Protein Engineering, Design & Selection vol. 20 no. 4 pp. 179–187, 2007
Published online April 12, 2007 doi:10.1093/protein/gzm010

Molecular modeling suggests induced fit of Family I carbohydrate-binding modules with a broken-chain cellulose surface

Mark R. Nimlos[1,6], James F. Matthews[2], Michael F. Crowley[3], Ross C. Walker[4], Giridhar Chukkapalli[4], John W. Brady[2], William S. Adney[1], Joseph M. Cleary[4], Linghao Zhong[5] and Michael E. Himmel[1]

[1]National Renewable Energy Laboratory, Golden, CO 80401, USA, [2]Department of Food Science, Cornell University, Ithaca, NY 14853-7201, USA, [3]The Scripps Research Institute, La Jolla, CA 92037, USA, [4]San Diego Supercomputer Center, La Jolla, CA 92093-0505, USA and [5]Pennsylvania State University, Mont Alto, PA 17237, USA

[6]To whom correspondence should be addressed.
E-mail: mark_nimlos@nrel.gov

Cellobiohydrolases are the most effective single component of fungal cellulase systems; however, their molecular mode of action on cellulose is not well understood. These enzymes act to detach and hydrolyze cellodextrin chains from crystalline cellulose in a processive manner, and the carbohydrate-binding module (CBM) is thought to play an important role in this process. Understanding the interactions between the CBM and cellulose at the molecular level can assist greatly in formulating selective mutagenesis experiments to confirm the function of the CBM. Computational molecular dynamics was used to investigate the interaction of the CBM from *Trichoderma reesei* cellobiohydrolase I with a model of the (1,0,0) cellulose surface modified to display a broken chain. Initially, the CBM was located in different positions relative to the reducing end of this break, and during the simulations it appeared to translate freely and randomly across the cellulose surface, which is consistent with its role in processivity. Another important finding is that the reducing end of a cellulose chain appears to induce a conformational change in the CBM. Simulations show that the tyrosine residues on the hydrophobic surface of the CBM, Y5, Y31 and Y32 align with the cellulose chain adjacent to the reducing end and, importantly, that the fourth tyrosine residue in the CBM (Y13) moves from its internal position to form van der Waals interactions with the cellulose surface. As a consequence of this induced change near the surface, the CBM straddles the reducing end of the broken chain. Interestingly, all four aromatic residues are highly conserved in Family I CBM, and thus this recognition mechanism may be universal to this family.

Keywords: biomass/cellulase/cellulose/induced fit/molecular dynamics

Introduction

Cellobiohydrolases (CBH-EC 4.2.1.91) are known to display a multi-domain structure (Teeri *et al.*, 1992) characterized by a catalytic domain (core) and a carbohydrate-binding module (CBM) separated by a linker peptide (Srisodsuk *et al.*, 1993) rich in proline, serine and threonine (Teeri *et al.*, 1987). The non-catalytic CBMs are recognized as being an essential component of effective cellulase action on cellulose substrate (Gilkes *et al.*, 1988; Reinikainen *et al.*, 1991; Reinikainen *et al.*, 1992; Kruus *et al.*, 1995) and are thought to have three primary functions including proximity effects, substrate targeting and microcrystallite disruption (Kuutti *et al.*, 1991; Linder and Teeri, 1997; Boraston *et al.*, 2004). CBMs function as a means to promote the association of the enzyme with the substrate and subsequently increase the effective cellulase concentration (proximity effect) (Reinikainen *et al.*, 1991, 1992). CBMs have been shown to have selective affinities for substrates including crystalline and amorphous celluloses, and various soluble and non-soluble polysaccharides (targeting function) (Lamed *et al.*, 1994; Tomme *et al.*, 1995; Creagh *et al.*, 1996; Tormo *et al.*, 1996; Linder and Teeri 1997; Henshaw *et al.*, 2004). In addition, some CBMs appear to disrupt the structures of the carbohydrate ligands to render the substrate more susceptible to enzymatic function (disruptive function) (Din *et al.*, 1994). Of particular interest to biomass conversion are CBMs that target crystalline cellulose (Phelps *et al.*, 1995; Teeri, 1997; Boraston *et al.*, 2004), which forms the core of carbohydrate microfibrils that provide structure and strength to plant cell walls. It has long been recognized that crystalline cellulose is recalcitrant to enzymatic hydrolysis, limiting its conversion rate in many processes based on the production of fermentable sugars from plant biomass.

Family I CBMs, which are entirely fungal, are perhaps the most interesting. Of particular interest is the CBH I CBM produced by *Trichoderma reesi*, the most common source of commercial cellulases today. CBH I is a 'molecular machine', that is thought to be processive (Rouvinen *et al.*, 1990; Vrsanska and Biely, 1992; Barr *et al.*, 1996), moving along a crystalline cellulose chain, 'pulling up' that chain and feeding it into the catalytic domain where cellobiose is formed by hydrolyzing alternate β-(1,4) glycosidic linkages. This exoglucanase is most efficient at hydrolyzing crystalline cellulose when mixed with endoglucanases, whose apparent role is to hydrolyze random single glycosidic linkages. CBH I is known to act upon the reducing end of an already broken chain of cellulose (Vrsanska and Biely, 1992). The processivity of these enzymes makes them highly active and attractive for bioprocessing crystalline cellulose found in plant cell walls. However, details of the functions of the different components of these enzymes during processivity remain unclear, if not controversial.

The structures of Family I CBMs are thought to display specificity for binding to crystalline cellulose through three aromatic amino acid side groups located on a relatively planar surface of the protein. These residues are nearly co-linear, and the distance between them is about the same as the cellobiose unit length (~1.1nm) in the cellulose

substrate. The surface containing these aromatic residues is generally regarded as hydrophobic and thought to match the (1,0,0) surface of crystalline cellulose 1β (Lehtio *et al.*, 2003), which exposes the faces of β-D-glucopyranose rings in the chair conformation. In this conformation, the α and β faces of glucopyranose rings have either two or three axial hydrogen atoms exposed, whereas the ring hydroxyl groups are in the equatorial position. As a result, the hydrophobic surfaces of these CBMs are well suited to bind to the hydrophobic (1,0,0) surface of crystalline cellulose. These hydrophobic interactions dominate the free energy of binding, which is primarily entropically driven with a small favorable enthalpic contribution (Creagh *et al.*, 1996). Experimentally, there is no evidence for strong hydrogen bonding (Boraston *et al.*, 2004). The increased entropy upon binding Family I CBMs to cellulose results from a liberation of structured water above the hydrophobic (1,0,0) surface of crystalline cellulose followed by the direct positioning of the CBM on the dehydrated cellulose surface (Boraston *et al.*, 2004).

Native crystalline cellulose, the substrate for CBH I, primarily exists in two forms: cellulose Iα and Iβ. These allomorphs are recalcitrant to hydrolysis because of their enhanced hydrogen-bonding networks (Nishiyama *et al.*, 2002, 2003). Higher plants targeted for biomass conversion typically have cellulose Iβ in the cell wall elementary microfibrils. According to crystallographic studies (Nishiyama *et al.*, 2002), cellulose Iβ has a monoclinic $P2_1$ structure with two cellobiose chains in each unit cell. The cellulose sheets are composed of linear cellulose chains bounded by the inter-chain hydrogen-bonding interactions between the glucosyl O6H (donor) in one chain and the O3 (acceptor) in the neighboring chain. These strong inter-chain hydrogen-bonding interactions are most likely responsible for the recalcitrant nature of the native crystalline cellulose.

Computational modeling of CBH I and its CBM allows an investigation of the chemical interactions that lead to processivity, binding and disruption of cellulose crystals. Earlier molecular dynamics (MD) modeling work was conducted on the CBM interacting with cellulose surface in vacuum (Kuutti *et al.*, 1991) and on the CBM alone in water (Hoffren *et al.*, 1995). More recently, docking calculations have been conducted with the CBM and a cellulose chain (Mulakala and Reilly, 2005). In the study reported here, the interactions of Family I CBM from *T. reesei* CBH I with a singly hydrolyzed cellulose surface (one in which a single chain of cellulose has been broken by hydrolysis) in a box of explicit water molecules is investigated using MD simulations. Four simulations are discussed here, in which the CBM was placed in close proximity to the reducing end of a broken chain of cellulose to capture molecular details of this interaction and to determine if the CBM aids decrystallization.

Materials and methods

The MD simulations in this study followed the temporal evolution of systems containing a slab of three sheets of cellulose and a CBM enclosed in a box of water molecules. The CHARMM suite of molecular mechanics software (Brooks *et al.*, 1983) (version 31b1) was used to set up, run and analyze the simulations. The MD simulations were run on IBM power-4-based compute nodes. Further analysis and visualization was conducted using the VMD program (Humphrey *et al.*, 1996). A standard force field was employed for the protein (MacKerell *et al.*, 1998), and water molecules were treated using the TIP3P model (Jorgensen *et al.*, 1983; Durell *et al.*, 1994). A carbohydrate force field developed by Brady and coworkers (Palma *et al.*, 2000; Kuttel *et al.*, 2002) was used for the cellulose. Simulations were conducted with the Verlet integration algorithm at a temperature of 300 K, a time step of 2 fs and a 14 Å cutoff of non-bonding interactions. In each simulation, the structure was prepared as described later and minimized for 500 steps. After minimization, a 1000-step heating MD simulation was conducted followed by a 10 000-step MD simulation to allow equilibration.

The initial structure of the cellulose was taken from neutron diffraction measurements of cellulose 1β (Nishiyama *et al.*, 2003). The cellulose slab used in this study was constructed of a sheet of crystalline origin chains sandwiched between two sheets composed of center chains. In each sheet, there were five cellulose chains with 12 glucose molecules in each chain. The initial structure of the CBH-I-binding domain was based on the results from NMR studies (Kraulis *et al.*, 1989). Simulations were started with the aromatic residues on the bottom of the CBM within 3 Å of the cellulose surface. A box of water molecules ($68 \times 48 \times 40$ Å^3) was added to the system so as to completely enclose the cellulose slab and the CBM. This box of water had an initial density of 0.033 molecules Å^{-1}. Figure 1 shows a representation of the components of a typical simulation, where the backbone of the protein is shown with a green ribbon, the three layers of cellulose are shown with a stick structure and the water molecules are shown as dots. The three tyrosine residues on the hydrophobic surface of the CBM are shown in orange. A series of 26 image boxes were placed around the simulation box to account for atoms that drift out of the boundary of the box. The SHAKE algorithm was used to constrain all bonds involving hydrogen atoms during MD simulations. In addition, the glucosyl residues on the periphery of the cellulose slab were fixed in space, whereas the interior glucosyl units were allowed to move. This prevented the sheets of cellulose from delaminating.

The results from five simulations are reported here, one without the CBM and four with it. To aid discussion, a numbering system was devised for the cellulose surface, and

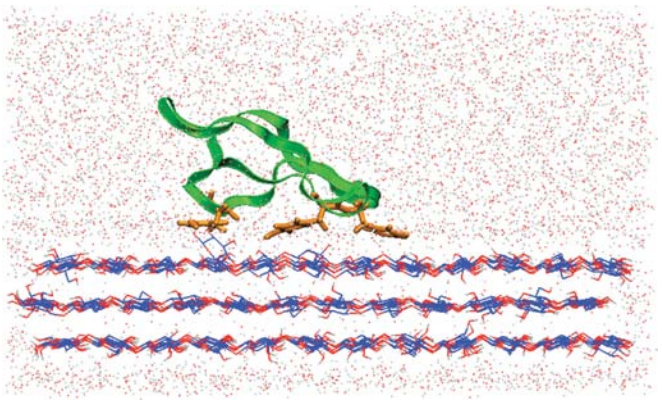

Fig. 1. View of the molecular components of a typical MD simulation of the CBM acting on cellulose. The binding domain, shown as a green ribbon with the tyrosine residues in orange, is placed ~3 Å above a three-sheet slab of cellulose in a box of water molecules (dots).

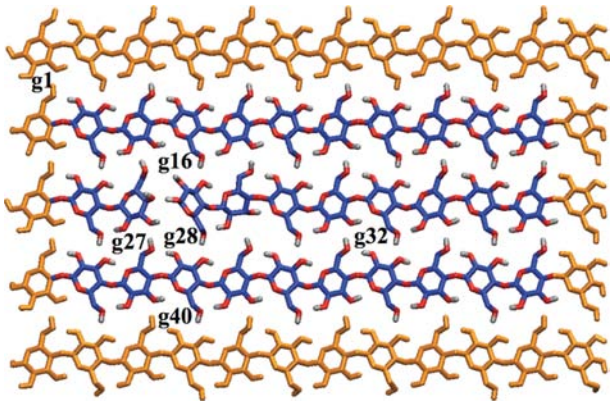

Fig. 2. View from top of cellulose surface showing the numbering scheme for the glucose residues. Starting from the top going to the left, they are numbered sequentially from 1 to 60. The orange glucose residues were restrained to their initial configuration during the simulation.

Fig. 2 shows the residue numbering system of a portion of the surface. In all the simulations, the center chain of the top sheet of the cellulose slab was hydrolyzed between the third (g27) and fourth (g28) glucans. The simulation without the CBM was run for 12 ns to allow the broken-chain model to equilibrate. Two hypotheses (Teeri *et al.*, 1992) have been proposed concerning the function and placement of the CBM from CBH I during processivity. The first hypothesis has the CBM riding on the cellulose surface in front of the cellulose chain being extracted. The second (Knowles *et al.*, 1987) has the CBM wedged under the cellulose chain, helping to remove it from the crystal. This hypothesis has been investigated recently (Mulakala and Reilly, 2005) using molecular docking. In the study reported here, the first hypothesis was investigated using MD modeling, and future studies of the second hypothesis are planned. The CBM was placed in front of the broken chain with the 'tall end' of the wedge shape (corresponding to the 24–36 amino acid loop) facing the break. Because of the size of the system (~13 000 atoms) and the recalcitrance of crystal cellulose, it was thought that long simulation times were required to identify interactions between the CBM and cellulose. In order to obtain meaningful results, dozens of simulations were conducted, with different substrate structures and different alignments of the CBM and typically short (1–3 ns) simulation times. Of this survey, the results from four representative simulations that were run for long simulation times are reported. The four simulations with the CBM, named Simulations I–IV, were run for at least 9 ns with Simulation II being run for a further 2.3 ns to allow further analysis of interesting behavior.

Results and discussion

Cellulose

Before investigating the interactions of the CBM with cellulose, simulations were conducted with a singly hydrolyzed cellulose substrate alone in water. This allowed the cellulose structure to equilibrate to a solvated conformation. After the glucosidic ether linkage was hydrolyzed, there were two hydroxyl groups located in the space that formerly contained an oxygen atom and this crowding resulted in a puckering of one or both of the terminal glucose residues. A potential energy minimization was conducted to allow relaxation of this packing frustration. This forced the glucose molecule at the reducing end slightly out of the plane of the crystal. After this initial minimization, a heating and equilibration simulation was conducted, followed by an MD simulation that ran for 6 000 000 steps or 12 ns. During this simulation, the reducing end and the non-reducing end created from hydrolysis twist out of the plane initially, but then the non-reducing end reannealed back into the plane. Figure 3 shows a side view of the top layer cellulose at different times during the simulation. Note that at the beginning of the simulation, the reducing end is slightly puckered out of the plane. At 4 ns, the non-reducing end is twisted out of the plane and the primary alcohol on the reducing end is pointing out of the plane. After 8 ns, the non-reducing end reannealed back into the plane and there was little change in the cellulose after this. This result suggests that the system had equilibrated.

The twisting of these two residues results in the breaking of the hydrogen bonds with adjacent chains. This can be seen in Fig. 4 which plots the inter-atomic distances between the hydrogen-bonding partners on the reducing end, g28, and the glucose residues on adjacent chains, g40 and g16. The hydroxyl group on C6 of g28 can form hydrogen bonds with the C2 and C3 hydroxyl groups on g40 and there are four possible hydrogen bonds, g28O6H—g40O3, g28O6H—g40O2, g28O6—g40O2H and g28O6—g40O3H. We have found that only the first three of these hydrogen bonds are observed in these simulations, and the distances between these atoms are plotted in Fig. 4a. The plot in Fig. 4b shows the distances between g16O6 and g28O2 and g28O3. Hydrogen bonds exist when these distances are ~2 Å. Notice that at the beginning of the simulation (0 ns), the reducing end is slightly puckered out of the plane. Figure 4b shows that O2 and O3 on the reducing end, g28, maintain hydrogen bonds to the primary alcohol on g16 throughout the simulation. Note that the predominant hydrogen bond throughout the simulation is g28O6H . . . g40O3. Interestingly, other hydrogen bonds can simultaneously form the g28O6 hydroxyl group during the simulation. This is due

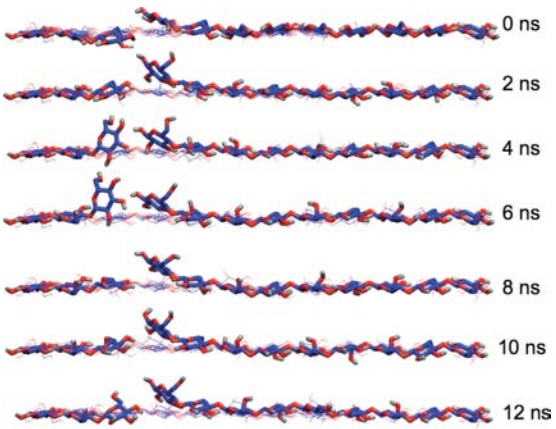

Fig. 3. Evolution of the broken chain of cellulose as a function of simulation time. This picture shows snapshots of the top layer of the cellulose slab captured during the 12 ns simulation. For simplicity, water molecules are not shown. The chain with the hydrolyzed bond (g25–g36 in Fig. 2) is shown in a licorice representation, whereas the other chains are shown in a thin line representation.

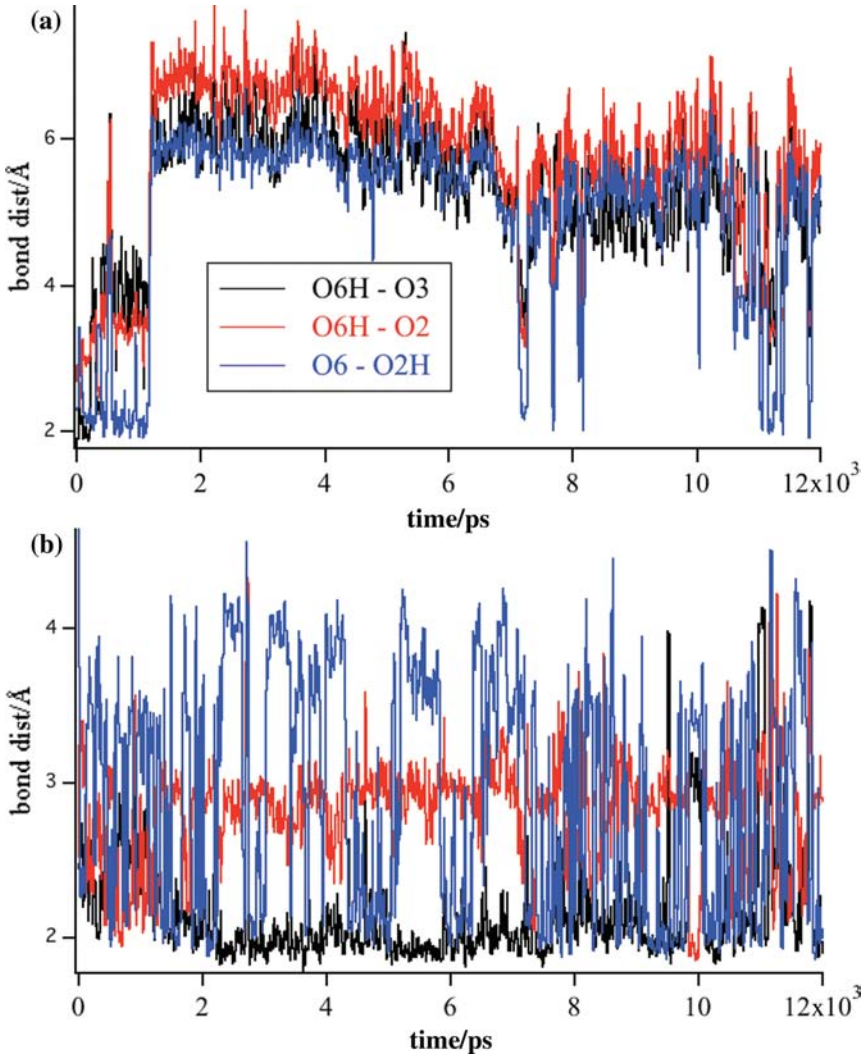

Fig. 4. Plots of the distances between potential hydrogen-bonding partners on (**a**) g28 and g40 (Fig. 2) and (**b**) g16 and g28. For instance, the black trace in (a) is the distance between the H atom of the O6 hydroxyl group on g28 and O3 on g40. Distances of ~2 Å indicate that a hydrogen bond is formed.

to the flexibility of the reducing end. Hydrogen bonds also exist between the primary alcohol on g28 and O2 and O3 on g16 as is shown in Fig. 4a. However, after ~1 ns, these bonds are broken and remain broken throughout the rest of the simulation. This is consistent with the primary alcohol on the reducing end twisting out of the plane.

As a consequence of the equilibration process, the surface glucose residues away from the break changed their hydrogen-bonding patterns with adjacent residues from O6H—O3 [the crystal-bonding pattern (Nishiyama *et al.*, 2002)] to O6H—O2. This type of rearrangement of hydrogen bonding of surface residues was also observed in an earlier, more comprehensive study of cellulose (Matthews *et al.*, 2006). In the middle sheet, the inter-chain hydrogen bonding essentially remained O6H—O3 throughout the simulation. The change in the hydrogen bonding on the surface was due primarily to a change in the conformation of the primary alcohol from TG (the C6—O6 bond was *trans* to the C5—O5 bond) to 80% GG (C6—O6 *gauche* to both C5—C4 and C5—O5) and 16% GT. In the middle sheet, the conformation remained primarily TG.

When the reducing end twisted out of the plane of the cellulose, the inter-chain hydrogen bonds were replaced with hydrogen bonds to water molecules. This observation was confirmed by counting the hydrogen bonds, which are defined as a donor–acceptor distance less than 3.4 Å and an donor–H–acceptor angle less than 150°. During the simulation, the reducing end (g28) had an average of 4.1 hydrogen bonds to water molecules and 2.0 hydrogen bonds to other glucose residues. Four glucose residues away from the break, the residue g32 had an average of 1.9 hydrogen bonds with water molecules and 2.4 hydrogen bonds with other glucose residues. Another consequence of the twisting of the reducing end was that water molecules had access to the space beneath it. At the start of the simulation, no water molecules were placed under the reducing end. After ~3 ns of simulation, an average of four water molecules were found in this location for the duration of the simulation.

Cellulose with CBM

Four simulations are discussed here, in which the CBM from CBH I was added to the equilibrated cellulose described earlier. These simulations were started with the tall end of the wedge near the break in the cellulose chain. The beginning alignment and the alignment after 9 ns of MD simulation are shown in Fig. 5 for the four simulations. The

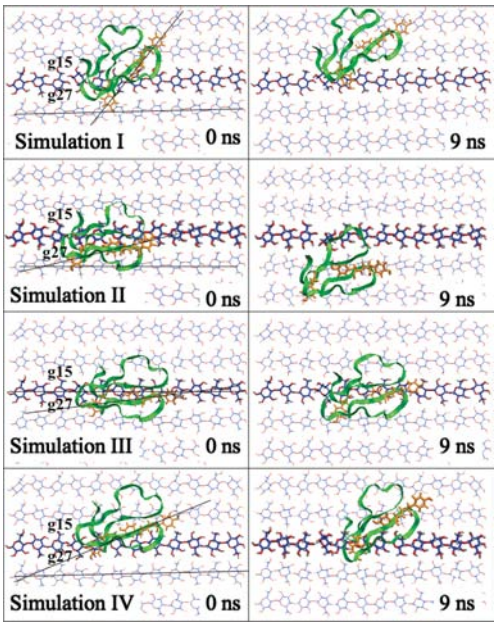

Fig. 5. Four simulations considered for this study at the start and after 9 ns. The backbone of the protein is shown as a green ribbon and the three tyrosine residues on the hydrophobic face are shown in orange. The angles made by these residues relative to the direction of the cellulose chains are shown by the black lines. The non-reducing end of the broken chain (g27) is also shown.

CBM was placed with the line of the three tyrosines at different angles with respect to the direction of the cellulose chains. Lines are drawn in the figure to indicate this angle. For Simulations I–IV, this angle was 49°, 7°, 15° and 20°, respectively. As shown in Fig. 1, during Simulations I and II the CBM underwent significant movement over the cellulose surface. In these simulations, the center of mass of the protein moved a maximum of 9 and 6 Å, respectively, whereas the center of mass of the protein only moved a maximum of 4 Å for Simulations III and IV. During all the simulations, the hydrophobic face defined by the three tyrosines maintained contact with the cellulose surface. These observations suggest that it is facile for the CBM to translate while remaining bonded on a crystalline cellulose surface. Processivity would necessarily require that the CBM remain fixed on the substrate surface, but be able to freely translate as the enzyme moves down a chain when sufficient force is applied to it.

Although the CBM appeared capable of significant translation on the cellulose surface, some of the simulations did not suggest a trend to this movement. For example, there did not appear to be a tendency in some of the simulations for the aromatic rings of the tyrosines to line up with the cellulose chains. If CBH I is to process along a cellulose chain, one would expect the CBM to also align with the chain to be hydrolyzed. However, Simulations I and IV showed no tendency for such alignment with the cellulose chains. On the other hand, Simulation III, which started out nearly parallel to the center cellulose chain, did roughly maintain this position. Simulation II also showed some tendency for the tyrosines to align with the chain adjacent to the broken chain, and after ~3.5 ns, the tyrosines were indeed aligned with this chain, so that the aromatic rings stacked on top of alternate sugar rings. The CBM maintained essentially this configuration throughout the remaining 9 ns of Simulation II. Future simulations are needed to characterize the association of the CBM with a single or multiple chains of cellulose.

Induced fit of CBM

In addition to aligning with the cellulose chain, Simulation II also showed that the CBM is capable of structural accommodation, possibly in response to the reducing end of a cellulose chain. This distortion is shown in Fig. 6. As mentioned above, within 3.5 ns the CBM moves so that the three tyrosines (Y5, Y31 and Y32) on the hydrophobic surface align with the cellulose chain adjacent to the broken chain. In this conformation, Y5 stacks on glucose residue g39, which is adjacent to the non-reducing end, Y31 stacks on g41 and Y32 stacks on g43. Figure 6b shows a snapshot of the simulation after 4.5 ns. Once the protein is in this position, the fourth tyrosine (Y13) unfolds from its internal location in the CBM and forms a hydrophobic or van der Waals interaction with the cellulose surface on the other side of the broken chain. The aromatic ring in Y13 stacks roughly above the aliphatic hydrogen atoms on C6 of residue g29. This transformation is complete after 5.8 ns, and Y13 remains in this position through the remaining 5.5 ns of the simulation. The surface-induced conformational structure is shown in Fig. 6c, which shows the CBM structure after 7.1 ns. As this figure shows, the CBM straddles the reducing end of the cellulose chain. This type of deformation and added van der Waals interaction with the cellulose was not observed in any of the other CBM simulations conducted for this study, but additional simulations starting with this distorted conformation retained this structure for several nanoseconds. Furthermore, Simulation III showed a similar distortion without binding between Y13 and the surface.

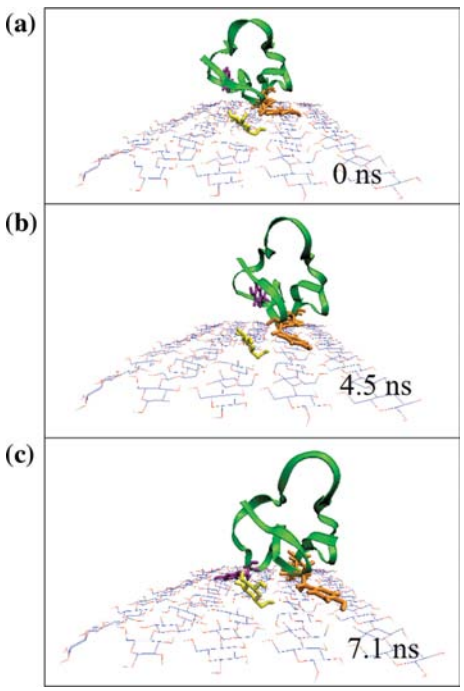

Fig. 6. Structures showing the rearrangement of the CBM during Simulation II. Notice tyrosine Y13 in purple is (**a**) faced into the protein at the start and (**b**) after the three other tyrosine residues align with the cellulose chain, but faces the cellulose in (**c**).

Note that in Simulation II the reducing end of the broken chain (g28) has twisted the opposite direction from the other simulations. In Simulation II, O6 on g28 remains hydrogen-bonded to the adjacent glucose residue (g40), whereas O2 and O3 of g28 have broken their hydrogen bonds with g16 and they point out of the plane. In the other CBM simulations and the simulation without the CBM, the O6 on g28 is not hydrogen-bonded to g40 and O2 and O3 on g28 remain hydrogen-bonded to g16.

The rearrangement of the CBM to encapsulate the reducing end of the broken chain strongly suggests that there may be an induced fit substrate/enzyme interaction. That is, a possible mechanism of recognition and specificity of the CBM is the rearrangement of the CBM to fit the reducing end of a cellulose chain. The odd twist of the cellulose chain further suggests that the substrate may also need to deform in order for this fit to occur. Induced-fit substrate–enzyme interactions are well-known phenomena (Koshland, 1958; Williamson, 2000; Gutteridge and Thornton, 2005), although this has not been suggested for Family I CBMs. Interestingly, the four aromatic residues in this study Y13, Y5, Y31 and Y32 are highly conserved in this family. Figure 7 shows a comparison of the sequences for a variety of Family I CBMs and shows that these four tyrosine residues match tyrosine, tryptophan or phenylalanine in the other family I CBMs. Thus, it seems likely that CBMs in this family have the ability to distort in a manner similar to that shown in Fig. 6. It is not clear why this rearrangement was only seen in one simulation. It is possible that this response was not observed in the other simulations due to insufficient conformational sampling (simulation time) required to bring the various initial orientations of CBM and cellulose chain into biologically relevant position. Ongoing work is aimed at further investigating this interaction and quantifying its lifetime and probability.

The structural changes induced in the CBM during Simulation II are consistent with the flexibility of the loop containing Y13 in this otherwise rigid protein. Figure 8 shows a cartoon representation of the protein structure with its three-strand β-sheet. In the β-sheet, strand β3 (residues 33–36) is hydrogen-bonded to strands β1 (residues 5–9) and β2 (residues 24–28). The rigidity of the β-sheet is further strengthened by a disulfide bridge between β1 (residue 8) and β2 (residue 25). A small loop between β2 and β3 is restricted because of hydrogen bonds formed between the backbone carbonyls and amides in this group and the side chain on an asparagine, N29. The β-sheet and this tight loop contain three tyrosine residues that make up the hydrophobic surface of CBM I. This rigidity allows the CBM to maintain the spacing of the tyrosine rings to match the cellobiose periodicity in cellulose. A second disulfide bridge between residues 19 and 35 adds rigidity to the loop between residue 19 and the start of β2 (residue 24). The remainder of this loop (residues 10–19) is more flexible.

Plots of the spatial deviation of each residue during Simulations I–IV show the rigidity of different sections of the CBM. During each step of the simulation, the backbone of the protein is aligned to the starting structure and the average of the deviation of each individual amino acid for each simulation was calculated, i.e. the root mean squared deviation (RMSD) over all atoms in each residue was determined for each time step in the simulation. Figure 9a and b show plots of the average RMSD for each residue during each simulation. We note that there was considerably more movement of the residues 11–17, which is in the loop between β1 and β2 and contains Y13. The flexibility of this section of the CBM allowed movement of the fourth tyrosine and enabled it to bind to the hydrophobic cellulose surface near the reducing end. The plot in Fig. 9c shows the RMSD for each residue for Simulation II averaged over the starting 2 ns (red) of the simulation and the last 2 ns of the simulation (black), after the protein has distorted. From this plot, one can clearly see that the fourth tyrosine (Y13) clearly underwent a large displacement after the protein distorts.

There was an indication of significant movement of Y13 in some of the other simulations. In the NMR study of this protein (Kraulis *et al.*, 1989) and at the start of each simulation, the OH group of Y13 formed a hydrogen bond to the O atom on residue 16, P16. This hydrogen bond appeared to help hold Y13 in its folded position and remained intact (∼2 Å) throughout Simulation IV and through most of Simulation I, where it periodically stretched to ∼4 Å. By comparison, this hydrogen bond stretched to over 15 Å in Simulation II after Y13 bonded to the cellulose surface.

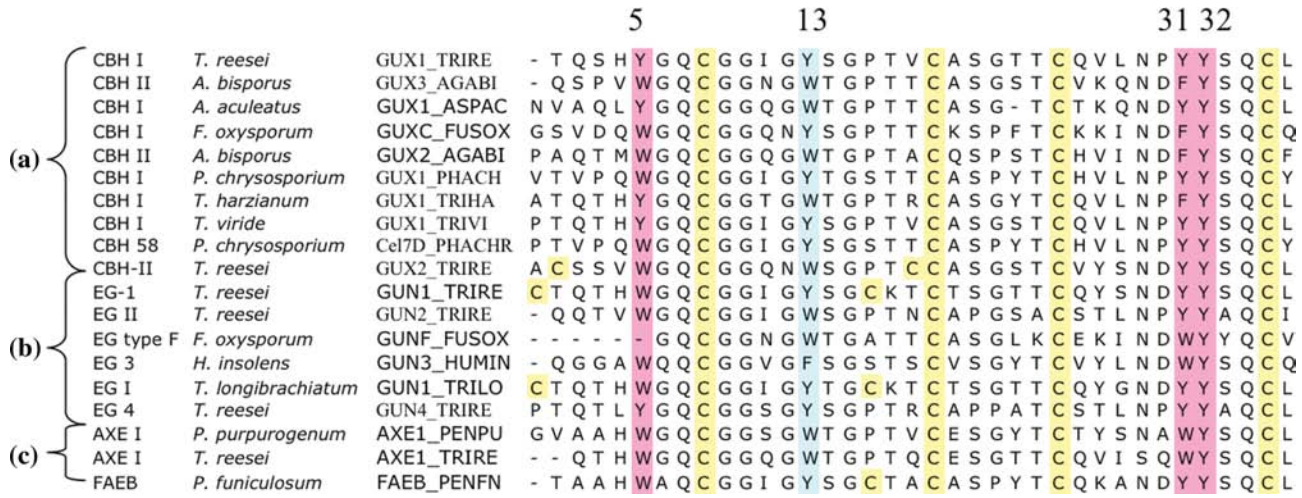

Fig. 7. Amino acid sequences for several Family I CBMs: (**a**) exoglucanases, (**b**) endoglucanases and (**c**) esterases.

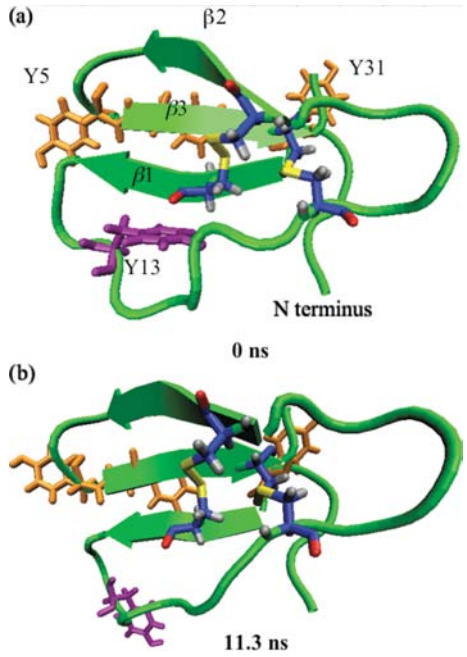

Fig. 8. A picture showing the beta sheets and the disulfide bonds (yellow) (**a**) before and (**b**) after (right) Y13 (purple) has unfolded. The three tyrosines on the hydrophobic face are shown in orange.

Simulation III shows a similar deformation of Y13 as was observed in Simulation II and this hydrogen bond is at times stretched to nearly 10 Å. However, in Simulation III, the fourth tyrosine did not appear to bond to the cellulose surface and after ~1 ns it returned to hydrogen bond with P16. These distortions are much larger than those reported in the NMR study (Kraulis *et al.*, 1989) of this CBM, which reported the largest RMSD distortions (1–2 Å) of residues 13–15. However, the hydrogen bond never stretches beyond 3 Å in this solution-phase study. The flexibility of this portion of the loop was also confirmed with MD simulations (Hoffren *et al.*, 1995) of the CBM in water, where structural variation was found in residues 12–16. Because the NMR study and the earlier MD simulations were conducted with the protein neat in water, the results reported here suggest a greater distortion of the protein in the presence of the cellulose substrate.

The hydrophobic interaction of Y13 with the cellulose surface can be seen by comparison of the density of water near the cellulose surface before and after this residue has shifted. Figure 10 shows all water molecules within 5 Å of the cellulose surface before and after this shift. At 4.5 ns, before Y13 has shifted (Fig. 10a), water molecules coat the cellulose surface except underneath the hydrophobic surface of the CBM. This exclusion of water is an indication of the interaction between the CBM and the cellulose surface. At this point in the simulation, the fourth tyrosine residue, Y13,

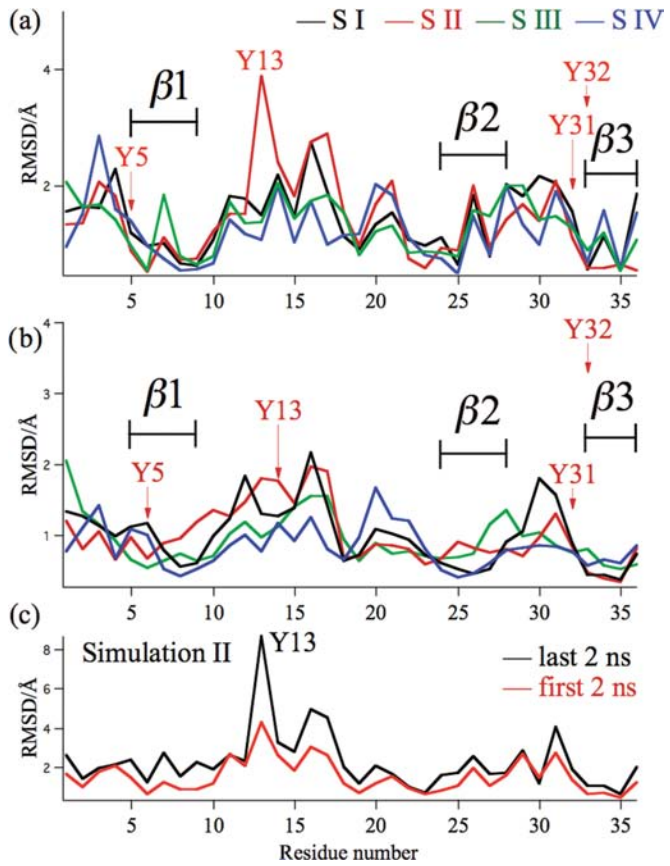

Fig. 9. Movement of individual residues in CBM during Simulations I–IV. The RMSD of (**a**) each residue, (**b**) the backbone and (**c**) each residue in Simulation II at the start (red) and towards the end (black), after the protein has distorted.

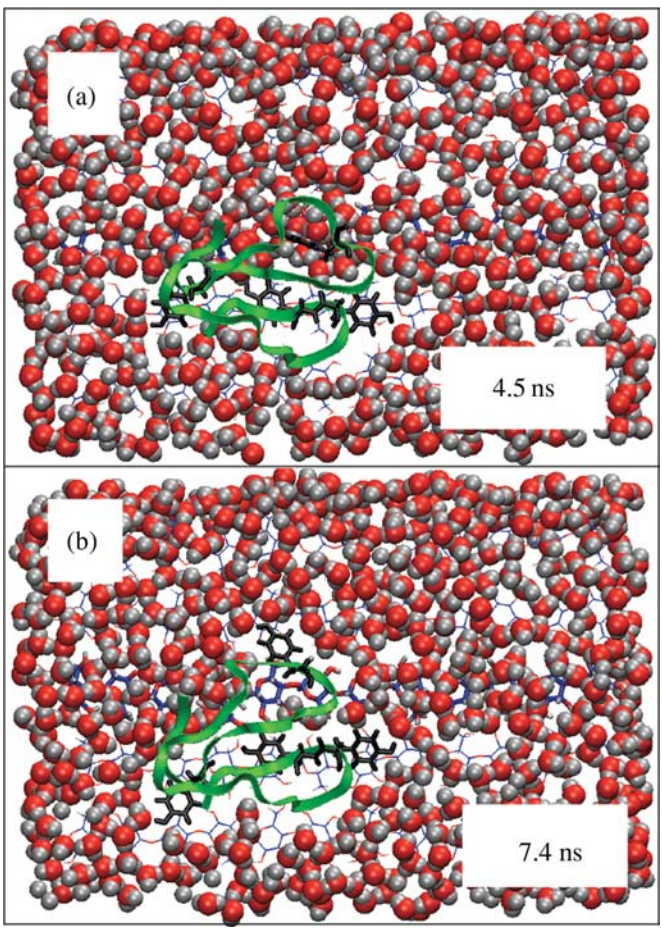

Fig. 10. A picture showing water molecules within 5 Å of the cellulose surface before (**a**) and after (**b**) the fourth tyrosine residue has unfolded.

is above this layer of water molecules. Figure 10b shows the water layer at 7.4 ns after tyrosine Y13 has shifted and is located above C6 on g29. Water is now excluded from above the glucose residues g39 to g41 as before, as well as g29 and g30. Thus, bonding of this open form of CBM to the cellulose surface appears to involve hydrophobic interactions between all four tyrosine residues and the cellulose. These hydrophobic interactions result in the exclusion of water molecules near the reducing end.

By encircling and enclosing the reducing end of the cellulose chain as shown in Figs. 6 and 10, the CBM may be in a position to aid the extraction of the cellulose chain. In the absence of water molecules, the hydroxyl groups on the reducing end could form hydrogen bonds with the amino acids in the fold of the protein. For example, occasional hydrogen bonding was observed between the Q7 glutamine of the protein and O3 on g28. There is also the possibility that the CBM could spontaneously remove the broken chain from the surface by forming multiple hydrogen bonds to the hydroxyl group on g28 and g29. This could be a mechanistic explanation for the disruptive function of the CBM (Din *et al.*, 1994) and is the basis of ongoing research.

Nature of CBM/cellulose binding

As discussed in the Introduction, the CBM from CBH I forms weak hydrophobic interactions with the cellulose substrate. The enthalpy of binding of the CBM to the cellulose surface is low (Creagh *et al.*, 1996) because there are few hydrogen bonds formed between the cellulose and the protein. During the four simulations discussed here, there were an average of between zero and two hydrogen bonds formed between the CBM and the cellulose surface. These hydrogen bonds add to the enthalpy of binding. This is consistent with experimental observations (Reinikainen *et al.*, 1992; Linder *et al.*, 1995; Reinikainen *et al.*, 1995) reported in the literature, which show that substitution of some of the amino acids on the CBM affect its binding to crystal cellulose. Our simulations showed that amino acids near the bottom of the CBM formed hydrogen bonds with the hydroxyl groups on the cellulose surface. For example, we found that glutamine Q34 formed transient hydrogen bonds to the surface which is consistent with experimental measurements (Reinikainen *et al.*, 1992, 1995; Linder *et al.*, 1995) of the binding isotherm of this CBM where Q34 was replaced with an alanine, and the binding affinity was subsequently decreased. We also observed hydrogen bonding between the hydroxyl groups of the cellulose and tyrosines Y5, Y31, Y32, glutamine Q7, asparagine N29, and after the rearrangement shown in Fig. 6, glycine G12. In addition to these hydrogen bonds formed with the flat surface of cellulose, hydrogen bonds also formed with the reducing end of the broken chain. Because this terminus extends out of the crystal, hydrogen bonds could readily be formed with histidine H4, serine S3 and glutamine Q2 on the 'tall end' of the wedge.

Table I shows the average number of hydrogen bonds that were formed between selected glucose residues on the cellulose surface and water molecules, other glucose residues and amino acids on the protein. Comparison between the simulation of cellulose without the CBM and Simulations I–IV shows that there were a similar number of hydrogen bonds between reducing terminal glucose residue, g28, and water molecules or g28 and other glucose residues. However, there were added hydrogen bonds to the protein. The exception is Simulation II, in which there were notably fewer hydrogen bonds to the other glucose residues. This could be an indication that the CBM enhances decrystallization. This table also shows the number of hydrogen bonds formed to glucose residues away from the reducing end (g29–g32) during Simulation III. The CBM remains positioned over these residues throughout this simulation. Comparison of the hydrogen bonding to these residues and to residue g32 in the simulation of cellulose without the CBM shows that the protein had only a slight affect on hydrogen bonding, with a small number of hydrogen bonds between the glucose residues and the protein. This is also consistent with a separate analysis that showed little effect of the protein on the conformation of the primary alcohol groups on the surface of the cellulose (data not shown).

Table I. Number of hydrogen bonds formed to glucose residues

	Residue[a]	Hydrogen-bonding partner		
		Water[b]	Glucose[c]	Protein[d]
Cellulose without CBM	g28	4.1	2.0	
	g32	1.9	2.4	
Simulation I	g28	3.7	2.0	0.4
Simulation II	g28	3.6	1.2	0.4
Simulation III	g28	4.0	2.0	0.4
	g29	0.7	2.7	0.0
	g30	1.6	2.6	0.1
	g31	0.8	2.9	0.0
	g32	1.0	2.8	0.1
Simulation IV	g28	4.3	1.9	0.1

[a]Residues are defined in Fig. 2.
[b]Number of hydrogen bonds between the residue and water molecules.
[c]Number of hydrogen bonds between the residues and all other glucose residues.
[d]Number of hydrogen bonds between the residue and the protein.

Conclusions

We believe that we are the first to show, using computational simulation, that Family I CBMs in proximity to cellulose are capable of distorting so that the fourth (and remaining) tyrosine can establish hydrophobic interactions with the cellulose surface. Now that this apparent induced fit mechanism has been observed for the CBH I CBM, one may conclude, considering the high homology demonstrated by other Family I CBMs, that all type 1 CBMs are capable of this conformational change. The interaction of four tyrosines with cellulose is entirely consistent with the CBMs found in bacterial enzymes; however, this interaction has not been reported for fungal CBMs. That this rearrangement only occurs in one simulation may be an indication that in the other simulations, the sampling was not long enough to correct for less productive starting orientations of CBM and cellulose. The observation that in Simulation II distortion occurred after the reducing end twisted in the opposite direction from the other simulations may provide an indication of the substrate conformations necessary to induce this conformational change.

In more than one of the MD simulations, significant motion of the CBM was observed. The observation that the

CBM translates freely and randomly on the cellulose surface is potentially of great interest for understanding the mechanism of processivity. This translation occurred in both perpendicular and parallel directions relative to the cellulose chains and appears to be undirected. This implies that the entire enzyme complex is required for processivity. The ease of two-dimensional translation is thus likely a requirement for efficient location of and interaction with reducing ends on the cellulose surface. Such translational freedom would be necessary in the event the enzyme adsorbs on the surface far from a reducing end.

The general nature of the interactions we observed for the Family 1 CBM with the cellulose surface using computer simulations reported here are consistent with traditionally held views of the role that hydrophobic interactions play in this process (see Introduction). The hydrophobic face of the three tyrosine rings stack on the hydrophobic (1,0,0) cellulose surface and there are only occasional hydrogen-bonding interactions between amino acids of the CBM and the hydroxyl groups of the unhydrolyzed glucose residues. The weakness of these individual van der Waals interactions is compensated for by the cumulative strength of the multiple hydrophobic coupling interactions and thus the CBM is tightly bound to the surface, while able to retain two-dimensional translational freedom. This critical property of translational freedom may be inhibited when the hydroxyl groups on a reducing end of a broken cellulose chain, which would now be puckered away from the surface, interact with the amino acids on the 1–5 loop of the CBM, or when the CBM distorts in response to the reducing end of the cellulose chain, enabling a new hydrophobic interaction with the cellulose surface using the fourth tyrosine residue. In this way, this CBM type may 'recognize' and respond productively to the reducing end of a cellulose chain.

In a recent CBM/cellulose computational docking study reported by Mulakala and Reilly (2005), the CBM was placed in a particular orientation relative to cellulose based on the premise that the CBM 'dives under' the top cellulose layer. This work illustrated possible sites for interaction, primarily H bonds, on the topside of the CBM with the displaced cellodextrin. In contrast, our present study permits the CBM to 'ride' near the broken cellodextrin lying in the cellulose surface and find, using MD, low-energy conformations. It is possible that both the surface decrystallization model proposed here and the cellulose reducing end entry model proposed by Mulakala and Reilly are consistent with biological function.

Acknowledgments

The authors thank Drs Mark Miller and Fran Berman for their support regarding the computational resources of the San Diego Supercomputer Center during the course of this work. Dr Charles Brooks and Dr Philip Mason also provided valuable comments and guidance to the authors. This research was supported in part by the National Science Foundation through the San Diego Supercomputer Center under grant SCI-0438741. The U.S. DOE Office of the Biomass Program also provided funding for this work.

References

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

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

Brooks,B.R., Bruccoleri,R.E., Olafson,B.D., States,D.J., Swaminathan,S. and Karplus,M. (1983) *J. Comput. Chem.*, **4**, 187–217.

Creagh,A.L., Ong,E., Jervis,E., Kilburn,D.G. and Haynes,C.A. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 12229–12234.

Din,N., Forsythe,I.J., Burtnick,L.D., Gilkes,N.R., Miller,R.C., Warren,R.A.J. and Kilburn,D.G. (1994) *Mol. Microbiol.*, **11**, 747–755.

Durell,S.R., Brooks,B.R. and Bennaim,A. (1994) *J. Phys. Chem.*, **98**, 2198–2202.

Gilkes,N.R., Warren,R.A.J., Miller,R.C. and Kilburn,D.G. (1988) *J. Biol. Chem.*, **263**, 10401–10407.

Gutteridge,A. and Thornton,J. (2005) *J. Mol. Biol.*, **346**, 21–28.

Henshaw,J.L., Bolam,D.N., Pires,V.M.R., Czjzek,M., Henrissat,B., Ferreira,L.M.A., Fontes,C. and Gilbert,H.J. (2004) *J. Biol. Chem.*, **279**, 21552–21559.

Hoffren,A.M., Teeri,T.T. and Teleman,O. (1995) *Protein Eng.*, **8**, 443–450.

Humphrey,W., Dalke,A. and Schulten,K. (1996) *J. Mol. Graph.*, **14**, 33–38.

Jorgensen,W.L., Chandrasekhar,J., Madura,J.D., Impey,R.W. and Klein,M.L. (1983) *J. Chem. Phys.*, **79**, 926–935.

Knowles,J., Lehtovaara,P. and Teeri,T. (1987) *Trends Biotechnol.*, **5**, 255–261.

Koshland,D.E. (1958) *Proc. Natl Acad. Sci. USA*, **44**, 98–104.

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

Kruus,K., Lua,A.C., Demain,A.L. and Wu,J.H.D. (1995) *Proc. Natl Acad. Sci. USA*, **92**, 9254–9258.

Kuttel,M., Brady,J.W. and Naidoo,K.J. (2002) *J. Comput. Chem.*, **23**, 1236–1243.

Kuutti,L., Laaksonen,L. and Teeri,T.T. (1991) *J. Chim. Phys. Chim. Biol.*, **88**, 2663–2667.

Lamed,R., Tormo,J., Chirino,A.J., Morag,E. and Bayer,E.A. (1994) *J. Mol. Biol.*, **244**, 236–237.

Lehtio,J., Sugiyama,J., Gustavsson,M., Fransson,L., Linder,M. and Teeri,T.T. (2003) *Proc. Natl Acad. Sci. USA*, **100**, 484–489.

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

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

MacKerell,A.D. *et al.* (1998) *J. Phys. Chem. B*, **102**, 3586–3616.

Matthews,J.F., Skopec,C.E., Mason,P.E., Zuccato,P., Torget,R.W., Sugiyama,J., Himmel,M.E. and Brady,J.W. (2006) *Carbohydr. Res.*, **341**, 138–152.

Mulakala,C. and Reilly,P.J. (2005) *Proteins Struct. Funct. Bioinform.*, **60**, 598–605.

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

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

Palma,R., Zuccato,P., Himmel,M.E., Liang,M.E. and Brady,J.W.(2000) In Himmel,M.E.(ed.), *Glycosyl Hydrolases in Biomass Conversion*. American Chemical Society, Washington, DC, pp. 112–130.

Phelps,M.R., Hobbs,J.B., Kilburn,D.G. and Turner,R.F.B (1995) *J. Bacteriol.*, **177**, 4356–4363.

Reinikainen,T.R., Ruohonen,L., Koivula,A., Srisodsuk,M., Jones,A., Knowles,J.K.C., Claeyssens,M. and Teeri,T.T. (1991) *Abstr. Pap. Am. Chem. Soc.*, **202**, 206–BIOT.

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

Reinikainen,T., Teleman,O. and Teeri,T.T. (1995) *Proteins*, **22**, 392–403.

Rouvinen,J., Bergfors,T., Teeri,T., Knowles,J.K.C. and Jones,T.A. (1990) *Science*, **249**, 380–386.

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

Teeri,T.T. (1997) *Trends Biotechnol.*, **15**, 160–167.

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

Teeri,T., Reinikainen,T., Ruohonen,L., Jones,T.A. and Knowles,J.K.C. (1992) *J. Biotechnol.*, **24**, 169–176.

Tomme,P., Driver,D.P., Amandoron,E.A., Miller,R.C., Antony,R., Warren,J. and Kilburn,D.G. (1995) *J. Bacteriol.*, **177**, 4356–4363.

Tormo,J., Lamed,R., Chirino,A.J., Morag,E., Bayer,E.A., Shoham,Y. and Steitz,T.A. (1996) *EMBO J.*, **15**, 5739–5751.

Vrsanska,M. and Biely,P. (1992) *Carbohydr. Res.*, **227**, 19–27.

Williamson,J.R. (2000) *Nat. Struct. Biol.*, **7**, 834–837.

Received November 14, 2006; revised; accepted January 23, 2007

Edited by Klaus Schulter