Interactions of the complete cellobiohydrolase I from *Trichodera reesei* with microcrystalline cellulose I*ß*

**Linghao Zhong · James F. Matthews · Michael F. Crowley · Tauna Rignall ·
César Talón · Joseph M. Cleary · Ross C. Walker · Giridhar Chukkapalli ·
Clare McCabe · Mark R. Nimlos · Charles L. Brooks III · Michael E. Himmel ·
John W. Brady**

Abstract We describe the construction of a model complex of the cellobiohydrolase I (CBH I) cellulase from *Trichoderma reesei* bound to a cellulose microfibril in an aqueous environment for use in molecular dynamics (MD) simulations. Preliminary characterization from the initial phases of an MD simulation of this complex is also described. The linker sequence between the two globular domains was found to be quite flexible, and the oligosaccharides bound to this linker were found to prefer to be splayed like the spokes in a wheel due to their hydration requirements. The overall conformations of the two globular domains remained stable in the simulations, although both underwent changes in their orientations.

Keywords Cellulase · Cellobiohydrolase I · Cellulose · Computer modeling · Molecular dynamics

L. Zhong · J. F. Matthews · C. Talón · J. W. Brady (✉)
Department of Food Science, Cornell University, Ithaca,
NY 14853, USA
e-mail: jwb7@cornell.edu

L. Zhong
Pennsylvania State University, Mont Alto, PA 17237,
USA

M. F. Crowley · C. L. Brooks III
The Scripps Research Institute, La Jolla, CA 92037, USA

T. Rignall · C. McCabe
Department of Chemical Engineering, Vanderbilt
University, Nashville, TN 37235-1604, USA

J. M. Cleary · R. C. Walker · G. Chukkapalli
San Diego Supercomputer Center, La Jolla,
CA 92093-0505, USA

M. R. Nimlos · M. E. Himmel
National Renewable Energy Laboratory, 1617 Cole
Boulevard, Golden, CO 80401-3393, USA

Introduction

Cellulose, the primary structural polysaccharide of plant cell walls, is the most abundant biopolymer in the biosphere (Ragauskas et al. 2006), and as such represents a significant energy reserve in the form of the chemical potential stored in its C–H and C–C bonds. With the continued growth in fossil fuel consumption, the need for practical renewable liquid fuels is becoming ever more critical (Farrell et al. 2006). Cellulose could potentially serve as such a renewable fuel source if it could be economically broken down into its component glucose for subsequent microbial fermentations. The principal limitation on the use of cellulose in such fermentations is the difficulty in hydrolyzing the glycosidic linkages between the monomers, a problem exacerbated by the insolubility of cellulose (Himmel et al. 2007) Unfortunately, the turnover rates for even the most efficient cellulase enzymes remain inadequate for commercial-scale bioethanol production (Sinnott 1990). Because of the high cost of producing and

using cellulases in proposed biorefineries, considerable attention has centered on the improvement of cellulase performance (Zhang et al. 2006) Although directed evolution and protein engineering methods have been applied to this problem, little improvement in specific activity has emerged (Himmel et al. 1999) Thus, a great need exists for acquiring a better understanding of how these enzymes function and the natural limitations of enzyme performance based on first principles. Protein engineers could then use such information to systematically enhance the performance of cellulases, if such improvements were possible.

Cellulases can be divided into two categories, exocellulases that hydrolyze cellulose chains from their termini and endocellulases that hydrolyze an interior glycoside linkage anywhere in the polysaccharide chain (Teeri 1997). The exocellulases can further be of two types, those that remove a single glucose or cellobiose unit from the chain terminus before dissociating and attacking another chain, and those that processively hydrolyze a single chain (Barr et al. 1996). Processive enzymes are perhaps the most interesting from a practical standpoint since they would seem to offer the greatest potential for efficiency. The fungal exocellulases, such as the cellobiohydrolase I (CBH I) from *Trichoderma reesei*, are complex multi-domain "molecular machines." Many questions remain about the functioning of these enzymes and a detailed understanding of their activity is lacking. However, the full picture of how these enzymes function will require not only a detailed knowledge of the hydrolysis mechanism, but also knowledge of how the enzymes interact with their cellulose substrate.

The collective structure of the natural cellulose substrate is surprisingly complex for such a simple homopolymer. Cellulose does not occur as a single chain, but is synthesized as a bundle of a number of parallel, oriented chains, organized into microfibrils as the fundamental structural unit (Doblin et al. 2002). The chain length (degree of polymerization, DP) for the individual cellulose chains ranges from about 2000 to more than 15000 glucose residues (Sjöström 1993; Kuga and Brown 1991). Cellulose can vary from the so-called elementary fibrils in plants, which contain approximately 36 cellodextrin chains (Doblin et al. 2002), to the large microfibrils and macrofibrils of cellulosic algae, which contain

more than 1200 chains (Sugiyama et al. 1985; Newman 1999; Koyama et al. 1997). The shape of a cellulose microfibril is determined by the geometry of the cellulose synthase complex and by the local environment (Doblin et al. 2002). A significant proportion of the cellulose in these various fibers is highly regular to the point of being locally crystalline, although it is not possible to produce single crystals of pure cellulose from fibers of any significant DP. Several possible crystalline forms for cellulose have been proposed and characterized by fiber diffraction (Langan et al. 1999; Nishiyama et al. 2002, 2003; Wada et al. 2004). Native cellulose fibers and fibrils are thought to consist primarily of two crystal forms, labeled Iβ and Iα, which differ in the relative packing of their hydrogen bonding sheets. In both, the equatorial–equatorial β-(1→4)-linkage of cellulose produces a relatively flat, ribbon-like conformation for the individual chains (Sjöström 1993). A typical fibril might have both of these crystal forms alternating along the same fibril, along with extensive regions of non-crystalline, amorphous packing, and can even have both crystalline packings coexisting in the same portion of the fibril. This complex structure is probably necessary for the primary structural function of cellulose fibrils and fibers, which must simultaneously exhibit both the great strength of the crystalline packing along with the flexibility presumably imparted by the amorphous regions. However, the low solubility and resistance to disentanglement is probably the single most important rate-limiting feature of this substrate making it difficult to depolymerize industrially. This resistance is further complicated by the fact that cellulose from plant sources is often intimately associated with significant amounts of hemicellulose and lignin (Coughlan and Hazlewood 1993; Tarchevsky and Marchenko 1991).

Given these difficulties, understanding how processive cellulase enzymes bind to and interact with microcrystalline cellulose could be of great practical utility and help elucidate the mechanism of their processivity. Unfortunately, conventional experimental methods are not able to probe this level of complexity on the molecular scale. We have undertaken molecular dynamics computer simulations of such a system as an alternate approach to understanding processivity. Here we report the construction of the first computer model of the processive exocellulase *T. reesei* CBH I interacting with a

microcrystalline fibril in an aqueous environment. We also describe some preliminary observations concerning the evolution of this model in molecular dynamics (MD) simulations. Because of its immense size and the many uncertainties about the structure of the individual parts, constructing the starting model for such a system and optimizing a molecular mechanics program to handle it on a massively parallel processor supercomputer are major undertakings. Portions of this system have been previously studied using molecular modeling (Kuutti et al. 1991), but no previous attempt has been made to model the entire enzyme-substrate complex due to the sheer size of the system. Because this simulation is so massive, it requires enormous amounts of computer time and resources. The present paper describes the nontrivial task of constructing the initial model for such a system and presents the preliminary observations from the first 1.5 ns of equilibration/ simulation time for this system.

Methods

The large size of the cellulase/fibril complex and the long timescale for significant events along the pathway for the overall deconstruction process mean that useful information from MD simulations of this system will ultimately require the largest supercomputer facilities available. However, in order to begin such a simulation it is necessary to construct a plausible model for the starting structure for the enzyme/substrate complex. This task is made more difficult by the fact that, while the structures of parts of the system have been individually studied, the structure of the overall complex is unknown. In addition, significant uncertainties remain even for those parts of the system that have been studied. For example, diffraction studies of cellulose have been reported (Nishiyama et al. 2002, 2003), but the fine details of these structures remain somewhat controversial (Matthews et al. 2006; Yui et al. 2006). Several crystal structures for the catalytic domain are

available (Divne et al. 1994, 1998; Ståhlberg et al. 1996), but they do not contain all of the coordinates for the glycosylating oligosaccharides or the conformation of the transition to the linker domain. The structure of the binding domain has been studied in solution by NMR (Kraulis et al. 1989), which produces a closely related family of conformations, from among which a single structure must be selected. The linker domain is flexible, and the presence of glycosylation may promote an extended structure, although there is some uncertainty about the most populated end-to-end distance in this type of cellulase (Receveur et al. 2002; von Ossowski et al. 2005). Finally, any changes that might result from linking all of these elements together and docking the protein onto a cellulose surface are also uncharacterized.

The initial structural model for this system was generated using several sources of information. The conformation of the enzyme's catalytic domain was taken from the crystal structure deposited with the Protein Data Bank (PDB), 2CEL (Divne et al. 1994). Since the crystal structure is for a double mutant, it was necessary to convert the residues Asp94 and Gln212 back to the wild-type residues Gly94 and Glu212. The calcium ion present in the active site of this crystal structure was changed to a water molecule. The conformation of the binding domain is known from NMR and was also obtained from the Protein Data Bank (Kraulis et al. 1989). Although the conformation of the 27 residue linker peptide is unknown, its sequence has been determined and is known to be heavily *O*-glycosylated with mannose residues at the serines and threonines (Harrison et al. 1998; Hui et al. 2001). Since crystallographic data is not available to determine the extent of glycosylation at each site, the suggestions of Nevalainen et al. (1997) were followed and are given in Fig. 1. For the C-terminus of this segment, the two proline residues were placed in the polyproline conformation. The other residues were initially given arbitrary conformations and the linker segment was separately relaxed using constrained MD simulations. These

```
        M  M  M                    M  M
     M  M  M  M                 M  M  M
     M  M  M  M                 M  M  M     M  M
     |  |  |  |                 |  |  |     |  |
Gly-Asn-Pro-Pro-Gly-Gly-Asn-Pro-Pro-Gly-Thr-Thr-Thr-Thr- Arg-Arg-Pro-Ala-Thr-Thr-Thr-Gly-Ser-Ser-Pro-Gly-Pro-Pro
```

Fig. 1 The CBH I linker peptide sequence showing the extent of glycosylation at each site used in the current work

relaxed coordinates were then "patched" to the terminal sequences for the two globular domains. The five terminal residues of this sequence at each end were manually adjusted to provide an unstressed transition between the linker segment and the globular domain. The glycine residues were arranged so as to make the binding domain approximately in the same plane as the entrance tunnel of the catalytic domain. Less manual adjustment was necessary at the N-terminal domain where this linker sequence joins with the catalytic domain because the terminal residue is a glycine. All atoms were present in the simulation and the coordinates for the hydrogen atoms not included in the PDB data set were constructed using the standard CHARMM algorithm (Brooks et al. 1983). The charges of the protein atoms were taken from the standard CHARMM set appropriate for a pH of 7 and resulted in a net charge on the protein of −17.

This enzyme complex was then manually docked onto the surface of a model microcrystalline cellulose fibril constructed from the proposed cellulose Iβ crystalline structure Nishiyama et al. (2002). Smaller crystals of this proposed structure have been previously modeled with two different force fields (Matthews et al. 2006; Yui et al. 2006). The (1,0,0) crystal face has been studied previously and was chosen for docking in this work since this is believed to be the target of the cellulose binding domain of CBH I (Lehtiö et al. 2003). The microfibril model that was constructed contained 108 individual cellulose chains each containing 40 glucose units, producing a microfibril 206 Å long with a diameter of approximately 60 Å. There are 91,044 atoms in these cellulose chains. This microfibril was 16 layers deep counting from the (1,0,0) face and 11 layers deep counting from the (1,1,0) face. In this model microfibril the (1,0,0) face is four chains wide and the (1,1,0) face is six chains wide. Although this model microfibril was larger than an actual plant microfibril, it was chosen to represent a partially hydrolyzed microfibril from *Halocynthia papillosa*, and contained extended faces of all of the most important crystal surfaces (Helbert et al. 1998). The arbitrary starting structure of this microfibril was then partially relaxed using 100 ps of unconstrained MD simulation in aqueous (TIP3P: Jorgensen et al. 1983; Durell et al. 1994, see below) solution in a periodic orthorhombic box similar to that used for the entire

protein-substrate complex (see below). During this simulation a pronounced right-handed twist developed in the crystalline fibril (see Fig. 2), as was seen in the smaller crystals of cellulose Iβ studied earlier (Matthews et al. 2006), consistent with experimental observations (Hanley et al. 1997).

The entire cellulase protein complex was docked onto the (1,0,0) surface of this cellulose microfibril. Docking was accomplished by placing the binding domain over the central two chains of the (1,0,0) surface on the fibril (of the four such chains in our microfibril). The binding domain was oriented such that the three Tyr residues on the binding surface were positioned parallel to the direction of the chains but placed between these two middle chains, at a distance of 3.5 Å above the cellulose crystal surface. Figure 2 shows two views of this enzyme-substrate complex. The resulting docked complex was surrounded with water at a density of 1 gm/cm³ under orthorhombic periodic boundary conditions, thermalized and equilibrated (Leach 1996).

All of the calculations reported here used the CHARMM molecular mechanics program (Brooks
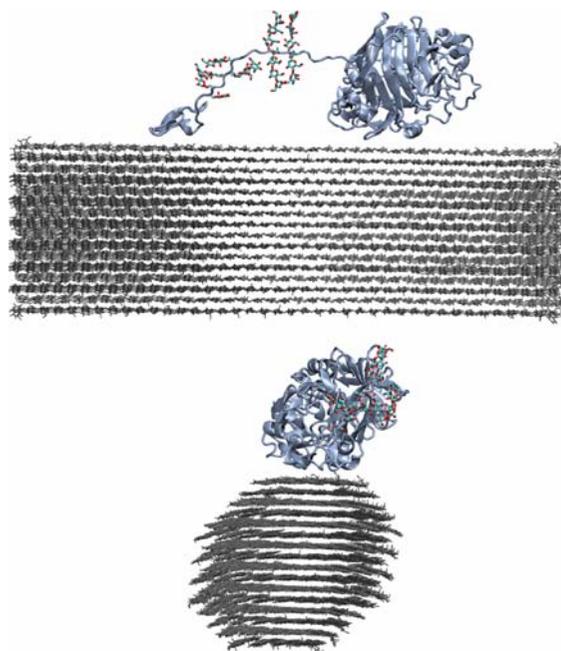
Fig. 2 Side and end-on views of the complex of CBH I docked onto the (1,0,0) surface of the 108-chain model cellulose microfibril. The twist that developed in the microfibril during its separate equilibration dynamics sequence is apparent in these views
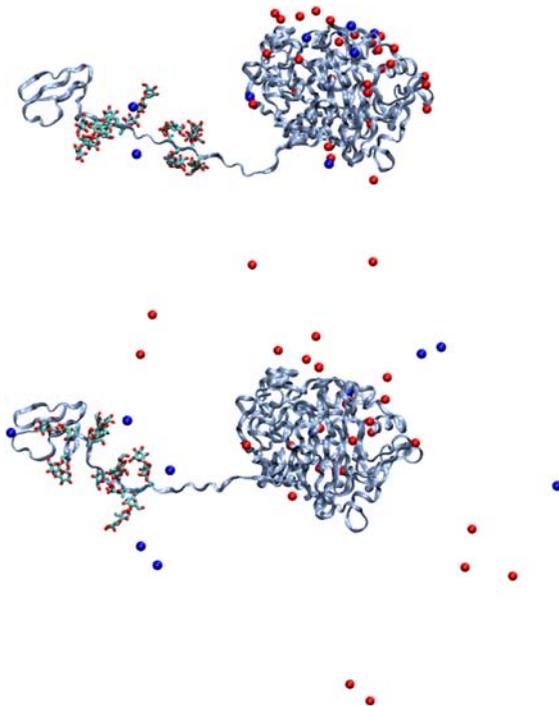
et al. 1983). The CHARMM27 force field was used to describe the amino acid residues of the linker polypeptide as well as the sugars (MacKerell et al. 1998). The sugar atoms were modeled using parameters specifically developed for carbohydrates (Palma et al. 2000; Kuttel et al. 2002). Water molecules were represented using the CHARMM implementation of the TIP3P force field (Jorgensen et al. 1983; Durell et al. 1994). The microfibril-cellulase complex was placed in an equilibrated rectangular box of water molecules with dimensions 280.0 Å by 202.0 Å by 124.0 Å. All those water molecules that overlapped with the carbohydrate or protein heavy atoms were deleted. Since the protein carries a net charge, a total of 28 sodium and 11 chloride counter ions were added to the system. Each counter ion was individually placed near each charged group to locally neutralize the charge. This was accomplished by taking the water molecule closest to each charged group and replacing it with a sodium cation in the case of negatively charged groups or a chloride anion in the case of positively charged groups. The resulting simulation contained 204,399 water molecules and 711,788 atoms in total.

Two hundred steps of steepest descent minimization, followed by 100 steps of conjugate gradient minimization were first applied to the system to relieve any serious strains resulting from the set-up procedure. MD simulations were then used to heat the system from 50 to 300 K in 50 K increments over a period of 10 ps, followed by an additional 190 ps of equilibration at 300 K. After this heating and equilibration stage the system velocities were not again adjusted, and the system was simulated in the NVE ensemble using a Verlet integrator with a step size of 2 fs. Long range electrostatic interactions were determined using the particle-mesh Ewald (PME) method (Darden et al. 1993) with a PME charge grid spacing of approximately 1.0 Å. A real-space Gaussian width (kappa) of 0.32 (1/Å) and fifth degree of B-spline interpolation were used. van der Waals interactions including image atoms were truncated at 10.0 Å using switching functions. In all calculations, a dielectric constant of 1 was used. Covalent bond lengths involving hydrogen atoms were kept fixed at their equilibrium lengths using the constraint algorithm SHAKE (van Gunsteren and Berendsen 1977). The MD simulations were run for a period of 1.5 ns.

A system of this size, with over 700,000 atoms, presents special computational challenges not generally encountered in smaller protein simulations. Only tightly integrated supercomputers can deal with calculations of this size and complexity, which necessarily implies the need for efficient parallelization to accelerate the integration rate in the time domain. Classical molecular mechanics codes are particularly difficult to parallelize and inherently fall short of the efficiencies achieved for other types of calculations. Nonetheless, in order for a calculation of the type attempted here to be practical it is necessary to adapt the MD code to be used such that it makes optimal use of the multiple processors available. After the performance of the system was benchmarked, we chose to use two IBM P690 compute nodes, provided by the San Diego Supercomputer Center, with 32 processors per node throughout our simulation. Work is currently under way to improve the parallel efficiency of the CHARMM software, without making any additional approximations, to enable scaling of MD simulations for this and comparable systems to hundreds and potentially thousands of processors.

Results and discussion

The protein-substrate complex constructed here was stable under the CHARMM-TIP3P force fields during energy minimization and MD simulations. During this relatively short initial MD simulation period, no major changes in the complex occurred. Thus, while the collective protein conformation and its positioning on the microfibril were obviously arbitrary, the construct was not unreasonable and did not lead to significant artifactual changes due to poor, high-energy placements of any portions of the system. Figure 3 presents two views looking "down" on the enzyme bound to the cellulose fibril surface, the top showing the starting configuration and the bottom showing the configuration after 1.10 ns. In the starting structure, counter-ions were place in positions adjacent to the charged groups they were intended to neutralize. As can be seen, as the simulation proceeds, these counter-ions solvate and diffuse away, with little tendency to remain bound to the protein. This result is reasonable and consistent with expectations, but demonstrates that ion–ion and
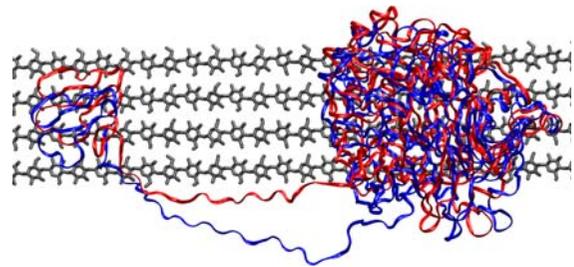
Fig. 4 The initial and final configurations of the protein, shown as alpha carbon traces, relative to the top layer of the (100) surface of the crystalline microfibril (only the first layer of cellulose is shown in grey; blue: initial structure, red: final conformation

Fig. 3 Two views of the CBH I cellulase complex with its neutralizing counterions as seen from "above". Chloride ions are shown as grey spheres and sodium ions as red spheres. The upper view is of the initial structure, and the lower view is of the last configuration of the simulation

ion–water interactions are appropriately balanced and produce a plausible ionic distribution. While the overall conformation of the protein remained stable, a number of moderate changes in the positions of the globular domains and in the conformation of the linker sequence developed during the simulation, as can be seen by comparing the two panels of Fig. 3. These conformational changes will be individually discussed in the following sections.

Linker flexibility

From the figures it is apparent that the CBH I complex undergoes conformational changes during the course of this short simulation. Figure 4 compares the protein conformation as a backbone trace at the beginning and end of the 1.5 ns simulation. All three domains of the protein underwent changes, but as can be seen from this figure, the greatest changes occurred in the linker domain. The overall

conformations of the two globular domains did not change significantly from the reported experimental conformations (rms change of 2.41 Å for the catalytic domain and 4.81 Å for the binding domain). All of the secondary structural elements of these two globular domains remained intact throughout the simulation, as did all of the significant features of the tertiary structure. These domains primarily fluctuated about their original conformations, but did change their orientations relative to one another and to the cellulose surface (discussed below). It can be seen from the simulation (see movie in Supporting Information) that the linker domain displays considerable flexibility. During the simulation the linker segment whips about between the two heavier domains. Given the apparent flexibility of this chain, it is not clear that it has the capacity to store energy in a manner similar to a compressed or stretched spring, as has been previously postulated in theories of processivity. In the course of these fluctuations the linker bowed up away from the substrate surface as the two globular domains drew more than 4 Å closer to one another. It is unclear whether they were drawn together by the change in the linker conformation or whether they simply diffused closer and the linker adopted the bowed conformation in response, but the latter possibility seems more probable. Further work is planned to more fully investigate this question.

The sequence of this linker domain (see Fig. 1) is fairly unusual, with a high proportion of threonine, proline and glycine residues and two repeating reverse collagen-like sequences of Pro-Pro-Gly, along with one collagen-like Gly-Pro-Pro sequence. In the middle of this linker chain are two adjacent Arg residues at R449 and R450 (residues 15 and 16 in

the linker alone). As a result of its collagen-like character, portions of the linker frequently adopted pseudo-helical conformations. Toward the end of the simulation, the linker developed a pronounced bend near to (but two residues away from) the two Arg residues (see Figs. 3, 4), with the sequences on either side having an overall linear conformation. In the final configuration seen in Fig. 4, the *C*-terminal portion of this chain (near the binding domain) exhibits one half turn of the pseudo-collagen-like helix. It remains to be seen in future simulations how persistent and rigid these extended regions are.

Oligosaccharide conformation

One of the more interesting effects of solvation of this complex was in the conformations of the oligosaccharide chains glycosylated to the linker domain of the cellulase complex. A significant change was observed in the conformations of these oligosaccharides as the simulation proceeded. In the starting structure, these sugar chains were arbitrarily placed in completely extended conformations pointing in directions determined by the local conformation of the polypeptide backbone. In practice this placement resulted in several of these chains being adjacent and almost parallel in the starting structure, as can be seen in Fig. 5, which focuses on just these residues and the linker backbone.

Interestingly, both sets of adjacent oligosaccharides remained quite extended as the simulation in solution proceeded, but the overall complex adjusted such that they pointed in different directions, radially extended like spokes on an axle. This conformational change has the effect of placing the chains essentially as far apart as they can be. This mutual avoidance apparently results from the hydration of these chains, because their hydration shells would otherwise interfere with one another. As shown below, in the splayed conformation, each oligosaccharide chain is independently solvated. Water molecules that directly bridge oligosaccharide chains by hydrogen bonds are rarely found. Only three such cases are observed at the end of simulation (Fig. 6). This separation did not occur in a parallel vacuum simulation. As the simulation progressed and the chains moved apart, the number of hydrogen bonds that each oligosaccharide made to water increased, leading to a large
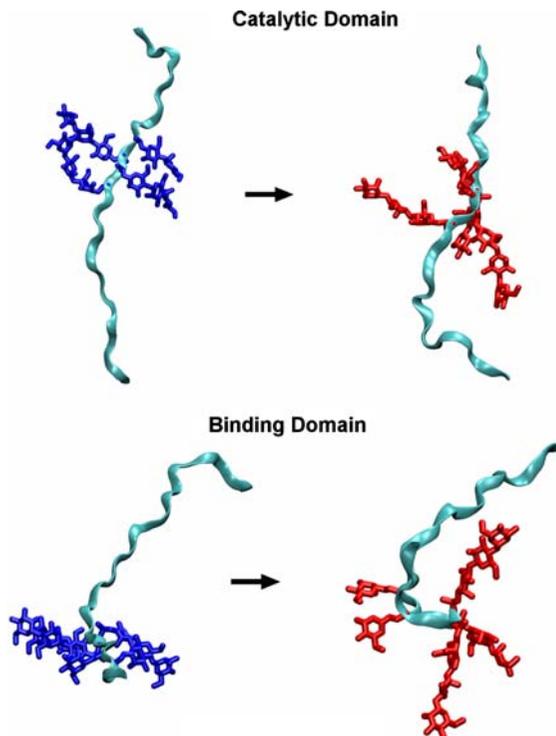
Fig. 5 Top: Left: The arbitrary parallel starting conformations for the oligosaccharide chains of the linker region (blue). Two stretches of the linker peptide backbone are shown for convenience, as alpha carbon traces. The nine oligosaccharide chains occur as two nearby groups, top and bottom. Right: The conformations of these chains at the end of the simulation. These chains have splayed out like the arms of a windmill or spokes in a wheel
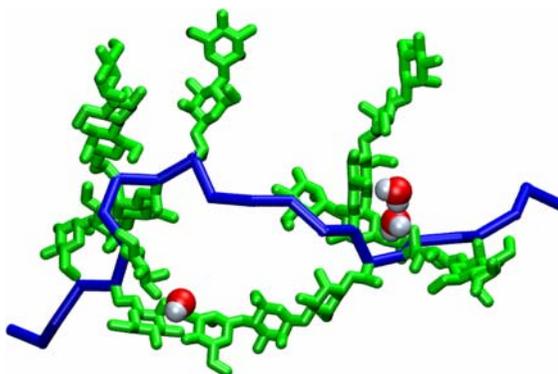
Fig. 6 Three water molecules that form H-bond bridges between oligosaccharide chains. Blue: backbone trace of linker domain; green: oligosaccharides

increase in the total number of oligosaccharide–water hydrogen bonds, from only 78 in the starting structure to approximately 160 in the solvated structure, with

fluctuations of approximately +/−5 hydrogen bonds on average. This change occurred very quickly, being essentially complete after only 100 ps of simulation time.

The internal conformations of these oligosaccharide chains were surprisingly rigid, remaining quite extended, as shown in the lower panel in Fig. 5. As a result, once these conformational shifts were completed, the local structure became stable enough to allow the solvent density relative to the oligosaccharides and polypeptide backbone to be contoured in the same fashions as has been done for the substrate surface and for individual monosaccharide rings (Matthews et al. 2006; Liu and Brady 1996; Brady 1993; Schmidt et al. 1996; Liu et al. 1997). The local solvent density at each point relative to the protein functional groups was calculated by dividing the region around the protein into small cubes and averaging how often water molecules occupied each cube relative to the occupancy expected in bulk liquid water. Contour maps were then prepared showing those regions with high and low water densities. Such solvent density mapping was applied to the present simulations using procedures developed in previous studies (Liu and Brady 1996; Brady 1993; Schmidt et al. 1996; Liu et al. 1997). For the purpose of

calculating solvent density distributions, the volume of the primary system was divided into small cubes 0.30 Å in length. Complete coordinate sets were saved every 10 fs during the simulation, and these coordinates were subsequently used to calculate the average density of water molecules in each indexed cubic box using programs developed "in house". The calculated densities were normalized relative to a uniform distribution in the same volume and were displayed graphically using VMD (Humphrey et al. 1996).

The calculated water density contours are shown in Fig. 7. As can be seen, there are well-defined bands of water density corresponding to water molecules hydrogen bonding to the chains, and in rare cases bridging between adjacent chains, or between one carbohydrate chain and the protein backbone. It is not yet known what the effects of these regions of sugar, protein, and water between the two globular domains are on the domain motions, and to what extent the fairly extensive conformational fluctuations of the linker domain itself are affected by the water structuring. It is possible that these regions of localized water structuring could define a gel-like zone "cushioning" the interactions of the globular domains.
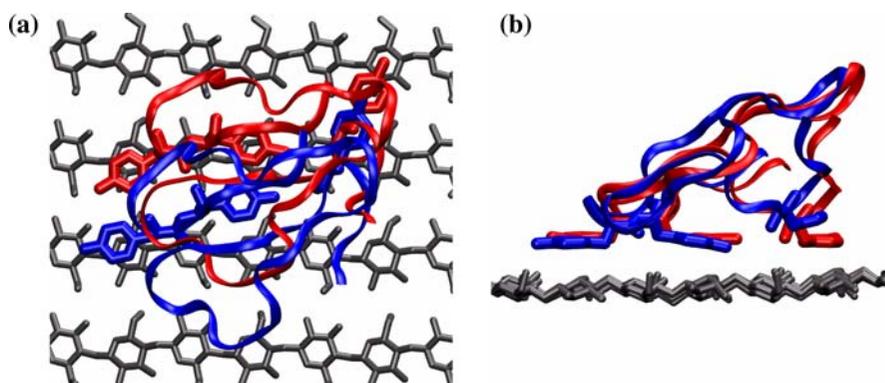
Fig. 7 Contour map of the average solvent density around the glycosylating oligosaccharides of the linker domain. The densities were computed separately for the two oligosaccharide groups relative to the alpha carbons of the short peptide to which the oligosaccharide groups are attached. The contour level shown corresponds to a density 2.0 times the bulk water density
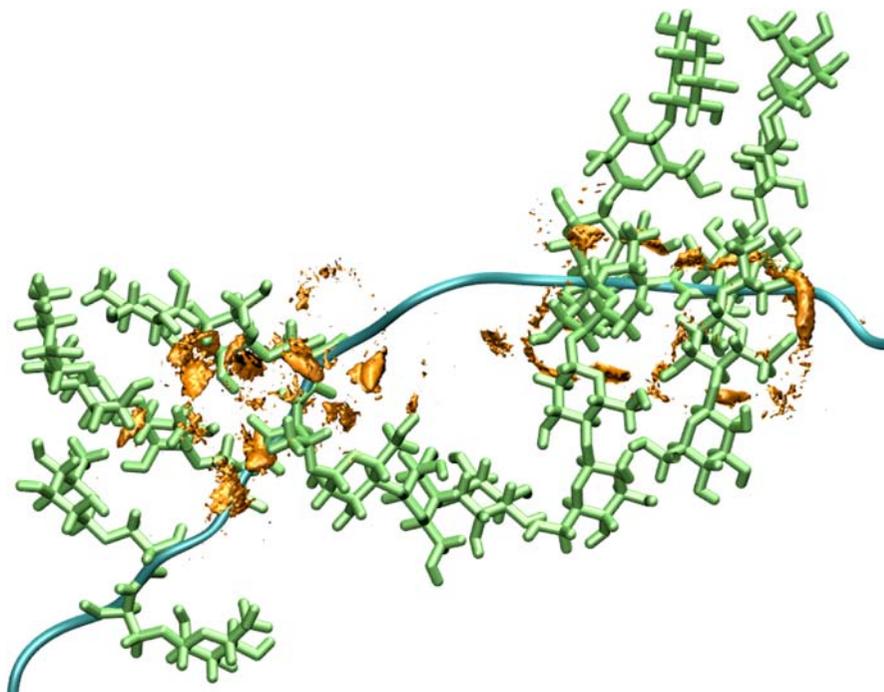
Cellulose binding domain

Figure 8 shows the changes in the position of the binding domain during the course of the simulation. The overall conformation of this globular domain does not change significantly, but it does re-orient somewhat on the cellulose surface. Three Tyr residues (Y466, Y492 and Y493) are suggested to play important roles in the docking of this domain to the microfribil by aligning their rings relative to the sugar monomers (Hoffrén et al. 1995). In the initial conformation, this domain was positioned by aligning these three rings across two sugar chains in order to avoid building in this assumed alignment. In this arbitrary starting structure, the Y492 ring stacked perfectly with one sugar ring, while the Y493 ring sat atop the groove between two sugar chains. Although the Y466 ring was located above another sugar chain, it did not stack with any sugar ring (Fig. 8a). Unlike previous simulations of this binding domain (Nimlos et al. 2007), at least during this initial simulation period, the orientation of this domain changed such that the three Tyr residues on its binding surface become less aligned with the cellulose surface chains. It is not yet clear to what extent the orientation of this domain is being perturbed by being tethered to the catalytic domain via the linker domain. By the end of the simulation, the binding domain reoriented itself to a conformation in which the rings of residues Y466 and Y492 were positioned above grooves between sugar chains, and the Y493 ring was located above the chain but did not stack with any sugar ring. However, it was observed that the plane of the Y466 ring aligned better with the microfibril surface at the end of simulation, whereas it was tilted in the initial conformation (Fig. 8b).
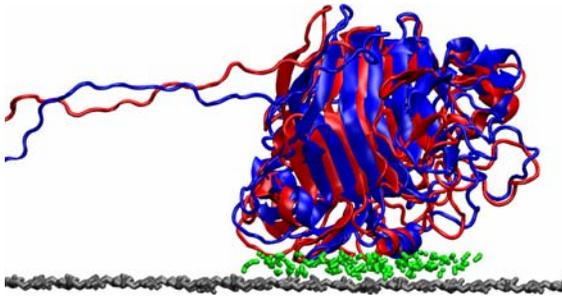
The translational motion of the binding domain is apparent from Fig. 4. As discussed above, the two globular domains approached one another about 4 Å closer than in their initial conformation during the 1.5 ns simulation. This result was mainly due to the translational motion of this smaller globular binding domain along the direction of the sugar chains. Translational movement across the cellulose surface perpendicular to the sugar chains was also observed in the simulation. However, this domain maintained its distance from the microfibril during the simulation, moving neither closer to nor farther away from the sugar surface. This result is consistent with recent studies that have investigated the interactions between the binding domain and the cellulose substrate (Nimlos et al. 2007).

Catalytic domain

During the simulation the catalytic domain did not exhibit any large conformational changes, and the active site tunnel did not collapse, instead remaining filled with water molecules. The catalytic domain did not touch the cellulose microfibril surface in the starting structure, but was separated from it by a layer of water molecules, as can be seen from Fig. 9. While this large globular domain randomly diffused about on its tether during the simulation, the layer of water between it and the surface remained approximately the same after 1.5 ns of dynamics. For a protein unit of this size, the motions exhibited in Fig. 9 are essentially what should be expected from undirected diffusion.

Protein motions in the catalytic domain were not uniform, as some portions of this domain exhibited

Fig. 8 (**a**) "Top" (left) and (**b**) "side" (right) views of the initial (blue) and final (red) positions of the binding domain (ribbon) relative to the top layer of the (1,0,0) surface of the microfibril. Three Tyr residues (from left to right Y492, Y493, and Y466) are indicated in a "licorice" representation

Fig. 9 The position of the catalytic domain relative to the surface of the microfibril. The initial structure is shown in blue and the final structure in red, with the layer of water separating the catalytic domain from the surface shown in green

greater deviations from the starting conformation than did others. The greatest conformational changes were found for residues adjacent to the linker peptide. However, these were not the only residues that underwent large changes. As shown in Fig. 10, all residues in the catalytic domain that face the binding domain show significant motions during the simulation (shown in green in Fig. 10). Many of these residues exhibited flexibility comparable to the residues in the linker chain. Motions within ordered secondary structures (α-helices and β-sheets) of a protein are generally limited, due to the constraints imposed by their hydrogen bonds. However, several of the α-helices and β-sheets in the region of the catalytic domain near the linker peptide and facing the binding domain also exhibited large motions
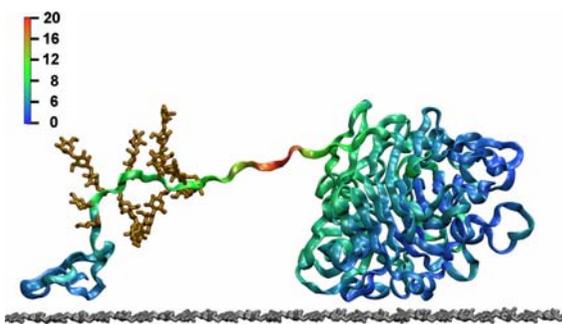
Fig. 10 The final conformation of the protein-substrate complex, seen from the "side", showing the flexibility of the linker sequence and its change in conformation during the simulation. The protein is illustrated as a Cα trace ribbon diagram showing secondary structure elements in the globular domains. Motions of the CBH I are mapped onto the final conformation (prepared with VMD: (Humphrey et al. 1996)), color coded by the backbone RMSD of each residue. The color scale is indicated on the left in Å^2

(Fig. 10). No collapse of the empty (but solvated) active site tunnel occurred during the simulation, and the residues composing the tunnel walls experienced only limited motions (blue in Fig. 10). Importantly, this result suggests that the catalytic tunnel is always "open", even in the absence of a cellulose strand. This has implications for the mechanism by which a cellulose chain initially enters the catalytic tunnel and is the subject of ongoing investigations.

Overall dynamics

As already noted, in general the overall structure and conformation constructed for this enzyme-substrate complex was stable during the initial equilibration stages of the MD simulation, but nevertheless underwent some conformational changes even on this short time scale. These changes can perhaps be best seen in Fig. 10, which displays the final frame of the MD simulation. The most significant change that can be seen in this figure is in the linker segment which as already noted has developed a significant bend near to the binding domain in the midst of the most heavily glycosylated region of the sequence. As this simulation proceeds it will be interesting to determine whether this change in the conformation of the linker domain affects the relative dynamics of the two globular domains.

Conclusions

A plausible model for the interaction of the CBH I cellulase protein with a cellulose microfibril has been constructed and has been shown to be stable under physiological simulation conditions. In the preliminary MD simulations used to "temper" this model of the complex, the linker domain between the two globular domains was found to be much more flexible than the globular domains and underwent the greatest conformational changes from its initial placement in the model. The final model from these simulations is currently being used to continue the study to much longer times to determine how slower relaxation process may alter the structure of the complex and to see whether these changes lead to insights into how the system functions. Among the most significant of the observations which can be made from our

preliminary MD study is that it does not appear that the very flexible linker domain chain can store energy in the manner of a spring so as to draw the CD closer to the CBD. However, the significant bend that developed in this region of the polypeptide by the end of the trajectory may signal a change to a different dynamical behavior, and the continuation of the simulations should indicate whether this change is a short-time fluctuation or a significant transition in behavior with mechanistic implications. The position of the bend in the linker sequence was in the highly glycosylated region of the chain, and the oligo-saccharides themselves underwent significant conformational changes away from the arbitrary starting structure, which could have important consequences. It should be noted that this change was largely due to specific interactions with water molecules, which demonstrates the importance of having explicit solvent molecules included in the simulation rather than using a continuum solvation model.

An enzyme evolved to processively hydrolyze a single cellulose chain in a microfibril might be presumed to interact with the crystalline substrate in such a way as to promote the removal of that chain from the surface as well as to possess some feature that makes successive hydrolysis more favorable than dissociation. Since conventional molecular mechanics simulations such as those reported here do not allow bond scission, many of the essential features of processivity presumably cannot be captured in such simulations, but the interactions of the full complex with the substrate could be revealing concerning how a single chain is disentangled from its fibril matrix. Unfortunately, such interactions are presumably slow on the molecular timescale, which is a severe problem in straightforward simulations of the type reported here, and more sophisticated methods will be needed to probe these interactions more deeply. Clearly, a system with a broken chain will be needed to examine how a chain is pulled up from the surface. Such a system has been prepared and will be described in a future report.

The stage is now set for application of longer MD simulation times on enhanced computational platforms to this and improved variants of this important functional cellulase model. The microfibril constructed for this simulation is actually larger in diameter than most natural cellulose microfibrils, and this size was selected in part to provide an extended

(1,0,0) planar surface on which to dock the enzyme that would be larger than the width of the protein (see Fig. 2). However, a smaller substrate (the actual diameter of experimental microfibrils) would not only be a more realistic model, but would also be less computationally "expensive," allowing the extension of simulations to much longer times. For this reason, a new enzyme-substrate complex using a more realistic 36-cellulose chain microfibril is currently being constructed and will also be reported in a future communication.

References

Barr BK, Hsieh Y-L, Ganem B, Wilson DB (1996) Identification of two functionally different classes of exocellulases. Biochemistry 35:586–592

Brady JW (1993) Molecular dynamics simulations of carbohydrates. Forefronts/Cornell Theory Center 9:7

Brooks BR, Bruccoleri RE, Olafson BD, Swaminathan S, Karplus M (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 4:187–217

Coughlan MP, Hazlewood GP (1993) Hemicellulose and hemicellulases. Portland, London

Darden T, York D, Pedersen L (1993) Particle mesh Ewald: an N log(N) method for Ewald sums in large systems. J Chem Phys 98:10089–10092

Divne C, Stahlberg J, Reinikainen T, Ruohonen L, Pettersson G, Knowles JKC, Teeri TT, Jones TA (1994) The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. Science 265:524–528

Divne C, Ståhlberg J, Teeri TT, Jones TA (1998) High-resolution crystal structures reveal how a cellulose chain is bound in the 50 Å long tunnel of cellobiohydrolase I from *Trichoderma Reesei*. J Mol Biol 275:309–325

Doblin MS, Kurek I, Jacob-Wilk D, Delmer DP (2002) Cellulose biosynthesis in plants: from genes to rosettes. Plant Cell Physiol 43:1407–1420

Durell SR, Brooks BR, Ben-Naim A (1994) Solvent-induced forces between two hydrophilic groups. J Phys Chem 98:2198–2202

Farrell AE, Plevin RJ, Turner BT, Jones AD, O'Hare M, Kammen DM (2006) Ethanol can contribute to energy and environmental goals. Science 311:506–508

Hanley SJ, Revol J-F, Godbout L, Gray DG (1997) Atomic force microscopy and transmission electron microscopy of cellulose from *Micrasterias denticulata*; evidence for a chiral helical microfibril twist. Cellulose 4:209–220

Harrison MJ, Nouwens AS, Jardine DR, Zachara NE, Gooley AA, Nevalainen H, Packer NH (1998) Modified glycosylation of cellobiohydrolase I from a high cellulose-producing mutant strain of *Trichoderma reesei*. Eur J Biochem 256:119–127

Helbert W, Nishiyama Y, Okano T, Sugiyama J (1998) Molecular imaging of *Halocynthia papillosa* cellulose. J Struct Biol 124:42–50

Himmel ME, Ruth MF, Wyman CE (1999) Cellulase for commodity products from cellulosic biomass. Curr Opin Biotechnol 10:358–364

Himmel ME, DIng S-Y, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD (2007) Biomass recalcitrance: engineering plants and enzymes for biofuels production. Science 315:804–807

Hoffrén A-M, Teeri TT, Teleman O (1995) Molecular dynamics simulation of fungal cellulose-binding domains: differences in molecular rigidity but a preserved cellulose binding surface. Protein Eng 8:443–450

Hui JPM, Lanthier P, White TC, McHugh SG, Yaguchi M, Roy R, Thibault P (2001) Characterization of cellobiohydrolase I (Cel7A) glycoforms from extracts of *Trichoderma reesei* using capillary isoelectric focusing and electrospray mass spectrometry. J Chromatogr B Anal Technol Biomed Life Sci 752:349–368

Humphrey W, Dalke A, Schulten K (1996) VMD-visual molecular dynamics. J Mol Graphics 14:33–38

Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79:926–935

Koyama M, Sugiyama J, Itoh T (1997) Systematic survey on crystalline features of algal celluloses. Cellulose 4:147–160

Kraulis PJ, Clore GM, Nilges M, Jones TA, Pettersson G, Knowles J, Gronenborn AM (1989) Determination of the three-dimensional solution structure of the C-terminal domain of cellobiohydrolase I from *Trichoderma reesei*. A study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing. Biochemistry 28:7241–7257

Kuga S, Brown RM (1991) Physical structure of cellulose microfibrils: implications for biogenesis. In: Haigler CH, Weiner PJ (eds) Biosynthesis and biodegradation of cellulose. Marcel Dekker, New York, pp 125–142

Kuttel M, Brady JW, Naidoo KJ (2002) Carbohydrate solution simulations: producing a force field with experimentally consistent primary alcohol rotational frequencies and populations. J Comput Chem 23:1236–1243

Kuutti L, Laaksonen L, Teeri TT (1991) Interaction studies of the tail domain of cellobiohydrolase I and crystalline cellulose using molecular modelling. J Chimie Physique et de Physico-Chimie Biologique 88:2663–2667

Langan P, Nishiyama Y, Chanzy H (1999) A revised structure and hydrogen-bonding system in cellulose II from a neutron diffraction analysis. J Am Chem Soc 121:9940–9946

Leach AR (1996) Molecular modelling: principles and applications. Longman, Harlow

Lehtiö J, Sugiyama J, Gustavsson M, Fransson L, Linder M, Teeri TT (2003) The binding specificity and affinity determinants of family 1 and family 3 cellulose binding modules. Proc Natl Acad Sci USA 100:484–489

Liu Q, Brady JW (1996) Anisotropic solvent structuring in aqueous sugar solutions. J Am Chem Soc 118:12276–12286

Liu Q, Schmidt RK, Teo B, Karplus PA, Brady JW (1997) Molecular dynamics studies of the hydration of α,α-trehalose. J Am Chem Soc 119:7851–7862

MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102:3586–3616

Matthews JF, Skopec CE, Mason PE, Zuccato P, Torget RW, Sugiyama J, Himmel ME, Brady JW (2006) Computer simulation studies of microcrystalline cellulose Iβ. Carbohydr Res 341:138–152

Nevalainen H, Harrison M, Jardine D, Zachara NE, Paloheimo M, Suominen P, Gooley AA, Packer NH (1997) Glycosylation of cellobiohydrolase I from *Trichoderma reesei*. In: *TRICEL 97 conference: carbohydrates from Trichoderma reesei and other microorganisms*. The Royal Society of Chemistry, Cambridge, UK; Ghent, Belgium

Newman RH (1999) Estimation of the lateral dimensions of cellulose crystallites using ^{13}C NMR signal strengths. Solid State Nucl Magn Reson 15:21–29

Nimlos MR, Matthews JF, Crowley MF, Walker RC, Chukkapalli G, Brady JW, Adney WS, Cleary JM, Zhong L, Himmel ME (2007) Molecular modeling suggests induced fit of family I carbohydrate binding modules with a broken chain cellulose surface. Protein Eng Des Sel 20(4):179–187

Nishiyama Y, Langan P, Chanzy H (2002) Crystal structure and hydrogen-bonding system in cellulose Ib from synchrotron X-ray and neutron fiber diffraction. J Am Chem Soc 124:9074–9082

Nishiyama Y, Sugiyama J, Chanzy H, Langan P (2003) Crystal structure and hydrogen bonding system in cellulose I$_\alpha$ from synchrotron X-ray and neutron fiber diffraction. J Am Chem Soc 125:14300–14306

Palma R, Zuccato P, Himmel ME, Liang G, Brady JW (2000) Molecular mechanics studies of cellulases. In: Himmel ME (eds) Glycosyl hydrolases in biomass conversion. American Chemical Society, Washington, pp 112–130

Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CAWJ, Frederick J, Hallett JP, Leak DJ, Liotta CL, Mielenz JR, Murphy R, Templer R, Tschaplinski T (2006) The path forward for biofuels and biomaterials. Science 311:484–489

Receveur V, Czjzek M, Schülein M, Panine P, Henrissat B (2002) Dimension, shape, and conformational flexibility of a two domain fungal cellulase in solution probed by

small angle X-ray scattering. J Biol Chem 277:40887–40892

Schmidt RK, Karplus M, Brady JW (1996) The anomeric equilibrium in D-xylose: free energy and the role of solvent structuring. J Am Chem Soc 118:541–546

Sinnott ML (1990) Catalytic mechanisms of enzymatic glycosyl transfer. Chem Rev 90:1171–1202

Sjöström E (1993) Wood chemistry. Academic Press, San Diego

Ståhlberg J, Divne C, Koivula A, Piens K, Claeyssens M, Teeri T, Jones T (1996) Activity studies and crystal structures of catalytically deficient mutants of cellobiohydrolase I from *Trichoderma reesei*. J Mol Biol 264:337–349

Sugiyama J, Harada H, Fujiyoshi Y, Uyeda N (1985) Lattice images from ultrathin sections of cellulose microfibrils in the cell wall of *Valonia macrophysa* Kutz. Planta 166:161–168

Tarchevsky IA, Marchenko GN (1991) Cellulose: biosynthesis and structure. Springer-Verlag, Berlin

Teeri TT (1997) Crystalline cellulose degradation: new insight into the function of cellobiohydrolases. TIBTECH 15:160–167

van Gunsteren WF, Berendsen HJC (1977) Algorithms for macromolecular dynamics and constraint dynamics. Mol Phys 34:1311–1327

von Ossowski I, Eaton JT, Czjzek M, Perkins SJ, Frandsen TP, Schülein M, Panine P, Henrissat B, Receveur-Brechot V (2005) Protein disorder: conformational distribution of the flexible linker in a chimeric double cellulase. Biophys J 88:2823–2832

Wada M, Chanzy H, Nishiyama Y, Langan P (2004) Cellulose III$_I$ crystal structure and hydrogen bonding by synchrotron X-ray and neutron fiber diffraction. Macromolecules 37:8548–8555

Yui T, Nishimura S, Akiba S, Hayashi S (2006) Swelling behavior of the cellulose Iβ crystal models by molecular dynamics. Carbohydr Res 341:2521–2530

Zhang Y-HP, Himmel ME, Mielenz JR (2006) Outlook for cellulase improvement: screening and selection strategies. Biotechnol Adv 24:452–481