Carbohydrate Research 344 (2009) 1984–1992

Contents lists available at ScienceDirect

Carbohydrate Research

journal homepage: www.elsevier.com/locate/carres

Computational simulations of the *Trichoderma reesei* cellobiohydrolase I acting on microcrystalline cellulose Iβ: the enzyme–substrate complex

Linghao Zhong [a,e], James F. Matthews [a], Peter I. Hansen [a], Michael F. Crowley [b], Joseph M. Cleary [b], Ross C. Walker [c], Mark R. Nimlos [b], Charles L. Brooks III [d], William S. Adney [b], Michael E. Himmel [b], John W. Brady [a,*]

[a] Department of Food Science, Cornell University, Ithaca, NY 14853, United States
[b] National Renewable Energy Laboratory, 1617 Cole Boulevard, Golden, CO 80401-3393, United States
[c] San Diego Supercomputer Center, La Jolla, CA 92093-0505, United States
[d] Department of Chemistry and Biophysics Program, University of Michigan, 930 North University Avenue, Ann Arbor, MI 48109, United States
[e] Pennsylvania State University, Mont Alto, Mont Alto, PA 17237, United States

ARTICLE INFO

Article history:
Received 6 December 2008
Received in revised form 13 July 2009
Accepted 14 July 2009
Available online 18 July 2009

Keywords:
Cellulase
Cellobiohydrolase I
Cellulose
Computer modeling
Molecular dynamics

ABSTRACT

Cellobiohydrolases are the dominant components of the commercially relevant *Trichoderma reesei* cellulase system. Although natural cellulases can totally hydrolyze crystalline cellulose to soluble sugars, the current enzyme loadings and long digestion times required render these enzymes less than cost effective for biomass conversion processes. It is clear that cellobiohydrolases must be improved via protein engineering to reduce processing costs. To better understand cellobiohydrolase function, new simulations have been conducted using CHARMM of cellobiohydrolase I (CBH I) from *T. reesei* interacting with a model segment (cellodextrin) of a cellulose microfibril in which one chain from the substrate has been placed into the active site tunnel mimicking the hypothesized configuration prior to final substrate docking (i.e., the +1 and +2 sites are unoccupied), which is also the structure following a catalytic bond scission. No tendency was found for the protein to dissociate from or translate along the substrate surface during this initial simulation, nor to align with the direction of the cellulose chains. However, a tendency for the decrystallized cellodextrin to partially re-anneal into the cellulose surface hints that the arbitrary starting configuration selected was not ideal.

1. Introduction

Recently we reported the results of molecular dynamics (MD) simulations of cellobiohydrolase I (CBH I) from *Trichoderma reesei* interacting with a model segment of a cellulose microfibril.[1] CBH I consists of three separate domains, a large globular catalytic domain (CD) containing the active site inside a tunnel formed by a loop of the protein chain, a smaller globular carbohydrate binding module (CBM) that docks the enzyme onto the surface of the substrate, and an unstructured linker segment binding the two globular domains together (see Fig. 1). This exocellulolytic enzyme is thought to bind to the hydrophobic faces of crystalline cellulose, and to processively hydrolyze cellulose chains from the reducing end by removing successive cellobiose units. In the previous simulation, the complete enzyme complex was docked onto the center of a model fibril, with no interaction between the substrate chains and the catalytic domain. Thus, the model was designed to examine the binding interaction of the complex with the substrate, but not necessarily to directly mimic any possible mechanism of the processivity of the enzyme.

We report here a new simulation of a similar system, almost identical to the previous one, with the exception that the cellulose chain directly under the complex in the fibril was 'broken', in the sense that it was shortened by 11 residues, to allow the chain to be pulled up from the crystalline fibril and to allow the terminal residues of the chain to be fed into the active site tunnel. This system more closely models the initial stage of processive cellulase activity. The cellulose chain does not span the catalytic site, but is positioned as it may reside after a hydrolysis event allows a product cellobiose to exit the catalytic tunnel. The simulation can be used to determine whether such a positioning of the complex is energetically favorable, or whether on a short timescale the system relaxes toward some alternate, more favored arrangement. While the advance of the chain into the active site, with one cellobiose unit occupying the +1 and +2 sugar binding sites beyond the scissile glycosidic bond, is likely to be an activated process requiring the surmounting of an energy barrier that would probably be unlikely on the timescale of a simple simulation, it might nevertheless be instructive to examine whether or not there appears to be

* Corresponding author. Tel.: +1 (607) 255 2897.
E-mail address: jwb7@cornell.edu (J.W. Brady).

0008-6215/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.carres.2009.07.005

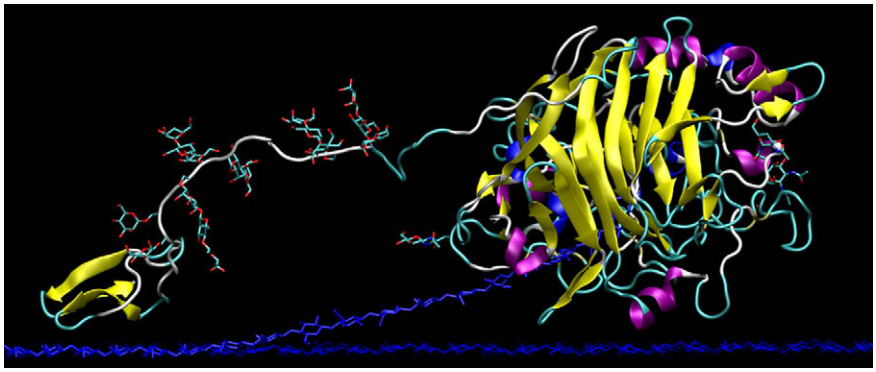

Figure 1. Side view of the complex of CBH I docked onto the (1 0 0) surface of the 108-chain model cellulose microfibril. For clarity, only the chains of the top layer of the cellulose substrate are indicated in blue, but with the chain from this layer bound into the active site tunnel shown in orange. Secondary structural elements in the globular domains are indicated using Richardson representations (purple: alpha helices; yellow: beta sheets).

any tendency for the chain to move in that direction. Furthermore, it has been suggested that the heavy glycosylation of the linker chain between the two globular domains might bind so much water, and thus inhibit and restrict their mobility, as to create a gel-like zone between these modules, inhibiting their relative motions and tending to push the CBM forward along the chain, away from the heavier CD, thus promoting processivity. This paper reports the results of the first 4 ns of MD simulation of this system in terms of the implications for these and other hypotheses about how the enzyme functions.

2. Methods

The simulation reported here was performed in a manner very similar to the previous simulation of this enzyme on a cellulose microfibril.[1] As before, the model cellulose Iβ microfibril contained 108 individual cellulose chains, with all but one having 40 glucose units. The principal difference in the present simulation was that one of the central four chains of the hydrophobic 1 0 0 face was shortened by 11 residues, and the reducing end of this dangling chain was manually placed into the active site tunnel of the catalytic domain. As in the previous simulation, the coordinates for the catalytic domain were taken from the reported crystal structure (7cel) with a cellohexose chain bound in the active site tunnel, and the coordinates of the CBM were taken from the reported NMR structure; for both of these globular modules, the actual coordinates were obtained from the Protein Data Bank.[2–4] The approximate initial distance between the center of mass of the catalytic and binding domains was 79 Å. In the crystallographic structure for the CD of CBH I, there are six substrate glucose residues in the −1 to −6 binding sites, and a cellobiose product molecule in the +1 and +2 sites. The crystallographic coordinates for the cellohexose chain were used as the terminal six-residue portion of the 29-residue substrate chain pulled up from the fiber surface. Another eight sugar residues were used to bridge between the point where these six residues emerge from the catalytic tunnel and the point where the remaining 15 sugar residues began, which were taken to remain 'flat' in the surface of the reported crystal structure of cellulose Iβ. Figure 1 displays a view of the starting conformation of the protein bound to the top layer of the substrate fibril.

As in the previous study, the glycosylating oligosaccharides of the linker domain were taken to be unbranched β(1→4)-linked D-mannose oligomers ranging from 1 to 3 monomers in length.[5] The CHARMM molecular mechanics program[6] was used to model the system using the CHARMM protein force field parameters[7] for the CBH I and previously reported CHARMM parameters for the carbohydrates.[8] The CHARMM program was required because other molecular mechanics programs are not currently able to handle the CMAP components of the CHARMM force field for the protein.[9] The enzyme–fibril complex was placed in a box of TIP3P water molecules[10] with dimensions 279.9 Å × 202.2 Å × 124.4 Å, and those water molecules that overlapped with protein or carbohydrate atoms were discarded. To produce a neutral system, 28 Na^+ and 11 Cl^- ions were introduced by transforming the water molecules closest to the charged amino acid residues into the appropriate counterion. The equations of motion were integrated in the NVE ensemble using a Verlet integrator scheme with a step size of 2 fs, and with long range electrostatic interactions treated using the particle mesh Ewald method.[11] Covalent bond lengths involving hydrogen atoms were kept fixed using the SHAKE algorithm.[12] The system was heated to 300 K over 30 ps and then further integrated for a total of 4 ns.

3. Results and discussion

The binding domain is presumably important in the mechanism of processivity of the CBH I enzyme because experimental evidence has shown that removal of this small domain abolishes activity on crystalline cellulose.[13–15] The exact function of binding domains in cellulases is still the subject of considerable debate.[16,17] It is possible that the only function is to prevent the catalytic domain from diffusing away from the surface of the substrate, thus increasing the local concentration. However, it has also been suggested that the linker segment might store energy, in the manner of a compressed spring, perhaps forcing the chain further into the active site after each bond scission and product escape, or driving the CBM to advance along the substrate chain, pulling the CD along after it.[18] Such a mechanism would presumably require the affinity of the chain terminus for the unoccupied +1 and +2 binding sites to be very strong, and that once bound, this binding would necessarily pull the two domains closer together. It might also be hypothesized that such a shortening of the inter-domain separation might store energy in the linker like a compressed spring, which could then drive the lighter CBM further down the substrate chain, promoting processivity.

To test this hypothesis, it would be of interest to monitor both the positioning of the substrate chain in the active site tunnel and the separation distance between the two domains as a function of time. Figure 2 displays the history of the component of the separation vector along the direction of the substrate chains. As can be seen from this figure, the separation distance fluctuates

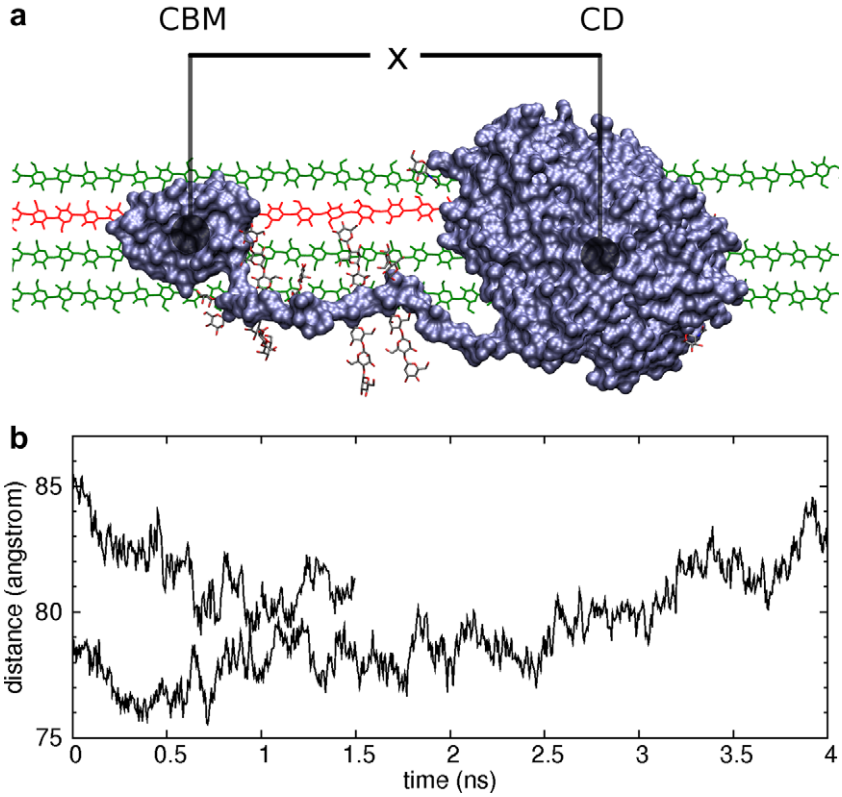

Figure 2. The *x*-axis component of the distance along the cellulose fibril direction (defined in (a)), between the centers of mass of the cellulose binding module and the catalytic domain over the first 4 ns of simulation time (b). Also shown in (b) is the same distance from the earlier (1.5 ns) simulation[1] of this protein without a substrate chain in the active site tunnel. The *x*-axis is chosen because it is aligned with the direction of the cellulose chains.

considerably on the timescale of the simulation, but gradually increases by about 2–4 Å over 4 ns. This would indeed suggest that there is a tendency for the globular domains to remain apart that would favor proccessivity. This observation is supported by comparing to the same distance as calculated from the previous simulation, with no substrate chain in the active site tunnel, which is also shown in Figure 2. Unfortunately, it is difficult to know whether this difference in trends is significant, however, as both simulations are too short to be adequately converged.

Also we observed two associated changes in the substrate chain as well. The first of these was a partial annealing of a portion of the chain back into the groove in the surface formerly occupied by this chain before it was pulled up and placed into the active site tunnel. Figure 3 illustrates this process. In the initial geometry, constructed manually, the approach of the chain to the surface was made very gradual, in the belief that such smaller perturbations would be more favorable energetically. This gradual approach can be seen both in Figure 1 and in the upper panel of Figure 3. During the course of the first 3 ns of the simulation, four residues of the chain annealed back into their original positions. All four of these residues reformed their original crystal structure hydrogen bonds upon settling back into their places in the groove in the surface of the crystal. Two more residues remained essentially parallel to the flat orientations of the other sugar rings of the surface, but remained slightly above their sites in the crystal. This more conventional conformation for all of these residues resulted in a much larger conformational shift in the first residue that pulled away from the surface, as can be seen in the lower panel of Figure 3. These next two residues rotated with respect to the orientations of the other glucose rings in the chain. The first residue to rotate significantly away from the orientations of the other residues, assumed (φ,ψ) values of approximately (+20, −40) and the next, one sugar closer to the CD, adopted the approximate conformation (+20, −40). Closer to the CD, however, the rings once again have approximately the same (φ,ψ) orientations as those in the surface.

The other important change in the substrate chain that occurred in this simulation was that it was partially pulled out of the active site tunnel as these monomer units in the chain annealed back into their crystal sites. Figure 4 illustrates this binding shift and partial substrate escape. Thus, during the course of this simulation, there was no tendency observed for the substrate chain to be spontaneously drawn back into the active site until the point that the +1 and +2 sites were occupied; in fact, they did not even remain in their initial positions.

As the substrate annealed back into the groove left when it was pulled up, it was found to displace the water molecules that had filled the positions previously occupied by these four glucose residues, which can be seen in Figure 5. These water molecules did not prevent the annealing process, apart from slowing the kinetics, presumably because they have a less favorable interaction free energy in the vacant groove than does the polysaccharide.

During the course of the simulation, the CBM showed no tendency to dissociate from the cellulose surface, but did move about slightly on the surface. This diffusion took place as a series of 'site-hops' between relatively stable positions on the substrate, relative to the cellulose chains. Figure 6 illustrates the motions of the CBM during the course of the simulation. The first four nanoseconds of this trajectory can be divided into three nearly equal segments corresponding to the residence of the CBM in three such successive stable positions. Whereas the simulation was initiated with the CBM set at an angle with respect to the direction of the chains, there was no tendency observed here for the CBM to align its three contact-surface tyrosine residues along this parallel direction. In fact, as the simulation proceeded, the CBM actually rotated on

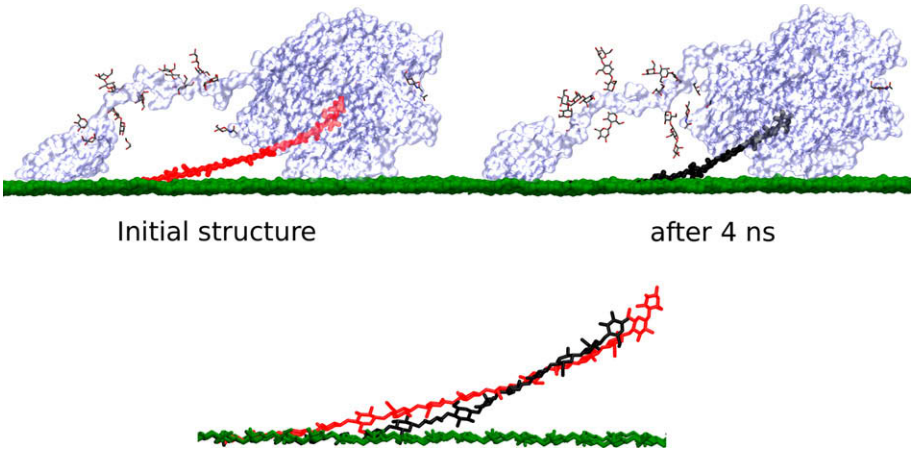

Figure 3. Top: left, the initial conformation of the substrate chain (shown in red); right, the final conformation (shown in black). Bottom: A closeup of the initial chain conformation, in red, overlain with the final conformation, in black.

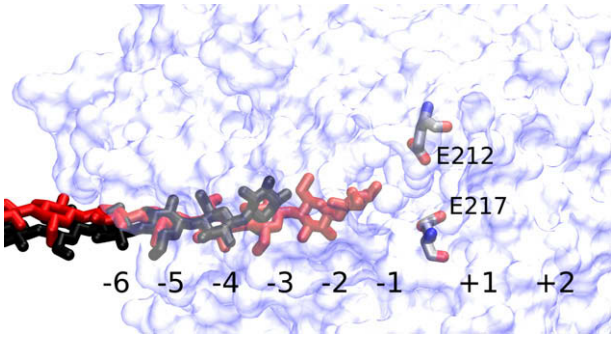

Figure 4. The initial (red) and final (black) positions of the reducing end of the substrate in the active site tunnel, with the catalytic residues 212 and 217 at the scission point indicated.

the surface such that the approximate line of these residues was almost perpendicular to the cellulose chain alignment. This change in orientation can be seen in Figure 6 in the panel on the upper right. Also, as the CBM moved about on the surface, the center of mass of the domain, indicated by small colored spheres in the upper right panel of Figure 6, was forced to move away from the surface of the crystal (trajectory segment shown in green) as the complex moved up over the C3 hydroxyl group of the lower chain shown in the upper left panel of Figure 6. When the CBM subsequently 'hopped back' to another stable position, shown in blue, the distance decreased again. The difference in the separation distances involved in these motions is small, changing by only a little more than 1 Å.

The degree of rotation of the CBM on the surface can be approximately quantified by defining a vector in the protein framework and calculating the angle that it makes with the direction of the chains in the crystalline substrate. Because of the approximate alignment of tyrosine residues along the binding surface of the CBM, these residues make a reasonable choice to define the internal vector, using the C^{α} positions of the Tyr 466 and Tyr 492 as shown in Figure 7. The approximate alignment of the substrate chains can be defined using the glycosidic linkage oxygen atoms of the chain. As shown in Figure 7, the choice used here was the O1 atoms of two cellobiose units separated by an intervening cellobiose. The projection of this angle onto the substrate surface plane was then used to follow the rotation of the CBM on the surface. The evolution of the orientation of the CBM defined using this angle is shown in Figure 8.

Although this definition of the orientation angle is somewhat arbitrary, it captures the qualitative description of the orientational angle. As can be seen from Figure 8, the CBM did not rotate significantly during the first three nanoseconds of the simulation, and then began to rotate fairly rapidly during the period from three nanoseconds to four nanoseconds into the simulation. It is not clear if there is any significance to this rotation, but it should be noted that it roughly corresponded to the period during which the two globular domains moved apart (Fig. 2b).

A previous study of the CBM alone bound to a flat 1 0 0 cellulose Iβ surface found that the loop containing the fourth tyrosine

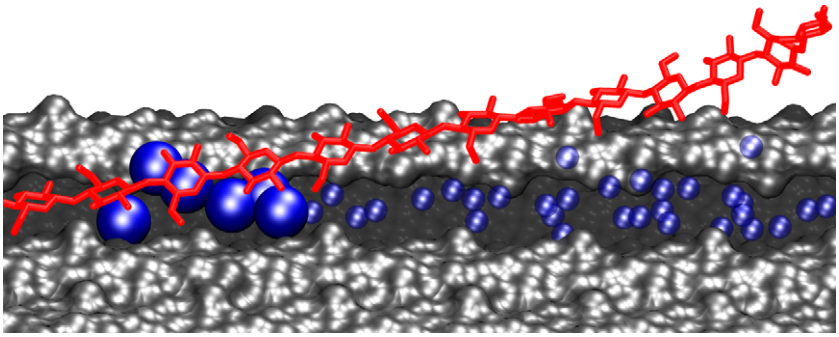

Figure 5. Blue spheres represent water molecules in the groove from which the cellodextrin chain was pulled up. The large blue spheres represent the water molecules which are displaced when four glucose monomers reanneal.

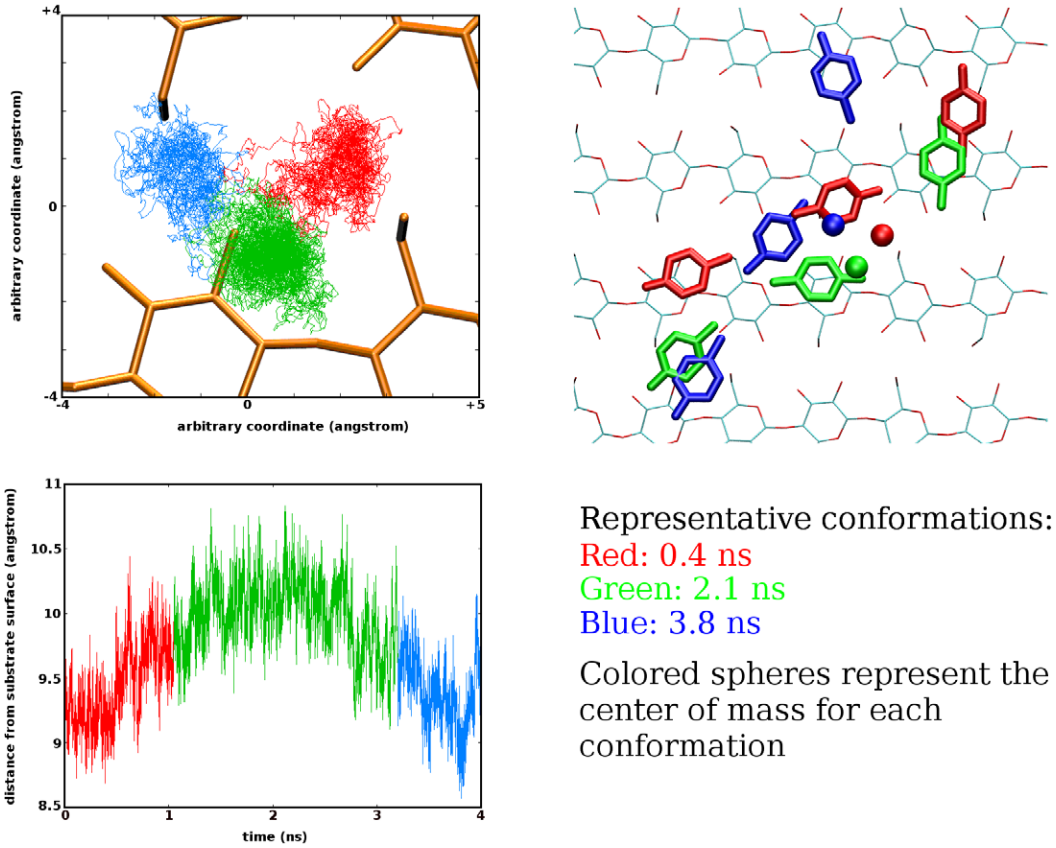

Figure 6. The diffusive motions of the binding domain over the (1 0 0) surface of crystalline Iβ cellulose. In the upper right panel, the three contact-surface tyrosine residues are shown at three different times during the simulation, along with the center of mass of the CBM, indicated by a small sphere. The positions of each of these residues are indicated at 0.4, 2.1, and 3.8 ns, using red, green, and blue, respectively. The CBM was initially oriented at an angle with respect to the direction of the chains to avoid arbitrary bias. During the simulation, this angle actually increased, until it was nearly perpendicular to the direction of the chains. The diffusive motion on the surface occurred as hops between three distinct positions, indicated by the same color scheme, as seen in the upper left. In the intermediate position, the central tyrosine residue was directly over one of the glucose residues, which forced the CBM to move further away from the surface, as can be seen in the lower left panel, which shows the distance of the center of mass of the CBM from the surface.

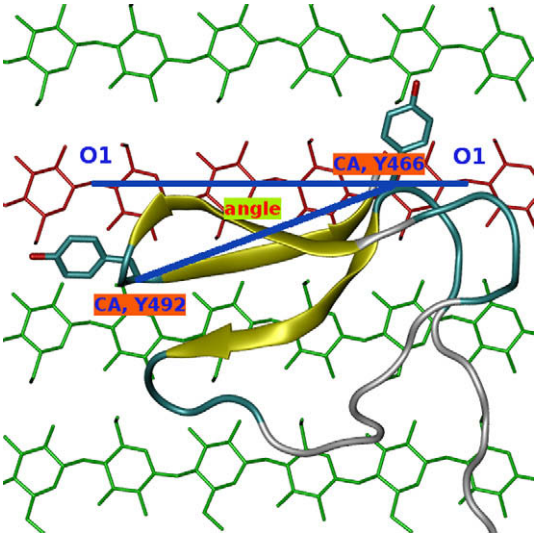

Figure 7. The definition of the vectors in the CBM and the crystalline substrate used to define the rotation angle of the CBM on the surface (see text).

residue, Tyr 474, underwent a significant conformational change that brought that residue down into contact with the cellulose surface, where it could establish a hydrophobic interaction with the cellulose surface.[19] No such transition took place in the present simulation. Figure 9 shows the superposition of the backbone of the CBM at 4 ns with the initial structure of this domain. Apart from minor side chain rotations, there were no significant conformational changes, and the backbone traces in particular are quite similar. The reason for this difference in behavior is possibly due to the briefness of the present simulation. The reported conformational change apparently requires the surmounting of a significant barrier and occurred in approximately 20% of the observed simulations on the timescale of this simulation (tens of nanoseconds; M. Nimlos, unpublished results).

As might be expected from the fact that the cellulose substrate chain partially pulls out of the active site tunnel, the linker segment showed no indication of being under significant tension or to be loaded with conformational free energy. Its conformation fluctuated significantly and apparently freely, as in the previous simulation.[1] It did end up in a somewhat 'more elevated' overall position at the end of the simulation (Fig. 10), without however bringing the globular domains closer together; in fact, they are slightly farther apart at the end of the simulation (see Fig. 2).

Previous studies have found that amino acids and sugars are able to impose considerable local structuring on the water molecules in their first hydration shells,[20,21] and that this structuring extends to a lesser extent to at least second solvation shells as well.[21] This structuring reduces both the translational and rotational freedom of the water molecules involved relative to the mobility experienced by pure bulk water. As can be appreciated from Figure 1, a significant portion of the water between the two globular domains is in the first or second hydration shell of either

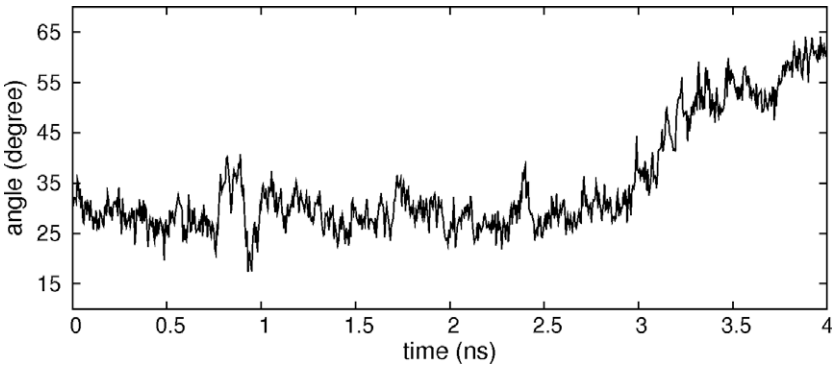

Figure 8. The evolution during the simulation of the vector defined in Figure 7 defining the orientation of the CBM on the substrate surface.

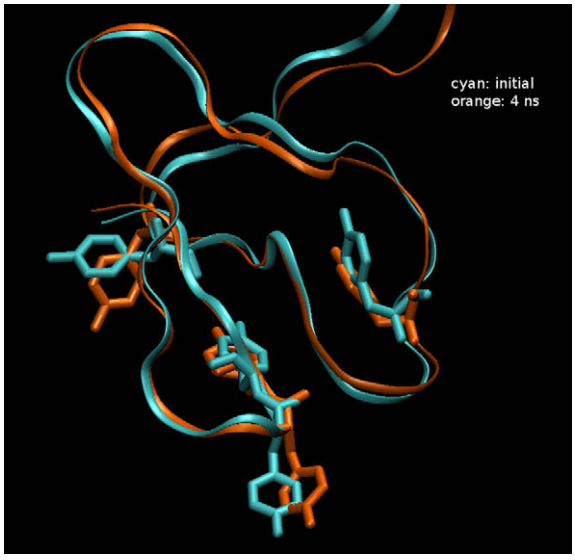

Figure 9. An overlay of the initial backbone trace for the CBM, shown in cyan, with that of the backbone trace at 4 ns into the simulation. The positions of the four tyrosine side chains at these two times are also displayed. As can be seen, virtually no changes occurred in the backbone conformation of the CBM over this period, and while the tyrosine side chains underwent small local rotations, there was no major conformational change involving the fourth residue as has been observed in previous simulations.

the globular domain surfaces themselves, the linker segment and its side chains, the glycosylated oligosaccharides, the cellulose surface, or the substrate chain being hydrolyzed. For this reason, it was hypothesized that the lessened mobility of these water molecules created a buffer region between the two globular domains with a gel-like character that would serve to keep these two domains apart.

In order to test for this possibility, the dynamics of these water molecules was examined in detail. Figure 11 displays the rotational correlation functions for water molecules in this buffer region between the two globular domains, compared to that for bulk water far from any other part of the system. For the bulk water molecules, as can be seen from the log plot in Figure 11b, the relaxation can be described as a single exponential decay with a relaxation correlation time of 1.75 ps. The region between the two domains was arbitrarily described as a box with walls tangent to the contact surfaces of the two globular domains in one direction, and with two other walls defined as parallel to the direction of the substrate chains and tangent to the excursions of the linker domain. The cellulose surface of course provided the floor of this region, and the upper bound was defined by the presence of the linker chain. A small spherical volume taken in the center of this region, shown in green in Figure 11 had relaxation behavior very similar to the bulk water, with a correlation time τ only slightly different from that of the bulk water (1.92 ps). For those water molecules adjacent to the linker chain and its attached sugars, the relaxation is more complicated, and cannot be described by a single exponential decay, as their motions are necessarily strongly coupled to those of the glycopeptide, and their correlation times are correspondingly longer. For the first shell water molecules (i.e., those in monolayer contact with the atoms of the linker or its sugars), the relaxation time was 3.55 ps. As can be seen from Figure 11, even those in the second solvation layer are substantially restricted in their rotational motions, with a correlation time of 2.36 ps.

This dynamical restriction does not seem to affect the relative motions of the two domains, however, as can be seen from Figure 12, which displays the distribution of displacement step sizes in each direction for both domains. In this figure, the *x*-axis steps are aligned along the substrate chain directions, and correspond to displacements that would bring the two globular domains closer together or carry them further apart. As can be seen, the distribution of these steps is essentially normal within the statistics of the simulation, while a sponge-like or gel-like layer intervening between the two globular domains would presumably favor steps that carried them further apart. Thus, there is little indication in these limited data to suggest that the heavy glycosylation of the linker results in sufficient differences in the fluidity of the region between the two globular domains to play a large role in the processivity of the enzyme by directing diffusion or pushing the binding domain further along the direction of the chain.

4. Conclusions

The exact role of the substrate binding module of cellulases has been much discussed but still remains incompletely understood. It has been argued that one possible role of the CBM is either to disrupt the crystalline structure of the cellulose or to facilitate the removal of the chain being hydrolyzed from the cellulose surface. During the course of this short simulation, no indications of these functions were observed, which does not necessarily indicate that no such functions would be found on longer simulations. In addition, as can be seen in Figure 3, the portions of the chain close to the catalytic domain actually became more regular and annealed back into the empty groove of the substrate surface, showing little if any influence from the CBM.

The simulation is far too short to make definitive statements about the alignment of the CBM on the cellulose surface, but during this and previous simulations of the CBM of this protein the motions of the CBM appear to be relatively undirected random diffusion. It is interesting, however, that neither the present simula-

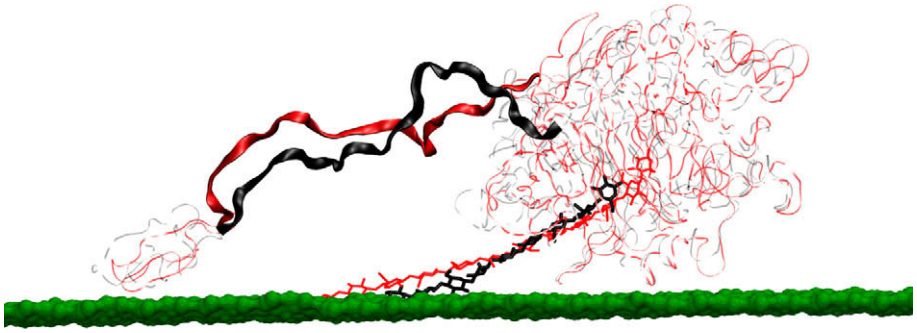

Figure 10. An overlay of the initial (red) and final (grey/black) backbone trace diagrams, showing how the linker domain fluctuates upward, away from the substrate surface.

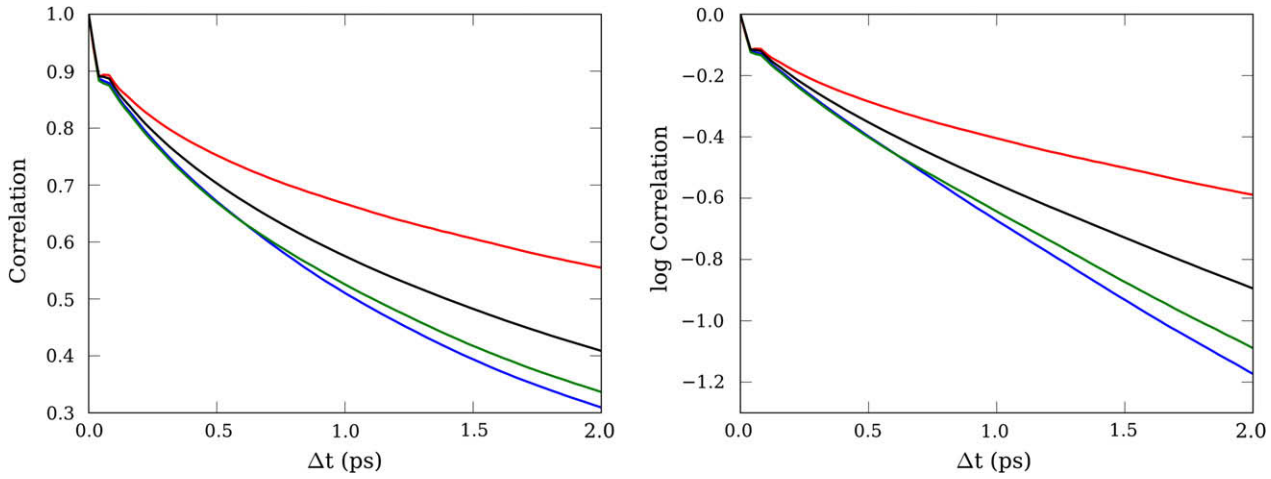

Figure 11. Left, the rotational correlation functions for water molecules in the region between the two globular domains; right, the log of the correlation functions. Color scheme: blue; bulk water molecules outside of this region; green, inside a 10 Å sphere in the middle of the box region; red, water molecules in the first hydration shell of the linker or the attached oligosaccharides; black, water molecules in the second hydration shells.

tion nor the previous simulation of this protein found any tendency for the CBM to dissociate from the cellulose surface. Throughout both simulations, the complex remained tightly bound to the substrate. The role of any possible conformational change in the CBM remains unclear. In a number of repeats of the previous simulation of just the CBM docked onto a regular surface, the conformational transition that brought the loop containing Tyr 474 into contact with the surface occurred in about 20% of the simulations of this length, suggesting a significant barrier to the transition that would require long simulation times to overcome.

The present simulation was not designed to directly test the hypothesis that energy is stored in the linker domain by stresses that arise from the binding of the substrate into the +1 and +2 binding sites prior to bond scission, because the system modeled represents the complex following the catalytic event and product escape. As such, it will serve as a baseline for comparison for the next stage of the simulation of this process, in which the substrate is fully placed into the active site, prior to bond scission. It has been suggested that such binding induces stress in the bound chain, inevitably raising the energy of the complex, which might well be absorbed by the linker domain, driving it further along the substrate, with any residual stress released upon bond scission. Even in the present simulation, the distance between the globular domains shortened slightly as the linker fluctuated upward before generally increasing by about 5–6 Å, or roughly the length of one glucose unit. While it is intriguing that the inter-domain distance increased in the present simulation with a substrate chain in the active site tunnel, while in the previous simulation with no substrate, the domains move closer together, caution is warranted in interpreting these results given the short duration of the simulations (and particularly the earlier calculation). It is entirely possible that this difference is simply the result of statistical accident in a short random walk. In particular, the data in Figure 12 indicate that the motions of the CBM approximate a random walk.

There is also little evidence from the analysis of the relaxational behavior of the water molecules localized between the two globular domains that this region has any pronounced gel-like behavior that might couple to the motions of the domains and keep them apart. Because there is some localization of these water molecules, and some measurable restriction on their dynamics, as seen in Figure 11, it is possible that this contributes to the tendency for the domains to stay apart, as seen in Figure 2b. If further trajectory averaging as these simulations are extended fails to find a significant effect of this sort, it should be possible to greatly simplify future simulations by eliminating the majority of the explicit water molecules, beyond a monolayer coverage, and replacing their effects with a continuum solvation model.[22–24]

The present simulation obviously contained many limitations and arbitrary approximations beyond its unfortunately brief duration. As was seen, the initial arbitrary placement of the substrate chain apparently was sub-optimal, and the same may be true for other parts of the complex. For example, the orientation of the catalytic domain relative to the substrate surface was also an arbitrary choice, and it is quite possible that the real system has another

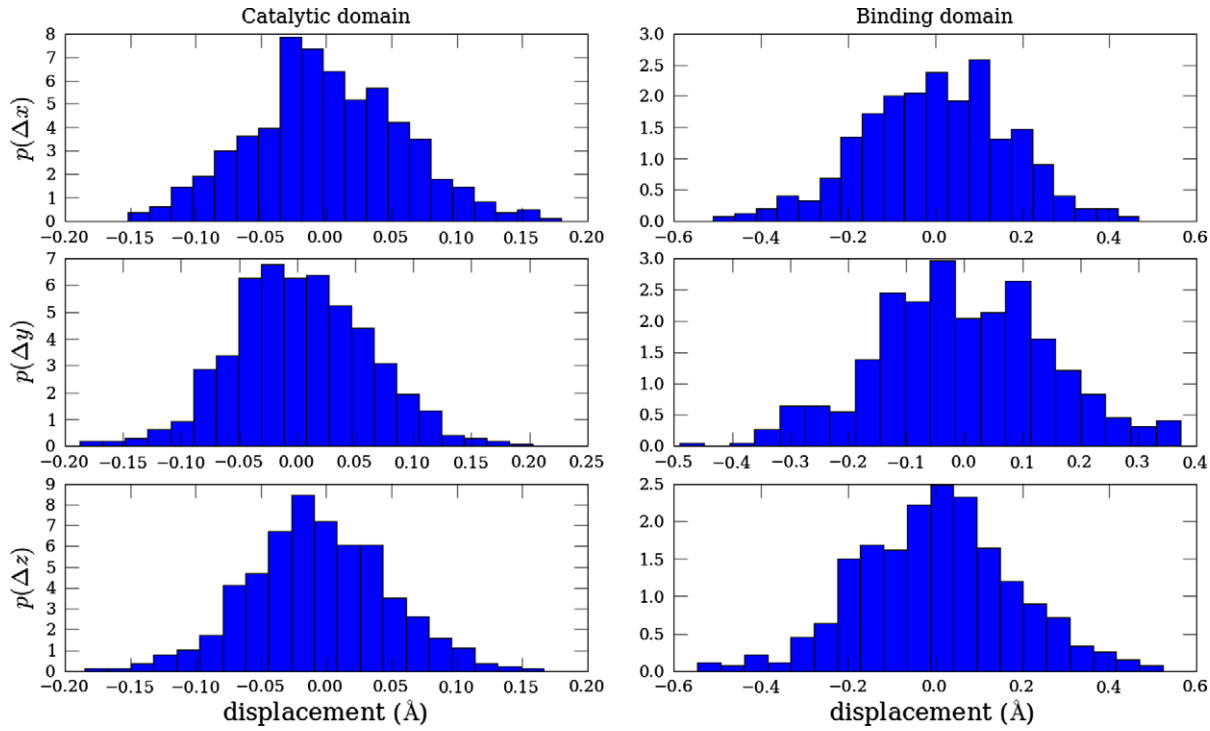

Figure 12. The distribution of steps of different sizes in both directions for *x* (top, the direction of the chain alignment in the substrate), *y* (middle), and *z* (bottom) for a short 0.5-ns interval of the simulation at the end of the trajectory (note different scales and bin sizes).

orientation, perhaps with the entrance of the catalytic tunnel more perpendicular to the surface. Further, while evidence indicates that binding takes place on the hydrophobic surfaces of fibrils,[25] a chain from the middle of such a surface may be an unlikely target for attack. Given the curvature of the microfibrils and the radius accessible to the CD with the long and floppy linker domain, it is entirely possible that a chain from another nearby crystal face is actually attacked. For small microfibrils, chains on the many exposed corners where faces terminate would seem like more favorable targets, with their greater exposure to solvent and fewer contacts with other crystal fibers.

Another important set of assumptions in these simulations was of course the choice of force fields used to model the various components of the complex. In general, the results of molecular mechanics simulations of proteins have been found not to be extremely sensitive to many of the details of the force fields used. However, the same is not necessarily true of cellulose, where the various available carbohydrate force fields produce different cellulose crystal structures.[26,27] This is particularly true with respect to the twist of the cellulose microfibrils, which has been observed experimentally,[28,29] but which is not reported in the crystal structure.[30] Unfortunately, it is not clear to what extent any artifacts arising from the choice of carbohydrate force field might affect the observed motions of the protein domains, or indeed even which force field might give the most realistic representation of actual cellulose.

It may well be the case that simulations of CBH I acting on cellulose must be conducted over much longer sampling times to demonstrate biologically relevant function. The primary impediment to this objective is the development of MD codes that scale well beyond the current limitations of CHARMM and AMBER, but which will still offer the unique carbohydrate and protein modeling capabilities available using these programs. It is likely, however, that the enzyme and substrate model presented in this study, although perhaps not a precise mimic of the actual biological complex, can serve as an ideal starting point for such future, long duration simulations.

Acknowledgments

This work was supported under the SciDAC Program by the U.S. DOE Office of Biological and Environmental Research and the Office of Advanced Scientific Computing Research. The authors also thank R. H. Atalla, J. Sugiyama, G. T. Beckham, and D. B. Wilson, for helpful discussions.

References

1. Zhong, L.; Matthews, J. F.; Crowley, M. F.; Rignall, T.; Talón, C.; Cleary, J. M.; Walker, R. C.; Chukkapalli, G.; McCabe, C.; Nimlos, M. R.; Brooks, C. L., III; Himmel, M. E.; Brady, J. W. *Cellulose* **2008**, *15*, 261–273.
2. Divne, C.; Stahlberg, J.; Teeri, T. T.; Jones, T. A. *J. Mol. Biol.* **1998**, *275*, 309–325.
3. Kraulis, P. J.; Clore, G. M.; Nilges, M.; Jones, T. A.; Pettersson, G.; Knowles, J.; Gronenborn, A. M. *Biochemistry* **1989**, *28*, 7241–7257.
4. Berman, H. M.; Battistuz, T.; Bhat, T. N.; Bluhm, W. F.; Bourne, P. E.; Burkhardt, K.; Feng, Z.; Gilliland, G. L.; Iype, L.; Jain, S.; Fagan, P.; Marvin, J.; Padilla, D.; Ravichandran, V.; Schneider, B.; Thanki, N.; Weissig, H.; Westbrook, J. D.; Zardecki, C. *Acta Crystallogr., Sect. D* **2002**, *58*, 899–907.
5. Nevalainen, H.; Harrison, M.; Jardine, D.; Zachara, N. E.; Paloheimo, M.; Suominen, P.; Gooley, A. A.; Packer, N. H. In *TRICEL 97 Conference: Carbohydrates from Trichoderma reesei and Other Microorganisms*; The Royal Society of Chemistry: Cambridge: UK; Ghent, Belgium, 1997.
6. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
7. 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.
8. Kuttel, M.; Brady, J. W.; Naidoo, K. J. *J. Comput. Chem.* **2002**, *23*, 1236–1243.
9. MacKerell, A. D.; Feig, M.; Brooks, C. L. *J. Am. Chem. Soc.* **2004**, *126*, 698–699.
10. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
11. Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
12. van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Phys.* **1977**, *34*, 1311–1327.
13. Gilkes, N. R.; Warren, R. A. J.; Miller, R. C.; Kilburn, D. G. *J. Biol. Chem.* **1988**, *263*, 10401–10407.
14. Reinikainen, T.; Ruohonen, L.; Nevanen, T.; Laaksonen, L.; Kraulis, P.; Jones, T. A.; Knowles, J. K. C.; Teeri, T. *Protein: Struct. Funct. Gen.* **1992**, *14*, 475–482.
15. Kruus, K.; Lua, A. C.; Demain, A. L.; Wu, J. H. D. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 9254–9258.

16. Kuutti, L.; Laaksonen, L.; Teeri, T. T. *J. Chim. Phys. Phys. Chim. Biol.* **1991**, *88*, 2663–2667.
17. Linder, M.; Teeri, T. T. *J. Biotechnol.* **1997**, *57*, 15–28.
18. Zhao, X.; Rignall, T. R.; McCabe, C.; Adney, W. S.; Himmel, M. E. *Chem. Phys. Lett.* **2008**, *460*, 284–288.
19. 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.
20. Schmidt, R. K.; Karplus, M.; Brady, J. W. *J. Am. Chem. Soc.* **1996**, *118*, 541–546.
21. Liu, Q.; Brady, J. W. *J. Am. Chem. Soc.* **1996**, *118*, 12276–12286.
22. Marrone, T. J.; Gilson, M. K.; McCammon, J. A. *J. Phys. Chem.* **1996**, *100*, 1439–1441.
23. Resat, H.; Marrone, T. J.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 522–532.
24. Smart, J. L.; Marrone, T. J.; McCammon, J. A. *J. Comput. Chem.* **1997**, *18*, 1750–1759.
25. Lehtiö, J.; Sugiyama, J.; Gustavsson, M.; Fransson, L.; Linder, M.; Teeri, T. T. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 484–489.
26. Nishiyama, Y.; Johnson, G. P.; French, A. D.; Forsyth, V. T.; Langan, P. *Biomacromolecules* **2008**, *9*, 3133–3140.
27. 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.
28. Hanley, S. J.; Revol, J.-F.; Godbout, L.; Gray, D. G. *Cellulose* **1997**, *4*, 209–220.
29. Hirai, A.; Tsuji, M.; Horii, F. *Senri Gakkai shi.* **1998**, *54*, 506–510.
30. Nishiyama, Y.; Langan, P.; Chanzy, H. *J. Am. Chem. Soc.* **2002**, *124*, 9074–9082.