The Energy Landscape for the Interaction of the Family 1 Carbohydrate-Binding Module and the Cellulose Surface is Altered by Hydrolyzed Glycosidic Bonds

Lintao Bu,[†,‡] Gregg T. Beckham,[†,‡] Michael F. Crowley,[§] Christopher H. Chang,[||] James F. Matthews,[§] Yannick J. Bomble,[§] William S. Adney,[§] Michael E. Himmel,[§] and Mark R. Nimlos*[,‡]

National Bioenergy Center, National Renewable Energy Laboratory, Golden, Colorado 80401, Chemical and Biosciences Center, National Renewable Energy Laboratory, Golden, Colorado 80401, Materials and Computational Sciences Center, National Renewable Energy Laboratory, Golden, Colorado 80401

Received: April 30, 2009; Revised Manuscript Received: June 11, 2009

A multiscale simulation model is used to construct potential and free energy surfaces for the carbohydrate-binding module [CBM] from an industrially important cellulase, *Trichoderma reesei* cellobiohydrolase I, on the hydrophobic face of a coarse-grained cellulose Iβ polymorph. We predict from computation that the CBM alone exhibits regions of stability on the hydrophobic face of cellulose every 5 and 10 Å, corresponding to a glucose unit and a cellobiose unit, respectively. In addition, we predict a new role for the CBM: specifically, that in the presence of hydrolyzed cellulose chain ends, the CBM exerts a thermodynamic driving force to translate away from the free cellulose chain ends. This suggests that the CBM is not only required for binding to cellulose, as has been known for two decades, but also that it has evolved to both assist the enzyme in recognizing a cellulose chain end and exert a driving force on the enzyme during processive hydrolysis of cellulose.

Introduction

Cellulose is the primary structural polysaccharide of plant cell walls and could potentially serve as a renewable energy source for ethanol or other transportation fuels derived from sugars.[1] However, due to the recalcitrance of biomass to deconstruction, the process of producing fuel from biomass sugars via fermentation has remained costly.[2,3] Because the high cost of cellulase enzymes is a key cost factor in this process, it is now widely accepted that enhanced understanding of the mechanisms of enzymatic degradation of cellulose remains critical to enabling a successful bioethanol industry.[4]

Trichoderma reesei cellobiohydrolase [CBH] I, referred to here as Cel7A, is an exoglucanase which acts on the available reducing ends of cellodextrins via a processive mechanism.[5,6] Cel7A from *T. reesei* (also known as *Hypocrea jecorina*) contains a catalytic domain and a family 1 carbohydrate-binding module [CBM] separated by a glycosylated linker peptide.[7,8] Cel7A is hypothesized to align on or near a hydrolyzed cellulose surface chain from the reducing end, decrystallizing the chain from the cellulose surface, and "directing it" to the entrance of the active site tunnel located in the catalytic domain where the cellulose chain is hydrolyzed to cellobiose. The characteristic of processivity makes Cel7A and other members of glycoside hydrolase family 7 among the most interesting and important group of cellulases known today. Among the many families of enzymes contributing to cellulose deconstruction (e.g., cellulose hydrolysis is accomplished in fungi by the synergistic action of exoglucanases and endoglucanases), Cel7A provides most of the hydrolytic potential. This has been confirmed from experimental work, where optimal mixtures of endoglucanases and CBHs invariably consist of 20% of the former and 80% of the latter.[9] However, many questions still remain about the detailed functions of the different domains in Cel7A during catalysis. Understanding the processive cycle of Cel7A and specifically the interactive mechanisms of the subdomains both with each other and with cellulose is essential to a complete characterization of cellulase function.

To that end, many studies have been conducted to probe the function of CBMs.[10−18] It has been proposed that CBMs have three primary functions: proximity effects (keeping the catalytic domain near cellulose), substrate targeting (binding to specific faces of cellulose), and nonhydrolytic cellulose disruption.[10] Proximity effects have been confirmed by multiple mutation studies[13,14,16] whereas substrate targeting has been shown on large crystals of cellulose with immuno-gold-labeled CBMs.[12] Cellulose disruption by some bacterial CBMs has been indirectly confirmed, but no mechanistic details are known regarding this function.[18,19]

Recently, Nimlos and co-workers used molecular dynamics (MD) to investigate the interaction of the CBM from *T. reesei* Cel7A with an atomistic model of the cellulose surface modified to display a broken cellulose chain.[17] The initial configuration in their simulations was such that the tyrosine residues on the hydrophobic surface of the CBM—specifically, Y5, Y31, and Y32—are in contact with the cellulose surface. From this configuration, a MD simulation shows that the fourth tyrosine residue in the CBM (Y13) moves from its internal position to form hydrophobic interactions with the cellulose surface. However, due to the limitation of the accessible time scales in all-atom simulations, the translation mechanism of the CBM along the cellulose chain was not examined.

To study more efficiently these types of systems at longer time scales than are accessible by fully atomistic simulations, coarse-grained models for a variety of biopolymers have been

* To whom correspondence should be addressed. E-mail: mark.nimlos@nrel.gov.
‡ National Bioenergy Center.
§ Chemical and Biosciences Center.
|| Materials and Computational Sciences Center.
† Equal contribution.

10.1021/jp904003z CCC: $40.75
Published on Web 07/13/2009

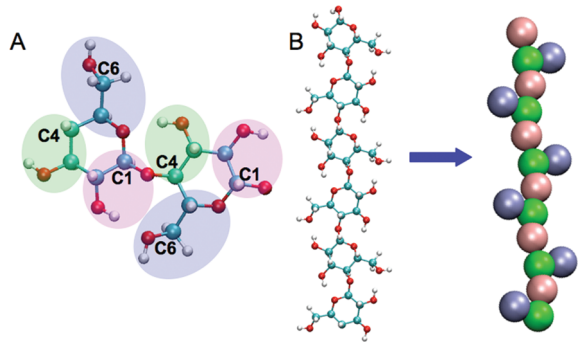

Figure 1. (A) Parsing of the atomistic cellulose residue model among coarse-grained beads. (B) The atomistic model and coarse-grained model of a cellodextrin chain demonstrate the reduction of the system size.

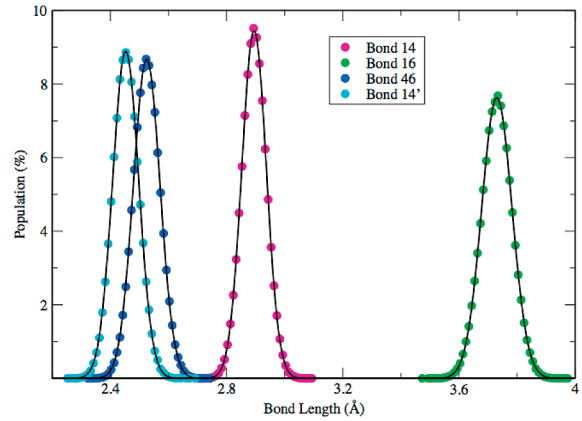

Figure 2. Bond length distribution for selected bonds from the atomistic simulation of the cellulose Iβ polymorph. Solid circles are sampled values of the distance between two atoms (C1, C4, or C6) in the atomistic simulation. The solid lines represent fitted Gaussian functions for the coarse-grained surface model.

developed.[20−23] In these models, a group of atoms are represented by a single bead to reduce the system size. Recently, Molinero and Goddard developed the first coarse-grained model (M3B) for carbohydrates using three beads to replace each glucose unit of oligosaccharides and one bead to represent each water molecule.[24,25] Subsequently, Liu and co-workers derived a coarse-grained model for monosaccharides in aqueous solution using a systematic, multiscale coarse-graining algorithm.[26] However, to our knowledge, coarse-grained models for crystalline cellulose have to date been limited to lattice models.[27]

Here, we present a thermodynamic study of the CBM from *T. reesei* Cel7A on a coarse-grained cellulose surface in implicit solvent. Specifically, we develop a coarse-grained model for the hydrophobic face of cellulose Iβ. From this model, we calculate the potential energy surface (PES) and free energy surface (FES) for the CBM (represented atomistically) on the hydrophobic face of a cellulose Iβ microfibril that is ~300 Å long, both in the case of a smooth surface (unbroken) and in the presence of a hydrolyzed bond (broken) in the center of the microfibril. In addition, we report several sets of MD simulations that corroborate the PES and FES results.

Methods

This section describes the development of the coarse-grained model for the hydrophobic surface of the cellulose Iβ polymorph, the construction of the PESs and FESs, and the parameters for the MD simulations.

Coarse-Grained Model. The approach taken to develop a surface model for the hydrophobic face of the cellulose Iβ polymorph was as follows:

1. Fit bonds, angles, and dihedral parameters to an atomistic simulation of a cellulose microfibril in explicit water,
2. scale the directionality of cellulose interactions according to dynamic self-assembly of individual cellulose chains in an implicit solvation model, and
3. validate the ability of the surface model to represent the lattice parameters of the cellulose Iβ microfibril with MD simulation.

Similar to the M3B model developed by Molinero and Goddard to represent malto-oligosaccharides in solution,[24] the coarse-grained model developed here also uses three neutral beads to represent each glucose unit, placing the beads at positions corresponding to the atoms C1, C4, and C6 in the atomistic model, as shown in Figure 1. The mass of each bead is the sum of the mass of the atoms that the specific bead replaces. The three beads in each glucose unit are connected by coarse-grained bonds. An additional bond representing the glycosidic linkage connects the glucose units in a cellulose chain. To parametrize the coarse-grained model for cellulose, an atomistic simulation of crystalline cellulose in water was conducted first to provide the detailed dynamics information of the positions of atoms C1, C4, and C6. This simulation is used to derive the force constants for coarse-grained beads. The C1 and C4 beads, shown in pink and green, respectively, in Figure 1 are referred to as "backbone" beads, whereas the C6 bead is referred to as a "side chain" bead.

The atomistic model of a microcrystallite of native cellulose Iβ was constructed using the coordinates reported by Nishiyama and co-workers.[28,29] The microcrystallite contains 32 cellulose chains with 14 glucans in each chain. The computational protocol follows that of Matthews and co-workers.[30] The CHARMM22 parameters[31] and the CSFF parameters[32,33] were used for the all-atom simulations. The crystal was placed in an equilibrated box of TIP3P[34] water molecules with initial dimensions of 62.2 Å by 62.2 Å by 93.3 Å. Water molecules that overlapped with the cellulose heavy atoms were removed. The system was relaxed in the *NPT* ensemble at 300 K with a Nosé−Hoover thermostat[35,36] and 1 atm for 500 ps with a step size of 1 fs and a 15.0 Å cutoff for nonbonded interactions, resulting in final box dimensions of 57.7 Å by 63.1 Å by 98.5 Å. The system was then simulated in the *NVE* ensemble for 1 ns at 300 K.

From the atomistic simulation, the intramolecular interactions (bonded terms) were fit. The interactions between coarse-grained beads can be expressed as a sum of the bonded and nonbonded terms:

$$E = \frac{1}{2}k(r - r_0)^2 + \frac{1}{2}k_\theta(\theta - \theta_0)^2 + B(1 + \cos(\varphi - \varphi_0)) + E_{\text{nbond}} \quad (1)$$

where r, θ, and φ are the distance, angle, and torsional angle between connected coarse-grained beads, respectively; r_0, θ_0, and φ_0 are the coarse-grained bond, angle, and torsional angle equilibrium values, respectively; and k, k_θ, and B are the force constants for the bond, angle, and torsional terms, respectively. The bonded parameters can be derived directly from Boltzmann inversion of the bond, angle, and torsional angle distributions of the all-atom model. Figure 2 demonstrates how the bond parameters of the coarse-grained beads are derived from the

TABLE 1: Bond Parameters

type	r_0 (Å)	k (kcal mol^{-1} Å^{-2})
14	2.89	335
16	3.73	218
46	2.52	282
14′	2.45	296

TABLE 2: Angle Bending Parameters

type	θ_0 (°)	k (kcal mol^{-1} rad^{-2})
641′	102	150
141′	171	148
614′	121	163
414′	162	150

1 ns atomistic simulation of crystalline cellulose in water. The distance distributions of four pairs of atoms, that is, C1−C4, C1−C6, C4−C6, and C4−C1′, were obtained from the atomistic simulation. Gaussian functions were then fit to the distance distributions to extract the equilibrium bond distances and force constants for the virtual bonds between two coarse-grained beads. The equivalent figures for the angles and dihedrals are shown in the Supporting Information.

The nonbonded interaction between coarse-grained beads is represented by the standard 6−12 Lennard-Jones potential. The depth of the potential energy well, D_0, was taken from the M3B model, and the distance, R_0, was determined by the distance between two corresponding atoms in the atomistic model. Because the nonbonded parameters in the M3B model were derived from the structure of amorphous α-D-glucose, they are not transferable to crystalline cellulose, in which strong directionality exists in the structure. For example, the nonbonded interactions between two adjacent cellulose chains in the same layer are dominated by strong hydrogen bond interactions, whereas there are primarily only hydrophobic interactions between two chains in different layers. To introduce directionality into the coarse-grained model, the interactions between different bead types were scaled by distinct rescaling factors.

These rescaling factors were chosen iteratively to enable the force field to govern best the self-assembly of dissociated cellulose chains. Fully radially dissociated cellulose structures with either 21 or 37 chains were chosen as the initial conformations in these simulations. The distance between two adjacent cellulose chains in the dissociated chain starting configuration is 20 Å, which is chosen so that no intermolecular interactions pre-exist between the chains. A one-sided cylindrical harmonic restraint with a 30 Å (for 21 chains) or a 40 Å (for 37 chains) radius and a force constant of 0.01 kcal mol^{-1} Å^{-2} was applied to prevent the cellulose chains from drifting away from the cylindrical core. Note that the radius of the restraining cylinder is much larger than the radius of the crystalline cellulose we studied here, the latter being 15−25 Å. Similar harmonic restraints were used to study membrane assembly of simple helix homo-oligomers[37] and helix association in *Mycobacterium tuberculosis* membrane proteins.[38]

In the self-assembly simulations and all other multiscale simulations conducted in this study, the generalized Born implicit solvation model with switching [GBSW][39,40] was used. The generalized Born radii for the cellulose model were assumed to be equal to the van der Waals radii of the coarse-grained beads. The GBSW smoothing length was 0.6 Å. A constant dielectric was used with a nonbonded cutoff distance of 10 Å applied to Lennard-Jones and electrostatic interactions. For the self-assembly simulations and all subsequent MD simulations, the time step was 1 fs, the simulations were conducted in the *NVT* ensemble, and the temperature was maintained at 300 K with a Nosé−Hoover thermostat.[35,36]

TABLE 4: Nonbonded Parameters

bead	mass (amu)	R_0 (Å)	D_0 (kcal mol^{-1})
B1	59	5.76	1.25
B4	43	6.73	1.22
B6	60	4.70	1.19

During the 100 ps self-assembly simulations, the dissociated individual cellulose chains coalesced to form sheets similar to those in cellulose Iα and Iβ. Even though the same interaction between backbone beads (C1 and C4) and side-chain beads (C6) exists between layers as within a layer, the layers still form and keep a larger separation than intralayer distances. Thus, the coarse-grain model appears to mimic hydrogen bonding well and is able to reproduce emergent layer formation with few imperfections in the crystal order. A close examination of the self-assembled structure suggests that the cellulose chains form individual layers, that is, the (100) surface of cellulose Iβ or the (110) surface of Iα, reasonably correctly. However, the layers do not stack together like the hydrogen bonded sheets in cellulose I, which is a limitation of the model. The finalized coarse-grained parameters for the cellulose Iβ polymorph (100) surface are listed in Tables 1, 2, 3, and 4. The final scaling factors are as follows: the repulsive components of the interactions are all scaled by a factor of 1.5, the attractive component between backbone beads and side chain beads was scaled by a factor of 1.5, the backbone−backbone interaction was scaled by a factor of 0.1, and the side-chain−side-chain bead interaction was scaled by a factor of 0.2. A CHARMM topology file for the coarse-grained model can be found in the Supporting Information.

To validate the coarse-grained model, we ran a 1 ns MD simulation of the coarse-grained cellulose Iβ polymorph starting from the crystal structure. The cellulose microfibril was 4 layers with 4 chains per layer; the cellulose chains were each 10 cellobiose units long. A snapshot of the crystal from two different views is shown in Figure 3. The lattice parameters and degree of twist in the microfibril from this validation simulation are shown in Table 5, which shows that both are in relatively good agreement with the crystal structure from the atomistic simulation, with the exception of the *a* lattice parameter. Because the objective of this study was to examine the interaction of the CBM with the hydrophobic face (100) of the cellulose Iβ polymorph, this discrepancy was considered to be unimportant for the intended application. However, we must therefore stress that the coarse-grained model presented here is only useful in studying the (100) face of cellulose Iβ and should not be used to study the physical behavior of other surfaces or bulk behavior of crystalline cellulose. Additionally, we point out that the model developed here could be improved further to incorporate behavior of the entire cellulose microfibril (and

TABLE 3: Torsion Angle Parameters

type	641′6′	641′4′	141′6′	141′4′	6141′	1641′	6414′	4614′	41′4′1″
B (kcal/mol)	33.6	10.4	4.3	2.8	6.0	148.3	13.5	95.7	4.6
φ_0 (°)	−2	−12	−161	−170	−23	4	−166	5	−8

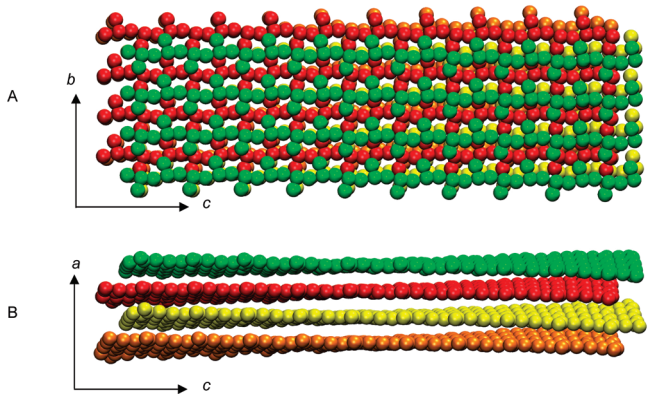

Figure 3. Snapshot of the MD simulation of a coarse-grained crystal used to validate the coarse-grained model for the (100) face of cellulose. (A) View from the (100) face and (B) view from the (010) face.

TABLE 5: Unit Cell Parameters of All-Atom Cellulose and the Coarse-Grained Cellulose Surface Model[a]

	native crystal[b]	all-atom model[c]	CG model, crystal simulation
a (Å)	7.8	8.5	10.8
b (Å)	8.2	8.1	8.1
c (Å)	10.4	10.5	10.5
γ (°)	96.5	≈ 90	90.0
twist (°)	0	1.4−1.7	2.5

[a] Note that the unit cell parameters b and c define the (100) surface. [b] Refs 28, 29. [c] Similar to ref 30.

other cellulose polymorphs) if more rigorous coarse-graining methods are used, such as that developed by Voth and co-workers.[26,41,42]

Potential Energy Surfaces. The potential energy surfaces were calculated for the CBM on the hydrophobic face of the cellulose microfibril by placing the CBM on an $X-Y$ grid facing toward the nonreducing end and minimizing the system at each grid point. The potential energy reported is the interaction energy between the cellulose microfibril and the CBM. Two PESs were generated: one with a flat, unbroken surface and another with a hydrolyzed cellulose chain in the center of the surface, as shown in Figure 4. In both cases, the cellulose microfibril is 4 chains across, 4 layers deep, and 30 cellobiose units long. All minimizations were conducted with the Adopted-Basis Newton−Raphson algorithm.

The starting configuration for the PES was generated by minimizing the coarse-grained cellulose microfibril without the CBM until convergence with harmonic restraints of 0.01 kcal mol^{-1} Å^{-2} along the microfibril perimeter. Separately, the CBM was dynamically equilibrated on a coarse-grained cellulose surface for 1 ns. As observed in previous work, Y13 unfolds from the body of the protein onto the cellulose surface.[17] After minimization of the microfibril, the cellulose was fixed, and the dynamically equilibrated CBM was introduced at the center of the hydrophobic face of the microfibril and minimized until convergence. This minimized configuration of both the CBM and the cellulose microfibril formed the starting point for all PES calculations.

The resolution of the PES grid on the hydrophobic face of the microfibril is 0.25 Å in both the X and Y directions. Preliminary testing showed that the center of mass of the CBM moves less than the grid resolution for each grid point. The microfibril is approximately 306 Å by 26 Å. To avoid edge effects in both the X and Y directions, the grid for the PES consisted of 993 by 13 points, which corresponds to 248 Å by 3 Å. For each minimization, the CBM in the starting configuration described above was translated to the relevant grid point and minimized until convergence, which typically occurred in less than 1000 steps at any given point. For the PES with the hydrolyzed microfibril, the bond between residues 15 and 16, shown in Figure 4 as a blue space-fill representation, was hydrolyzed at the start of each minimization run. The hydrolyzed bond was placed at an approximate microfibril length of 125 Å. During each minimization for the PES, the cellulose microfibril was harmonically restrained with a force constant of 0.01 kcal mol^{-1} Å^{-2} along the microfibril perimeter. The location of the hydrolyzed bond was chosen to be in the interior of the hydrophobic face to avoid edge effects, but it is possible that cellulose chains on exposed corners are more susceptible to attack by endoglucanases than are chains in the interior of this face.

In each case, the PES was collapsed into a 1-dimensional curve by averaging the values for every slice in the Y direction (across the width of the microfibril), corresponding to an average over 13 potential energy values per point along the X direction (down the length of the microfibril). To smooth the noise in the data, a moving average was taken ±2 grid points in the X direction.

Free Energy Surfaces. The free energy surfaces were generated at 300 K using the equilibrium path sampling (EPS) method,[43] which is based on a hybrid Monte Carlo−MD scheme, BOLAS.[44] The free energy simulations were split into 165 windows of 2 Å in the X direction, each with a 0.5 Å overlap on each side for a total overlap of 1 Å per window. For the FES, trajectories were 0.5 ps long, and the EPS simulation was run in each window until 1000 trajectories were accepted. The reaction coordinate is the X coordinate of the center of mass of the CBM. The cellulose microfibril was restrained with force constant of 1.0 kcal mol^{-1} Å^{-2}. No constraints were placed on the CBM center of mass in the Y or Z directions. Starting configurations for each window were the minimized configuration in the center grid point of the relevant PES axial slice. The weighted histogram analysis method (WHAM) was used to construct the FES.[45] To minimize the computational cost, the FES for the broken chain was constructed only in the region directly around the hydrolyzed glycosidic linkage.

Results and Discussion

The PESs along the entire microfibril are shown in Figure 5. Snapshots of the PESs for the unbroken and hydrolyzed microfibril are shown in Figure 6 for a length of 80 Å along the microfibril. Both PESs exhibit deeper regions of stability for the CBM every 10 Å, which corresponds to the length of a cellobiose unit and a shallower minimum every other 5 Å, corresponding to a glucose unit. Interestingly, Cel7A produces a cellobiose unit as the hydrolysis product, so the distance between the CBM "global" minima correspond to the length scale of the catalytic cycle of the entire enzyme. In addition, the unbroken microfibril PES exhibits a uniform periodicity along the entire length of the microfibril, as expected. This indicates that the CBM will undergo random diffusion in either direction along the hydrophobic face in the presence of a uniform surface at equilibrium conditions.

However, the PES for the cellulose surface with the hydrolyzed bond exhibits markedly different behavior in the region around the hydrolyzed bond as shown in Figures 5B and 6B. Specifically, there is a significant barrier to translate over the hydrolyzed cellulose chain on the order of 10 kcal/mol, and there is a significant downhill trend in the potential energy for a length of approximately 25 Å after the CBM center of mass

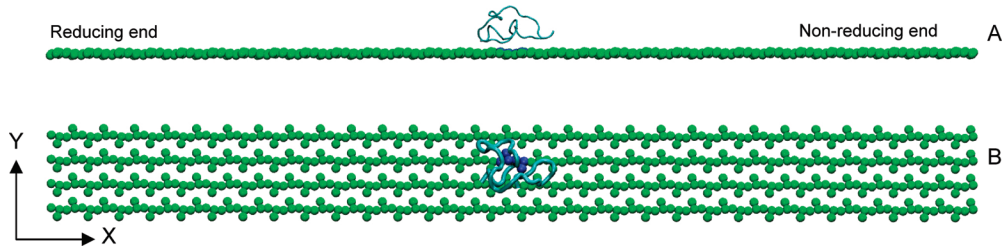

Figure 4. System setup for CBM on the (100) hydrophobic face of cellulose Iβ. For clarity, only one cellulose layer is shown. (A) Side view. The Cel7A enzyme processes from the reducing end to the nonreducing end, or from left to right in this diagram. (B) Top view in which the hydrolyzed bond is shown in blue space fill under the CBM. The CBM coordinates shown are the starting coordinates for the PES calculations.

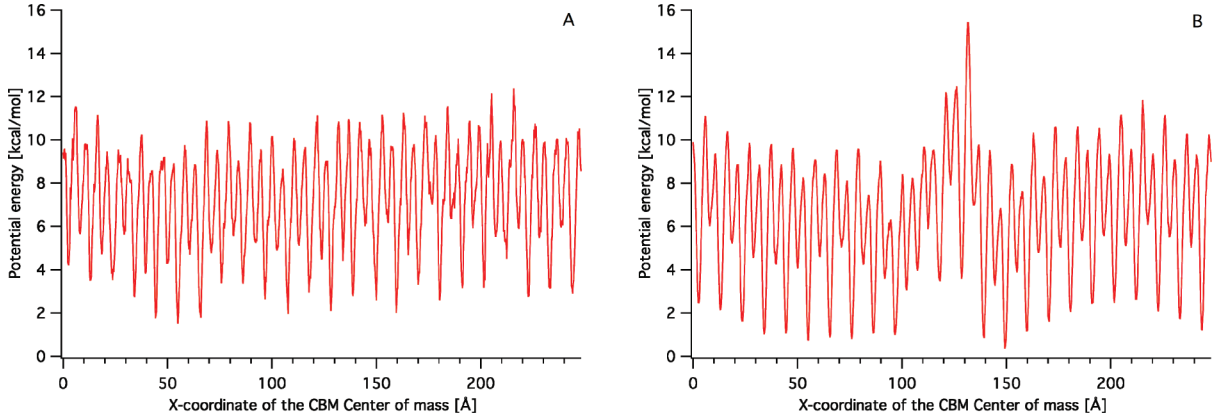

Figure 5. Relative potential energy as a function of the cellulose microfibril length shown in Figure 4 for (A) the unbroken surface and (B) the broken surface along the entire microfibril length of 248 Å. A length of 0 corresponds to the reducing end and a microfibril length of 248 Å corresponds to the nonreducing end.

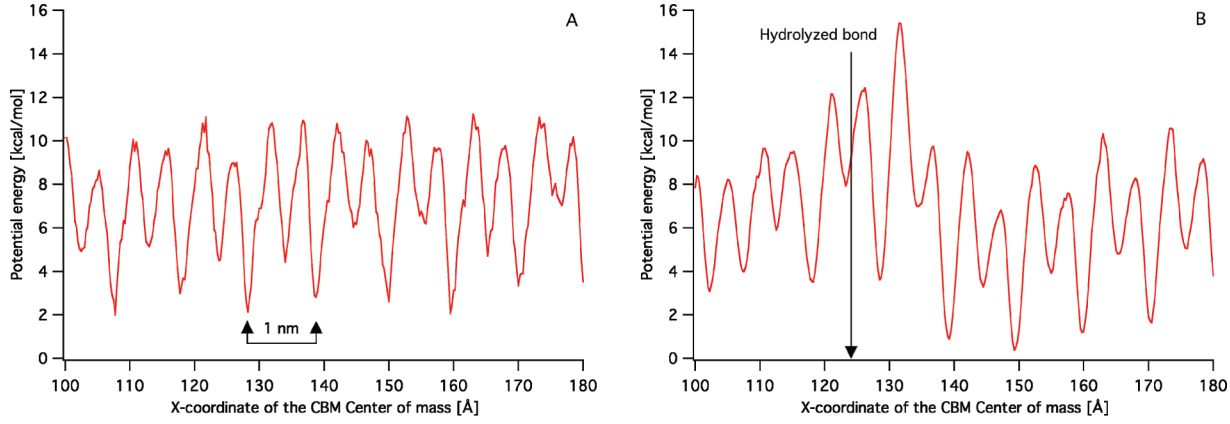

Figure 6. A section of the PES for (A) the unbroken cellulose microfibril and (B) the broken cellulose microfibril for a 80 Å section in the middle of the microfibril. The position of the break is labeled in (B). The coordinate for the PES calculations is the center of mass of the CBM.

passes over the break, as highlighted in Figure 6B. This distance is greater than the distance from the CBM center of mass to the CBM edge, which means that the driving force is exerted on the CBM even after the subdomain is not overlapping the hydrolyzed bond.

The corresponding FESs are shown for the 50 Å around the midpoint of the microfibril in Figure 7. The entire FES for the unbroken microfibril is shown in the Supporting Information. The FES for the unbroken microfibril exhibits a small free energy barrier for processivity of approximately 1−2 kcal/mol, which is within the typical resolution of free energy calculations. The processivity unit for the FES is 5 Å. Since the FES is significantly damped relative to the PES, the deeper well is not seen every 10 Å as is shown by the PES. In the case of the hydrolyzed chain, the CBM exhibits two significant free energy barriers of approximately 5 kcal/mol to translation over the broken strand. The downhill trend to translate away from the break is not as apparent in the FES, most probably because the free energy results are significantly noisier than the PES results and the resolution of free energy calculations is typically on the order of 1−2 kcal/mol, which is the resolution observed in the FES construction for this model. It should also be noted that because the model is constructed with implicit solvent and a coarse representation of cellulose, the absolute potential energy and free energy barriers are most likely damped. The relative magnitudes of the barriers, however, serve to demonstrate the differences in the unbroken surface and the surface with the hydrolyzed cellulose chain. The Fourier transforms of the unbroken PES and the corresponding unbroken FES are shown in Figure 8. As shown in the frequency domain, there are significant peaks in the PES every 10 and 5 Å (0.1 and 0.2 $Å^{-1}$, respectively). The Fourier transform for the FES shows a significant peak every 5 Å, but noisier results for 10 Å, as expected.

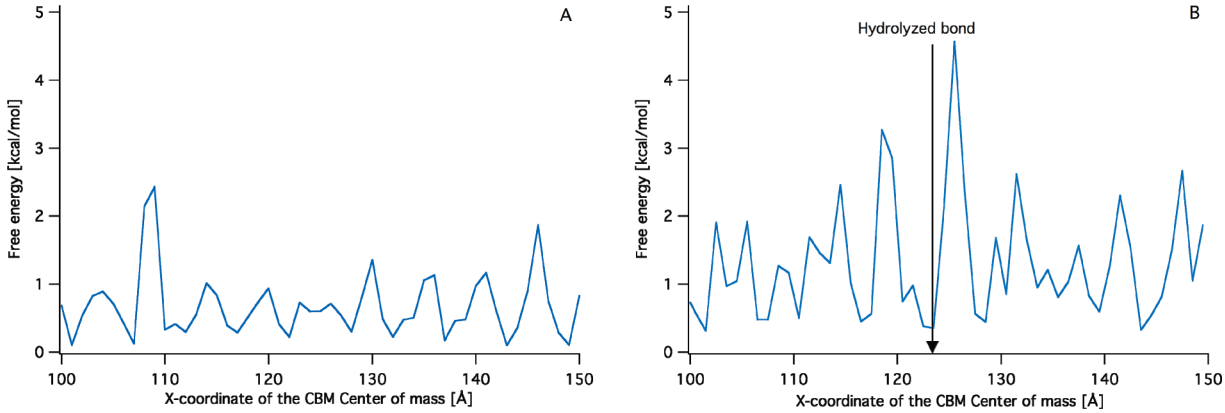

Figure 7. Free energy surface for (A) a 50 Å section of the unbroken cellulose microfibril and (B) a 50 Å section for the broken cellulose microfibril.

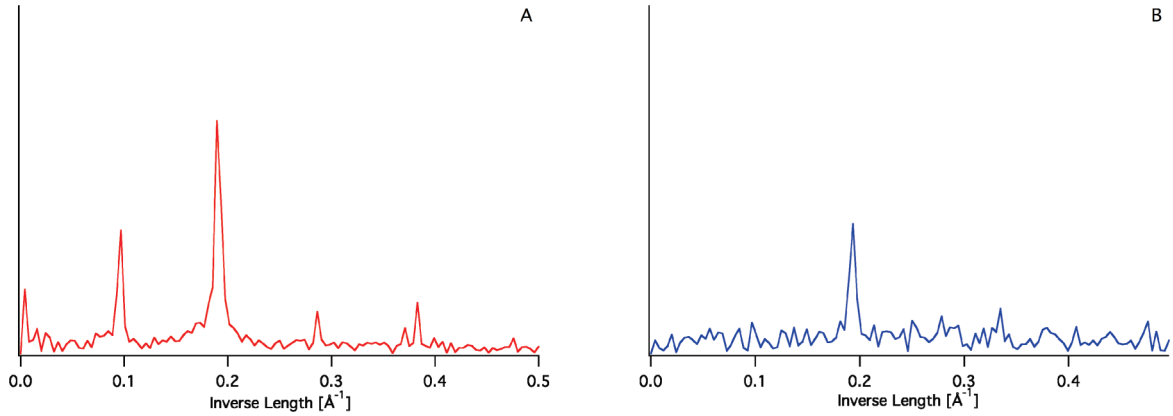

Figure 8. Frequency domain Fourier transforms for (A) the unbroken PES and (B) the unbroken FES. The PES shows significant peaks at 5 and 10 Å, and the FES shows a significant peak every 5 Å.

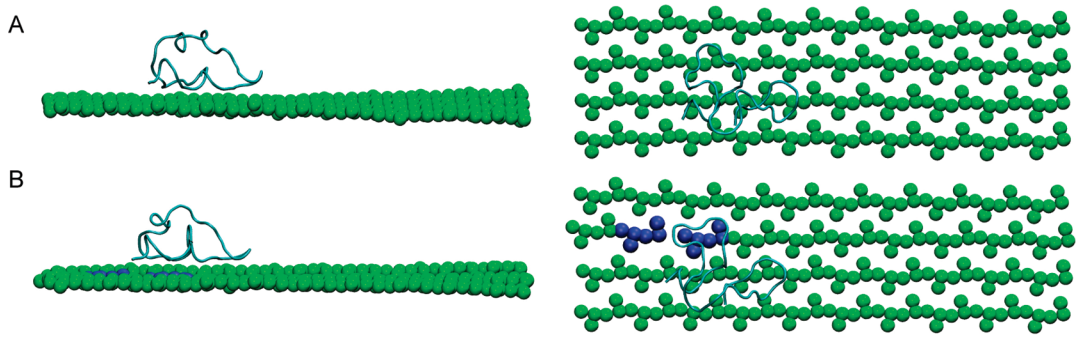

Figure 9. Starting configurations for two sets of MD simulations. The microfibril used in this simulation is approximately 100 Å long. For clarity, only the top layer of cellulose is shown. (A) Unbroken microfibril and (B) the microfibril with the hydrolyzed bond shown in blue space fill.

In addition, we conducted several sets of MD simulations to corroborate and clarify the disparity between the PES (where a driving force for translation away from a broken strand is present) and FES results (where the driving force for translation away from the broken strand is not observed). Two sets of randomly seeded MD simulations consisting of 30 simulations of 20 ns each were collected. To reduce computational costs, a shorter microfibril was used, which was ~100 Å long. The other two dimensions of the microfibril are the same as those used in the PES calculations and FES simulations. Figure 9 shows the configurations for the two MD simulation sets. In Figure 9A, the CBM is on a smooth surface facing the nonreducing end of the microfibril, which is the direction of processivity for the entire Cel7A enzyme, whereas the configuration in Figure 9B is approximately the same setup with a hydrolyzed bond. It should be noted that a single starting configuration for the CBM along the cellulose chain of interest was used for each of the two simulation sets. Initiating the trajectories from different CBM positions on other cellulose chains may affect the results; however, these additional scenarios were not examined because they are outside the scope of the present study. In addition, a third set of MD simulations was conducted to ensure that significant end effects were not seen in the MD simulations with the smaller microfibril; these results are included in the Supporting Information. The center of mass of the CBM was monitored as a function of the trajectory length. The histograms of the initial and final positions for the two simulation configurations (from Figure 9) are shown in Figure 10. In the case of the simulation of the CBM on the unbroken microfibril, the mean center of mass *X*-coordinate for the CBM changes by 2.1 Å after 20 ns. The standard deviation of the CBM center of mass *X*-coordinate is 5.7 Å. The mean and standard deviation

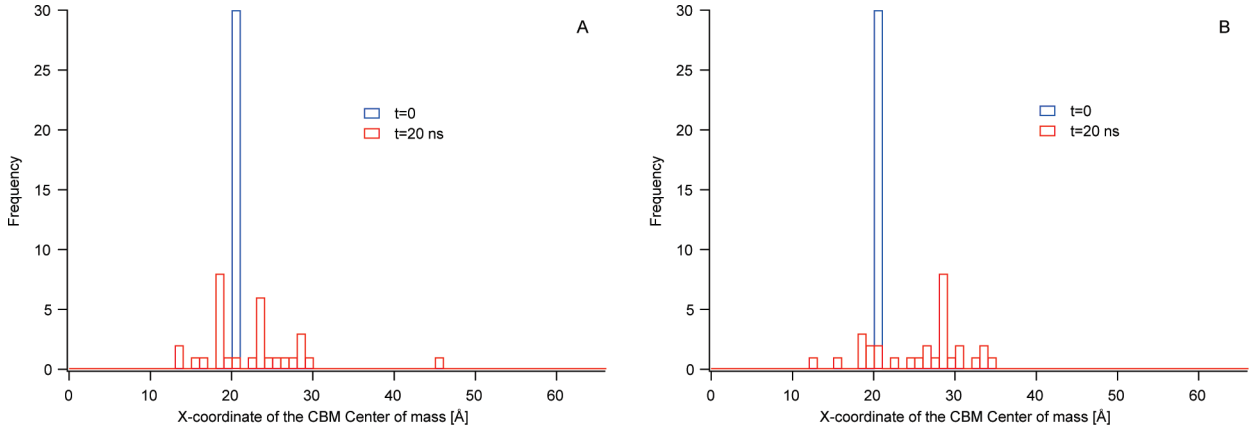

Figure 10. Histograms of the CBM center of mass as a function of time for 2 sets of 30 MD simulations of 20 ns each. (A) Unbroken microfibril with the starting configuration shown in Figure 9A. $\mu(t = 0\ \text{ns}) = 20.5$ Å, $\mu(t = 20\ \text{ns}) = 22.6$ Å, $\sigma(t = 20\ \text{ns}) = 5.7$ Å. The mean is slightly skewed to be higher than the $t = 0$ value because of the single simulation in which the CBM translated 25 Å. (B) Microfibril containing the hydrolyzed bond at the end of the microfibril as shown in Figure 9B. $\mu(t = 0\ \text{ns}) = 20.3$ Å, $\mu(t = 20\ \text{ns}) = 25.6$ Å, $\sigma(t = 20\ \text{ns}) = 6.0$ Å.

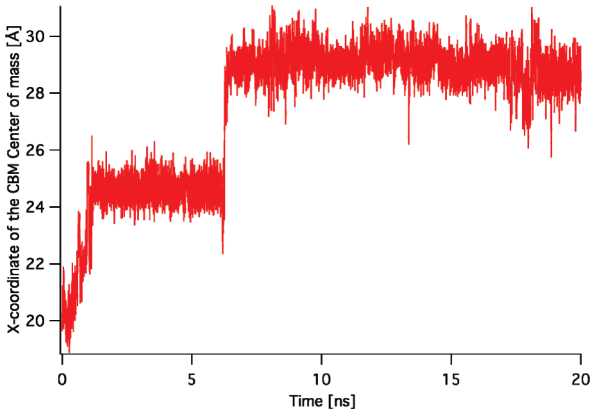

Figure 11. Translation of the CBM along the *X* direction of the cellulose microfibril in a 20 ns MD simulation. The CBM translates twice during this simulation in increments of 5 Å.

are slightly skewed because of the single simulation in which the CBM translated approximately 25 Å. However, as expected with an unbroken microfibril surface and in a system that obeys detailed balance, the CBM is able to translate in both directions. For the surface with the hydrolyzed bond, the CBM diffusion is biased in a single direction, as predicted from the PES results, and the majority of the CBM simulations translate approximately 1 cellobiose unit, as predicted by the PES and FES results. The mean value for the *X*-coordinate of the CBM center of mass for the MD simulation set with the hydrolyzed bond is 25.6 Å after 20 ns, and the standard deviation is 6.0 Å. Using these results, we can clarify the disparity between the PES and the FES by comparing the drift in the mean displacement within the set of 20 ns MD simulations, which is markedly different in these two cases. Specifically, the 1-dimensional diffusion coefficient along the cellulose surface in the cellodextrin chain direction is approximately 1 Å^2/ns, computed from both the unbroken and broken histograms. From the diffusion coefficient and combining data from Figures 10a and 10b, we can estimate the standard deviation (σ) in the displacement on 30 outcomes per case (unbroken and broken surfaces) as approximately 1 Å, or ($\sim 6/N^{1/2}$) Å where $N = 30$ trials. The drift in the mean at $t = 20$ ns for the unbroken case is then approximately 2σ, a large fraction of which is from the single trajectory that translates ~25 Å. The drift in the mean in the case of the broken cellulose surface is about 6σ, which is significantly higher than the expected result from an unbiased, random walk. This result corroborates the observation of a driving force for translation away from a broken strand seen in the PES, but not seen in the FES, most probably because the FES results for this model are at the resolution of typical free energy calculations, as discussed above.

The third MD simulation set, which was an unbroken microfibril with the CBM started at the midpoint of the microfibril confirmed the results shown in Figure 10A, for which no significant edge effects were observed. These results are shown in the Supporting Information. In addition, a single trace for a representative trajectory is shown in Figure 11. As expected from the PES and the FES, the CBM translates in 5 and 10 Å increments with a minimal amount of time spent for transitions between the regions of stability.

The computational results presented in this study show that the CBM from *T. reesei* Cel7A diffuses processively on the hydrophobic surface of cellulose on the length scale of both a

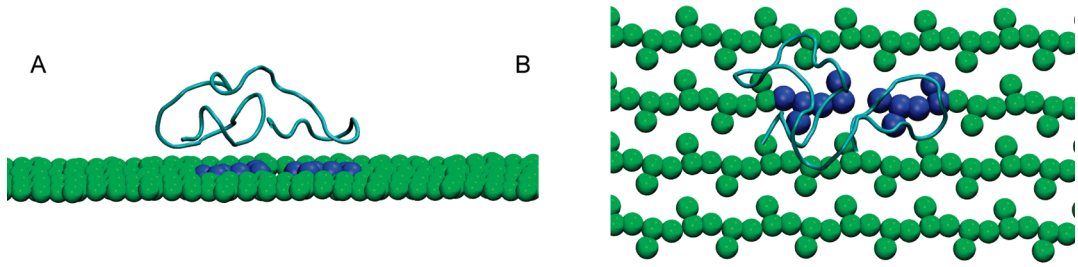

Figure 12. Snapshots from the FES calculations with the hydrolyzed bond highlighted in blue. (A) Side view. (B) Top view in which the hydrolyzed bond is shown in blue space fill under the CBM. Note that the hydrolyzed cellulose chain does not lift out of the cellulose microfibril but that there is a substantial gap across the hydrolyzed bond.

glucose unit and a cellobiose unit. The PES results show that the cellobiose processivity unit is a lower energy state than the glucose unit. This suggests that the CBM alone diffuses on a length scale equivalent to the catalytic cycle of the Cel7A enzyme, which to our knowledge was not known for the CBM before this study. Because this model was developed to examine the long-time behavior and processivity of the CBM, it is unclear if the finer resolution details of the cellulose will play a significant role in the processivity of the CBM or which particular residues play a role in the processivity mechanism. This question will be directly addressed in future work with a more appropriate model for probing atomistic-level interactions between the CBM and the finer resolution features of the cellulose surface.

Another conclusion we can draw from this study is that the CBM not only aids in binding to cellulose, which has been known for many years, but that it can exert a thermodynamic driving force on the Cel7A enzyme during processive cellulose hydrolysis. This is supported by the PES results, in which the CBM exhibits a downhill potential energy landscape when moving away from the hydrolyzed microfibril, and corroborated by the subsequent MD simulations. As shown in Figure 12, the cellulose chain is not decrystallized (i.e., it is not pulled out and it does not come out of the plane of the hydrophobic face of the cellulose microfibril) during the MD or FES simulations or the PES calculations, yet there is still a driving force for the CBM to translate away from the hydrolyzed bond. As the cellulose chain threads into the active site tunnel of the catalytic domain of Cel7A, as must happen during processive cellulose hydrolysis, the cellulose chain will be pulled out of the surface. In this case, where the cellulose strand may be somewhat suspended above the planar surface of the hydrophobic face, we expect that the driving force for the CBM to translate may be even stronger than in the case of a hydrolyzed bond still residing in the plane of the (100) surface. Therefore, we predict that the CBM will aid in processivity of the individual cellulose chain into the tunnel of the Cel7A catalytic domain. In addition, this study shows that the CBM exhibits an energetic barrier to translate over a hydrolyzed bond. This result agrees with an assumption by Wang and co-workers, in which an elegant kinetic model for processive cellulase action was derived and the CBM diffusion direction was biased by a broken chain of cellulose being threaded into a cellulase tunnel.[46]

Conclusions

Using a mixed resolution molecular simulation model, we predict that the CBM from *T. reesei* Cel7A exhibits a thermodynamic driving force to translate away from a hydrolyzed cellulose chain on the hydrophobic face of the cellulose Iβ polymorph. In addition, with this model, the CBM exhibits an energy minimum every 5 Å and a deeper minimum approximately every 10 Å, the latter of which corresponds to the length of cellobiose which is the product of the Cel7A during processive action on cellulose. In the future, we plan to address the mechanism of the translation of the *T. reesei* Cel7A CBM and other CBMs on cellulose surfaces using models of different resolution.

Acknowledgment. The work to develop the coarse-grained cellulose surface model was supported by the National Renewable Energy Laboratory Directed Research & Development program and by the U.S. Department of Energy under contract No. DE-AC36-99GO10337 with the National Renewable Energy Laboratory. The work to model *T. reesei* CBM acting on the cellulose surface model was supported by the DOE Office of the Biomass Program. Computational resources for this research were supported in part by the Golden Energy Computing Organization at the Colorado School of Mines using resources acquired with financial assistance from the National Science Foundation and the National Renewable Energy Laboratory. We thank Professor Baron Peters of UCSB for helpful discussions.

Supporting Information Available: Angle and dihedral parameter fits from the atomistic simulation of cellulose, a CHARMM topology file for the coarse-grained model of the hydrophobic face of cellulose Iβ, and the entire FES along the unbroken microfibril. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Ragauskas, A. J.; Williams, C. K.; Davison, B. H.; Britovsek, G.; Cairney, J.; Eckert, C. A.; Frederick, W. J.; Hallett, J. P.; Leak, D. J.; Liotta, C. L.; Mielenz, J. R.; Murphy, R.; Templer, R.; Tschaplinski, T. *Science* **2006**, *311*, 484–489.

(2) 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.

(3) Himmel, M. E.; Ruth, M. F.; Wyman, C. E. *Curr. Opin. Biotechnol.* **1999**, *10*, 358–364.

(4) Himmel, M. E. ; Picattago, S. K. *Biomass Recalcitrance: Deconstructing the Plant Cell Wall for Bioenergy*; Blackwell Publishing: London, UK, 2008; pp 1−6.

(5) Barr, B. K.; Hsieh, Y. L.; Ganem, B.; Wilson, D. B. *Biochemistry* **1996**, *35*, 586–592.

(6) Vršanská, M.; Biely, P. *Carbohydr. Res.* **1992**, *227*, 19–27.

(7) Srisodsuk, M.; Reinikainen, T.; Penttila, M.; Teeri, T. T. *J. Biol. Chem.* **1993**, *268*, 20756–20761.

(8) Teeri, T. T.; Lehtovaara, P.; Kauppinen, S.; Salovuori, I.; Knowles, J. *Gene* **1987**, *51*, 43–52.

(9) Baker, J. O.; Ehrman, C. I.; Adney, W. S.; Thomas, S. R.; Himmel, M. E. *Appl. Biochem. Biotechnol.* **1998**, *70/72*, 395–403.

(10) Boraston, A. B.; Bolam, D. N.; Gilbert, H. J.; Davies, G. J. *Biochem. J.* **2004**, *382*, 769–781.

(11) Kraulis, J.; Clore, G. M.; Nilges, M.; Jones, T. A.; Pettersson, G.; Knowles, J.; Gronenborn, A. M. *Biochemistry* **1989**, *28*, 7241–7257.

(12) Lehtio, J.; Sugiyama, J.; Gustavsson, M.; Fransson, L.; Linder, M.; Teeri, T. T. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 484–489.

(13) Linder, M.; Lindeberg, G.; Reinikainen, T.; Teeri, T. T.; Pettersson, G. *FEBS Lett.* **1995**, *372*, 96–98.

(14) Linder, M.; Mattinen, M. L.; Kontteli, M.; Lindeberg, G.; Stahlberg, J.; Drakenberg, T.; Reinikainen, T.; Pettersson, G.; Annila, A. *Protein Sci.* **1995**, *4*, 1056–1064.

(15) Linder, M.; Teeri, T. T. *J. Biotechnol.* **1997**, *57*, 15–28.

(16) Reinikainen, T.; Ruohonen, L.; Nevanen, T.; Laaksonen, L.; Kraulis, P.; Jones, T. A.; Knowles, J. K.; Teeri, T. T. *Proteins* **1992**, *14*, 475–482.

(17) 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.

(18) Din, N.; Gilkes, N. R.; Tekant, B.; Miller, R. C.; Warren, R. A. J.; Kilburn, D. G. *Biotechnology* **1991**, *9*, 1096–1099.

(19) Kataeva, I. A.; Seidel, R. D.; Shah, A.; West, L. T.; Li, X. L.; Ljungdahl, L. G. *Appl. Environ. Microbiol.* **2002**, *68*, 4292–4300.

(20) Ayton, G. S.; Noid, W. G.; Voth, G. A. *MRS Bull.* **2007**, *32*, 929–934.

(21) Head-Gordon, T.; Brown, S. *Curr. Opin. Struct. Biol.* **2003**, *13*, 160–167.

(22) Lazaridis, T.; Karplus, M. *Curr. Opin. Struct. Biol.* **2000**, *10*, 139–145.

(23) Tozzini, V. *Curr. Opin. Struct. Biol.* **2005**, *15*, 144–150.

(24) Molinero, V.; Goddard, W. A. *J. Phys. Chem. B* **2004**, *108*, 1414–1427.

(25) Molinero, V.; Goddard, W. A. *Phys. Rev. Lett.* **2005**, *95*, 45701.

(26) Liu, P.; Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2007**, *111*, 11566–11575.

(27) Sild, V.; Stahlberg, J.; Pettersson, G.; Johansson, G. *FEBS Lett.* **1996**, *378*, 51–56.

(28) Nishiyama, Y.; Langan, P.; Chanzy, H. *J. Am. Chem. Soc.* **2002**, *124*, 9074–9082.

(29) Nishiyama, Y.; Sugiyama, J.; Chanzy, H.; Langan, P. *J. Am. Chem. Soc.* **2003**, *125*, 14300–14306.

(30) 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.
(31) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
(32) Kuttel, M.; Brady, J. W.; Naidoo, K. J. *J. Comput. Chem.* **2002**, *23*, 1236–1243.
(33) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
(34) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
(35) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695.
(36) Nose, S. *J. Chem. Phys.* **1984**, *81*, 511.
(37) Bu, L.; Im, W.; Brooks, C. L. *Biophys. J.* **2007**, *92*, 854–863.
(38) Bu, L.; Brooks, C. L. *J. Am. Chem. Soc.* **2008**, *130*, 5384–5385.
(39) Im, W.; Feig, M.; Brooks, C. L. *Biophys. J.* **2003**, *85*, 2900–2918.
(40) Im, W.; Lee, M. S.; Brooks, C. L. *J. Comput. Chem.* **2003**, *24*, 1691–1702.
(41) Noid, W. G.; Chu, J. W.; Ayton, G. S.; Krishna, V.; Izvekov, S.; Voth, G. A.; Das, A.; Andersen, H. C. *J. Chem. Phys.* **2008**, *128*, 11.
(42) Noid, W. G.; Liu, P.; Wang, Y.; Chu, J. W.; Ayton, G. S.; Izvekov, S.; Andersen, H. C.; Voth, G. A. *J. Chem. Phys.* **2008**, *128*, 20.
(43) Peters, B.; Zimmermann, N. E. R.; Beckham, G. T.; Tester, J. W.; Trout, B. L. *J. Am. Chem. Soc.* **2008**, *130*, 17342–17350.
(44) Radhakrishnan, R.; Schlick, T. *J. Chem. Phys.* **2004**, *121*, 2436–2444.
(45) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011.
(46) Ting, C. L.; Makarov, D. E.; Wang, Z. G. *J. Phys. Chem. B* **2009**, *113*, 4970–4977.

JP904003Z