

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/227394218>

# Coarse-Grained Model for the Interconversion between Native and Liquid Ammonia-Treated Crystalline Cellulose

ARTICLE in THE JOURNAL OF PHYSICAL CHEMISTRY B · JUNE 2012

Impact Factor: 3.3 · DOI: 10.1021/jp300354q · Source: PubMed

CITATIONS

9

READS

28

## 6 AUTHORS, INCLUDING:



**Giovanni Bellesia**

University of California, Santa Barbara

26 PUBLICATIONS 567 CITATIONS

SEE PROFILE



**Shishir Chundawat**

Rutgers, The State University of New Jersey

66 PUBLICATIONS 2,078 CITATIONS

SEE PROFILE



**Paul Langan**

Oak Ridge National Laboratory

123 PUBLICATIONS 3,052 CITATIONS

SEE PROFILE



**Bruce E Dale**

Michigan State University

333 PUBLICATIONS 11,020 CITATIONS

SEE PROFILE

# Coarse-Grained Model for the Interconversion between Native and Liquid Ammonia-Treated Crystalline Cellulose

Giovanni Bellesia,<sup>†</sup> Shishir P. S. Chundawat,<sup>‡</sup> Paul Langan,<sup>§</sup> Antonio Redondo,<sup>†</sup> Bruce E. Dale,<sup>‡</sup> and S. Gnanakaran<sup>\*,†</sup>

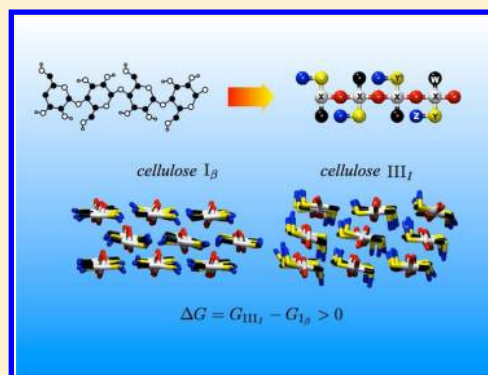
<sup>†</sup>Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, United States

<sup>‡</sup>Great Lakes Bioenergy Research Center (GLBRC), Biomass Conversion Research Laboratory, Department of Chemical Engineering and Material Science, Michigan State University, Lansing, Michigan 48824, United States

<sup>§</sup>Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831-6475, United States

## S Supporting Information

**ABSTRACT:** We present the results of Langevin dynamics simulations on a coarse-grained model for a structural transition in crystalline cellulose pertinent to the cellulose degradation problem. We analyze two different cellulose crystalline forms: cellulose  $I_\beta$  (the natural form of cellulose) and cellulose  $III_1$  (obtained after cellulose  $I_\beta$  is treated with anhydrous liquid ammonia). Cellulose  $III_1$  has been the focus of wide interest in the field of cellulosic biofuels, as it can be efficiently hydrolyzed to readily fermentable glucose (its enzymatic degradation rates are up to 5-fold higher than those of cellulose  $I_\beta$ ). The coarse-grained model presented in this study is based on a simplified geometry and on an effective potential mimicking the changes in both intracrystalline hydrogen bonds and stacking interactions during the transition from cellulose  $I_\beta$  to cellulose  $III_1$ . The model reproduces both structural and thermomechanical properties of cellulose  $I_\beta$  and  $III_1$ . The work presented herein describes the structural transition from cellulose  $I_\beta$  to cellulose  $III_1$  as driven by the change in the equilibrium state of two degrees of freedom in the cellulose chains. The structural transition from cellulose  $I_\beta$  to cellulose  $III_1$  is essentially reduced to a search for optimal spatial arrangement of the cellulose chains.



## INTRODUCTION

Efficient enzyme-catalyzed degradation of crystalline cellulose to glucose is one of the main scientific roadblocks to the production of biofuels from lignocellulosic biomass.<sup>1</sup> Cellulose is synthesized at the plasma membrane of growing plants as crystalline nanofibers stabilized by an extended network of hydrogen bonds and by stacking interactions. Both intracrystalline hydrogen bonds and stacking interactions make crystalline cellulose recalcitrant to enzymatic hydrolysis, therefore hampering the efficiency of the cellulose-to-glucose degradation process.

A novel approach to improve the enzyme deconstruction of crystalline cellulose relies on the use of chemical treatments for altering the cellulose crystal structure, rewiring its internal hydrogen bond network and increasing its depolymerization rate.<sup>2–8</sup> In particular, anhydrous liquid ammonia (an inexpensive and easily recoverable chemical) has been used to convert the naturally occurring crystalline form of cellulose (cellulose  $I_\beta$ ) to a less recalcitrant form called cellulose  $III_1$ . This ammonia-based chemical treatment leads to a rearrangement of the internal cellulose hydrogen bond network and consequently to a change in the cellulose crystalline structure. The relevance of the less recalcitrant cellulose  $III_1$  crystalline form in biofuel production has been confirmed by recent experiments that

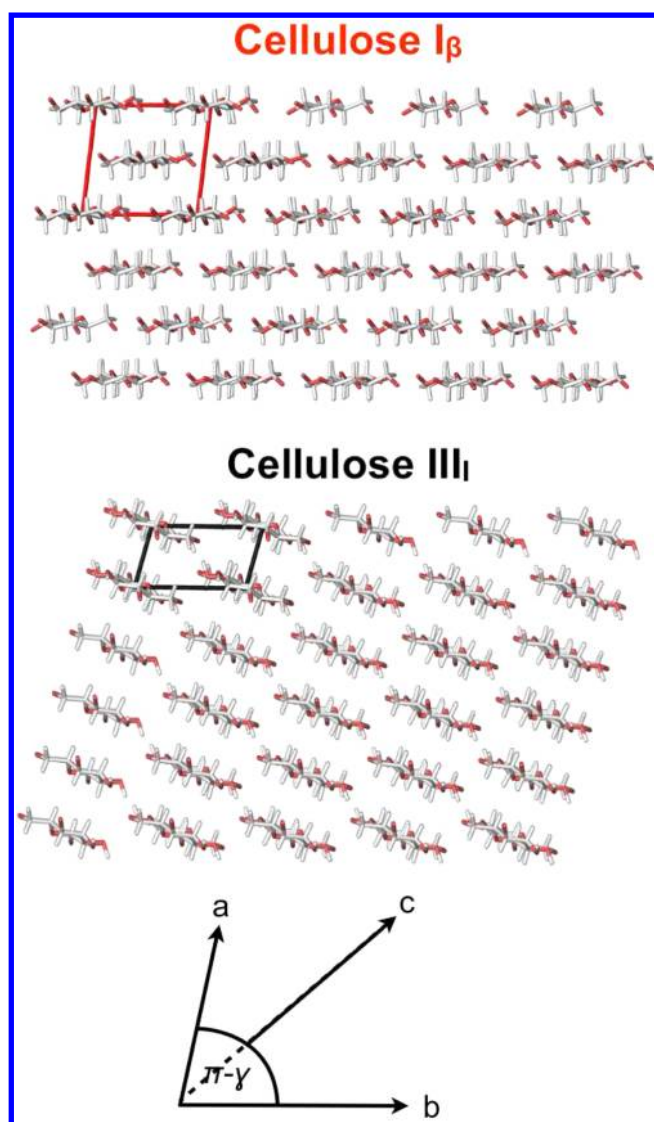
showed up to 5 fold higher enzymatic degradation rates for cellulose  $III_1$  compared to cellulose  $I_\beta$ .<sup>9</sup> The crystalline structures of both cellulose  $I_\beta$  and cellulose  $III_1$  have been resolved in recent years via synchrotron X-ray and neutron diffraction experiments<sup>10,11</sup> and are shown in Figure 1. A physical understanding of how the main structural and thermodynamic differences between these two crystalline cellulose allomorphs affect their different enzyme hydrolysis rates could lead to the design of novel cellulose chemical pretreatments and more efficient hydrolytic enzymes.

Results from recent molecular dynamics (MD) and quantum chemical studies revealed several key microscopic insights on the difference between cellulose  $I_\beta$  and cellulose  $III_1$  and provided mechanistic details on the ammonia penetration into cellulose  $I_\beta$ .<sup>9,12</sup> A comparative MD study on water-solvated fibrils of cellulose  $I_\beta$  and  $III_1$ <sup>9</sup> shows that thermal fluctuations within the crystalline fibril core as well as on the solvent-exposed fibril surface are systematically larger in cellulose  $III_1$  than in cellulose  $I_\beta$ . Also, the cellulose  $III_1$  fibril appears to be hydrated to a greater degree compared to cellulose  $I_\beta$ .

**Received:** January 11, 2012

**Revised:** June 17, 2012

**Published:** June 19, 2012



**Figure 1.** Experimentally-based crystal structure schematics for cellulose  $I_{\beta}$  and cellulose  $III_I$ .  $a$ ,  $b$ ,  $c$ , and  $\gamma$  define the crystal unit parameters. The vector  $c$  has magnitude equal to the cellobiose length and is perpendicular to the plane defined by the figure.  $a \sin(\gamma)$  defines the intersheet direction, whereas  $b$  defines the distance between cellulose chains within the same horizontal sheet. Cellulose structures were generated using Pymol.<sup>35</sup>

Accordingly, all-atom MD simulations capture the formation of a larger number of cellulose–water hydrogen bonds in cellulose  $III_I$  with respect to cellulose  $I_{\beta}$ . This hydration trend was recently confirmed by quantum chemical calculations.<sup>13</sup>

Furthermore, MD simulations confirm the presence of subtle differences in the conformation of single cellulose chains within the cellulose  $I_{\beta}$  and cellulose  $III_I$  fibrils.<sup>9</sup> The most relevant of these conformational differences is related to the different rotational state of the glucose side-chain hydroxymethyl group. In cellulose  $I_{\beta}$  the hydroxymethyl group is mainly found in the so-called *tg* (trans–gauche) rotational state, whereas in cellulose  $III_I$  the dominant rotational state is *gt* (gauche–trans). Another MD simulation study on the interaction between liquid ammonia and a cellulose  $I_{\beta}$  fibril elucidates the importance of the hydroxymethyl group rotational state in driving the structural crossover to an intermediate crystalline structure similar to cellulose  $III_I$ .<sup>12</sup> In more detail, that study

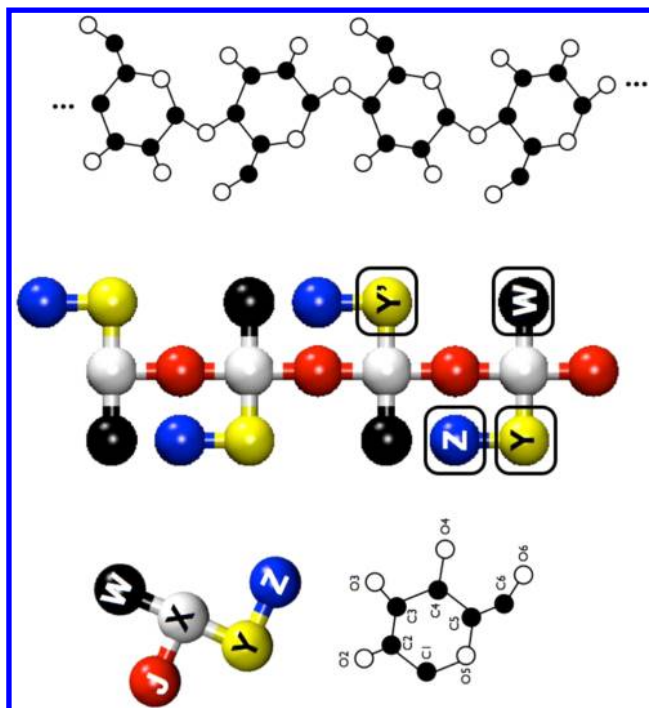
shows that a change from the *tg* to *gt* rotational state in the hydroxymethyl group initiates a rearrangement of the hydrogen bond network within the cellulose fibril and eventually leads to a crossover toward a new structure resembling cellulose  $III_I$ .

In this paper, we introduce a simplified molecular model for crystalline cellulose that captures the interconversion between cellulose  $I_{\beta}$  and cellulose  $III_I$ . Our approach to modeling is similar in nature to the one typically employed in generic, beads-and-springs protein models<sup>14</sup> built considering the most basic features of a polymer chain (chain connectivity, short-range excluded volume and van der Waals attraction) together with a limited set of critical degrees of freedom pertinent to the system of interest and to the problem under consideration. In detail, our model is based on the following physical considerations. First, experimental results and previous all-atom molecular dynamics simulations<sup>9–12,15</sup> highlight the importance of the hydroxymethyl group rotational state in determining the crystalline cellulose hydrogen bond pattern and eventually the cellulose crystal structure. Our model must hence incorporate a coarse-grained representation of the hydroxymethyl group that is realistic enough to include the relevant rotational states. Second, the stabilization of cellulose in its crystalline and aggregated states involves hydrogen bonds as well as stacking interactions (mostly van der Waals interactions) between glucose rings on neighboring sheets within the cellulose crystal. The importance of intersheet spacing in the interconversion between different crystalline cellulose phases has been confirmed by a recent ab initio computational study.<sup>16</sup> Third, in all crystalline cellulose allomorphs, bond lengths, bond angles, and dihedral angles do not vary significantly from their equilibrium values, with the only exception being the hydroxymethyl dihedral angle. Our model includes all these interactions with appropriate relative strengths. Essentially, we introduce an effective potential based on two energy terms: a Lennard–Jones term affecting the intersheet equilibrium distance (along direction  $a$  in Figure 1), and a dihedral constraint mimicking the switch of the hydroxymethyl group between *tg* and *gt* rotational states in a single cellulose chain. We find that both energy terms are necessary to describe the changes in the intersheet interactions (direction  $a$  in Figure 1) between cellulose  $I_{\beta}$  and cellulose  $III_I$ <sup>10,11</sup> and to reproduce the correct packing for both crystalline allomorphs.

In our model, most of the interactions are purely repulsive and therefore we essentially reduce the cellulose  $I_{\beta}$ –cellulose  $III_I$  structural transition to a search for optimal spatial arrangement of the cellulose chains. Our model augments recent coarse-grained modeling studies focusing on cellulose-solvated oligomers as well as on crystalline cellulose fibrils.<sup>17–22</sup> The novelty of our model is its straightforward description of a solid–solid phase transition, pertinent to cellulose degradation, within a simple physical framework. The thermodynamics of the transition has been analyzed using free energy perturbation theory. The results obtained from our modeling study suggest that chemical treatments aimed at disrupting cellulose intersheet interactions should be able to interfere with both the native intracrystalline hydrogen bond network and stacking interactions and therefore decrease cellulose recalcitrance to enzymatic degradation.

## ■ MATERIALS AND METHODS

**Coarse-Grained Model.** The coarse-grained geometry used in our model is shown in Figure 2. Each glucose unit is



**Figure 2.** Geometry of the coarse-grained model. Each glucose monomer (left) is represented by five beads: two for the backbone (X and J; white and red, respectively) and three for the side chains (Y, W, and Z; yellow, black, and blue, respectively). A more detailed mapping between the beads and the glucose heavy atoms can be defined as follows: for the backbone, X represents the atoms C1, C2, C3, C4, C5, and O5, and J represents the glycosidic oxygen O4 (or O1). For the side chain, Y and Z represent C6 and O6, whereas W corresponds to the two oxygens O2 and O3. Atomistic representations of both a glucose monomer (left) and a cellulose oligomer are shown for comparison. Black circles = carbon atoms. White circles = oxygen atoms. Hydrogen atoms are not shown. The highlighted beads ZYWW' are the ones affected by a critical dihedral potential term in our Hamiltonian. Details on that potential term can be found in the Materials and Methods.

represented by five beads (J, X, Y, W, Z) connected by harmonic springs. The mapping between the beads and the glucose heavy atoms can be defined as follows: for the backbone, X represents the atoms C1, C2, C3, C4, C5, and O5, and J represents the glycosidic oxygen O4 (or O1). For the side groups, Y and Z represent the hydroxymethyl dihedral switch, whereas W corresponds to the two oxygens O2 and O3. The Hamiltonian used in our study has the following functional form:

$$\begin{aligned}
 H(r, \theta, \omega) = & \sum_{i \neq j} 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \Lambda_{ij} \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \\
 & + \sum_{\text{bonds}} \frac{K_b(ij)}{2} (r_{ij} - r_{ij}^0)^2 \\
 & + \sum_{\text{angles}} \frac{K_\theta(ijk)}{2} (\theta_{ijk} - \theta_{ijk}^0)^2 \\
 & + \sum_{\text{dihe}} C(ijkl) \cos(\omega_{ijkl} - \omega_{ijkl}^0)
 \end{aligned} \quad (1)$$

In reduced units  $\epsilon^*$ ,  $\sigma^*$ , and  $m^*$  define the energy, length, and mass units, respectively. For the Lennard–Jones energy

term,  $\epsilon_{ij} = 1.0\epsilon^*$ ,  $\forall ij$  except for XX and ZW pairs where  $\epsilon_{XX} = \epsilon_{ZW} = 2.0\epsilon^*$ ,  $\sigma_{ij} = 1.0\sigma^*$ ,  $\forall ij$  except for XX pairs where the equilibrium distance  $\sigma_{XX}$  takes on the values 1.50, 1.60... $2.0\sigma^*$  and  $\Lambda_{ij} = 0.0$  except for XX and ZW pairs where  $\Lambda_{XX} = \Lambda_{ZW} = 1.0$ . The bonded interactions energy terms are parametrized as follows:  $K_b(ij) = 800.0\epsilon^*(\sigma^*)^{-2}$ ,  $r_{ij}^0 = \sigma^*$ ,  $K_\theta(ijk) = 40.0\epsilon^*$ ,  $\theta_{jxy}^0 = \theta_{pxw}^0 = \theta_{xyz}^0 = 90^\circ$ ,  $\theta_{jxj}^0 = 180^\circ$ ,  $\theta_{xjx}^0 = 120^\circ$ ,  $C(ijkl) = 13.0\epsilon^*$ ,  $\omega_{jjjj}^0 = \omega_{pxxj}^0 = 0.0^\circ$  and  $\omega_{zywy'}^0 = 130^\circ$ ,  $135^\circ$ , ...,  $180^\circ$ . The four beads affected by the equilibrium dihedral phase  $\omega_{zywy'}$  are highlighted in Figure 2. The mass is  $1.0m^*$  for all five bead types. The energy coefficients were initially chosen as follows: (1) for the bonded interactions, the goal was to maintain the various bond lengths and angles almost rigid (energy coefficients  $\sim 10^1$ – $10^2 k_B T$ ). (2) For the nonbonded interactions, the range of the repulsive part is simply  $\sigma^*$ . The XX attractive part is what changes in our model going from cellulose I $_\beta$  to III $_I$  and was chosen to be  $\sim 3$ – $4k_B T$ .

**Free Energy Calculation.** The free energy perturbation (FEP) method under isothermal–isobaric conditions (NPT)<sup>23,24</sup> is used to estimate the relative change in free energy  $\Delta G_{\text{FEP}}$  as the two parameters  $\omega_{zywy'}^0$  and  $\sigma_{XX}$  change (see previous section). In the FEP approach  $\Delta G_{\text{FEP}}$  between two states A and B is calculated as follows:

$$\begin{aligned}
 \Delta G_{\text{FEP}} = G_B - G_A = & \sum_{i=1}^{N_w} \Delta G(\lambda_i) = -k_B T \\
 & \sum_{\lambda=0}^1 \ln \left\langle \frac{V \exp \left[ -\frac{H(r, \lambda + \delta\lambda) - H(r, \lambda)}{k_B T} \right]}{\langle V \rangle} \right\rangle_{\lambda}
 \end{aligned} \quad (2)$$

where  $N_w$  is the number of the intermediate states defined by the value of the perturbation parameter  $\lambda$ ,  $\langle \rangle$  denotes an average in the NPT ensemble,  $k_B T = 0.793745\epsilon^*$  ( $k_B = 1.0$ ),  $V$  is the volume of the simulation box,  $\delta\lambda = 0.1$  and  $\lambda$  varies from 0.0 to 1.0. In order to apply the FEP method we need to explicitly show the dependence of the Hamiltonian from the perturbation parameter  $\lambda$  and connect this dependence to the change of the parameters  $\sigma_{XX}$  and  $\omega_{zywy'}^0$ . In the simple case where  $\sigma_{XX}$  and  $\omega_{zywy'}^0$  vary simultaneously, the Lennard–Jones energy term for the XX pair and the torsional potential term for the quadruplet ZYWW' (Figure 2) can be rewritten as:

$$4\epsilon_{ij} \left[ \left( \frac{2.0 - 0.5\lambda_{\text{LJ}}}{r_{XX}} \right)^{12} - \left( \frac{2.0 - 0.5\lambda_{\text{LJ}}}{r_{XX}} \right)^6 \right] \quad (3)$$

and

$$C(\text{ZYWW}') \cos(\omega_{zywy'} + 180^\circ - 50^\circ \lambda_{\text{rot}}) \quad (4)$$

respectively, and  $\lambda_{\text{LJ}} = \lambda_{\text{rot}} = \lambda$ . We associate  $\lambda_{\text{LJ}} = \lambda_{\text{rot}} = \lambda = 0.0$ , and therefore  $\omega_{zywy'}^0 = 180^\circ$  and  $\sigma_{XX} = 2.0\sigma^*$ , with the *tg* hydroxymethyl rotational state typical of cellulose I $_\beta$ . The *gt* rotational state typical of cellulose III $_I$  is associated with  $\lambda_{\text{LJ}} = \lambda_{\text{rot}} = \lambda = 1.0$  and hence  $\omega_{zywy'}^0 = 130^\circ$  and  $\sigma_{XX} = 1.5\sigma^*$ . The change in  $\sigma_{XX}$  from 2.0 to 1.5 going from cellulose I $_\beta$  to III $_I$  can be associated with increased intersheet interactions (*a* direction in Figure 1) between glucose rings in cellulose III $_I$  with respect to cellulose I $_\beta$ .<sup>10,11</sup> The decrease of the equilibrium distance  $\sigma_{XX}$  does not affect the intrasheet interactions (*b* direction in Figure 1) between X beads, as these are counterbalanced by the interactions between the side chain beads Y, W, and Z. It is worth mentioning that the main goal of the free energy



perturbation calculations is to obtain a qualitative picture of the relative stability of the two allomorphs. It is, in fact, difficult to assess the model's accuracy, as there are no detailed experimental results on the free energy difference between cellulose I<sub>β</sub> and III<sub>I</sub>.

**Langevin Dynamics Simulations.** Langevin dynamics (LD) simulations under NPT ensemble conditions have been carried out using the NAMD software package<sup>25</sup> [with  $\lambda_{LJ}$  and  $\lambda_{rot}$  taking the values 0.0, 0.1, 0.2...1.0]. The time step was  $t_s = 0.0145t^*$  where  $t^* = \sigma^*/\epsilon^*m^*$  and the simulation time was  $10^7t_s$  for each of the 242 independent LD simulations in our study. The temperature was fixed at  $T = 0.7944\epsilon^*k_B$  via a Langevin thermostat, whereas the pressure was fixed at  $p = 1.2 \times 10^{-4}p^*$  where  $p^* = \epsilon^*(\sigma^*)^{-3}$  using a Nosé–Hover Langevin barostat.<sup>26,27</sup> The barostat allowed the three dimensions of the periodic box to change independently. The cutoff for the Lennard–Jones potential was fixed at  $4\sigma^*$ , the damping coefficient in the Langevin integrator was set to  $\zeta = 1.45(\sigma^*)^{-1}$ , the Langevin piston period was  $2.89t^*$ , and the Langevin piston decay was  $1.45t^*$ . An infinite crystal with a structure qualitatively similar to cellulose I<sub>β</sub> was generated, applying periodic boundary conditions on the central simulation box containing 36 coarse-grained, octameric cellulose chains (see Figure S1, Supporting Information). The initial crystal conformation underwent a short energy minimization, and it was used to start a first set of 11 LD simulations. In this set, we varied  $\lambda_{LJ}$  and  $\lambda_{rot}$  simultaneously from 0.0 to 1.0, and the final conformation for the  $i$ th LD run ( $i = 1, 2, 3...10$ ) was used as the starting conformation for the  $i+1$ th LD run. The final conformation of the 11th LD simulation run from this set (where  $\lambda_{LJ} = \lambda_{rot} = \lambda = 1.0$ ) was then used to start a second set of 11 LD simulations where we varied  $\lambda_{LJ}$  and  $\lambda_{rot}$  simultaneously from 1.0 back down to 0.0. In addition to these first two simulation sets, we started a number of simulations where  $\lambda_{LJ}$  and  $\lambda_{rot}$  vary independently between 0 and 1. Figure S2 (Supporting Information) shows in a grid-like form all the 242 simulations performed in this study and how the final and initial configurations of neighboring simulations in the  $(\lambda_{LJ}, \lambda_{rot})$  two-dimensional space are connected. Each cell in the grid represents an independent LD run.

**Additional Details on the Calculation of the Free Energy Surface.** We focus here, for the sake of simplicity, on the simulation scheme shown at the top of Figure S2. Using a matrix notation, we denote the two LD simulations where  $\sigma_{XX} = 2.0$  and  $\omega_{ZYWY}^0 = 180^\circ$ , and  $\sigma_{XX} = 1.5$  and  $\omega_{ZYWY}^0 = 130^\circ$  as (1,1) and (1,1), respectively. We use a free energy perturbation equation to calculate  $\Delta G$  between two neighboring LD simulations in the grid shown. As an example, for LD simulations (1,1) and (1,2):

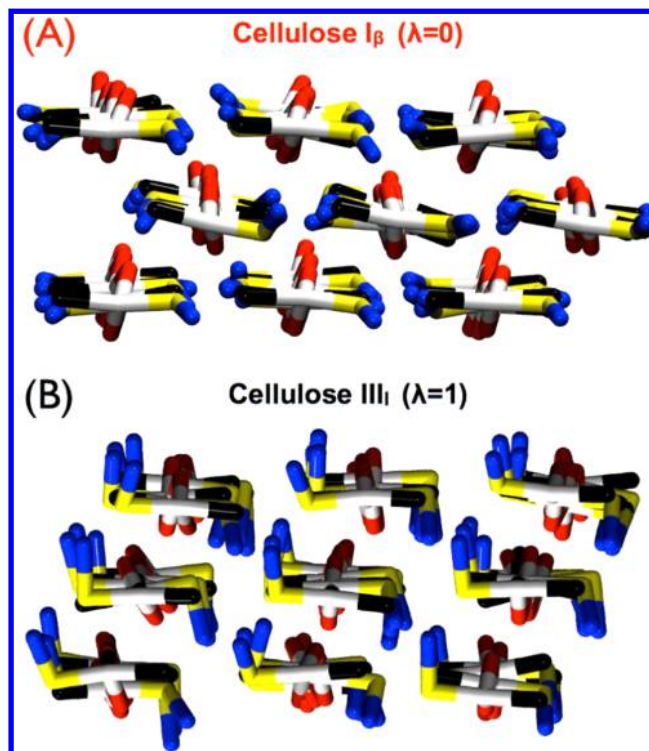
$$\Delta G[(1,2), (1,1)] = -k_B T \ln \left\langle \frac{V \exp \left[ -\frac{H(r,(1,2)) - H(r,(1,1))}{k_B T} \right]}{\langle V \rangle} \right\rangle \quad (5)$$

where  $H(r,(1,2))$  and  $H(r,(1,1))$  are the Hamiltonians for LD simulations (1,1) and (1,2), respectively. We set the LD simulation (1,1) as the zero free energy point and calculate the free energy difference between (1,1) and  $(i,j)$  ( $i,j > 1$ ) as the average over a number of possible “paths” starting at (1,1) and ending at  $(i,j)$ . The paths are generated considering the portion of the grid comprised between (1,1) and  $(i,j)$  and considering

only “forward” steps. The difference in free energy  $\Delta G[(i,j), (1,1)]$  for each path is calculated as the sum of all the free energy steps between neighboring LD simulations (nodes on the grid) composing the path. The number of paths for each  $\Delta G[(i,j), (1,1)]$  varies between  $\sim 80$  and  $\sim 400$ .

## RESULTS AND DISCUSSION

Two representative snapshots obtained from our LD simulations where  $\lambda_{LJ} = \lambda_{rot} = \lambda = 0.0$  (cellulose I<sub>β</sub>) and  $\lambda_{LJ} = \lambda_{rot} = \lambda = 1.0$  (cellulose III<sub>I</sub>) are shown in Figure 3A and



**Figure 3.** (A) Representative snapshot for the LD simulation where  $\lambda_{rot} = \lambda_{LJ} = 0.0$ . The crystal structure closely resembles cellulose I<sub>β</sub>. (B) Representative snapshot for the LD simulation where  $\lambda_{rot} = \lambda_{LJ} = 1.0$ . The crystal structure closely resembles cellulose III<sub>I</sub>. Only nine coarse-grained cellulose chains of the infinite crystal are shown.

Figure 3B, respectively (see also Figures S3 and S4 in Supporting Information for additional snapshots of the cellulose I<sub>β</sub> and cellulose III<sub>I</sub> structures). A comparison of Figure 3 and Figure S3 with the crystal structures in Figure 1 clearly shows that our simple molecular model reproduces the correct molecular packing for both the cellulose I<sub>β</sub> and cellulose III<sub>I</sub> allomorphs.

A quantitative structural analysis of the two crystal structures obtained from our LD simulations is given in Table 1 where we compare our results with experimental data<sup>10,11</sup> and results from previous all atom molecular dynamics simulations.<sup>9</sup> The first four columns on the left of Table 1 show the values for the different ratios of the time-averaged distances ( $a$ ,  $b$ , and  $c$ ) and for the angle  $\gamma$  (for the definitions of  $a$ ,  $b$ ,  $c$ , and  $\gamma$  see Figure 1). The four values  $\langle a \rangle / \langle b \rangle$ ,  $\langle a \rangle / \langle c \rangle$ ,  $\langle b \rangle / \langle c \rangle$ , and  $\langle \gamma \rangle$  uniquely define the shape of the monoclinic crystals. The two remaining angles defining the monoclinic crystal unit in cellulose I<sub>β</sub> and III<sub>I</sub> are not considered in our analysis because they are both equal to  $90^\circ$  and therefore identical to the experimental values.

**Table 1. Quantitative Comparison between the Average Properties Obtained from our LD Simulations with  $\lambda = 0.0$  (cellulose I <sub>$\beta$</sub> ) and  $\lambda = 1.0$  (cellulose III<sub>I</sub>)<sup>a</sup>**

| cellulose I <sub><math>\beta</math></sub> | $\langle a \rangle / \langle b \rangle$ | $\langle a \rangle / \langle c \rangle$ | $\langle b \rangle / \langle c \rangle$ | $\langle \gamma \rangle$ | $\sigma_{\%}(a)$ | $\sigma_{\%}(b)$ | $\sigma_{\%}(a) / \sigma_{\%}(b)$ |
|---|---|---|---|--------------------------|------------------|------------------|-----------------------------------|
| Exp                                       | 0.95                                    | 0.76                                    | 0.79                                    | 96.5                     |                  |                  |                                   |
| MD  | 0.93                                    | 0.67                                    | 0.72                                    | 98.2                     | 2.4              | 2.0              | 1.2                               |
| CG  | 0.93                                    | 0.75                                    | 0.81                                    | 91.1                     | 3.5              | 2.8              | 1.3                               |
| cellulose III <sub>I</sub>                | $\langle a \rangle / \langle b \rangle$ | $\langle a \rangle / \langle c \rangle$ | $\langle b \rangle / \langle c \rangle$ | $\langle \gamma \rangle$ | $\sigma_{\%}(a)$ | $\sigma_{\%}(b)$ | $\sigma_{\%}(a) / \sigma_{\%}(b)$ |
| Exp                                       | 0.57                                    | 0.43                                    | 0.76                                    | 105.1                    |                  |                  |                                   |
| MD  | 0.53                                    | 0.39                                    | 0.74                                    | 114.5                    | 3.6              | 3.0              | 1.2                               |
| CG  | 0.52                                    | 0.44                                    | 0.85                                    | 111.6                    | 5.9              | 4.4              | 1.3                               |

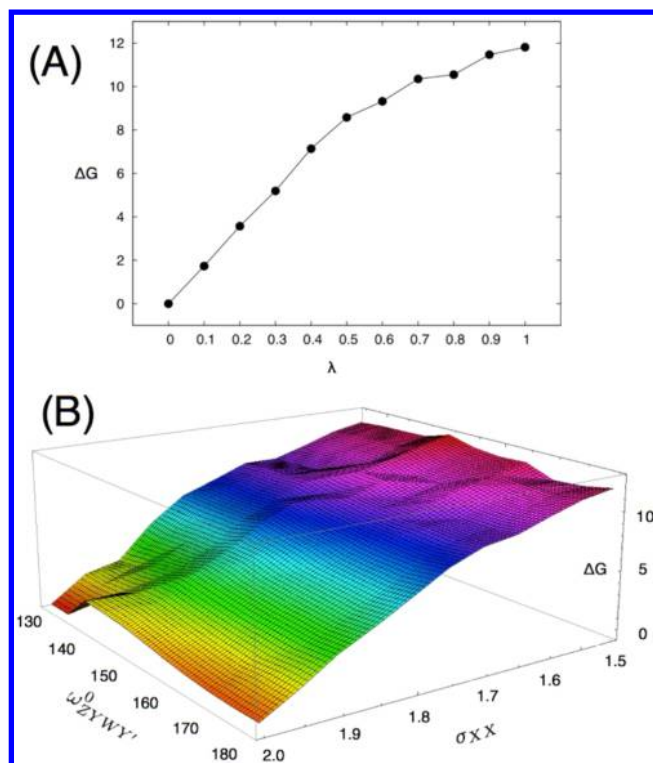
<sup>a</sup> $a$ ,  $b$ ,  $c$ , and  $\gamma$  define the crystal unit parameters, whereas  $\sigma_{\%}$  is the relative standard deviation (expressed as a percentage) of the time-averaged values denoted with  $\langle \rangle$  and calculated from our LD simulations. Exp: experimental results from synchrotron X-ray and neutron diffraction. MD: all atom molecular dynamics simulations. CG: coarse-grained model. The relative errors on  $\langle a \rangle / \langle b \rangle$ ,  $\langle a \rangle / \langle c \rangle$ ,  $\langle b \rangle / \langle c \rangle$ , and  $\langle \gamma \rangle$  vary between 1% and 4%.

Our model shows an excellent agreement with the structural data from neutron and X-ray measurements.<sup>10,11</sup> In fact, its degree of structural accuracy is comparable or slightly better to the one observed in all atom MD simulations. The largest deviations from the experimental values have been observed in the cellulose III<sub>I</sub> crystal and specifically for  $\langle a \rangle / \langle b \rangle$  and  $\langle b \rangle / \langle c \rangle$  (8.8% and 11.8%, respectively), whereas the ratio  $\langle a \rangle / \langle c \rangle$  and the angle  $\gamma$  are reproduced with a similar degree of accuracy in both the cellulose I <sub>$\beta$</sub>  and III<sub>I</sub> crystals (Table 1). If we consider  $\sigma^* = 1.26$  Å, we observe that the crystal unit cells dimensions for the two allomorphs are reproduced with a level of accuracy comparable to the one observed in all atom MD simulations. Given the generic, phenomenological nature of the model, we believe that the ratios in Table 1 are more informative than the single values of the unit cell dimensions.

Another important structural parameter is the crystal unit volume. Experiments show that cellulose chains are more densely packed in cellulose I <sub>$\beta$</sub>  with a crystal unit volume ratio between cellulose I <sub>$\beta$</sub>  and III<sub>I</sub>,  $V(I_{\beta})/V(III_I) = 0.95$ .<sup>10,11</sup> Our model reproduces correctly the denser packing in cellulose I <sub>$\beta$</sub>  with a volume ratio equal to  $0.86 \pm 0.09$ .

The coarse-grained model also captures the differences in structural flexibility between the two cellulose allomorphs. On the three remaining (right) columns of Table 1 we show the relative standard deviations  $\sigma_{\%}(a)$  and  $\sigma_{\%}(b)$  along the intersheet and intrasheet directions, respectively ( $a$  and  $b$  axes in Figure 1), together with their respective ratios. The relative standard deviations  $\sigma_{\%}(a)$  and  $\sigma_{\%}(b)$  give a measure of the structural flexibility of the crystals along the respective directions. The coarse-grained model reproduces with good accuracy the trends observed in all atom MD simulations. The reproduced structural flexibility trends are as follows: (i) the overall larger structural flexibility of the cellulose III<sub>I</sub> crystal, (ii) the largest deviation observed along the intersheet direction ( $a$ ) in cellulose III<sub>I</sub>; (iii) the ratios between the structural fluctuations along the intersheet and intrasheet directions systematically larger than 1.

Our simplified model qualitatively reproduces the expected relative stability between the crystalline cellulose allomorphs. Figure 4A shows the change in free energy (normalized over the number of cellobiose units in the simulation box) between cellulose I <sub>$\beta$</sub>  and cellulose III<sub>I</sub> as a function of  $\lambda = \lambda_{IJ} = \lambda_{rot}$ . The



**Figure 4.** (A) Free energy profile showing the path from cellulose I <sub>$\beta$</sub>  to cellulose III<sub>I</sub> when  $\lambda_{rot} = \lambda_{IJ} = \lambda$ . (B) Free energy surface plot in the  $\omega_{ZYWY}^0, \sigma_{XX}$  space.  $\omega_{ZYWY}^0 = 180^\circ$  and  $\sigma_{XX} = 2.0$  (bottom right corner) correspond to cellulose I <sub>$\beta$</sub> , while  $\omega_{ZYWY}^0 = 130^\circ$  and  $\sigma_{XX} = 1.5$  (top left corner) correspond to cellulose III<sub>I</sub>. The vertical axis is the relative free energy in units of  $\epsilon^*$ . Both plots show that cellulose III<sub>I</sub> in our model has a higher free energy than cellulose I <sub>$\beta$</sub> . In both plots the errors vary from 0.1 to  $0.7\epsilon^*$ .

change in free energy is defined as  $\Delta G = G_{III} - G_I$  and is expressed in units of  $\epsilon^*$ . The free energy minimum (rescaled to 0.0) for  $\omega_{ZYWY}^0 = 180^\circ$  and  $\sigma_{XX} = 2.0$  corresponds to cellulose I <sub>$\beta$</sub>  (Figure 3A), the higher free energy value ( $+11.81 \pm 0.40\epsilon^*$ ) for  $\omega_{ZYWY}^0 = 130^\circ$ , and  $\sigma_{XX} = 1.5$  corresponds to cellulose III<sub>I</sub>. The change in free energy between cellulose I <sub>$\beta$</sub>  and cellulose III<sub>I</sub> is mostly due to a change in enthalpy ( $\Delta H = 13.54 \pm 0.30\epsilon^*$ ), while the entropy term  $T\Delta S = 1.73 \pm 0.40\epsilon^*$ . The enthalpy is calculated as  $H = \langle U \rangle + P\langle V \rangle$  where  $U$  is the potential energy,  $P$  is the pressure, and  $V$  is the volume of the simulation box.  $\langle \rangle$  indicates an average over simulation time. We also observe that  $\Delta S > 0$ , indicating that entropy is higher in cellulose III<sub>I</sub> than in cellulose I <sub>$\beta$</sub> . The higher free energy and the enhanced structural flexibility provide a molecular-level explanation for the lower thermal stability<sup>3</sup> of cellulose III<sub>I</sub>.

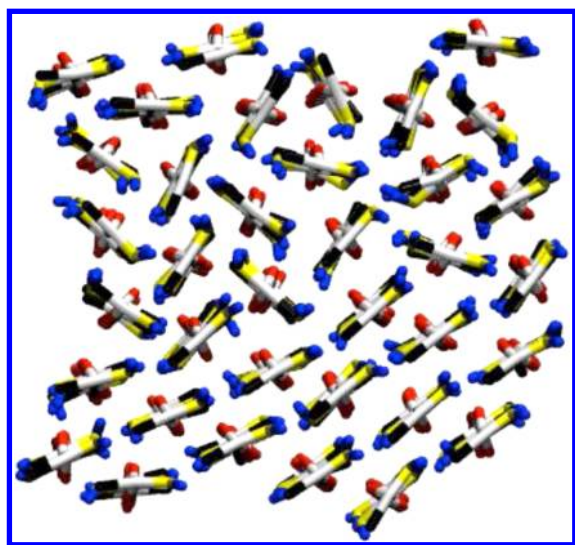
Furthermore, we explore the free energy dependence from the two key structural parameters  $\omega_{ZYWY}^0$  and  $\sigma_{XX}$  varying independently from  $180^\circ$  to  $130^\circ$ , and from 2.0 to 1.5, respectively (or alternatively from  $\lambda_{rot}$  and  $\lambda_{IJ}$  varying independently from 0.0 to 1.0). The resulting free energy surface is shown in Figure 4B. It highlights different paths along these two key parameters whereas, as mentioned before, Figure 4A describes the structural transition when  $\omega_{ZYWY}^0$  and  $\sigma_{XX}$  are varied simultaneously to model the changes in the intersheet interactions between cellulose I <sub>$\beta$</sub>  and cellulose III<sub>I</sub>.<sup>10,11</sup> The details of the calculation of the free energy surface are given in the Supporting Information. The free energy surface shows that changes in the equilibrium distance  $\sigma_{XX}$  are associated to large



changes in the relative free energy. However, the relevance of the change in  $\omega_{ZYWY}^0$  cannot be underestimated. Changes in the equilibrium state of our effective dihedral potential are, indeed, required to obtain a proper cellulose III<sub>I</sub> structure. This has been confirmed by analyzing LD simulations where energy changes associated only with changes in  $\sigma_{XX}$  do not lead to the proper conversion to cellulose III<sub>I</sub> (data not shown).

It is worth mentioning at this point of our analysis that the use of an infinite crystal, as opposed to a finite microfibril, was motivated by the goal of showing that the transformation from cellulose I <sub>$\beta$</sub>  to III<sub>I</sub> can be reproduced even in the absence of surface effects and structural defects. In other words, we wanted to show an idealized and yet faithful picture of this transition which mostly relies on optimal three-dimensional packing of the cellulose chains. In addition, large microcrystals of cellulose I <sub>$\beta$</sub>  to III<sub>I</sub> have been generated by Langan and collaborators<sup>10,11</sup> and used in their X-ray and neutron diffraction studies. These large microcrystals are well approximated in silico by infinite crystals. In addition, the model can be easily extended to finite systems. We observe that it reproduces stable 36-chain, rhomboid and hexagonal fibrils of cellulose I <sub>$\beta$</sub>  and III<sub>I</sub> with a degree of polymerization greater than 100. A detailed study of those fibrils is currently in progress and will be the subject of a future publication.

Finally, we fail to observe an exact reverse structural crossover from cellulose III<sub>I</sub> to I <sub>$\beta$</sub> . This is somewhat consistent with measurements that considered the conversion of cellulose III<sub>I</sub> to I <sub>$\beta$</sub>  at high temperatures. A set of 121 LD simulations (bottom part of Figure S2) starting from a crystalline structure resembling cellulose III<sub>I</sub> was used to monitor the cellulose III<sub>I</sub> to I <sub>$\beta$</sub>  transformation. We observed that the final structure obtained (where  $\omega_{ZYWY}^0 = 180^\circ$  and  $\sigma_{XX} = 2.0$ ) only partially retains the crystal structure of cellulose I <sub>$\beta$</sub>  and exhibits a significant number of structural defects (Figure 5). These structural defects appear to be independent of any artifacts due to the size of the chosen periodic box in our LD simulations.



**Figure 5.** Snapshot of the crystal structure obtained at the end of 21 sequential equilibrium LD simulations on a periodic box containing 36 octameric chains. We started this set of simulations from the cellulose I <sub>$\beta$</sub>  crystal structure ( $\lambda = \lambda_{IJ} = \lambda_{rot} = 0.0$ ) and varied simultaneously  $\lambda_{IJ}$  and  $\lambda_{rot}$  to 1.0 and then back to 0.0 (eq 4 for details on the simulations protocol). The structure only partially retains the crystal order typical of cellulose I <sub>$\beta$</sub>  and presents some degree of structural disorder.

We observe similar defects in LD simulations with a periodic box containing 144 octameric chains (Figure S5, Supporting Information) with the same simulations reproducing the same structural parameters and the free energy difference obtained with 36 octameric chains (data not shown). Experiments have shown that the distinction between native cellulose I <sub>$\beta$</sub>  and the reverted one (called cellulose I<sub>III</sub>) is that the latter has greater lateral disorder than the former.<sup>28–31</sup> In more detail, Chanzy et al. and Sugiyama et al. observed a reduction in the cellulose crystallite size due to fibrillation/cracking of the cellulose microcrystals (splitting into smaller fibers) along the 110 plane when cellulose III<sub>I</sub> is converted back to I <sub>$\beta$</sub> .<sup>31,32</sup> In a more recent study,<sup>33</sup> it is suggested that the cellulose fibrillation is caused by the creation of a large number of defects between neighboring cellulose sheets due to imperfect intersheet translation during the transformation. Similarly, we observe the presence of the structural defects as well as higher free energy (+1.09 e\*) and a slightly larger volume ( $V(I_{\beta})/V(I_{III}) = 0.97 \pm 0.01$ ) in the reverted cellulose I<sub>III</sub> model crystal compared to native cellulose I <sub>$\beta$</sub> . These properties imply that the reverted form of cellulose I<sub>III</sub> (derived from cellulose III<sub>I</sub>) would be more digestible by cellulases than native cellulose I <sub>$\beta$</sub> .

## CONCLUSION

In this study we presented the results of LD simulations on a coarse-grained model for crystalline cellulose. In particular, we analyzed two different cellulose crystalline forms: cellulose I <sub>$\beta$</sub>  (the natural form of cellulose found in higher plants) and cellulose III<sub>I</sub> (obtained after cellulose I <sub>$\beta$</sub>  is treated with anhydrous liquid ammonia). Cellulose III<sub>I</sub> is of wide interest in the field of cellulosic biofuels lately because it can be more easily hydrolyzed to glucose that eventually can be fermented to desirable fuels or chemicals. Our model accurately reproduces both the structural and thermomechanical properties of cellulose I <sub>$\beta$</sub>  and III<sub>I</sub>. The model in its current implementation has been thoroughly tested only on cellulose I <sub>$\beta$</sub>  and cellulose III<sub>I</sub> allomorphs. Nevertheless, we believe that our model also represents a simple general framework for modeling cellulose and that the key parameters we identify could be relevant also for describing general mechanisms in cellulose degradation processes. Indeed, a recent molecular dynamics study reveals a strict association between the hydroxymethyl rotational state and the dissolution process of cellulose driven by ionic liquids.<sup>34</sup>

Indeed, our study provides a simple explanatory framework for structural transitions in crystalline cellulose, in general, and for the transition from cellulose I <sub>$\beta$</sub>  to cellulose III<sub>I</sub> in particular, where the search for optimal spatial arrangement of the cellulose chains appears to be more important than the chemical details of the cellulose crystals. In terms of enzymatic cellulose deconstruction for biofuels production, our model connects the lower recalcitrance of cellulose III<sub>I</sub> (when compared to cellulose I <sub>$\beta$</sub> ) to higher free energy and to overall larger structural flexibility. Our simplified model suggests two key structural parameters that chemical treatments aiming at decreasing cellulose recalcitrance should consider. Namely, the model suggests that experimental strategies based on chemical compounds that specifically interfere with the energetics of the intersheet distance and expand cellulose I <sub>$\beta$</sub>  via alteration of its hydroxymethyl dihedral angles have the most chance of succeeding as effective chemical treatments for cellulose.

## ■ ASSOCIATED CONTENT

### ■ Supporting Information

Additional figures referenced in the main manuscript. This material is available free of charge via the Internet at <http://pubs.acs.org/>

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [gnana@lanl.gov](mailto:gnana@lanl.gov).

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This work was partly funded by the LDRD Program at Los Alamos National Laboratory and NABC; computational resources were provided by CNLS and LANL Institutional Computing. G.B. thanks the Center for Nonlinear Studies at LANL for support. S.P.S.C. and B.E.D. thank the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494) for funding and Dr. Christopher Bianchetti for assistance with Pymol.<sup>35</sup> We thank Jennifer Macke for critical reading of the manuscript.

## ■ REFERENCES

- (1) Chundawat, S. P. S.; Beckham, G. T.; Himmel, M.; Dale, B. E. *Annu. Rev. Chem. Biomol. Eng.* **2011**, *2*, 1–25.
- (2) Swatloski, R. P.; Spear, S. K.; Holbrey, J. D.; Rogers, R. D. *J. Am. Chem. Soc.* **2002**, *124*, 4974–4975.
- (3) Wada, M. *J. Polym. Sci., Part B: Polym. Phys.* **2002**, *40*, 1095–1102.
- (4) Ohno, H.; Fukaya, Y. *Chem. Lett.* **2009**, *38*, 2–7.
- (5) Liu, H.; Sale, K. L.; Holmes, B. M.; Simmons, B. A.; Seema, S. *J. Phys. Chem. B* **2010**, *114*, 4293–4301.
- (6) Perez, S.; Samain, D. *Adv. Carbohydr. Chem. Biochem.* **2010**, *64*, 25–116.
- (7) Kahlen, J.; Masuch, K.; Leonhard, K. *Green Chem.* **2010**, *12*, 2172–2181.
- (8) van de Vyver, S.; Geboers, J.; Jacobs, P. A.; Sels, B. F. *ChemCatChem* **2011**, *3*, 82–94.
- (9) Chundawat, S. P.; Bellesia, G.; Uppugundla, N.; da Costa Sousa, L.; Gao, D.; Cheh, A.; Agarwal, U. P.; Bianchetti, C. M.; Phillips, G. N. J.; Langan, P.; Balan, V.; Gnanakaran, S.; Dale, B. E. *J. Am. Chem. Soc.* **2011**, *133*, 11163–11174.
- (10) Nishiyama, Y.; Langan, P.; Chanzy, H. *J. Am. Chem. Soc.* **2002**, *124*, 9074–9082.
- (11) Wada, M.; Chanzy, H.; Nishiyama, Y.; Langan, P. *Macromolecules* **2004**, *37*, 8548–8555.
- (12) Bellesia, G.; Chundawat, S. P. S.; Langan, P.; Dale, B. E.; Gnanakaran, S. *J. Phys. Chem. B* **2011**, *115*, 9782–9788.
- (13) Parthasarathi, R.; Bellesia, G.; Chundawat, S. P. S.; Dale, B. E.; Langan, P.; Gnanakaran, S. *J. Phys. Chem. A* **2011**, *115*, 14191–14202.
- (14) Tozzini, V. *Curr. Opin. Struct. Biol.* **2005**, *15*, 144–150.
- (15) Chen, P.; Nishiyama, Y.; Mazeau, K. *Macromolecules* **2012**, *45*, 362–368.
- (16) Bucko, T.; Tunega, D.; Angyan, J.; Hafner, J. *J. Phys. Chem. A* **2011**, *115*, 10097–10105.
- (17) Molinero, V.; Goddard, W. A. I. *J. Phys. Chem. B* **2004**, *108*, 1414–1427.
- (18) Lopez, C. A.; Rzepiela, A.; de Vries, A. H.; Dijkhuizen, L.; Huenenberger, P. H.; Marrink, S. J. *J. Chem. Theory Comput.* **2009**, *5*, 3195–3210.
- (19) Bu, L.; Beckham, G. T.; Crowley, M. F.; Chang, C. H.; Matthews, J. F.; Bomble, Y. J.; Adney, W. S.; Himmel, M. E.; Nimlos, M. R. *J. Phys. Chem. B* **2009**, *113*, 10994–11002.
- (20) Hynninen, A.-P.; J., R. M.; Beckham, G. T.; Crowley, M. F.; Nimlos, M. R. *J. Chem. Theory Comput.* **2011**, *7*, 2137–2150.
- (21) Srinivas, G.; Cheng, X.; Smith, J. C. *J. Chem. Theory Comput.* **2011**, *7*, 2539–2548.
- (22) Wohler, J.; Berglund, L. A. *J. Chem. Theory Comput.* **2011**, *7*, 753–760.
- (23) Shing, K. S.; Chung, S. T. *J. Phys. Chem.* **1987**, *91*, 1674–1681.
- (24) Ghoufi, A.; Bonal, C.; Morel, J.; Morel-Desrosiers, N.; Malfreyt, P. *J. Phys. Chem. B* **2004**, *108*, 11744–11752.
- (25) Phillips, J.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
- (26) Feller, S.; Zhang, Y.; Pastor, R.; Brooks, B. *J. Chem. Phys.* **1995**, *103*, 4613.
- (27) Martyna, G. J.; Tobias, D. J.; L., K. M. *J. Chem. Phys.* **1994**, *101*, 4177.
- (28) Wada, M.; Heux, L.; Sugiyama, J. *Biomacromolecules* **2004**, *5*, 1385–1391.
- (29) Segal, L. *J. Polym. Sci.* **1961**, 395–409.
- (30) Roche, E.; Chanzy, H. *Int. J. Biol. Macromol.* **1981**, *3*, 201–206.
- (31) Sugiyama, J.; Harada, H.; Saiki, H. *Int. J. Biol. Macromol.* **1987**, *9*, 122–130.
- (32) Chanzy, H.; Henrissat, B.; Vincendon, M.; Tanner, S. F.; Belton, P. S. *Carbohydr. Res.* **1987**, *160*, 1–11.
- (33) Wada, M. *Macromolecules* **2001**, *34*, 3271–3275.
- (34) Liu, H.; Chen, G.; Kent, M. S.; Stavila, V.; Simmons, B. A.; Sale, K. L.; Seema, S. *J. Phys. Chem. B* **2012**.
- (35) DeLano, W. *The PYMOL Molecular Graphics System*, 2002.