JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
ARTICLE
pubs.acs.org/JACS

Restructuring the Crystalline Cellulose Hydrogen Bond Network Enhances Its Depolymerization Rate

Shishir P. S. Chundawat,[*,†,‡,▽] Giovanni Bellesia,[⊗,¶,▽] Nirmal Uppugundla,[‡] Leonardo da Costa Sousa,[†] Dahai Gao,[†] Albert M. Cheh,[§] Umesh P. Agarwal,[∥] Christopher M. Bianchetti,[⊥] George N. Phillips, Jr.,[⊥] Paul Langan,[#,○] Venkatesh Balan,[†,‡] S. Gnanakaran,[*,⊗] and Bruce E. Dale[†,‡]

[†]Biomass Conversion Research Laboratory, Department of Chemical Engineering and Materials Science, and [‡]DOE Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, Michigan 48824, United States

[⊗]Theoretical Biology and Biophysics, T-6, [¶]Center for Nonlinear Studies (CNLS), and [#]Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87525, United States

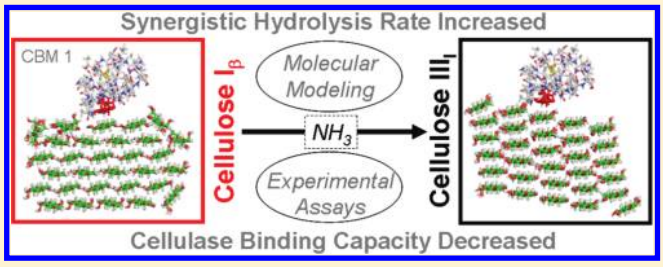
[§]Departments of Environmental Science and Chemistry, American University, Washington, D.C. 20016, United States

[∥]Forest Product Laboratory, USDA Forest Service, Madison, Wisconsin 53726, United States

[⊥]Department of Biochemistry and DOE Great Lakes Bioenergy Research Center, University of Wisconsin, Madison, Wisconsin 53706, United States

S *Supporting Information*

ABSTRACT: Conversion of lignocellulose to biofuels is partly inefficient due to the deleterious impact of cellulose crystallinity on enzymatic saccharification. We demonstrate how the synergistic activity of cellulases was enhanced by altering the hydrogen bond network within crystalline cellulose fibrils. We provide a molecular-scale explanation of these phenomena through molecular dynamics (MD) simulations and enzymatic assays. Ammonia transformed the naturally occurring crystalline allomorph I_β to III_I, which led to a decrease in the number of cellulose intrasheet hydrogen bonds and an increase in the number of intersheet hydrogen bonds. This rearrangement of the hydrogen bond network within cellulose III_I, which increased the number of solvent-exposed glucan chain hydrogen bonds with water by ~50%, was accompanied by enhanced saccharification rates by up to 5-fold (closest to amorphous cellulose) and 60−70% lower maximum surface-bound cellulase capacity. The enhancement in apparent cellulase activity was attributed to the "amorphous-like" nature of the cellulose III_I fibril surface that facilitated easier glucan chain extraction. Unrestricted substrate accessibility to active-site clefts of certain endocellulase families further accelerated deconstruction of cellulose III_I. Structural and dynamical features of cellulose III_I, revealed by MD simulations, gave additional insights into the role of cellulose crystal structure on fibril surface hydration that influences interfacial enzyme binding. Subtle alterations within the cellulose hydrogen bond network provide an attractive way to enhance its deconstruction and offer unique insight into the nature of cellulose recalcitrance. This approach can lead to unconventional pathways for development of novel pretreatments and engineered cellulases for cost-effective biofuels production.



INTRODUCTION

Lignocellulose, a renewable feedstock, can be processed into a variety of fuels and chemicals, reducing our current dependence on petroleum crude. Cellulose, an abundant component of lignocellulosic biomass, is self-assembled in plant cell walls as crystalline nanofibers from the linear β-(1,4) linked D-glucose polymers. Cellulose can be hydrolyzed to monomeric glucose, which can eventually be converted to biofuels (e.g., alcohols, hydrocarbons) via microbial fermentation or chemical catalysis.[1,2]

Several different approaches are being developed to pretreat lignocellulose so that its complex architecture can be disrupted, thereby making its cellulosic component more accessible to water and enzymes for an accelerated conversion to glucose.[3,4] These approaches can be classified into two major categories of thermochemical pretreatments:[4] the first type facilitates removal of hemicellulose and/or lignin and increases enzyme accessibility to embedded crystalline cellulose fibrils (with little or no major change to cellulose crystallinity),[3,4] whereas the second type proceeds further to disrupt cellulose crystallinity and hence increase glycosidic bond accessibility.[5−7] Most thermochemical pretreatments fall within the first category (e.g., dilute acid, ammonia fiber expansion (AFEX), steam explosion, hot water, organosolv), except the ones that use ionic liquids or concentrated acids (like 85% phosphoric acid). However, complexities associated with both the cost-effective utilization and recycling of these chemicals have so far prevented rapid commercialization of either of these approaches.

Received: February 16, 2011
Published: June 10, 2011

ACS Publications © 2011 American Chemical Society

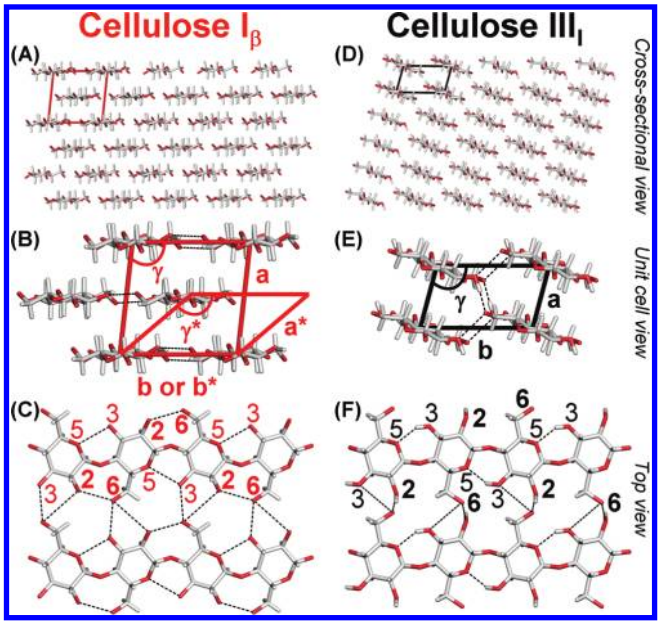


Figure 1. Schematic representation of glucan chain organization and hydrogen bond network (depicted by dotted lines) within cellulose I_β (A−C) and cellulose III_I (D−F) crystalline allomorphs in cross-sectional (A,D), equatorial (B,E), and axial (C,F) views. Definitions of unit cell dimensions for cellulose I_β and III_I are provided (B,E). Oxygen atoms are highlighted in red and numbered on the basis of adjoining carbon atoms (C,F). This figure was generated using Pymol.[45]

Despite improved accessibility to cellulose fibers upon pretreatment, crystalline cellulose is inherently difficult to hydrolyze into its constituent glucose monomers. It is mostly thought that crystalline cellulose recalcitrance to hydrolysis is due to the presence of strong intracrystalline hydrogen-bonding and stacking forces (among other factors like accessible surface area and microfibril shape), giving rise to the extraordinary stability of crystalline cellulose nanofibers that strongly resist chemically or biologically catalyzed depolymerization.[1,2,8,9]

A route to improve cellulose conversion is to engineer enzymes with enhanced specific activities for crystalline cellulose hydrolysis.[10−14] However, despite significant advances in our understanding of the molecular-scale enzymatic mechanisms driving cellulose deconstruction,[15] engineering highly efficient cellulases remains a major challenge. An alternative approach relies on the use of inexpensive and easily recoverable chemicals to alter the cellulose crystal structure in order to increase its rate of depolymerization.[16,17] Cellulose I_α and cellulose I_β are the predominant allomorphic forms of cellulose found in primitive microorganisms and higher plants, respectively.[18] In addition to these natural forms, there are several allomorphs that can be produced by various thermochemical pretreatments.[19,20] In particular, both native cellulose allomorphs can be converted irreversibly into cellulose II during caustic mercerization or regeneration (e.g., from ionic liquids)[21,16] and into cellulose III_I by treatment with liquid ammonia and other amines (Figure 1).[20,22] Taking advantage of such cellulose allomorphs in biomass-to-biofuels conversion processes is limited by the lack of understanding of the impact of intra- and intermolecular hydrogen bonding within crystalline cellulose fibrils upon their interaction with glycosyl hydrolases (GHs).[16−23]

In this study we investigate an alternative approach to pretreatment that does not involve cellulose decrystallization to amorphous cellulose using expensive chemicals that are difficult to recycle, but rather a subtle structural conversion between crystalline forms catalyzed by ammonia to enhance cellulose depolymerization kinetics. This ammonia-based pretreatment produces cellulose III_I (without any relevant loss of crystallinity) with enzymatic hydrolysis rates the closest to those of amorphous cellulose among all reported cellulose allomorphs. We study the impact of this structural modification of cellulose on both cellulase binding and synergistic activity using enzymes isolated from a well-known cellulose-degrading fungus, *Trichoderma reesei*. Cellulase binding and hydrolytic activity measurements were carried out using several distinctive GH families that provided insight into enzyme/substrate structure−function relationships. We also complement our experimental study with extensive molecular dynamics (MD) simulations on model fibrils of cellulose I_β and cellulose III_I. Our MD simulations reveal how the differences between these two allomorphs, in terms of hydrogen-bonding patterns and internal rotational degrees of freedom, impact their structural flexibility and surface hydration properties and eventually influence their interactions with cellulases. The fundamental insights gained from this combined experimental−theoretical approach will be critical to guide the development of improved ammonia pretreatment processes and novel engineered cellulases that are optimized for rapid and efficient hydrolysis of cellulosic biomass.

METHODS

Molecular Dynamics Simulations. We performed MD simulations on model fibrils of cellulose I_β and cellulose III_I to reveal structural and dynamical features of both allomorphs. X-ray and neutron diffraction coordinates[19,20] have been used to generate one rhomboid cellulose I_β fibril and one rhomboid cellulose III_I fibril (Figure 1). The rhomboid shape of the cellulose I_β fibril[24] was chosen because of its wide hydrophobic surfaces, which have been shown to be the preferential binding site for different cellulose binding modules (CBMs).[25] The rhomboid shape was chosen also for the cellulose III_I fibril as it retains both a hydrophobic/hydrophilic surface ratio and a solvent-accessible surface area (SASA) similar to those of the rhomboid cellulose I_β fibril. Moreover, since the two-dimensional crystal unit of cellulose III_I (in the plane perpendicular to the cellulose chains' main axes) contains one cellulose chain and has a rhomboid shape, a rhomboid cellulose III_I fibril maximizes the number of nearest neighbors for each cellulose chain. Each fibril was composed of 30 octameric glucan chains consistent with the size of elementary cellulose fibrils derived from plant cell walls that were used for all subsequent experimental studies. The choice of using 30 chains was motivated by a general consensus on the lowest possible limit for the number of glucan chains within an elementary cellulose fibril.[26−30]

MD simulations were performed under NPT conditions (isothermal−isobaric ensemble) using the NAMD software[31] with the GLYCAMO6[32] force field and the TIP3P explicit water model.[33] A Langevin thermostat and Nosé−Hoover Langevin barostat with a stochastic component were used to control the temperature and the pressure, respectively.[34,35] The damping coefficient for the Langevin integrator was set to 1.0 ps^{-1}, while for the Nosé−Hoover Langevin barostat we applied an oscillation period of 200 fs and a damping period of 100 fs. The cutoff for the nonbonded interactions in the coordinate space was fixed at 10.0 Å. All the simulations were performed under periodic boundary conditions, and the long-range electrostatic interactions were calculated by using the Ewald summation method with the particle mesh Ewald algorithm.[36] The particle mesh Ewald accuracy was fixed at 10^{-6}, the order of the interpolation functions on the grid was set to 4 (cubic), and the grid

spacing was ~1.0 Å. The fibrils were solvated in a rectangular water box, and each octameric cellulose chain was covalently connected to its periodic image along its main axis in order to mimic an infinitely long cellulose fibril. The solvated cellulose fibrils underwent first a local optimization, followed by a short (~0.5 ns) NPT-MD simulation, where the temperature was gradually increased from 100 to 298 K, and by a 200 ns long NPT-MD simulation at T = 298 K and P = 1.01325 bar. The first 20 ns of the 200 ns run was considered as the initial equilibration time and the remaining 180 ns as the production time. The time step was fixed at 2.0 fs. The covalent bonds involving hydrogen atoms were constrained by means of the SHAKE algorithm.[37] We also performed an additional 50 ns MD simulation of a solvated octameric cellulose chain under the same thermodynamic conditions used for the cellulose fibrils. This additional MD simulation was used to calculate the hydroxymethyl rotational state population for a solvated cellulose chain.

Cellulosic Substrates and Their Characterization by X-ray Diffraction (XRD). Avicel microcrystalline cellulose (PH-101, Sigma-Aldrich, St. Louis, MO), cotton linters (long fibers, Sigma-Aldrich), and cotton ball fibers (scoured and bleached, courtesy of Dr. Alfred French, USDA) were used for all enzyme assays. Premilled corn stover, passed through a 10 mm screen, was procured from National Renewable Energy Laboratory (NREL, Golden, CO) and was further milled to 100 μm as described previously,[38] prior to pretreatment. The corn stover feedstock was grown from Pioneer Hybrid seed variety (33A14) and harvested in 2002 from the Kramer farm in Wray, CO.

Cellulose III_I (from Avicel, cotton linter, and cotton fiber) was prepared by soaking dry biomass in anhydrous liquid ammonia (~5−10 g of ammonia per gram of biomass) at 90−100 °C for 120 min. The samples were dried under nitrogen and purged overnight to remove residual ammonia. Cellulose II was prepared by mixing Avicel with 25% NaOH at 4 °C, at a loading of 1:10 (gram of substrate per milliliter of solution) for 60 min. The slurry was then centrifuged, filtered, and washed with deionized water until neutral pH. Avicel was also soaked in 28−30% ammonium hydroxide solution for 60 min at 4 °C prior to filtration and washing until neutral pH. Regenerated amorphous cellulose from Avicel was prepared using 83% phosphoric acid (at 4 °C for 60 min) based on a previously published protocol.[5] Regenerated amorphous cellulose samples were freeze-dried prior to analysis by XRD and subsequent enzymatic assays. Conventional AFEX pretreatment was carried out on corn stover at 130 °C, 60% moisture loading, 1:1 NH_3-to-biomass loading (w/w) for 45 min total residence time.[39]

The cellulose crystallinity index was estimated based on the XRD spectral peak deconvolution and amorphous spectra subtraction methods (see Supporting Information).[40] Dry biomass samples were packed into 1.5 mm diameter quartz capillaries (Charles Supply Co., Natick MA) and mounted on a Bruker AXS (Madison, WI) three-circle goniometer. The quartz capillaries are for all practical purposes transparent to Cu Kα X-rays and did not contribute any appreciable background signal. All XRD patterns were collected on a Smart6000 CCD area detector that was positioned 50 mm from the sample with an exposure time of 90 s. Cu Kα radiation was generated by a Microstar rotating Cu anode generator at 45 kV and 45 mA (Bruker AXS) with a source size of ~0.3 mm in diameter. Each diffraction pattern image contained data ranging from 1° to 37° 2θ. The images were unwarped and converted using the Apex2 v2009.3-0 (Bruker AXS) software package. The converted images were integrated from 3 to 37° 2θ in 0.04° steps using the XRD2Eval module in Apex2 v2009.3-0 (Bruker AXS).

Cellulase Source, Enzymatic Hydrolysis, and Adsorption Assays. Spezyme CP (88 mg/mL) and Novozyme 188 (150 mg/mL) were procured from Genencor (Danisco Inc., Genencor Division, Rochester, NY) and Sigma-Aldrich, respectively. Total protein concentration was based on Kjeldahl analysis. *T. reesei*-based cellobiohydrolases (Cel7A, Cel6A) and endoglucanase (Cel7B) were purified from Spezyme CP, while β-glucosidase was purified from Novozyme 188.[38] Other *T. reesei* endoglucanases (Cel5A, Cel12A, Cel61A, and Cel61B) were cloned, expressed, and purified as described elsewhere (courtesy of Dr. Jonathan Walton, Michigan State University).[41] *Acidothermus cellulolyticus* endocellulase (Cel5A_ac) was obtained as described elsewhere (courtesy of Dr. Sandra Austin-Phillips, University of Wisconsin, Madison).[42] The catalytic cores for endocellulases, based on information provided in the Carbohydrate Active Enzymes Database (CAZY, http://www.cazy.org),[43] are referred to when analyzing their sequences and X-ray structures (RCSB Protein Data Bank).[44] Pymol was used to analyze and generate all protein structures.[45] Enzymatic hydrolysis and reducing sugar analyses were carried out as described previously.[38]

Cellulase adsorption assays were performed using a microplate-based method as highlighted here and adsorption parameters determined using the Langmuir model ($B/S = B^* = A_{max}K_dF/(1 + K_dF)$).[46] A 2.2 mL deep-well microplate (Lot No. 780271, Greiner, Monroe, NC) was used to load 200 μL of 0.5% (w/v) Avicel-derived cellulose I_β or III_I. Twenty-five microliters of 1 M citrate buffer (pH 4.5) and 275 μL of enzymes (cellulase loading ranged from 1 to 300 mg/g glucan) and water mixtures were added in each well. Each well contained 50 mM citrate buffer and 1 mg of glucan. After incubation for 2 h at 4 °C, samples containing enzymes mixture and insoluble biomass were filtered through a 0.45 μm microplate filter (Lot No. R6PN00144, Millipore, Ireland), and 350 μL of the supernatant was loaded into a microplate to measure UV 280 nm absorbance and hence determine unbound protein concentration in the supernatant. Molecular weights of Cel7A, Cel6A, and Cel7B used for estimating protein molar concentrations were 55, 50, and 48 kDa, respectively. Bound enzyme concentration was calculated by the depletion method (i.e., subtracting the unbound enzymes present in supernatant from the total added enzyme concentration). The bound (B/S or B^*; μM protein/g substrate) and free (F; μM protein) enzyme concentrations were fit to a Langmuir single-site adsorption model to determine maximum cellulase binding capacities (A_{max}; mg protein/g substrate) for each cellulose allomorph and cellulase combination using the minimum least-squares method in Excel (Microsoft, Redmond, WA) with adjusted R^2 values always >0.98.

RESULTS

Crystal Unit Cell Dimensions of Cellulose Fibrils from MD Simulations. The computational work presented here builds on previous cellulose I_β and cellulose III_I synchrotron X-ray and neutron diffraction studies to determine the detailed atomic resolution crystal structures of both crystalline allomorphs.[19,20] Cellulose I_β has space group $P2_1$, with a monoclinic unit cell composed of two parallel chains and with dimensions a = 7.784 Å, b = 8.201 Å, c = 10.38 Å, and γ = 96.5° (Figure 1). In the I_β crystalline phase, the cellulose chains are hydrogen-bonded within sheets (b direction) that stack on top of each other through C−H···O hydrogen bonding and stacking interactions. Cellulose III_I has space group $P2_1$, with a monoclinic unit cell composed of one chain and with dimensions a = 4.45 Å, b = 7.85 Å, c = 10.31 Å, and γ = 105.1° (Figure 1). The III_I crystalline phase is stabilized by hydrogen bonds (HBs) within the sheets as well as by an extended "zigzag" network of intersheet O2−O6 HBs (~$a\sin(\gamma)$ direction, Figure 1).

The crystal unit parameters $a = \vec{a}/\hat{a}$, $b = \vec{b}/\hat{b}$, together with the angle $\gamma = \hat{b}\cdot\hat{a}$, define the crystal plane perpendicular to the cellulose chain main axis, while $c = \vec{c}/\hat{c}$ is the length of cellobiose unit, $\alpha = \hat{b}\cdot\hat{c}$, and $\beta = \hat{a}\cdot\hat{c}$. The values of the crystal unit parameters obtained from MD simulations are shown in Table 1. The parameters a, b, and γ, which define the plane perpendicular to the cellulose chain's main axis, have small relative deviations (Δ < 4%) from the experimental values reported in cellulose I_β.

Table 1. Crystal Unit Cell Parameters for Cellulose I_β and Cellulose III_I[a]

	Cellulose I_β						
	a	b	c	α	β	γ	$a^* \sin(\gamma^*)$
exptl[19]	7.784	8.201	10.38	90	90	96.5	3.867
avg	7.498	8.095	11.13	91.4	89.2	98.2	3.705
σ_N, %	2.429	1.965	0.83	12.0	10.1	2.0	2.522
Δ, %	3.674	1.293	7.23	1.6	0.9	1.8	4.189
	Cellulose III_I						
	a	b	c	α	β	γ	$a \sin(\gamma)$
exptl[20]	4.45	7.85	10.31	90	90	105.1	4.30
avg	4.34	8.23	11.18	91.7	87.8	114.5	3.92
σ_N, %	3.62	3.02	0.87	12.2	16.4	4.5	5.06
Δ, %	2.47	4.84	8.44	1.9	2.4	8.9	8.84

[a] exptl, experimental values from refs 19 and 20; avg, average from MD simulations; σ_N, relative standard deviation (%); Δ, relative deviation from experimental value (%). For a correct comparison of the thermal fluctuations along the intersheet direction, we consider $a^* \sin(\gamma^*)$, since the cellulose I_β crystal unit is composed of two cellulose chains (origin and center chains, see Figure 1).

Table 2. Relevant Internal Dihedral Degrees of Freedom for Cellulose I_β and Cellulose III_I[a]

	Cellulose I_β			
	χ	χ'	φ	ψ
		Origin		
exptl[19]	170.0	−70.0	−98.5	−142.3
avg	166.2	−70.3	−97.0	−142.1
σ_N, %	4.5	11.5	6.3	4.1
Δ, %	2.2	0.4	1.5	0.1
		Center		
exptl[19]	158.0	−79.0	−88.7	−147.1
avg	171.5	−53.6	−95.5	−141.1
σ_N, %	4.5	20.0	6.6	4.1
Δ, %	8.5	32.1	7.7	4.1
	Cellulose III_I			
	χ	χ'	φ	ψ
exptl[20]	44.0	163.0	−92.0	−146.0
avg	54.1	171.5	−106.3	−139.2
σ_N, %	19.6	4.1	8.7	5.2
Δ, %	22.5	5.2	15.5	4.7

[a] exptl, experimental values from refs 19 and 20; avg, average from MD simulations; σ_N, relative standard deviation (%); Δ, relative deviation from experimental value (%). The average values for χ and χ' in cellulose I_β and cellulose III_I are calculated considering only the values of the *tg* and *gt* populations, respectively.

In cellulose III_I we observe similar relative deviations for a and b, while the angle γ appears to have a larger deviation (Δ = +8.9%) from the experimental value. The slightly larger deviations (Δ = +7.23% and +8.44%) of the c value (cellobiose unit length) in both cellulose I_β fibril and cellulose III_I are due to an increased chain stiffness originating from the particular boundary conditions used in our simulations, i.e., from the "infinite" length of the cellulose chains. Indeed, MD simulations of finite length fibrils composed of cellulose octamer chains show, for the parameter c, a relative deviation Δ ≈ +4.0%. The overall deviation from the experimental values is also within the expected margin of uncertainty in measuring both cellulose allomorphs' unit cell dimensions. The unit cell parameters were calculated from MD simulations considering only the glucan chains in the crystalline core of the fibrils, therefore excluding the glucan chains on the fibril surface.

Internal Degrees of Freedom of Cellulose Fibrils from MD Simulations. Table 2 shows the MD simulation results for some of the relevant internal dihedral degrees of freedom for cellulose I_β and cellulose III_I fibrils. The primary dihedral angles considered in our analysis are defined as follows: χ, O5−C5−O6−C6; χ', C4−C5−O6−C6; φ, $O5_{i+1}-C1_{i+1}-O4_i-C4_i$; and ψ, $C1_{i+1}-O4_i-C4_i-C5_i$. The dihedral angles χ and χ' are related to the conformation of the hydroxymethyl group, while φ and ψ are related to the conformation of the glycosidic linkage between subsequent glucose monomers and have similar values in cellulose I_β and cellulose III_I. Our results show good agreement with experiments for both cellulose allomorphs.[19,20]

In terms of internal dihedral degrees of freedom, the only difference between glucan chains in cellulose I_β and cellulose III_I fibrils is in the torsional-state occupancy of the hydroxymethyl group. For the cellulose I_β fibril, the torsional-state occupancy of the hydroxymethyl group (defined by the angles χ and χ') is *tg* = 92.6%, *gt* = 4.7%, and *gg* = 2.7%. The dominance of the *tg* rotational state in cellulose I_β is in agreement with experimental results[19] and is consistent with the formation of HBs (O6−Ox, x = 2, 3) within the fibril layers (intrasheet HBs). In cellulose III_I, the torsional-state occupancy of the hydroxymethyl group is *tg* = 11.8%, *gt* = 63.6%, and *gg* = 24.6%. These results reproduce the dominance of the *gt* rotational state seen in experiments,[20] although this dominance is not as straightforward as the one observed for the rotational-state *tg* in cellulose I_β (alternative rotational states *tg* and *gg*, were previously not observed in experiments[20]) and is consistent with the formation of HBs (O6−O2) between neighboring cellulose fibril sheets (intersheet HBs).

Structural Flexibility of Cellulose Fibrils from MD Simulations. In our analysis, the unit cell dimensions are calculated as thermodynamic averages along the MD trajectories. In addition to the analysis of the average values, we consider the relative standard deviation σ_N (%) for the values $a^* \sin(\gamma^*)$, b, and c in cellulose I_β and $a \sin(\gamma)$, b, and c in cellulose III_I (Table 1) as a measure of the thermal fluctuations within the fibril crystalline cores (cellulose chains not exposed to solvent) and, consequently, as an estimate of their relative structural flexibility[47] (See Figure 1 for the definitions of a^* and γ^*). The quantities $a^* \sin(\gamma^*)$ or $a \sin(\gamma)$, b, and c are considered for measuring the thermal fluctuations of the fibril crystalline core along the "intersheet direction" (Figure 1B,E), the "intrasheet direction" (Figure 1C,F), and the glucan chain's main axis, respectively. The crystalline core of the cellulose III_I fibril showed larger thermal fluctuations along the inter- and intrasheet directions when compared with that of cellulose I_β. The largest thermal fluctuations are observed along the intersheet direction in cellulose III_I (i.e., thermal fluctuations of $a \sin(\gamma)$). We have also captured the difference in structural flexibility between cellulose I_β and cellulose III_I using an alternate approach based on gyration tensors. The eigenvalues of the gyration tensor quantify the change in fibril shape along the three main geometric axes.[48] The relative

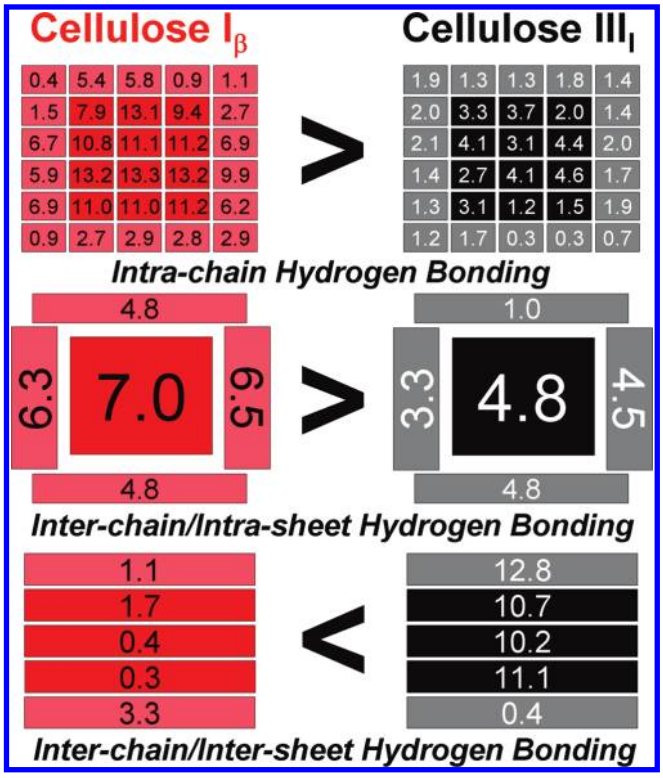


Figure 2. Hydrogen bond network in cellulose I_β (left) and cellulose III_I (right). The two 6×5 grids on the top row show the average number of intrachain HBs per chain. The central row shows the average number of "interchain/intrasheet" HBs (HBs between glucan chains within the same fibril horizontal layer), while the bottom row shows the average number of "interchain/intersheet" HBs (HBs between neighboring layers within the cellulose fibril). The crystalline core (dark red or black) and solvent-exposed (pink or gray) cellulose chains are highlighted.

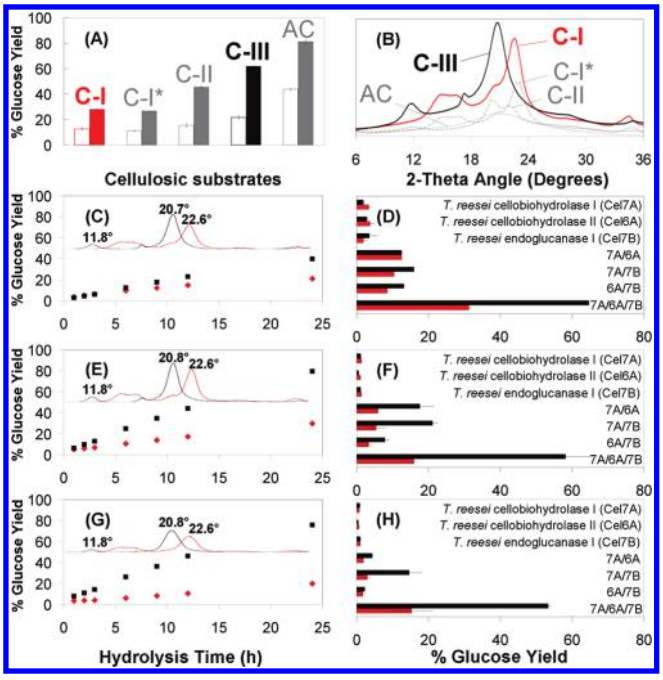


Figure 3. Enzymatic digestibility of cellulose as a function of its crystalline state and cellulases employed. (A) Native cellulose (cellulose I_β, or C-I), derived from Avicel, was treated with liquid ammonia (cellulose III_I, or C-III), sodium hydroxide (cellulose II, or C-II), 28% ammonium hydroxide (C-I*), and concentrated phosphoric acid (amorphous cellulose, or AC). Equivalent enzyme loadings were added for each assay (1.5 FPU Spezyme CP cellulase/g glucan + 6.4 p-NPGU Novo188 β-glucosidase/g glucan). The left and right bars represent 6 and 24 h data, respectively. (B) XRD spectra for cellulose allomorphs (2θ range 6°−36°). (C−H) Hydrolysis time course for cellulose I_β (red diamonds) and cellulose III_I (black squares) derived from Avicel (C), cotton linters (E), and cotton fibers (G). Inset depicts XRD spectra for respective substrates. Avicel was hydrolyzed by ∼1.5 FPU Spezyme CP cellulase/g glucan, while the cotton-derived substrates were hydrolyzed by ∼15 FPU Spezyme CP cellulase/g glucan. Enzymatic digestibility of cellulose I_β (red bars on bottom) and cellulose III_I (black bars on top) derived from Avicel (D), cotton linters (F), and cotton fibers (H) for various combinations of *T. reesei*-based exo- (Cel7A, Cel6A) and endocellulases (Cel7B). Equivalent enzyme loadings were added for each assay (2.5 mg of each cellulase/g Avicel or 10 mg of each cellulase/g cotton-derived substrates). In each case, 10% additional β-glucosidase relative to total cellulase was added to prevent cellobiose buildup.

fluctuations of the average eigenvalues have been considered in our analysis (see Supporting Information). Consistent with the results obtained from the analysis of the crystal unit dimensions, we observe the largest fluctuations of the crystalline core in cellulose III_I along both the inter- and intrasheet directions.

Hydrogen Bond Network of Cellulose Fibrils from MD Simulations. Hydrogen bonds play an important role in stabilizing the ordered crystalline structures of cellulose I_β and III_I; however, the total number and type of HBs differ between the two allomorphs. The schemes on the top line of Figure 2 shows the average number of intrachain HBs for each cellulose chain in the fibril. The crystalline core of the I_β fibril overall possesses a larger number of intrachain HBs when compared to the crystalline core of the III_I fibril. In both fibrils, the number of intrachain HBs drops considerably in the cellulose chains exposed to the solvent, although it remains larger in the I_β fibril than in the III_I fibril. The schemes on the middle line of Figure 2 show the average number of interchain/intrasheet HBs for both the crystalline core and the solvent-exposed cellulose chains. The results are similar to what was observed for intrachain HBs: a larger number of HBs in the cellulose I_β fibril both in the crystalline core and in the solvent-exposed cellulose chains. Intersheet HBs display an opposite trend, with the cellulose III_I fibril possessing a larger number of these HBs. In summary, we observe that the hydrogen bond network in cellulose I_β is dominated by intrasheet (intra- and interchain) O2−O6, O3−O6, and O3−O5 HBs. Conversely, cellulose III_I is mainly stabilized by intersheet O2−O6 HBs that are entirely missing in cellulose I_β. It is important to note that intrachain O2−O6 and O3−O5 HBs stabilize the two-fold helical conformation of cellulose chains, while interchain O6−O2 and O6−O3 HBs participate in intrasheet glucan chains' cohesion. Intrachain O2−O6 HBs stabilizing the two-fold helical chain conformation are weakened in cellulose III_I. We also observe a smaller number of HBs in solvent-exposed cellulose chains with respect to the chains in the crystalline core.

Cellulose Crystalline Structure Impacts Its Overall Enzymatic Digestibility Rate. Various cellulose allomorphs were generated from crystalline cellulose (Avicel) to investigate differences in the hydrolysis rates for cellulose I_β, cellulose II, and cellulose III_I (Figure 3A). Mercerization of cellulose I_β with NaOH and NH_3 resulted in distinct XRD spectra (Figure 3B), typical of cellulose II and cellulose III_I, respectively. Conversely, samples treated with aqueous ammonia (28% NH_4OH) resulted in a

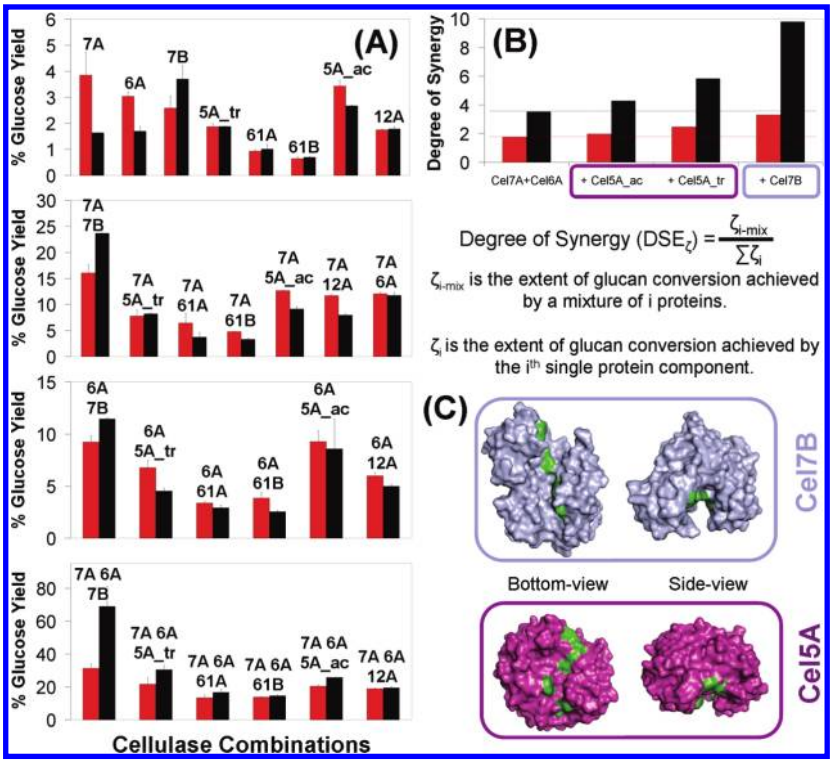


Figure 4. (A) Purified cellulase (Cel7A, Cel6A, Cel7B, Cel5A_ac, Cel5A_tr, Cel12A, Cel61A, Cel61B) 24 h glucan hydrolysis yields of Avicel-derived cellulose I_β (red bars, left) and cellulose III_I (black bars, right) for single, binary, or ternary enzyme combinations. Each purified enzyme was loaded at 2.5 mg each per gram of glucan, and additional β-glucosidase (10% of total cellulase added) was supplemented in each assay to prevent inhibition by cellobiose. Standard deviations were within ±20% of the reported mean values. (B) Influence of endocellulase active-site structure on degree of synergistic effect (DSE) with exocellulases (Cel7A + Cel6A) during hydrolysis of crystalline cellulose I_β (red bars, left) and cellulose III_I (black bars, right). Cel7B and Cel5A_tr are from *T. reesei*, while Cel5A_ac is from *A. cellulolyticum*. (C) Cel7B and Cel5A catalytic domains bottom (left) and side views (right) (PDB codes: 1EG1 and 1H1N), with aromatic residues within catalytic cleft (highlighted in green) that facilitate formation of catalytically active complex with individual cellulose chain, shown to highlight structural differences between endocellulase glycosyl hydrolase families 7 and 5. Protein structures were generated using Pymol.[45]

spectrum similar to that of native cellulose I_β. The percent crystallinity values of cellulose I_β and cellulose III_I were comparable (see Supporting Information). Modification of cellulose crystal structure by NaOH and NH_3 enhanced the hydrolysis yield by 1.5- and 2-fold, respectively, compared to that of cellulose I_β. There was also a 3−4-fold increase in hydrolysis yield of amorphous cellulose that is consistent with previous findings.[5]

The % crystallinity values for cellulose III_I derived from Avicel, linters, and cotton fibers were also comparable to their untreated controls (Figure 3C,E,G XRD spectra insets). Formation of crystalline cellulose III_I derived from Avicel, linters, and cotton fibers resulted in a −20%, 20%, and 120% increase in initial hydrolysis rate (<5% conversion within 1 h) with respect to untreated controls, respectively (Figure 3C,E,G). However, the overall glucan hydrolysis rate (>5% conversion within 24 h) of cellulose III_I derived from Avicel, linters, and cotton fibers was 2.2-, 2.9-, and 4.2-fold higher than that of native celluloses, respectively. The improvement in digestibility for cellulose III_I was strongly correlated to initial substrate crystallinity. Higher crystallinity cotton fiber cellulose gave 4−5-fold increase in hydrolysis rate for cellulose III_I versus cellulose I_β. However, lower crystallinity celluloses (Avicel, linters) gave only a corresponding 2−3-fold increase. It was also necessary to hydrolyze at least 5−10% of the lower crystallinity cellulose allomorphs to see any significant increase in hydrolysis yield for cellulose III_I, unlike the more crystalline cellulose allomorphs. The overall differential hydrolysis rates for cellulose allomorphs can be arranged in the following order: amorphous cellulose > cellulose III_I > cellulose II > cellulose I_β.

Endo- and Exocellulases Show Reduced Binding but Enhanced Synergistic Activity on Crystalline Cellulose III_I. Two exocellulases, Cel7A (cellobiohydrolase I, or CBH-I) and Cel6A (CBH-II), and an endocellulase, Cel7B (EG I), from *T. reesei* were added in varying combinations to hydrolyze cellulose I_β and cellulose III_I derived from Avicel, cotton fibers, and linters. No significant improvement in apparent specific activity for any of the cellulases was seen when added alone on cellulose III_I derived from Avicel, linters, or cotton fibers (Figure 3D,F,H). Interestingly, the only significant enhancement in the hydrolysis yield for cellulose III_I versus I_β was observed for several binary and ternary combinations of exo- and endocellulases. The binary degree of synergistic effect (DSE_2; definition provided in Figure 4B) between Cel7A/Cel6A and Cel7A/Cel7B was 1.7 and 2.5 on cellulose I_β, compared to 3.5 and 4.4 on cellulose III_I, respectively. There was a 30−100% and 40−370% increase in the binary DSEs observed for cellulase mixtures hydrolyzing Avicel- and cotton-derived cellulose III_I, respectively. The highest glucan conversion was seen for an equimass ternary combination of Cel7A/Cel6A/Cel7B, which yielded results comparable to those obtained for crude, unfractionated cellulase mixtures. The optimal ratio of Cel7A/Cel6A/Cel7B (15 mg/g glucan total enzyme loading) that maximizes the 24 h hydrolysis yield of Avicel-derived cellulose I_β and cellulose III_I had a marginally higher percentage

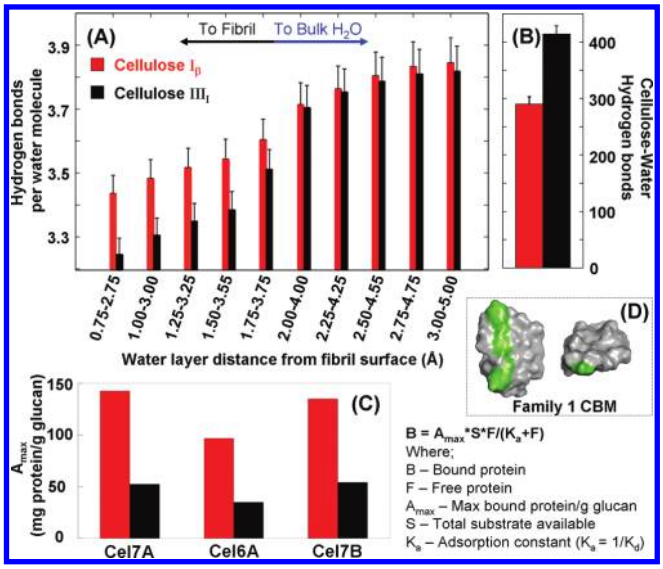


Figure 5. Influence of cellulose I_β (red bars, left) and cellulose III_I (black bars, right) crystal structures on fibril hydration and cellulase binding capacity. Average number of hydrogen bonds between water molecules (A) and between cellulose–water molecules (B) in the vicinity of cellulose I_β and III_I fibril surfaces as predicted by MD simulations. (C) Maximum *T. reesei* exo- and endocellulases' binding capacity determined from Langmuir adsorption model for 277 K equilibrium adsorption to cellulose I_β and cellulose III_I. (D) Family 1 cellulose binding module bottom (left) and side views (right) (PDB code: 1CBH), with aromatic residues highlighted in green. Protein structure generated using Pymol.[45]

of Cel7B in the latter case (32 vs 35 wt % Cel7B of total protein mass; mixture optimization data not shown). Other endocellulases from homologous and distinct GH families were also tested in combination with Cel7A and Cel6A, namely, Cel5A_tr, Cel12A, Cel61A, and Cel61B from *T. reesei* and Cel5A_ac from *A. cellulolyticus*, for their hydrolytic potential on Avicel-derived cellulose I_β and cellulose III_I (Figure 4A). A marginal decrease in hydrolytic yield was noticed for individual exocellulases (Cel7A, Cel6A) on cellulose III_I as compared to individual endocellulases that resulted in equivalent glucan conversions on both substrates. For binary cellulase mixtures, glucan conversions for both substrates rarely exceeded 10%, except for certain exo- and endocellulase combinations. In contrast, for ternary cellulase combinations, most mixtures resulted in glucan conversions in the ranges 10–30% and 15–70% for cellulose I_β and cellulose III_I, respectively. However, only GH family 5 and family 7 endocellulases resulted in significantly higher conversions for cellulose III_I (>25% increase in hydrolysis yield with respect to control) compared to other endocellulase families (e.g., GH families 12 and 61).

Enzyme–substrate binding affinity was experimentally determined by fitting a Langmuir adsorption model to the *T. reesei* Cel7A, Cel6A, and Cel7B binding isotherms for Avicel-derived cellulose I_β and cellulose III_I (Figure 5) (Note: Cel7A, Cel6A, and Cel7B are the most abundant *T. reesei* cellulases necessary for efficient cellulose hydrolysis[38] and are thus ideal candidates for enzyme binding studies). All three cellulases showed an overall reduced maximum binding affinity by 60–70% (based on A_{max}; see Figure 5C for details) for cellulose III_I compared to native cellulose. In contrast to what has been reported in the literature previously, i.e., that the enzymatic hydrolysis rate was found to be directly correlated to the extent of cellulase adsorbed to cellulose (i.e., mg protein bound per gram of substrate),[4,9,11] we find that there is reduced cellulase binding to the more readily digestible cellulose III_I allomorph. Cellulose I_β and cellulose III_I fibrils were closely examined by MD simulations to provide insight into the role of fibril surface hydration in the enzyme–substrate interaction.

Table 3. Rotational-State Occupancy (%) for the Hydroxymethyl Group in the Fibril Crystalline Core and on the Surface Cellulose Chains in Cellulose I_β and Cellulose III_I Fibrils, and for a Solvated Cellulose Octamer Chain[a]

	tg	*gt*	*gg*
	Cellulose I_β		
crystalline core	**92.6**	4.7	2.7
surface chains	25.8 (−72.1)	28.1 (+83.3)	**46.1 (+94.1)**
	Cellulose III_I		
crystalline core	11.8	**63.6**	24.6
surface chains	4.2 (−64.4)	**52.3 (−17.8)**	43.5 (+43.5)
solvated cellulose octamer	3.0	27.2	**69.8**

[a] In parentheses, for the surface chains, the percentage increase/decrease for the different rotational states with respect to the values for the crystalline core is indicated. The largest value for each considered system is indicated in bold font.

Surface Properties of Cellulose Fibrils from MD Simulations. The flexibility of surface chains of both cellulose fibrils are estimated via a root-mean-square deviation (rmsd) analysis. We find that for cellulose III_I, rmsd = 1.0 ± 0.1 Å, whereas for cellulose I_β, rmsd = 0.7 ± 0.1 Å. These results confirm the trends observed for the crystalline core, where larger structural fluctuations were observed in cellulose III_I. The largest values for the rmsd (i.e., largest structural fluctuations) have been observed for the corner chains in both cellulose I_β and cellulose III_I fibrils. Also, glucan chains on the top and bottom sides (horizontal) of the fibrils tend to have a larger rmsd with respect to the ones composing the lateral sides of the fibrils (see Figure 1).

The results in the previous sections also show that there are substantial differences in the torsional-state occupancy of the hydroxymethyl group between glucan chains in the crystalline core of the cellulose I_β and cellulose III_I fibrils. In Table 3 we extend our analysis, showing the hydroxymethyl rotational-state occupancy for the glucan chains (i) in the fibril crystalline core, (ii) on the solvent-exposed surface of the fibril (surface chains), and (iii) in bulk water (single cellulose chain in water). We also show the percentage increase/decrease for the different rotational states in the surface chains, with respect to the correspondent values for the crystalline core (in parentheses). We observe a large drop and a large increase in the occupancy of the *tg* and *gg* rotational states, respectively, in both cellulose I_β and cellulose III_I upon "moving" from the chains in the crystalline core to the ones on the surface and to a single cellulose chain in aqueous solution. It is interesting to note that cellulose I_β undergoes a drastic change, with the *tg* rotational state, dominant in the crystalline core, being the least populated in the surface chains, whereas in cellulose III_I the *gt* rotational state remains the dominant one in the crystalline core and on the fibril surface.

The structural ordering of water in the vicinity of the fibril was estimated by calculating the average number of HBs per water molecule for water layers (2.0 Å thick) at increasing distances from the cellulose fibril surface (Figure 5A). In our calculations,

Table 4. Percentage Contribution to the Water−Cellulose Hydrogen Bond Network of Oxygens O6 (Hydroxymethyl Group), O5 (on the Glucose Ring), O3 and O2 (Bonded to C3 and C2 on the Glucose Ring, Respectively), and O4 (Glycosidic Bond)[a]

	O6	O5	O4	O3	O2
cellulose I_β	14.7, 20.2	0.0, 4.6	0.0, 4.8	10.8, 13.5	14.6, 15.7
cellulose III_I	14.5, 13.2	0.0, 6.0	0.0, 2.0	17.0, 17.9	15.0, 13.7
$R(III_I/I_\beta)$	1.1	1.9	0.6	2.1	1.4

[a] The two values listed for each oxygen atom refer to "donor" and "acceptor" percentage contributions, respectively. $R(III_I/I_\beta)$ is the ratio between the total number of HBs in cellulose III_I and cellulose I_β. The O4 oxygen atom is the only one forming a higher number of HBs with water (as an acceptor) in cellulose I_β with respect to cellulose III_I ($R < 1$).

the HBs within the cellulose fibrils, between cellulose and water, and between water molecules have been defined by a cutoff distance of 3.3 Å and a cutoff angle of 30°.[49−51] In Figure 5A, the water layers are 2.0 Å thick and are defined by the pair of numbers along the *x*-axis. For example, the data related to the label 0.75−2.75 refer to the number of HBs per water molecule considering all the water molecules between distances 0.75 and 2.75 Å from the cellulose fibril surface (distances are measured considering the positions of the water oxygen atoms). The value of the average number of HBs per water molecule in bulk water at room temperature and pressure is ∼3.8−4.0. Thus, our analysis shows that the water−water hydrogen bond network is strongly perturbed by cellulose III_I in the vicinity of fibril surface.

Second, we calculated the average number of cellulose−water HBs between surface glucan chains and water (Figure 5B). The number of HBs has been normalized by the ratio of the SASAs of cellulose I_β and cellulose III_I ($SASA(III_I)/SASA(I_\beta) = 0.94$). Consistent with the data shown in Figure 5A, we find a ∼1.5−fold increase in the number of cellulose−water HBs for cellulose III_I with respect to cellulose I_β (Figure 5B). In both cellulose I_β and cellulose III_I, the oxygen atoms O6 (hydroxymethyl group), O3, and O2 (bonded to carbon atoms C3 and C2 in the glucose ring, respectively) participate, as both donors and acceptors, in about 90% of the water−cellulose HBs (see Table 4). The oxygen atoms O4 (glycosidic linkage) and O5 (oxygen within the glucose ring) are acceptors in the remaining ∼10% of water−cellulose HBs. The average lifetime of the HBs formed by the O6, O3, and O2 atoms with water varies in the interval 0.3−1.3 ps. The longer average lifetimes in this interval have been observed in cellulose III_I when O6, O2, and O3 act as hydrogen donors.

Given the importance of the hydrophobic face of the cellulose fibril surface in the cellulase binding process,[25] we also calculated the number of cellulose−water HBs considering only the glucan chains that constitute the hydrophobic face. The data show the same trend observed for the fibrils as a whole, i.e., more HBs are formed in cellulose III_I than in cellulose I_β. The relative increase in the number of cellulose−water HBs in cellulose III_I with respect to cellulose I_β, for the hydrophobic face alone, was found to be ∼1.7.

DISCUSSION

In this work, we have sought alternative ways to improve the efficiency of crystalline cellulose digestion by modifying the substrate characteristics rather than through engineering better enzymes. Microbes have evolved complex protein machineries that can hydrolyze native insoluble polysaccharides to soluble sugars through mechanisms that are still far from being understood. Efficient enzymatic catalysis of insoluble crystalline cellulose I_β is thought to involve at least the following mechanistic steps, as also highlighted elsewhere:[4] (i) enhancing substrate-bound enzyme concentration driven by interaction between CBM and surface fibril glucan chains,[52] (ii) recognition of glucan chain reducing end (typically in the case of processive enzymes like Cel7A) and binding to substrate in a productive orientation,[53] (iii) disruption of cellulose crystal structure by CBM to facilitate release of glucan strands,[54,55] (iv) capturing and directing single glucan chains into the enzyme catalytic active site,[56,57] and (v) processive threading of glucan chains upon successive hydrolytic cleavage.[15,56] In such heterogeneous interfacial catalysis, the efficiency of the cellulase action is significantly limited by the crystalline nature of the polysaccharide substrate as also suggested by recent work on chitin, a close structural analog of cellulose.[58,59] Accordingly, our results show that modifying the crystalline properties of cellulose using cost-effective pretreatment options is a viable strategy to improving apparent cellulase activity and providing avenues for future cellulase engineering efforts.

Disruption of HBs within crystalline cellulose, using either concentrated acids or ionic liquids to solvate cellulose followed by its precipitation with water, producing mostly amorphous cellulose (and/or cellulose II), can help enhance the enzymatic hydrolytic rate by several-fold. Amorphous cellulose has a higher enzymatic hydrolysis rate largely due to its increased enzyme-accessible surface area and reduced crystallinity (i.e., reduced HBs between glucan chains). Since industrial-scale production of amorphous cellulose within a cellulosic biorefinery is still in its research and development phase, other alternative routes to enhance crystalline cellulose reactivity should be explored as well. There are currently not many studies that explain why disruption of the cellulose intra- and intersheet hydrogen bond network enhances cellulase activity.[5] Additional insight into the mechanism of cellulase action can be obtained by selectively disrupting and reorganizing the hydrogen bond network within crystalline cellulose. One simple way to selectively alter the hydrogen bond network from its naturally occurring crystalline state, called I_β, to an activated phase, called III_I, is by treating cellulose with anhydrous liquid ammonia. The interaction between anhydrous liquid ammonia and cellulose I_β crystals leads to a change in the pattern of intrachain and intrasheet hydrogen-bonding and to the formation of new intersheet HBs without any significant loss in overall crystallinity. In the conversion from cellulose I_β to cellulose III_I, ammonia molecules penetrate the cellulose I_β crystal and form a crystalline complex called ammonia−cellulose I,[22] where each ammonia molecule sits in a distorted box defined by the edges of four neighboring glucan chains. X-ray and neutron diffraction studies have indicated that ammonia molecules within the crystal interact with the neighboring chains via multiple HBs.[22,60] Following ammonia evaporation, cellulose does not revert to its initial crystal form (cellulose I_β) but adopts the cellulose III_I form. This new crystalline form of cellulose now allows us to investigate the role of intra- and intersheet HBs on its interaction with cellulases as compared to cellulose I_β and amorphous cellulose.

We found that this subtle structural alteration within the cellulose III_I hydrogen bond network enhanced its overall enzymatic hydrolysis yield by up to 5-fold compared to native cellulose I_β. We used all-atom simulations to identify key structural differences that distinguish cellulose III_I from cellulose I_β and to relate these differences to the improvement in hydrolysis rates for

the former. Comparative MD simulations reveal that cellulose III_I is characterized by a marginally lower packing density as confirmed by the values of the mean squared radius of gyration. The number of HBs within a single horizontal sheet is higher in cellulose I_β fibrils, while the number of HBs between the neighboring sheets is higher in cellulose III_I. These tendencies are common to both the chains in the crystalline core and the chains exposed to the aqueous solvent (i.e., surface chains). Intrachain HBs within increasing chain-length cellulose oligomers have been previously implicated in the increased flexural rigidity and reduced flexibility of the oligomers,[8] which are likely to present a thermodynamic barrier to the formation of a catalytically active complex with cellulases. Furthermore, the crystalline core of the cellulose III_I fibril is characterized by larger structural fluctuations along both the intra- and intersheet directions, when compared with cellulose I_β. Cellulose chains on the surface of cellulose III_I also show larger structural fluctuations as quantified via rmsd's from their position at the beginning of the MD equilibrated trajectory. These larger structural fluctuations in the crystalline core, as well as in the surface chains, may provide a molecular-level explanation for the lower thermal stability reported for cellulose III_I.[61] The larger structural fluctuations of the surface chains along the inter- and intrasheet directions suggest that individual glucan chains for cellulose III_I would be more readily accessible by cellulases compared to cellulose I_β.

The enhanced enzymatic digestion of cellulose III_I may be related to the "amorphous-like" nature of its surface chains. The hydroxymethyl group rotational state occupancy within glucan polymer chains is a critical measure for understanding the interplay between intersheet/intrasheet and cellulose−water hydrogen-bonding in solvated crystalline cellulose fibrils. The calculated redistribution of the inter-/intrasheet hydrogen bond network upon going from cellulose I_β to cellulose III_I is consistent with experimental results.[19,20] In more detail, upon going from cellulose I_β to cellulose III_I, we observe a decrease in the number of intrasheet (intra- and interchain) HBs and an increase in the number of intersheet HBs. The hydroxymethyl rotational-state occupancy of cellulose chains on the surface of I_β and III_I fibrils differs only in the values of *tg* and *gt* states. Our results show that, for the solvent-exposed chains, $tg(I_\beta) > tg(III_I)$ and $gt(I_\beta) < gt(III_I)$, while $gg(I_\beta) \approx gg(III_I)$. In particular, the low percentage (0−4%) for the *tg* rotational state seen for cellulose III_I has also been observed in both solvated and amorphous cellulose.[62] Thus, the surface chains in cellulose I_β have a higher tendency to form intrasheet HBs with the neighboring chains, while the surface chains in cellulose III_I are more inclined to form HBs with bulk water molecules. Indeed, we do observe that cellulose III_I has a greater tendency to form HBs with water. This is consistent with the results on the hydroxymethyl group rotational state at the cellulose−water interface. When compared with cellulose I_β, the solvent-exposed glucan chains in the cellulose III_I fibril display more flexibility, higher hydration, and a lower percentage of the *tg* rotational state for the hydroxymethyl group. These structural attributes are typically associated with amorphous cellulose. However, amorphous cellulose is known to lack long-range order and well-defined molecular conformation that is typical of both crystalline cellulose allomorphs.[62] Also, both cellulose I_β and cellulose III_I were found to adopt classical chair-like pyranoid ring conformations, unlike amorphous cellulose, which was reported to have 27% rings in less stable conformations.[62] This suggests that high-energy distorted pyranoid ring puckering stabilization within a cellulase active site was an unlikely reason for the enhanced hydrolysis rates seen for cellulose III_I.

Both secreted fungal cellulase and cell-complexed bacterial cellulosome paradigms have been explored as potential routes to deconstruct lignocellulose for biofuel applications.[63] However, here we have focused on using secreted fungal enzymes that are of greater commercial interest due to higher protein titers and effective hydrolytic activity reported for *T. reesei* enzymes.[4] As suggested recently,[4,64] it is likely that the cellulose allomorph type determines the relative extent of work fungal cellulases must perform to extract (or decrystallize) cellobiose units from the crystal surface prior to the formation of a catalytically active complex between the enzyme−substrate for subsequent hydrolysis of the glycosidic linkage. The increased flexibility of the glucan chains within cellulose III_I was shown to be dependent on its crystal structure and resulted in enhanced overall hydrolysis yields. However, not all observed enhancements can be attributed to the restructuring of cellulose structure alone, as a significant gain in digestion efficiency is seen only for certain mixtures of exo- and endocellulases. Crude fungal cellulase mixtures are typically composed of 50−80% exocellulases (40−60% Cel7A, 10−20% Cel6A) and 10−25% endocellulases (5−10% Cel7B, 1−10% Cel5A, 1−5% Cel12A, < 1% Cel61A).[65] Addition of endoglucanases significantly accelerated the depolymerization of cellulose III_I compared to that using exocellulases alone. We found that Cel7B was the most effective degrader of cellulose, followed by Cel5A, Cel12A, and Cel61A. Endocellulases create endocuts randomly along glucan polymer chains for cellulose crystals that synergistically enhance apparent exocellulase activity. The synergistic activity of endo- and exocellulases allowed near-theoretical glucan conversions to glucose at industrially relevant low enzyme loadings for cellulose III_I, unlike recent work that reported using 10−20-fold higher Cel7A loading alone (see Supporting Information) with no endocellulase addition to achieve significantly less than theoretical conversion.[17]

Surprisingly, the maximum surface-bound cellulase capacity (i.e., mg bound cellulase per unit weight substrate) for cellulose III_I was 60−70% lower compared to that for native cellulose I_β. Hence, these differences in the hydrolytic activity cannot be explained in terms of the respective binding capacities, unlike what has been reported earlier for amorphous cellulose or cellulose I_β.[4,9,11] Even though all three major *T. reesei* cellulases exhibited similar maximum binding capacities for each of the cellulose allomorphs (Cel7A ∼ Cel7B > Cel6A), their overall binding capacities are 2−3-fold greater for cellulose I_β than for cellulose III_I. Previous studies have indicated that cellulases preferentially bind to the axial cellulose I_β crystal surface possibly due to van der Waals and aromatic ring polarization interactions involving the aromatic residues of the CBMs (Figure 5D) and the pyranose rings.[15,25] Most *T. reesei* cellulases are comprised of highly homologous family 1 CBMs,[15] and that could explain their similar binding affinities for each cellulose allomorph. Recent MD simulation work has shown that family 1 CBMs have greater affinity to the hydrophobic face of cellulose I_β compared to its relatively more hydrophilic surfaces.[66] This could explain why CBM family 1 cellulases have a lower total binding capacity to cellulose III_I, which we have shown to have more hydrophilic fibril surfaces than native cellulose. The relevant question, however, is why these fungal cellulases with overall reduced binding capacity to cellulose III_I are able to achieve higher hydrolysis rates, further highlighting that the crystalline cellulose hydrogen bond network has a more subtle effect on cellulase binding and activity than previously thought. Though the binding of cellulases to native cellulose is mostly

driven via the interaction with the aromatic planar residues found in family 1 CBMs,[46,67,68] its processivity has been found to be driven via hydrogen-bonding.[15] It is currently unclear if the increased tendency for cellulose III_I to allow HBs with water molecules for surface-exposed glucan chains also facilitates its interaction with CBM processivity (and hence increased cellulase efficiency) despite the reduced overall cellulase binding capacity. This has significant implications for rational protein engineering and directed evolutionary strategies for enhancing cellulase activity (and enhancing productive binding affinity for CBMs)[11] toward more efficient cellulose III_I-rich pretreated lignocellulose saccharification.

In addition to the structural and molecular properties of cellulose III_I and CBMs discussed earlier, the interaction of the catalytic domain with the substrate through its active site cleft must play an important role as well. We found that the DSE on cellulose III_I for a endo/exo mixture comprising of Cel7B was at least 2 times greater than for mixtures comprising GH family 5 endocellulases (Cel5A_ac and Cel5A_tr). Comparison of the active-site clefts of these enzymes reveals that Cel7B has a long, deep and relatively unrestricted cleft compared to Cel5A. This open active-site cleft of Cel7B likely plays a critical role in providing an easily accessible platform for more-productive glucan chain binding and efficient catalysis of cellulose III_I. Accordingly, Cel61A, which has no discernible open active-site cleft, shows minimal improvement in DSE on cellulose III_I. Based on the protein structures and activity assays, it appears that a cellulase such as Cel7B with a more open and unrestricted active-site cleft may further accelerate the degradation of cellulose III_I compared to other endocellulases. Furthermore, it is likely that the glucan chain binding free energy for endocellulases is less than that of exocellulases, since threading by processive enzymes should additionally contribute to lowering the thermodynamic barrier to cellulose decrystallization prior to hydrolysis.[4,64] This suggests that altering the cellulose crystal structure (from I_β to III_I) would reduce the decrystallization barrier for cellulases and hence enhance their apparent activity. Even though our combined study of experiments and theory highlights the critical role of cellulose crystal structure in cellulase binding and their hydrolytic activity, several other factors, such as the shape of the microfibril (e.g., hexagonal, rhomboid), relative surface chain population (e.g., edge, middle), total available substrate binding sites, and their relative hydrophobicity to hydrophilicity, would differentially impact the productive/nonproductive interaction of cellulase−cellulose system and the ensuing hydrolysis kinetics. More research is necessary to better understand these otherwise poorly characterized factors, as also highlighted elsewhere.[4,64]

Finally, AFEX, the leading ammonia-based pretreatment,[4] was carried out to determine the impact of forming cellulose III_I using a realistic lignocellulosic feedstock such as corn stover (see Supporting Information). Conventional AFEX pretreatment was recently shown to enhance plant cell wall saccharification rate by delocalizing lignin/hemicellulose to increase cellulase access to embedded crystalline cellulose fibrils.[69] However, there was no major cellulose decrystallization or formation of cellulose III_I detectable during this process. Altering pretreatment conditions to maximize cellulose accessibility and simultaneously produce cellulose III_I would be a significant improvement to the existing AFEX process. We found that there was an 80% increase in glucan digestibility within 6 h of enzymatic hydrolysis for cellulose III_I-rich AFEX corn stover compared to cellulose I_β-rich AFEX corn stover, suggesting that formation of cellulose III_I within a realistic lignocellulosic biomass can indeed improve its hydrolysis rates. There is scope for further improvements to the modified AFEX process that maximize cellulose III_I formation while simultaneously co-extracting valuable lignin and/or hemicellulose co-products to provide additional benefits to downstream enzymatic hydrolysis and microbial fermentation processes.[70] Cellulose II has a crystal structure comparable to that of cellulose III_I and has been reported to also enhance apparent cellulase activity,[23] though marginally lesser than what was seen for cellulose III_I. However, it should be relatively easier to adapt AFEX to produce cellulose III_I and recover ammonia than using sodium hydroxide as pretreatment catalyst to produce cellulose II from a commercial perspective. Using anhydrous liquid ammonia can also minimize water usage in the biorefinery, further improving process economics. Designing economical pretreatments that maximize accessible cellulose surface area for enzyme attack while minimizing the thermodynamic cost for solvent-exposed glucan chain decrystallization by cellulases will lead to enhancing biomass polysaccharides' hydrolysis rates and subsequent development of cost-effective cellulosic biorefineries.[4]

CONCLUSIONS

Two common approaches to improve the enzymatic conversion of polymeric cellulose to glucose are reducing cellulose crystallinity (and thus increasing glycosidic bond accessibility to cellulases) by thermochemical treatments[4−7] and/or improving cellulase activity by protein engineering (and thus increasing cellulase catalytic efficiency).[4,10,11,13,14] However, we demonstrate here an alternative approach, based on reorganizing the hydrogen bond network within crystalline cellulose to produce a cellulose III_I allomorph that can synergistically enhance apparent cellulase activity. Our integrated experimental and theoretical efforts have identified key structural and molecular features of cellulose III_I that make this strategy feasible and, importantly, provide an avenue for further improvements (e.g., engineering cellulases and enzyme cocktails for improved binding and reactivity on cellulose III_I). Compared to previous published reports, AFEX pretreatment was also adapted here to produce restructured cellulose III_I under significantly milder pretreatment and reduced enzyme loading conditions to achieve near-theoretical glucan conversion. This approach can significantly reduce the cellulase loading necessary to achieve economical conversion of lignocellulosic biomass to fuels and chemicals, while also providing fundamental insights into the nature of cellulose recalcitrance.

ASSOCIATED CONTENT

Supporting Information. Additional results and details of methods discussed briefly within the manuscript; initial and final MD simulation configurations for both cellulose polymorphs (.pdb format). This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author
chundawa@msu.edu; gnana@lanl.gov

Present Addresses
○Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, United States.

Author Contributions

[▽]These authors contributed equally to this work.

■ ACKNOWLEDGMENT

This work was funded in part by the DOE Great Lakes Bioenergy Research Center (DOE Office of Science BER DE-FC02-07ER64494). Additional funding was provided from the LANL LDRD-X98U program. Computing resources were provided by LANL institutional computing. We thank Jonathan Walton, Sandra Austin-Phillips, Alfred French, Masahisa Wada, Jennifer Macke, Parthasarathi Ramakrishnan, Ezhiylmurugan Rangasamy, and Robert Smith for their timely assistance.

■ REFERENCES

(1) Himmel, M. E.; Ding, S.-Y.; Johnson, D. K.; Adney, W. S.; Nimlos, M. R.; Brady, J. W.; Foust, T. D. *Science* **2007**, *315*, 804–807.
(2) Binder, J. B.; Raines, R. T. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 4516–4521.
(3) da Costa Sousa, L.; Chundawat, S. P. S.; Balan, V.; Dale, B. E. *Curr. Opin. Biotechnol.* **2009**, *20*, 339–347.
(4) Chundawat, S. P. S.; Beckham, G. T.; Himmel, M.; Dale, B. E. *Annu. Rev. Chem. Biomol. Eng.* **2011**, *2*, 121–145.
(5) Zhang, Y.-H. P.; Cui, J.; Lynd, L. R.; Kuang, L. R. *Biomacromolecules* **2006**, *7*, 644–648.
(6) Swatloski, R. P.; Spear, S. K.; Holbrey, J. D.; Rogers, R. D. *J. Am. Chem. Soc.* **2002**, *124*, 4974–4975.
(7) Rollin, J. A.; Zhu, Z.; Sathitsuksanoh, N.; P, Z. Y. H. *Biotechnol. Bioeng.* **2011**, *108*, 22–30.
(8) Shen, T.; Gnanakaran, S. *Biophys. J.* **2009**, *96*, 3032–3040.
(9) Hall, M.; Bansal, P.; Lee, J. H.; Realff, M. J.; Bommarius, A. S. *FEBS J.* **2010**, *277*, 1571–1582.
(10) Heinzelman, P.; Snow, C. D.; Wu, I.; Nguyen, C.; Villalobos, A.; Govindarajan, S.; Minshull, J.; Arnold, F. H. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 5610–5615.
(11) Carrard, G.; Koivula, A.; Söderlund, H.; Béguin, P. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10342–10347.
(12) Zhang, Y.; Himmel, M.; Mielenz, J. *Biotechnol. Adv.* **2006**, *24*, 452–481.
(13) Zhang, X.; Zhang, Y. *Microbial Biotech.* **2011**, *4*, 98–105.
(14) Li, Y.; Irwin, D.; Wilson, D. *Appl. Environ. Microbiol.* **2011**, *76*.
(15) Beckham, G. T.; Matthews, J. F.; Bomble, Y. J.; Bu, L.; Adney, W. S.; Himmel, M. E.; Nimlos, M. R.; Crowley, M. F. *J. Phys. Chem. B* **2010**, *114*, 1447–1453.
(16) Weimer, P. J.; French, A. D.; Calamari, T. A. *Appl. Environ. Microbiol.* **1991**, *57*, 3101–3106.
(17) Igarashi, K.; Wada, M.; Samejima, M. *FEBS J.* **2007**, *274*, 1785–1792.
(18) Hayashi, N.; Sugiyama, J.; Okano, T.; Ishihara, M. *Carbohydr. Res.* **1997**, *305*, 109–116.
(19) Nishiyama, Y.; Langan, P.; Chanzy, H. *J. Am. Chem. Soc.* **2002**, *124*, 9074–9082.
(20) Wada, M.; Chanzy, H.; Nishiyama, Y.; Langan, P. *Macromolecules* **2004**, *37*, 8548–8555.
(21) Zugenmaier, P. *Prog. Polym. Sci.* **2001**, *26*, 1341–1417.
(22) Wada, M.; Nishiyama, Y.; Langan, P. *Macromolecules* **2006**, *39*, 2947–2952.
(23) Chen, Y.; Stipanovic, A.; Winter, W.; Wilson, D.; Kim, Y.-J. *Cellulose* **2007**, *14*, 283–293.
(24) Matthews, J. F.; Skopec, C. E.; Mason, P. E.; Zuccato, P.; Torget, R. W.; Sugiyama, J.; Himmel, M. E.; Brady, J. W. *Carbohydr. Res.* **2006**, *341*, 138–152.
(25) Lehtiö, J.; Sugiyama, J.; Gustavsson, M.; Fransson, L.; Linder, M.; Teeri, T. T. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 484–9.
(26) Dimitriu, S., Ed. *Polysaccharides: structural diversity and functional versatility*, 2nd ed.; Marcel Dekker: New York, 2005.
(27) O'Sullivan, A. *Cellulose* **1997**, *4*, 173–207.
(28) Somerville, C.; Bauer, S.; Brininstool, G.; Facette, M.; Hamann, T.; Milne, J.; Osborne, E.; Paredez, A.; Persson, S.; Raab, T.; Vorwerk, S.; Youngs, H. *Science* **2004**, *306*, 2206–2211.
(29) Arantes, V.; Saddler, J. N. *Biotechnol. Biofuels* **2010**, *3*, 4.
(30) Lodish, H. *Molecular Cell Biology*, 6th ed.; W.H. Freeman and Co.: Baltimore, MD, 2010.
(31) 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.
(32) Kirschner, K.; Yongye, A.; Tschampel, S.; González-Outeiriño, J.; Daniels, C.; Foley, L.; Woods, R. *J. Comput. Chem.* **2008**, *29*, 622–655.
(33) Jorgensen, W. L.; Chandrasekhar, J.; D, M. J.; Impey, R. W.; L, K. M. *J. Chem. Phys.* **1983**, *79*, 926–935.
(34) Feller, S.; Zhang, Y.; Pastor, R.; Brooks, B. *J. Chem. Phys.* **1995**, *103*, 4613.
(35) Martyna, G. J.; Tobias, D. J.; Klein, M. L. *J. Chem. Phys.* **1994**, *101*, 4177.
(36) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
(37) Ryckaert, J.; Ciccotti, G.; Berendsen, H. *J. Comput. Phys.* **1977**, *23*, 327–341.
(38) Gao, D.; Chundawat, S. P. S.; Krishnan, C.; Balan, V.; Dale, B. E. *Bioresour. Technol.* **2010**, *101*, 2770–2781.
(39) Lau, M. W.; Gunawan, C.; Dale, B. E. *Biotechnol. Biofuels* **2009**, *2*, 30.
(40) Park, S.; Baker, J.; Himmel, M.; Parilla, P.; Johnson, D. *Biotechnol. Biofuels* **2010**, *3*, 1–10.
(41) Banerjee, G.; Car, S.; Scott-Craig, J. S.; Borrusch, M. S.; Bongers, M.; Walton, J. D. *Bioresour. Technol.* **2010**, *101*, 9097–9105.
(42) Ziegelhoffer, T.; Raasch, J. A.; Austin-Phillips, S. *Plant Biotechnol. J.* **2009**, *7*, 527–536.
(43) Cantarel, B. L.; Coutinho, P. M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B. *Nucleic Acids Res.* **2009**, *37*, D233–D238.
(44) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235–242.
(45) DeLano, W. *The PYMOL Molecular Graphics System*, 2002.
(46) Creagh, A. L.; Ong, E.; Jervis, E.; Kilburn, D. G.; Haynes, C. A. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 12229–12234.
(47) Allen, T.; Andersen, O.; Roux, B. *J. Gen. Physiol.* **2004**, *124*, 679–690.
(48) Theodorou, D. N.; Suter, U. W. *Macromolecules* **1985**, *18*, 1206–1214.
(49) Mills, J. E.; Dean, P. M. *J. Comput. Aided Mol. Des.* **1996**, *10*, 607–622.
(50) Luzar, A.; Chandler, D. *Nature* **1996**, *379*, 55–57.
(51) Reddy, C. K.; Das, A.; Jayaram, B. *J. Mol. Biol.* **2001**, *314*, 619–632.
(52) Linder, M.; Teeri, T. *J. Biotechnol.* **1997**, *57*, 15–28.
(53) Nimlos, M. R.; Matthews, J. F.; Crowley, M. F.; Walker, R. C.; Chukkapalli, G.; Brady, J. W.; Adney, W. S.; Cleary, J. M.; Zhong, L.; Himmel, M. E. *Protein Eng. Des. Sel.* **2007**, *20*, 179–187.
(54) Gilkes, N. R.; Kilburn, D. G.; Miller, R. C.; Warren, R. A.; Sugiyama, J.; Chanzy, H.; Henrissat, B. *Int. J. Biol. Macromol.* **1993**, *15*, 347–351.
(55) Mulakala, C.; Reilly, P. *J. Proteins* **2005**, *60*, 598–605.
(56) Igarashi, K.; Koivula, A.; Wada, M.; Kimura, S.; Penttilä, M.; Samejima, M. *J. Biol. Chem.* **2009**, *284*, 36186–36190.
(57) von Ossowski, I.; Ståhlberg, J.; Koivula, A.; Piens, K.; Becker, D.; Boer, H.; Harle, R.; Harris, M.; Divne, C.; Mahdi, S.; Zhao, Y.; Driguez, H.; Claeyssens, M.; Sinnott, M. L.; Teeri, T. T. *J. Mol. Biol.* **2003**, *333*, 817–829.
(58) Horn, S. J.; Sikorski, P.; Cederkvist, J. B.; Vaaje-Kolstad, G.; Sørlie, M.; Synstad, B.; Vriend, G.; Vårum, K. M.; Eijsink, V. G. H. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 18089–18094.
(59) Vaaje-Kolstad, G.; Westereng, B.; Horn, S.; Liu, Z.; Zhai, H.; Sorlie, M.; Eijsink, V. *Science* **2010**, *330*, 219–222.

(60) Wada, M.; Nishiyama, Y.; Bellesia, G.; Forsyth, T.; Gnanakaran, S.; Langan, P. *Cellulose* **2011**, *18*, 191–206.
(61) Wada, M. *J. Polym. Sci., Part B* **2002**, *40*, 1095–1102.
(62) Mazeau, K.; Heux, L. *J. Phys. Chem. B* **2003**, *107*, 2394–2403.
(63) Bayer, E. A.; Chanzy, H.; Lamed, R.; Shoham, Y. *Curr. Opin. Struct. Biol.* **1998**, *8*, 548–557.
(64) Beckham, G.; Matthews, J.; Peters, B.; Bomble, Y.; Himmel, M.; Crowley, M. *J. Phys. Chem. B* **2011**, *115*, 4118–4127.
(65) Rosgaard, L.; Pedersen, S.; Langston, J.; Akerhielm, D.; Cherry, J. R.; Meyer, A. S. *Biotechnol. Prog.* **2007**, *23*, 1270–1276.
(66) Yui, T.; Shiiba, H.; Tsutsumi, Y.; Hayashi, S.; Miyata, T.; Hirata, F. *J. Phys. Chem. B* **2010**, *114*, 49–58.
(67) Guo, J.; Catchmark, J. M. Thermodynamics of Family 1 Cellulose-Binding Modules from T. reesei Cel7A and Cel6A. Presented at the American Society of Agricultural and Biological Engineers Annual International Meeting, Pittsburgh, PA, June 20−23, 2010; Paper No. 1009577.
(68) Mohamed, M.; Watts, H.; Guo, J. *Carbohydr. Res.* **2010**, *345*, 1741–1751.
(69) Chundawat, S. P. S.; Donohoe, B.; Sousa, L.; Elder, T.; Agarwal, U.; Lu, F.; Ralph, J.; Himmel, M.; Balan, V.; Dale, B. *Energy Environ. Sci.* **2011**, *4*, 973–984.
(70) Chundawat, S. P. S.; Sousa, L.; Cheh, A.; Balan, V.; Dale, B. E. Digestible lignocellulosic biomass and extractives and methods for producing same. U.S. Patent Appl. PCT/US2011/033079 filed on 04/19/2011 with United States receiving office.