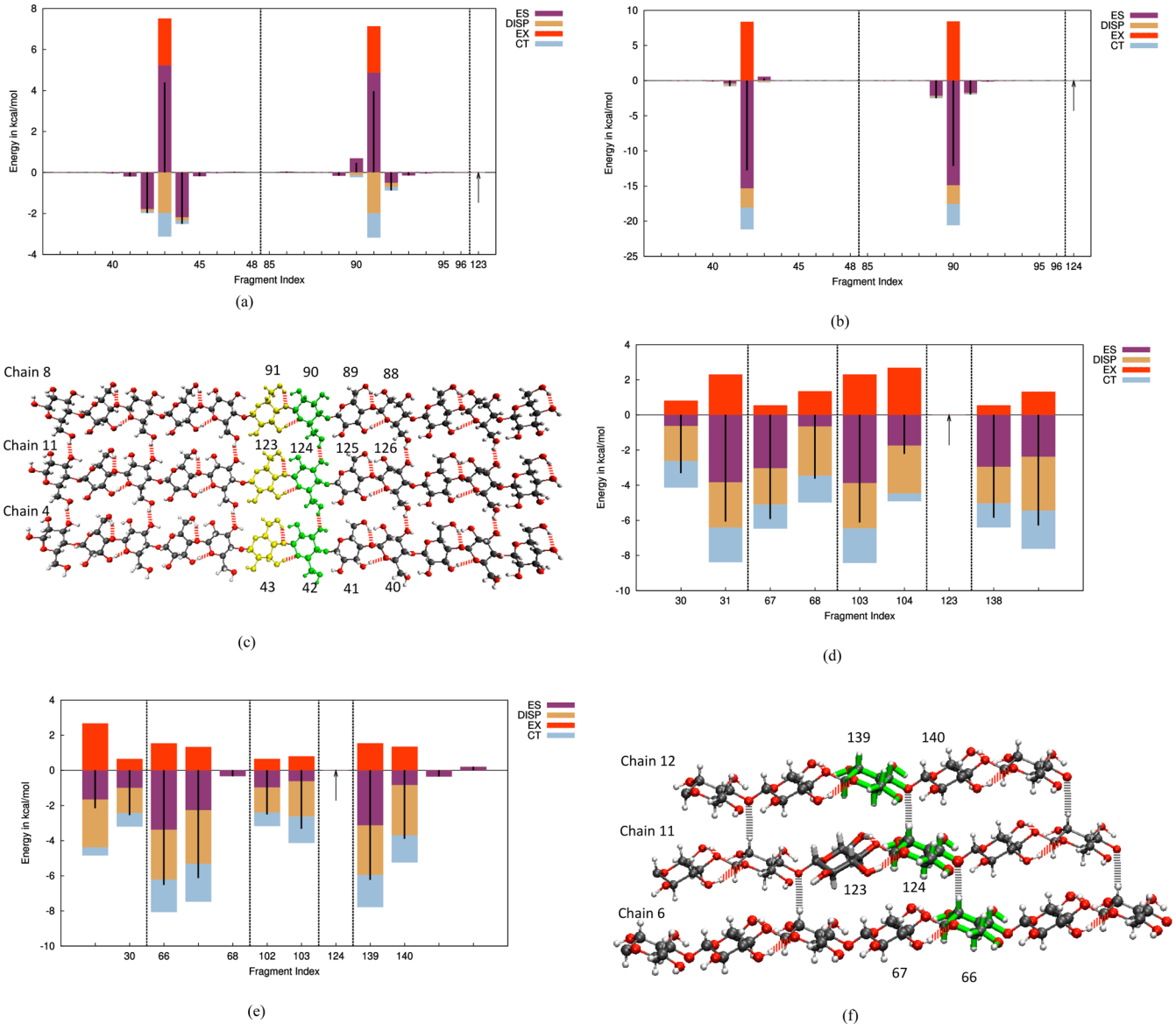


Figure 4. (a) Intrasheet fragment PIEDA for residue 123 showing the interaction between residue 123 and all residues on neighboring chains in the same sheet. (b) Intrasheet fragment PIEDA for residue 124 showing the interaction between residue 124 and all residues on neighboring chains in the same sheet. All interactions are in kcal/mol. (Similar notation is used in all relevant figures.) (c) Hydrogen bond network of the innermost sheet of crystalline cellulose. Dashed red lines indicate intrachain and intrasheet hydrogen bonds. (d) Intersheet fragment PIEDA map for residue 123 showing the interaction between residue 123 and all the residues on neighboring sheets. (e) Intersheet fragment PIEDA map for residue 124 showing the interaction between residue 124 and all the residues on neighboring sheets. (f) Nonconventional C−H···O hydrogen bonds in the crystalline cellulose Iα are indicated with gray dashed lines.

are examined in detail. The chain numbering for the crystalline and relaxed structures is given in Figures 2g and 3f, respectively.

C.1. Crystalline Cellulose Structure. Intrasheet Interaction between Residues with Identical Primary Alcohol Conformations (TG). To examine representative intrasheet residue interactions in the interior of a crystalline cellulose structure, the PIEDA was examined for the two innermost residues in the inner sheet containing chains 8, 11, and 4 (Figure 4a, b). Figure 2f shows the location of the inner sheet as the sheet formed by chains 8, 11, and 4 in the crystalline cellulose structure. The hydrogen bond network in the inner sheet is shown in Figure 4c, with chain 11 as a middle chain and chains 8 and 4 as upper and lower chains, respectively. The residues are numbered as indicated in Figure 4c.

In the crystal structure, all of the residue conformations are TG (see Figure 1). In addition to the intrachain O3−H···O5 and O2−H···O6 hydrogen bonds discussed above, there are O2−H···O6 hydrogen bonds formed between an O2 of a glucose residue in one chain and the hydrogen attached to an O6 of the glucose residue on the neighboring chain in the same sheet. The intrasheet hydrogen bonding along an inner chain in the inner sheet (chain 11) alternates such that one glucose residue in the chain forms two hydrogen bonds with neighboring chains in the sheet and the adjacent glucose residue in the chain forms no hydrogen bonds with neighboring chains. Since the hydrogen bond interactions are periodic for every two neighboring residues, only intrasheet interactions with the two innermost residues (123 and 124) on inner chain 11 are shown in Figure 4a and b, respectively.

dx.doi.org/10.1021/jp406266u | J. Phys. Chem. B 2013, 117, 10430−10443

In all of the PIEDA figures, an arrow indicates which residue the interactions are taken with, each vertical colored bar shows the contribution from electrostatic, dispersion, exchange, and charge transfer, and the vertical black line within the bar shows the total fragment pair interaction energy, and where most of the interactions are zero. In addition, vertical dotted black lines indicate breaks in the fragment numbering. The legend represents interactions as ES for electrostatic, CT for charge transfer, DISP for dispersion, and EX for exchange repulsion. A similar figure for residue 125 is included in the Supporting Information to show the periodicity of the interactions. Thus, residue 123 has no intrasheet hydrogen bond, while the neighboring residue 124 has two intrasheet hydrogen bonds with neighboring chains. Residue 124 has one intrasheet hydrogen bond with residues 42 and 90. This is consistent with the attractive total fragment PIE interaction (−12.8 kcal/mol) associated with residue 124 and residue 42. A similar attractive total fragment PIE interaction (−12.2 kcal/mol) is also seen for the residue 124 and residue 90 on chain 8. An attractive interaction for residue 123 (a residue with no intrasheet hydrogen bond) is seen with residues 42 (−2.0 kcal/mol) and 44 (−2.5 kcal/mol) on chain 4, with the total fragment PIE of less than −3 kcal/mol for each. The total fragment PIE between residue 123 and residues 43 on chain 4 and residue 91 on chain 8 is repulsive.

The largest difference between the fragment PIEDA maps of residue 123 (Figure 4a) and residue 124 (Figure 4b) is from the electrostatic contribution that changes from attractive to repulsive. The stronger attractive interaction seen for residue 124 that forms two intrasheet hydrogen bonds (one with residue 43 on chain 4 and one with residue 91 on chain 8) is expected, as electrostatics is one of the primary fundamental originating interactions ascribed to hydrogen bond interactions. Dispersion (usually ascribed to stacking interactions) and charge transfer (usually ascribed to hydrogen bond interactions) contributions are comparable for residues 123 and 124 suggesting that stacking and hydrogen bond interactions work in a cooperative manner to strengthen the intrasheet interactions.

Intersheet Interaction between Residues with Identical Primary Alcohol Conformations (TG). Figure 2g shows that inner chain 11 is surrounded by chains 3, 6, 7, 9, and 12 that are located on neighboring sheets. The PIEDA for the two innermost residues in chain 11 (123 and 124) are shown in Figure 4d and e. There are more intersheet attractive interactions for both residues than the related intrasheet interactions (Figure 4a and b). For residue 123, there are five intersheet interaction energies between ∼4 and 6 kcal/mol and three between ∼2 and 4 kcal/mol. For residue 124, there are four interaction energies between ∼4−6 kcal/mol and four that are ∼2−4 kcal/mol. Each individual intersheet interaction energy is less than each individual intrasheet interaction energy. However, because there are more intersheet than intrasheet interactions, the cumulative intersheet interaction is stronger than the cumulative intrasheet interaction.

The intersheet crystalline interactions show no conventional X−H···Y hydrogen bonds, where X and Y are electronegative atoms. However, nonconventional hydrogen bonds of the C−H···O type can be found as shown in Figure 4f by the gray dashed lines. A nonconventional C−H−O hydrogen bond is formed between a hydrogen attached to C3 in one glucose residue with the O4 oxygen on the glucose residue on the neighboring sheet chain (C3−H···O4). The H···O distance in this C−H···O hydrogen bond is 3.4 Å, which is longer than the conventional hydrogen bond distance. The intersheet C−H···O hydrogen bonds alternate such that one glucose residue forms two intersheet hydrogen bonds and the adjacent glucose residue in the chain forms no C−H···O bonds with neighboring sheet chains (Figure 4f). Hence the C−H···O bond interactions are periodic for every two neighboring residues.

For example, two intersheet C3−H···O4 hydrogen bonds are seen for residue 124; one formed between C3−H of residue 124 and O4 of residue 139 and the other formed between C3−H of residue 66 and O4 of residue 124 (Figure 4f). This is consistent with the largest attractive interaction seen for residue 124 with residues 66 (−6.5 kcal/mol) and residue 139 (−6.2 kcal/mol). Interestingly, by comparing residue 124 (Figure 4e) with residue 123 (Figure 4d), it emerges that there is very little difference between residues involved in nonconventional hydrogen bonds and those without such bonds. The intersheet fragment PIEDA results for residues 123 and 124 show that dispersion and electrostatics are both major contributors for both residues. The main difference between the individual intrasheet and intersheet fragment PIEDA is that the electrostatic contributions are significantly less in the intersheet interactions.

C.2. Relaxed Cellulose Structure. Primary Alcohol Conformations and Hydrogen Bond Network. Using the relaxed structure shown in Figure 3g, the glucose residues with GG (red), GT (green), and TG (purple) primary alcohol conformations are shown in Figure 5. The relaxed structure

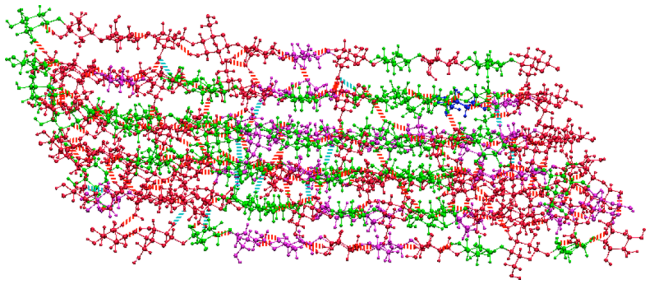


Figure 5. TG (purple), GG (red), and GT (green) primary alcohol conformations in the relaxed structure of the cellulose structure.

has 47% GG, 40% GT, and 13% TG and differs from the average MD structure by 1% GG, 5% GT, and 4% TG. These results are in general agreement with previous simulation studies initiated with cellulose oligomers,[49] microcrystalline cellulose Iβ,[48,50] and microcrystalline cellulose Iα.[51]

To characterize the role of the primary alcohol groups in strengthening the intrasheet hydrogen bond network in the relaxed cellulose structure without edge effects, the inner sheet (chains 8, 11 and 4) of the relaxed structure was analyzed (Figure 6). In the inner sheet (consisting of 36 residues), the percentage contributions of the GG, GT, and TG conformations are 33%, 44%, and 23% respectively. A total of 43 hydrogen bonds are observed for the 36 residues in Figure 6, with 23 interchain hydrogen bonds and 20 intrachain hydrogen bonds. An analysis of the conformations involved in the hydrogen bonding reveals that there are 3 hydrogen bonds for each of the GG-GG and GT-GT pairs, 19 for the GG-GT pairs, 4 for the GG-TG pairs, and 14 for the TG-GT pairs. Counting all of the hydrogen bonds for each conformer type and then dividing this count by the total number of conformers of that type gives the average contribution to hydrogen bonding per conformation. This analysis shows that each of the maximally occurring GT and the minimally occurring TG conformers contributes to about one and a half hydrogen

dx.doi.org/10.1021/jp406266u | J. Phys. Chem. B 2013, 117, 10430−10443

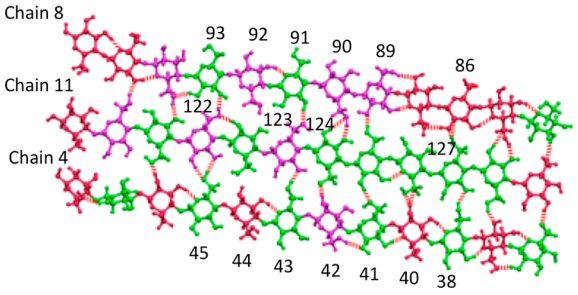


Figure 6. Intrasheet hydrogen bond network among GT (green), TG (purple), and GG (red) primary alcohol conformations in the inner sheet of the relaxed cellulose structure. Residue indices for residues in the inner sheet are in the following range: upper chain 85−96 (chain 8), middle chain 121−132 (chain 11), and lower chain 37−48 (chain 4). Numbering is shown for a few selected residues.

bonds and each GG conformer contributes to one hydrogen bond. Thus, even though the percentage of TG conformations in the inner sheet is much less than those of the GG and GT conformations, each GT and TG conformer contributes equally to the stability of the interchain hydrogen bonded network.

For the inner sheet, it is clear that the intrachain hydrogen bonding network also changes between the crystalline and relaxed structures. For the crystalline structure, each chain has eleven intrachain hydrogen bonds (Figure 4c), while in the relaxed structure there are eight intrachain hydrogen bonds for chain 8, two for chain 11, and nine for chain 4 (Figure 6). These changes are not unexpected given the increased number of intersheet hydrogen bonds in the relaxed structure.

Intrasheet Interaction between Residues with Different Primary Alcohol Conformations. To characterize the residue interactions that bind the chains in cellulose together, fragment PIEs between residues with different primary alcohol conformations in the inner sheet of the relaxed structure were calculated. To minimize edge effects, the interactions with residues 122 (TG), 124(GT), and 127(GT) (see Figure 6) were chosen since their residue environments involve different primary alcohol conformations and participate in one or more intrasheet hydrogen bonds. To evaluate the roles of relative orientations, of intrasheet hydrogen bonds, and of the relaxed environment on the intrasheet interaction between the residues, the fragment PIEs and PIEDAs between the above three selected residues (122, 124, and 127) with all the intrasheet residues on the neighboring chains were analyzed.

Fragment PIE and PIEDA Analysis for Residue 122 (TG). The residue environment including the central 122 (TG) residue and some GT residues that participate in one or more intrasheet hydrogen bonds is considered first. The fragment PIEs between residue 122 (a TG residue with 3 intrasheet hydrogen bonds in the inner sheet) and residues in the neighboring chains on an inner sheet were analyzed and are shown in Figure 7a. The interaction of residue 122 with residues 45 (GT) and 93 (GT) are the two most attractive interactions. The fragment PIE between residues 122 and 45 is −12.2 kcal/mol (2.3 kcal/mol more attractive than the fragment PIE between residues 122 and 93 (−9.9 kcal/mol)). The difference between the fragment PIEDA for residue 122 with residues 45 and 93 is that the electrostatic contributions are greater for the 122−45 interaction. This is consistent with the additional intrasheet hydrogen bond formed between residues 122 and 45 (there are two hydrogen bonds for the 122−45 interaction and one hydrogen bond for the 122−93 interaction). The contributions to the 122−45 and 122−93 interactions from dispersion and charge transfer are comparable.

Fragment PIE and PIEDA Analysis for Residue 124 (GT). The environment of residue 124 (GT) is considered next. Residues 124 (GT) and 122 (TG) differ in that they are in different relaxed environments. As illustrated in Figure 6, the environment of residue 122 (TG) includes interactions with two GT residues that are above and below, but the environment for residue 124 (GT) include interactions with two TG residues that are above and below. The fragment PIE and PIEDA between residue 124 (GT) and the residues in the neighboring chains in the inner sheet are shown in Figure 7b. The interactions of residue 124 with residues 42 (TG) and 90 (TG) are the two most attractive intrasheet interactions for residue 124. The total fragment PIE between residues 124 and 42 is −13.2 kcal/mol, 2.9 kcal/mol more attractive than the fragment PIE between residues 124 and 90 (Figure 7b). There are two intrasheet hydrogen bonds between residues 124 and 90 (see Figure 6) and one intrasheet hydrogen bond between residues 124 and 42 (see Figure 6). The additional attraction predicted for the 124−42 interaction is attributed to the relative orientation of fragments 124 and 42. Based on the fragment PIEDA chart for residue 124 in Figure 7b, the increased electrostatic contribution for the residue pair 124−42 is greater than the increased charge transfer contribution observed for the residue pair 124−90, suggesting a stronger hydrogen bond for the 123−42 pair.

Fragment PIE and PIEDA Analysis for Residue 127 (GT). The fragment PIE and PIEDA between residue 127 (GT) and all of the neighboring chain residues in the inner sheet are shown in Figure 7c. The most attractive interactions for residue 127 are found with residues 38 (GT) and 86 (GG). In addition, there are interactions with other GG residues in its local environment. Residue 127 participates in three intrasheet hydrogen bonds: one with residue 38 and two with residue 86. The fragment PIE between residues 127 (GT) and 86 (GG) is −12.0 kcal/mol, 4.5 kcal/mol more attractive than the fragment PIE between residues 127 (GT) and 38 (GT). Since there is a relatively large increase in the attraction when a second hydrogen bond is involved in the interactions of residues 127 and 86, this suggests that both the additional hydrogen bond and the relative conformations of the glucose residues contribute to the increased attractive interaction predicted to occur between residues 127 and 86.

Based on the intrasheet interactions between the residues in different environments in the inner sheet of the relaxed structure, the GT-TG pair with one intrasheet hydrogen bond exhibits a marginally more attractive fragment PIE than GT-GG and GT-GT pairs. Thus, in addition to the hydrogen bond interactions, the relative orientation of the glucose residues also plays a significant role in strengthening the intrasheet interactions in relaxed cellulose.

Intersheet Interaction between Residues with Different Primary Alcohol Conformations. Unlike the crystal structure where conventional hydrogen bonds are seen only for the intrasheet interactions, the relaxed structure has both intrasheet and intersheet hydrogen bonds, as shown in Figure 8. Only a few representative intersheet interactions are discussed and are taken from the innermost residues on chain 11.

The intersheet fragment PIEDA for the interactions of the innermost residue in chain 11, residue 124 (GT) with all residues on the chains in the neighboring sheets, is shown in Figure 9a. The largest total interaction is with fragment 29 on chain 3 (see Figure 3f for chain numbering) and represents a GT-GG interaction that includes two hydrogen bonds with a total

dx.doi.org/10.1021/jp406266u | J. Phys. Chem. B 2013, 117, 10430−10443

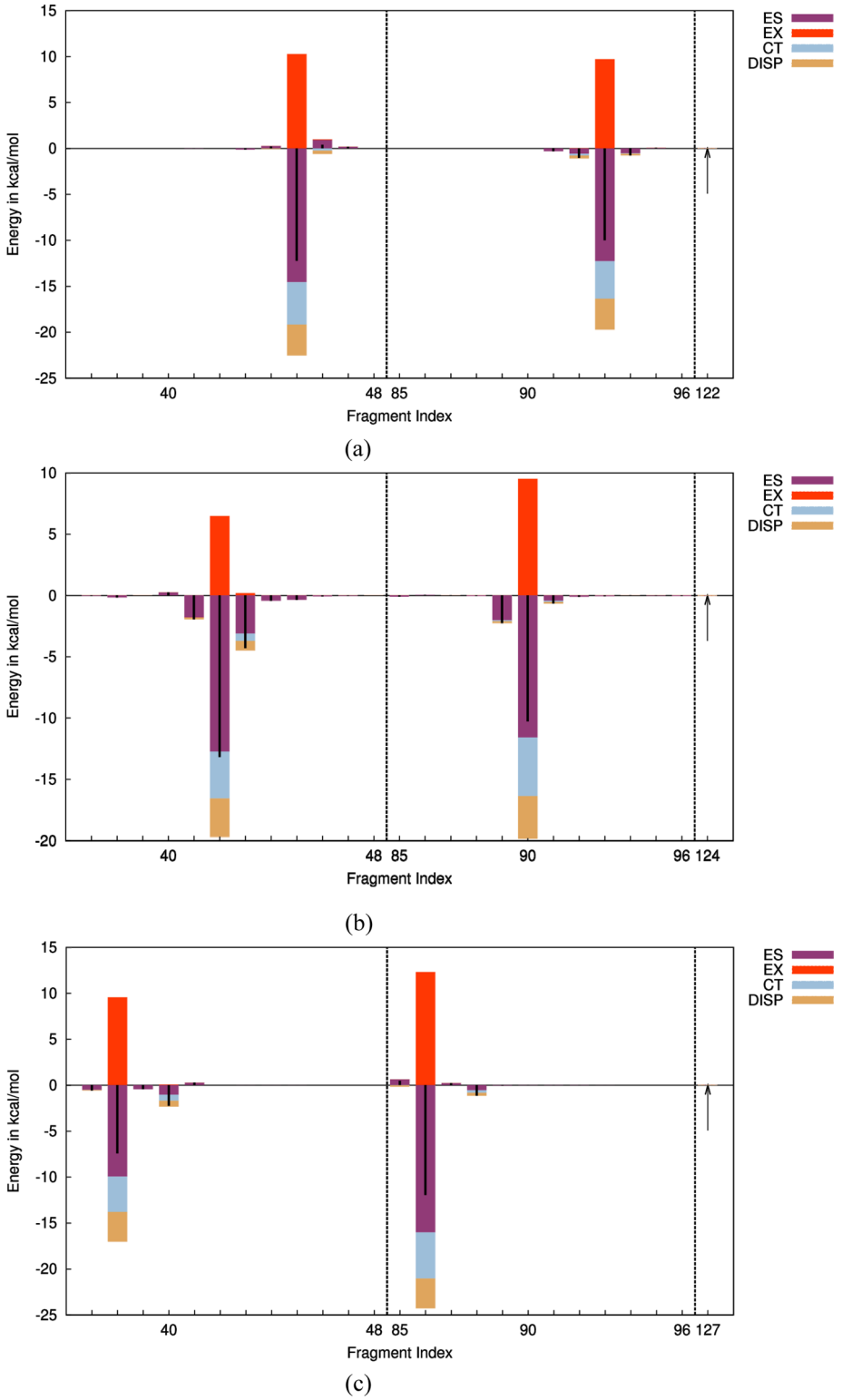


Figure 7. Intrasheet fragment PIEDA in kcal/mol (a) between residue 122 (TG) and intrasheet residues on the neighboring chains, (b) between residue 124 (GT) and intrasheet residues on the neighboring chains, and (c) between residue 127 (GT) and intrasheet residues on the neighboring chains.

interaction energy of −6.0 kcal/mol. This interaction energy is about half of that found for the similar GT-GG interaction within the sheets (residues 127 and 86, shown in Figure 7c). The PIEDA contributions for the 124−29 residue interaction (Figure 9a) show that the major contribution is from electrostatic interactions. There is a small contribution from dispersion and almost no charge transfer contribution. These PIEDA contributions are influenced by the relative orientations and distances between the glucose residues and contribute to the (absolute) low overall interaction energy compared with the intrasheet interactions. This is seen by comparing the intersheet fragment PIEDA with chain 3 in the relaxed structure residue 124 (GT) with the PIEDA for the same residue in the TG conformation in the crystalline structure. The fragment 29 with

dx.doi.org/10.1021/jp406266u | J. Phys. Chem. B 2013, 117, 10430−10443

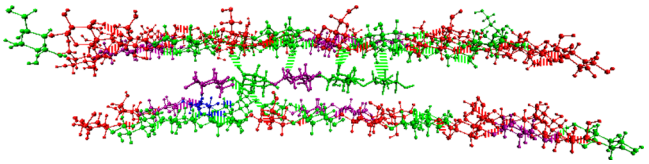
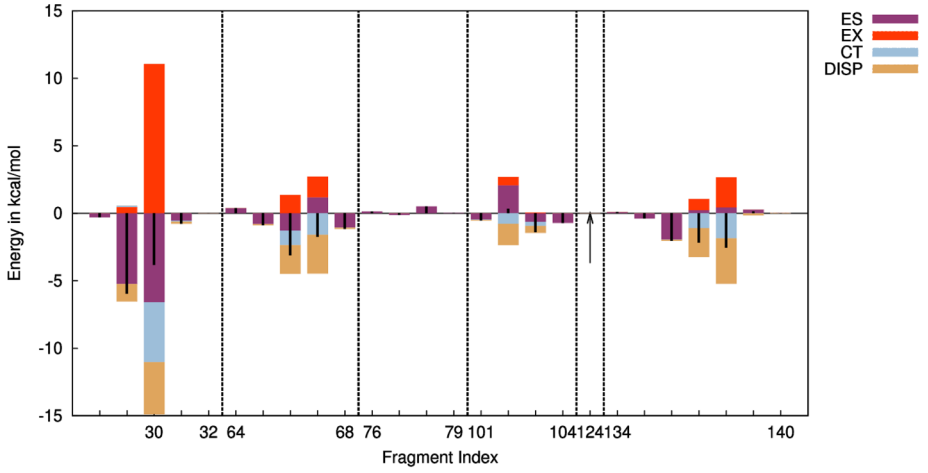


Figure 8. View along the chain for the relaxed structure. The middle chain is a truncated chain 11 that focuses on the inner residues. The top chains are chains 7, 12, and 3, and the bottom chains are chains 9 and 6. The intersheet hydrogen bonds are shown as green dashed lines.

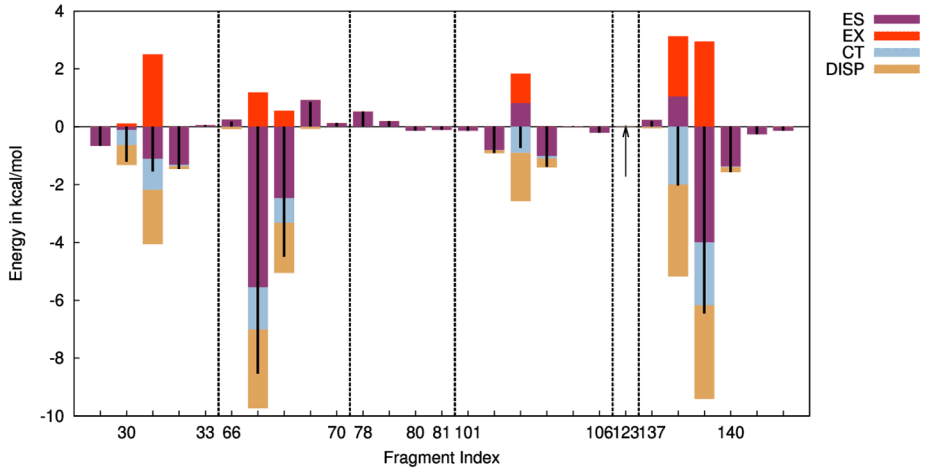
chain 3 total fragment PIE contributions change from −5.0 kcal/mol in the crystal structure to −10.0 kcal/mol in the relaxed structure. This increase in the attraction is attributed to hydrogen bonding, relative orientation of the residues, and distance changes in the relaxed structure.

The fragment PIEDA for residue 123 (TG) is shown in Figure 9b and has more interactions than does residue 124. For residue 123, the sum of the fragment PIE contributions due to interactions with all residues on chain 3 (residues 25 through 36) is −5.1 kcal/mol. The leading intersheet contribution is from residues 67 of chain 6 (−8.6 kcal/mol) and from residue 139 of chain 12 (−6.5 kcal/mol). The PIEDA contributions for the 123−67 interaction (Figure 9b) show that the largest contribution is from electrostatic interactions. The contribution from the dispersion interaction is larger than the contribution from the charge transfer interactions. The PIEDA for both residues 123 and 124 in the relaxed structure include fewer strong interactions with different residues compared to those in the crystal.

C.3. Total Fragment PIE for the Two Innermost Residues (residue 123 and 124). To understand the overall changes in the energetics between the crystalline and relaxed structures, the total fragment PIE for residues 123 and 124 with all other intrasheet and intersheet residues was calculated (Table 2). It is clear that the environments of the two residues are quite different. For residue 123, there is a large difference in the total fragment PIE for the relaxed environment (−57.5 kcal/mol) compared to the crystalline environment (−33.6 kcal/mol). This difference is primarily due to the increased attractive intrasheet interactions in the relaxed structure (−26.1 kcal/mol). For residue 124, there is only a relatively small difference between the crystal and relaxed structure total fragment PIE. In addition, the intrasheet and intersheet interactions are similar in energy for



(a)



(b)

Figure 9. Intersheet fragment PIEDA in kcal/mol (a) between residue 124 (GT) and all the residues in the chains on the neighboring sheets and (b) between residue 123 (TG) and all the residues in the chains on the neighboring sheets.

dx.doi.org/10.1021/jp406266u | J. Phys. Chem. B 2013, 117, 10430−10443

Table 2. Total Fragment PIE for Innermost Residues 123 and 124

	residue 123 FMO fragment PIE (kcal/mol)		residue 124 FMO fragment PIE (kcal/mol)	
	crystal structure (TG)	relaxed structure (TG)	crystal structure (TG)	relaxed structure (GT)
total	−36.7	−57.4	−63.8	−65.5
intrasheet	+2.8	−26.1	−30.1	−33.7
neighboring intersheet	−40.3	−31.2	−35.1	−28.9
all intersheet	−39.5	−31.3	−33.7	−31.8

both residue 124 structures. The difference between the total intersheet contribution and the neighboring intersheet contribution is less than 2 kcal/mol for all structures and residues, suggesting that the contribution from non-neighboring intersheet interactions is negligible. All of these differences between the crystal and the relaxed structure are attributed to (i) glucose conformational changes (changes in relative orientation) in the residues with respect to their crystalline TG conformation and (ii) a slight twist seen in the relaxed structure.

IV. CONCLUSIONS

The interactions within large molecular assemblies, such as cellulose fibrils, can be both subtle and critical for determining physical properties. A major challenge is to understand these interactions in a quantitative manner. Whereas molecular mechanics force fields have furnished much information relevant to cellulose deconstruction, a finer understanding of the interactions involved is likely to prove pivotal in the determination of the physical processes involved. The use of accurate quantum chemical techniques is a means of obtaining this understanding. Scaling quantum chemical methods to large systems is problematic, but this is overcome by the fragment molecular orbital method. In this paper, the ab initio FMO/MP2 method was used for the first time on large cellulose Iα structures to characterize the chain and residue scale interactions. In general, the total intersheet interactions are larger than the total intrasheet interactions in the crystal structure. The chain PIEDA shows that all of the contributions, electrostatic, charge transfer, dispersion, and exchange repulsion, must be included to account for the interactions between the chains. The FMO analysis for the relaxed structure shows that the intersheet and intrasheet interactions become more similar to each other than in the crystal structure, although some "corner" interactions are still significantly stronger.

Examination of the inner sheet of the relaxed structure, in which edge effects are minimized, reveals that, on average, each GT and TG primary alcohol conformation contributes to approximately one and a half intrasheet hydrogen bonds and each GG conformation contributes to one intrasheet hydrogen bond. The FMO fragment PIE for different inner sheet environments with residues that exhibit different primary alcohol conformations and form one or more intrasheet hydrogen bonds were analyzed. The maximum attractive intrasheet interaction for the relaxed structure was found to occur for the GT-TG residue pair with one intrasheet hydrogen bond, underlining the importance of the relative orientation of the residues in the relaxed structure. The FMO/MP2 chain PIEDA reveals that all interaction types must be included to explain the total chain PIE. The fragment PIEDA of selected TG, GT, and GG environments in the inner sheet of the relaxed structure reveals that inclusion of charge transfer, dispersion, and exchange repulsion is essential to predict the total fragment pair interaction.

The fragment PIEDA of the interaction of the innermost residues in the cellulose structure with residues on neighboring sheets was analyzed for the crystalline and relaxed structures. This analysis reveals that the increased total fragment PIE for the central residue 123 in the relaxed structure compared to the crystalline structure is mainly due to increased intrasheet interactions. In addition, the number of intrachain hydrogen bonds decreases in the relaxed structure relative to the crystalline structure.

The results of the current work are in general agreement with previous computations. However, direct comparisons are difficult due to the different types of cellulose modeled and the different computational models used in the simulations. In general, the current model is a better representation of the inner core of the cellulose structure since only the chains are capped and not cellobiose units. The current work shows that total intersheet interactions for the crystal structure are stronger than those for the intrasheet interactions in general agreement with previous MD studies.[48,49] However, individual residue−residue intersheet interactions are weaker than those for individual intrasheet interactions, in contrast to what was previously calculated in a M06-2X cellobiose model of cellulose Iβ.[14] Also, the normalized cellobiose interactions calculated in this work are stronger than those predicted by the CHARMM MD simulations,[50] but weaker than those predicted by the M06-2X study for the stacked cellobiose model.[14] The current work also shows that the main difference between what are conventionally called hydrogen bond interactions (intrasheet) and stacking interactions (intersheet) is that electrostatics make a much larger contribution to the intrasheet interactions. Dispersion, charge transfer, and exchange interactions are similar in magnitude for both intersheet and intrasheet interactions. In general, the sum of the dispersion and charge transfer contributions is equal to or more than the electrostatic contribution for the intersheet residue−residue interactions. Using the more realistic model of cellulose in the current work, the strength of the residue−residue interaction in cellulose is found to be determined by the cooperative influence of the hydrogen bond network, the glucose conformations, and the relative orientation of the glucose residues in cellulose.

It is important to recognize that the current work only examines the enthalpic contributions to the chemical system interactions. As has been shown by others, entropic contributions also play an important role.[50,53,54] However, examination of the entropic contributions to this large a system is outside the scope of the current work.

■ ASSOCIATED CONTENT

Supporting Information

Complete pair interaction energies in kcal/mol for Table 1 (where only unique interactions are given), additional information for the MD simulation, and the intrasheet fragment PIEDA showing the interaction of residue 125 with the other fragments on the neighboring chains in the same sheet in the crystal structure are available. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: twindus@iastate.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research is sponsored by U.S Department of Energy (USDOE) Scientific Discovery through Advanced Computing (SciDAC) program, under the joint auspices of the Offices of Advanced Scientific Computing Research (ASCR) and Biological and Environmental Research (BER). The work reported here was performed at the Ames Laboratory, FWP AL-08-330-039. Ames Laboratory is managed by Iowa State University for the U.S. Department of Energy under contract DE-AC02-07CH11358. An award of computer time was provided by the Innovative and Novel Computational Impact on Theory and Experiment (INCITE) program. This research used resources of the Argonne Leadership Computing Facility at Argonne National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under contract DE-AC02-06CH11357.

■ REFERENCES

(1) Himmel, M. E.; Ding, S. Y.; Johnson, D. K.; Adney, W. S.; Nimlos, M. R.; Brady, J. W.; Foust, T. D. Biomass Recalcitrance: Engineering Plants and Enzymes for Biofuels Production. *Science* **2007**, *315*, 804−807.

(2) Stephanopoulos, G. Challenges in Engineering Microbes for Biofuels Production. *Science* **2007**, *315*, 801−804.

(3) Hu, S.-Q.; Gao, Y.-G.; Tajima, K.; Sunagawa, N.; Zhou, Y.; Kavano, S.; Fujiwara, T.; Yoda, T.; Shimura, D.; Satoh, Y.; et al. Structure of bacterial cellulose synthase subunit D octamer with four inner passageways. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *107*, 17957−61.

(4) Zhu, J. Y.; Verril, S. P.; Liu, H.; Herian, V. L.; Pan, X.; Rockwood, D. L. On Polydispersity of Plant Biomass Recalcitrance and Its Effects on Pretreatment Optimization for Sugar Production. et al. *BioEnergy Res.* **2011**, *4*, 201−210.

(5) Matthews, J. F.; Skopec, C. E.; Mason, P. E.; Zuccato, P.; Torget, R. W.; Sugiyama, J.; Himmel, M. E.; Brady, J. W. Computer Simulation Studies of Microcrystalline Cellulose Iβ. *Carbohydr. Res.* **2006**, *341*, 138−152.

(6) Bergenstråhle, M.; Wohlert, J.; Himmel, M. E.; Brady, J. W. Simulation Studies of the Insolubility of Cellulose. *Carbohydr. Res.* **2010**, *345*, 2060−2066.

(7) Shen, T.; Langan, P.; French, A. D.; Johnson, G. P.; Gnanakaran, S. Conformational Flexibility of Soluble Cellulose Oligomers: Chain Length and Temperature Dependence. *J. Am. Chem. Soc.* **2009**, *131*, 14786−14794.

(8) Shen, T.; Gnanakaran, S. The Stability of Cellulose: A Statistical Perspective from a Coarse-Grained Model of Hydrogen-Bond Networks. *J. Biophys.* **2009**, *8*, 3032−3040.

(9) Nishiyama, Y.; Sugiyama, J.; Chanzy, H.; Langan, P. Crystal Structure and Hydrogen Bonding System in Cellulose Iα from Synchrotron X-ray and Neutron Fiber Diffraction. *J. Am. Chem. Soc.* **2003**, *125*, 14300−14306.

(10) Matthews, J. F.; Beckham, G. T.; Bergenstrahle, M.; Brady, J. W.; Himmel, M. E.; Crowley, M. F. Comparison of Cellulose Iβ Simulations with Three Carbohydrate Force Fields. *J. Chem. Theory Comput.* **2012**, *8*, 735−748.

(11) Guvench, O.; Greene, S. N.; Kamath, G.; Brady, J. W.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D., Jr. Additive Empirical Force Field for Hexopyranose Monosaccharides. *J. Comput. Chem.* **2008**, *29*, 2543−2564.

(12) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Gonzalez-Outeirino, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. GLYCAM06: A Generalizable Biomolecular Force Field. Carbohydrates. *J. Comput. Chem.* **2008**, *29*, 622−655.

(13) Lins, R. D.; Hunenberger, P. H. A New GROMOS Force Field for Hexopyranose-based Carbohydrates. *J. Comput. Chem.* **2005**, *26*, 1400−1412.

(14) Parthasarathi, R.; Bellesia, G.; Chundawat, S. P.; Dale, B. E.; Langan, P.; Gnanakaran, S. Insights into Hydrogen Bonding and Stacking Interactions in Cellulose. *J. Phys. Chem. A* **2011**, *115*, 14191−14200.

(15) Gordon, M. S.; Mullin, J. M.; Pruitt, S. R.; Roskop, L. B.; Slipchenko, L. V.; Boatz, J. A. Accurate Methods for Large Molecular Systems. *J. Phys. Chem. B* **2009**, *113*, 9646−9663 (Invited Centennial Feature Article).

(16) Pranami, G.; Slipchenko, L. V.; Lamm, M. H.; Gordon, M. S. Coarse-Grained Intermolecular Potentials Derived from the Effective Fragment Potential: Application to Water, Benzene, and Carbon Tetrachloride. In *Multi-scale Quantum Models for Biocatalysis: Modern Techniques and Applications*; York, D.M., Lee, T.-S., Eds.; Springer-Verlag: London, 2009.

(17) Devarajan, A.; Windus, T. L.; Gordon, M. S. Implementation of Dynamical Nucleation Theory Effective Fragment Potentials Method for Modeling Aerosol Chemistry. *J. Phys. Chem. A* **2011**, *115*, 13987−13996.

(18) Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. Fragmentation Methods: A Route to Accurate Calculations on Large Systems. *Chem. Rev.* **2012**, *112*, 632−672.

(19) Fedorov, D. G.; Kitaura, K. Second order Moeller-Plesset Perturbation Theory Based Upon the Fragment Molecular Orbital Method. *J. Chem. Phys.* **2004**, *121*, 2483−2490.

(20) Fedorov, D. G.; Olson, R. M.; Kitaura, K.; Gordon, M. S.; Koseki, S. A New Hierarchical Parallelization Scheme: Generalized Distributed Data Interface (GDDI), and an Application to the Fragment Molecular Orbital Method (FMO). *J. Comput. Chem.* **2004**, *25*, 872−80.

(21) Bekker, H.; Berendsen, H. J. C.; Dijkstra, E. J.; Achterop, S.; van Drunen, R.; van der Spoel, D.; Sijbers, A.; Keegstra, H.; Reitsma, B.; Renardus, M. K. R. Gromacs: A parallel computer for molecular dynamics simulations. In *Physics Computing 92*; de Groot, R. A., Nadrchal, J., Eds.; World Scientific: Singapore, 1993.

(22) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.

(23) Guvench, O.; Hatcher, E. R.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D. CHARMM Additive All-Atom Force Field for Glycosidic Linkages between Hexopyranoses. *J. Chem. Theory Comput.* **2009**, *5*, 2353−2370.

(24) Allen, M. P. ; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon: Oxford, 1987.

(25) Tironi, I. G.; Sperb, R.; Smith, P. E.; van Gunsteren, W. F. A Generalized Reaction Field Method for Molecular Dynamics Simulations. *J. Chem. Phys.* **1995**, *102*, 5451−5459.

(26) Schulz, R.; Lindner, B.; Petridis, L.; Smith, J. C. Scaling of Multimillion-Atom Biological Molecular Dynamics Simulation on a Petascale Supercomputer. *J. Chem. Theory Comput.* **2009**, *5*, 2798−2808.

(27) Nose, S. A Molecular Dynamics Method for Simulations in the Canonical Ensemble. *Mol. Phys.* **1984**, *52*, 255−268.

(28) Hanley, S. J.; Revol, J. F.; Godbout, L.; Gray, D. G. Atomic Force Microscopy and Transmission Electron Microscopy of Cellulose from Micrasterias Denticulata; Evidence for a Chiral Helical Microfibril Twist. *Cellulose* **1997**, *4*, 209−220.

(29) Zhao, Z.; Shklyaev, O. E.; Nili, A.; Mohamed, M. N. A. Cellulose Microfibril Twist, Mechanics, and Implication for Cellulose Biosynthesis. *J. Phys. Chem. A* **2013**, *117*, 2580−2589.

(30) Matthews, J. F.; Bergenstrahle, M.; Beckham, G. T.; Himmel, M. E.; Nimlos, M. R.; Brady, J. W.; Crowley, M. F. High-Temperature Behavior of Cellulose I. *J. Phys. Chem. B* **2011**, *115*, 2155−2166.

(31) Benziman, M.; Haigler, C. H.; Brown, R. M.; White, A. R.; Cooper, K. M. Cellulose Biogenesis: Polymerization and Crystallization are Coupled Processes in Acetobacter Xylinum. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 6678−6682.

(32) Haigler, C. H.; White, A. R.; Brown, R. M.; Cooper, K. M. Alteration of In Vivo Cellulose Ribbon Assembly by Carboxymethylcellulose and Other Cellulose Derivatives. *J. Cell Biol.* **1982**, *94*, 64−69.
(33) Bowling, A. J.; Amano, Y.; Lindstrom, R.; Brown, R. M. Rotation of Cellulose Ribbons During Degradation with Fungal Cellulase. *Cellulose* **2001**, *8*, 91−97.
(34) Koyama, M.; Helbert, W.; Imai, T.; Sugiyama, J.; Henrissat, B. Parallel-up Structure Evidences the Molecular Directionality During Biosynthesis of Bacterial Cellulose. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 9091−9095.
(35) Koyama, M.; Sugiyama, J.; Itoh, T. Systematic Survey on Crystalline Features of Algal Celluloses. *Cellulose* **1997**, *4*, 147−160.
(36) Lee, H.; Brown, R. J. A Comparative Structural Characterization of Two Cellobiohydrolases from Trichoderma Reesei: a High Resolution Electron Microscopy Study. *Biotechnol.* **1997**, *57*, 127−136.
(37) Brown, R., Jr. *Cellulose and wood: chemistry and technology*; John Wiley and Sons: New York, 1995; pp 639−657.
(38) Haigler, C. H.; Chanzy, H. J. Electron Diffraction Analysis of the Altered Cellulose Synthesized by Acetobacter Xylinum in the Presence of Fluorescent Brightening Agents and Direct Dyes. *J. Ultrastruct. Mol. Struct. Res.* **1988**, *98*, 299−311.
(39) Shibazaki, H.; Kuga, S.; Okano, T. Mercerization and Acid Hydrolysis of Bacterial Cellulose. *Cellulose* **1997**, *4*, 75−87.
(40) Kondo, T.; Nojiri, M.; Hishikawa, Y.; Togawa, E.; Romanovicz, D.; Brown, R. M. Biodirected Epitaxial Nanodeposition of Polymers on Oriented Macromolecular Templates. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 14008−14013.
(41) Nobles, D.; Romanovicz, D.; Brown, R., Jr. Cellulose in Cyanobacteria. Origin of Vascular Plant Cellulose Synthase? *Plant Physiol.* **2001**, *127*, 529−542.
(42) Humphrey, W.; Dalke, A.; Schulten, K. VMD - Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(43) Morokuma, K. Molecular Orbital Studies of Hydrogen Bonds. III. C═O···H−O Hydrogen Bond in $H_2CO···H_2O$ and $H_2CO···2H_2O$. *J. Chem. Phys.* **1971**, *55*, 1236−1244.
(44) Kitaura, K.; Morokuma, K. A New Energy Decomposition Scheme for Molecular Interactions within the Hartree-Fock Approximation. *Int. J. Quantum Chem.* **1976**, *10*, 325−340.
(45) Chen, W.; Gordon, M. S. Energy Decomposition Analyses for Many-Body Interaction and Applications to Water Complexes. *J. Chem. Phys.* **1996**, *100*, 14316−14328.
(46) Fedorov, D. G.; Kitaura, K. Pair Interaction Energy Decomposition Analysis. *J. Comput. Chem.* **2007**, *28*, 222−237.
(47) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; et al. General atomic and molecular electronic structure system. *J. Comput. Chem.* **1993**, *14*, 1347−1363. Gordon, M. S.; Schmidt, M. W. In *Theory and Applications of Computational Chemistry*; Dykstra, C. E., Frenking, G., Kim, K. S., Scuseria, G. E.; Eds.; Elsevier: New York, 2005; Ch. 41.
(48) Gross, A. S.; Chu, J. W. On the Molecular Origins of Biomass Recalcitrance: The Interaction network and solvation Structures of Cellulose Microfibrils. *J. Phys. Chem. B* **2010**, *114*, 13333−13341.
(49) French, A. D.; Miller, D. P.; Aabloo, A. Miniature Crystal Models of Cellulose Polymorphs and Other Carbohydrates. *Int. J. Biol. Macromol.* **1993**, *15*, 30−36.
(50) Beckham, G. T.; Matthews, J. F.; Peters, B.; Bomble, Y. J.; Himmel, M. E.; Crowley, M. F. Molecular-level Origins of Biomass Recalcitrance: Decrystallization Free Energies for Four Common Cellulose Polymorphs. *J. Phys. Chem. B* **2011**, *115*, 4118−4127.
(51) Klein, H. C.; Cheng, X.; Smith, J. C.; Shen, T. Transfer Matrix Approach to the Hydrogen-bonding in Cellulose I_α Fibrils Describes the Recalcitrance to Thermal Deconstruction. *J. Chem. Phys.* **2011**, *135*, 085106.
(52) Yui, T.; Hayashi, S. Molecular Dynamics Simulations of Solvated Crystal Models of Cellulose I_α and III_I. *Biomacromolecules* **2007**, *8*, 817−824.
(53) Cho, H. M.; Gross, A. S.; Chu, J.-W. *J. Am. Chem. Soc.* **2011**, *133*, 14033−14041.
(54) Chang, R.; Gross, A. S.; J-W.. *J. Phys. Chem. B* **2012**, *116*, 8074−8083.