

Identification of Amino Acids Responsible for Processivity in a Family 1 Carbohydrate-Binding Module from a Fungal Cellulase

Gregg T. Beckham,^{*†} James F. Matthews,[‡] Yannick J. Bomble,[‡] Lintao Bu,[†] William S. Adney,[‡] Michael E. Himmel,[‡] Mark R. Nimlos,[†] and Michael F. Crowley^{*‡}

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

Received: September 11, 2009; Revised Manuscript Received: November 12, 2009

We probe the molecular-level behavior of the Family 1 carbohydrate-binding module (CBM) from a commonly studied fungal cellulase, the Family 7 cellobiohydrolase (Cel7A) from *Trichoderma reesei*, on the hydrophobic face of crystalline cellulose. With a fully atomistic model, we predict that the CBM alone exhibits regions of thermodynamic stability along a cellulose chain corresponding to a cellobiose unit, which is the catalytic product of the entire Cel7A enzyme. In addition, we determine which residues and the types of interactions that are responsible for the observed processivity length scale of the CBM: Y5, Q7, N29, and Y32. These results imply that the CBM can anchor the Cel7A enzyme at discrete points along a cellulose chain and thus aid in both recognizing cellulose chain ends for initial attachment to cellulose as well as aid in enzymatic catalysis by diffusing between stable wells on a length scale commensurate with the catalytic, processive cycle of Cel7A during cellulose hydrolysis. Comparison of other Family 1 CBMs show high functional homology to the four amino acids responsible for the processivity length scale on the surface of crystalline cellulose, which suggests that Family 1 CBMs may generally employ this type of approach for translation on the cellulose surface. Overall, this work provides further insight into the molecular-level mechanisms by which a CBM recognizes and interacts with cellulose.

Introduction

Cellulose is the most abundant biopolymer on Earth and thus represents both a vast carbon source for organisms in the biosphere and a renewable feedstock for commodity scale fuels and products. In nature, cellulose is degraded into monomeric sugars by enzyme cocktails secreted by bacteria and fungi. Understanding of the molecular-level, mechanistic details of the processes by which natural enzymatic systems recognize and degrade cellulose will improve our understanding of the coevolution of plants, bacteria, and fungi, aid in the understanding of protein-carbohydrate recognition, and enable the development of engineered enzyme and synthetic catalysts for more efficient, cheaper bioprocesses for production of renewable biofuels and bioproducts from lignocellulosic feedstocks.^{1–5}

One of the most thoroughly studied cellulose degrading organisms, the fungus *Trichoderma reesei* (or *Hypocrea jecorina*), secretes a cocktail of enzymes that act synergistically to degrade plant cell walls.⁶ The most abundantly produced cellulase enzyme from *T. reesei* is the Family 7 cellobiohydrolase, Cel7A. The *T. reesei* Cel7A enzyme consists of a carbohydrate-binding module (CBM) and a large catalytic domain (CD), connected by an *O*-glycosylated linker peptide, as shown in Figure 1. The catalytic cycle of the enzyme is hypothesized to consist of the following steps: the enzyme recognizes and decrystallizes the reducing end of a single cellulose chain from a cellulose microfibril; the chain is threaded into the CD tunnel, where cellulose is hydrolyzed to cellobiose by a pair of acidic residues; expulsion of the cellobiose product;

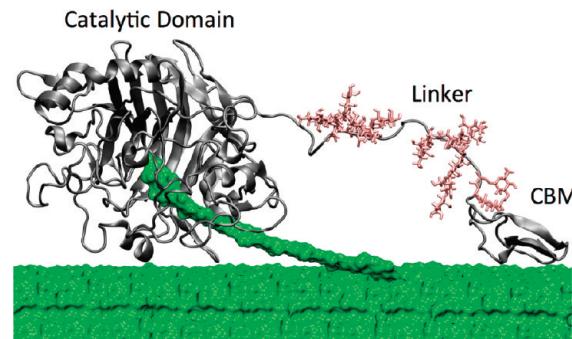


Figure 1. Cel7A from *T. reesei* on a cellulose surface with a bound cellulose chain. The *O*-glycosylation is shown in pink on the linker.

and processive threading of the chain into the transition state ensemble again for catalysis. Cel7A is hypothesized to repeat this cycle in a processive manner down the entire length of a cellulose chain from a microfibril.^{7–9} The crystal structure of the CD has been solved by X-ray diffraction,^{7,9} and the CBM structure has been solved by solution NMR coupled with simulated annealing.¹⁰

As the name suggests, CBMs bind to carbohydrate substrates, thereby increasing the surface concentration of enzymatic active sites for catalysis, which is an essentially universal function of CBMs in carbohydrate-degrading enzymes.² Indeed, experiments with intact Cel7A and Cel7A without a CBM demonstrate that the CBM is required for both binding affinity and efficient degradation of crystalline cellulose.^{11,12} The NMR structure of the Cel7A CBM reveals three aligned aromatic residues that are hypothesized to form the binding face to cellulose (Y5, Y31, and Y32) as well as several polar residues (among them, Q7 and N29) that potentially form hydrogen bonds to cellulose, as

* To whom correspondence should be addressed. E-mail: Gregg.Beacham@nrel.gov; Michael.Crowley@nrel.gov.

† National Bioenergy Center.

‡ Chemical and Biosciences Center.

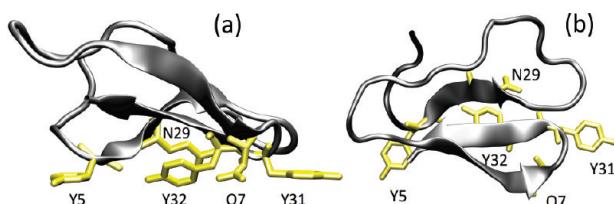


Figure 2. Carbohydrate-binding module from *T. reesei* Cel7A with Y5, Q7, N29, and Y31, Y32 shown. (a) Side view. (b) Top view. Structure taken from ref 10.

highlighted in Figure 2. Multiple studies have demonstrated that removal of either the aromatic or the polar residues reduces both the CBM binding affinity and the ability of Cel7A to degrade crystalline cellulose.^{13–17} Additionally, in a landmark study, Lehtio and co-workers used immuno-gold labeled CBMs visualized with transmission electron microscopy to demonstrate that the Cel7A CBM binds to the hydrophobic surface of cellulose I α .¹⁸ Because the hydrophobic face of cellulose I α and I β are similar at the molecular level, it is anticipated that the Cel7A CBM will bind to the hydrophobic face of cellulose I β as well. This assumption was used in previous computational studies of the Cel7A CBM on cellulose.^{19,20}

Despite the significant biological and technical relevance of Cel7A and the intense research efforts in the last two decades, the molecular-level, mechanistic details of the *T. reesei* Cel7A action on crystalline cellulose remain elusive. Toward that eventual objective, we conducted a thermodynamic study of the CBM from *T. reesei* Cel7A on cellulose. Specifically, we calculate the potential energy surface (PES) for the CBM on the hydrophobic face of a small cellulose I β slab. From the results, we present two major findings: (1) the Cel7A CBM alone exhibits regions of thermodynamic stability corresponding to a cellobiose unit, which as mentioned previously is the catalytic product of the processive cycle of the entire Cel7A enzyme and (2) four residues that hydrogen bond (and also forms a ring stack in one case) are responsible for the observed “processivity” distance of the CBM. These results presented here suggest that the CBM may aid in both the recognition and the processive action of the intact Cel7A enzyme.

Simulation Methods

We employ the same basic system setup as in ref 19 with slight modifications, as detailed below. The cellulose surface is 3 layers deep with 6 chains in each layer; the cellulose chains are 10 glucose units long, as shown in Figure 3. The cellulose I β crystal structure from Nishiyama and co-workers is used as the basis for generating the cellulose slab.²¹

The CBM is modeled with the CHARMM27 force field with the CMAP correction^{22–24} and the carbohydrate solution force field (CSFF) model for carbohydrates is applied to cellulose.²⁵ Water is modeled with the TIP3P force field.^{26,27} Particle mesh Ewald summation is used for electrostatics²⁸ with a sixth order b-spline interpolation, a Gaussian distribution with a width of 0.320 Å, and a mesh size of 60 × 60 × 48. The nonbonded interaction cutoff is 13 Å. The system was initially equilibrated for 20 ns in a molecular dynamics (MD) simulation at 300 K in the NVE ensemble with velocity rescaling in CHARMM. Covalent bonds to hydrogen atoms are fixed using the SHAKE algorithm²⁹ and the MD time step is 2 fs. The perimeter of the cellulose slab is harmonically restrained in the starting configuration with a force constant of 2.0 kcal/mol/Å² on the C1 and C4 carbon atoms. The system size is approximately 18 000 atoms with dimensions of approximately 62 Å × 62 Å × 47

Å, which allows a small, but sufficient space around the edges of the cellulose slab for the MD simulation and minimization calculations. The CBM is aligned in the hypothesized direction that the Cel7A enzyme would process along a cellulose chain; that is, the proximal end of the CBM points toward the nonreducing end of cellulose.^{7–9} This assumption was verified by collecting a PES obtained by rotating the CBM 360° in the equilibrium position on the cellulose surface. The grid size for the rotation was 10°, and the results showed a significant energy minimum in the chosen direction of the CBM (see the Supporting Information for the θ-PES). The initial structure for the PES is the average structure from the last 1 ns of the MD trajectory, which is shown in Figure 3.

The potential energy surface (PES) was constructed for the CBM on the hydrophobic face of the cellulose slab by placing the CBM on an x–y grid and minimizing the system at each grid point, similar to ref 20. The grid is 30 Å × 10 Å, and the grid range for the PES calculation is shown in Figure 4 with the four CBMs marking the limits of the grid. The resolution of the PES grid on the hydrophobic face of the cellulose slab is 0.25 Å in both the x and y directions, for a total of 121 × 41 = 4961 grid points, which is amenable to a massively parallel computational scheme, similar in spirit to large scale protein docking calculations, in this case with explicit solvent. Preliminary calculations demonstrated that the center of mass of the CBM moves less than the grid resolution for each grid point during minimization.

The minimization scheme for the PES is as follows: Each structure is solvated with an equal number of waters, and the CBM and cellulose are fixed while the water is minimized for 2500 steps with the Steepest-Descent (SD) algorithm followed by 2500 steps of the adopted basis Newton–Raphson (ABNR) algorithm. The CBM, cellulose, and solvent are then simultaneously relaxed, with the same restraints on the cellulose perimeter as during the MD simulation. The entire system is minimized with the ABNR method until convergence ($ΔE < 1 \times 10^{-6}$), which typically occurs in less than 10 000 steps. The PES is constructed by computing the cellulose-CBM interaction energy at each grid point. Three cycles of image convolution are conducted to smooth the interaction energy contour maps for visual clarity.

Results and Discussion

The PES for the CBM-cellulose interaction is shown in Figure 5a. As highlighted with circles, there are periodic energy minima in the x-direction along a cellobextrin chain approximately every 1 nm, which is about the length of a cellobiose (disaccharide) unit. As mentioned previously, cellobiose is the catalytic product of Cel7A; thus, this result suggests that the Cel7A CBM has evolved to translate along a given cellulose chain on the hydrophobic cellulose surface with the same critical length scale as the processive Cel7A catalytic cycle. It should be noted that the region of higher energy to the far right-hand side of Figure 5a is present because of minor edge effects; in those grid points, the front of the CBM is at the far edge of the cellulose slab. A snapshot of the CBM in the equilibrium position near the middle of Figure 5a is shown in Figure 5b. The blue, transparent shading represents the molecular “footprint” of the CBM on the cellulose surface. Figure 5c shows the CBM in a higher energy location translated approximately 5 Å to the right of the configuration shown in Figure 5b.

An interesting feature of the CBM–cellulose interaction is observed when separating the components of the PES into electrostatics and van der Waals interactions, shown in Figure

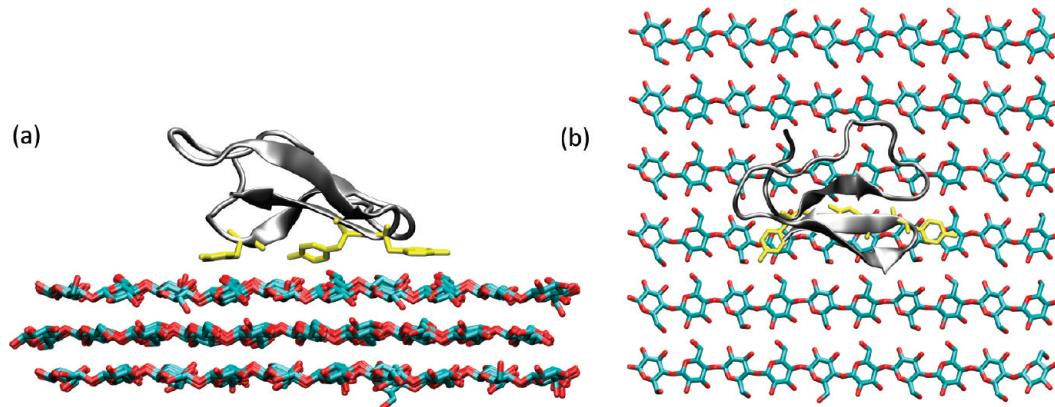


Figure 3. Equilibrated cellulose slab and CBM used for this study is similar to that in reference.¹⁹ The cellobextrin chains are 10 glucose units long with 6 chains per layer, and 3 layers total. The CBM is aligned in the putative direction of the Cel7A enzyme, and the alignment does not change significantly during the MD equilibration. (a) Side view. (b) Top view with only the topmost layer of cellulose shown for clarity.

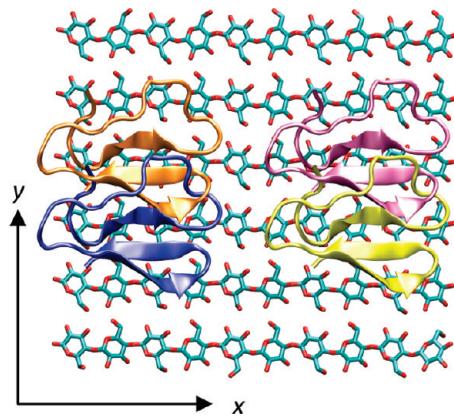


Figure 4. Extent of the x - y grid on the hydrophobic face of the cellulose I β slab. The four CBMs shown in different colors on the surface mark the four corners of the PES grid.

6. As seen in Figure 6a, the electrostatic interactions dominate the large changes in the energetic contributions to the Cel7A CBM processivity, and the electrostatics map exhibits the same trends as that seen in Figure 5. These results suggest that hydrogen bonding interactions will likely dominate the CBM processivity. For the van der Waals interactions shown in Figure 6b, the overall values for the energy are low, while the range in energy is small relative to the electrostatics PES. These types of interactions have been shown to be important for CBM binding to cellulose, which is driven by the hydrophobic effect³⁰ but are not as important as electrostatics in the CBM function after binding, i.e., the CBM processivity along a given chain.

The residues responsible for the observation shown in Figures 5 and 6a can be readily extracted from the same PES calculation. The residue-cellulose interaction energies were computed from the minimized structures, and PESs for each residue interacting with cellulose were generated. Of the 36 residues in the Cel7A CBM, four amino acid-cellulose interactions exhibit the same general behavior as that shown in Figures 5 and 6a: tyrosine 5 (Y5), glutamine 7 (Q7), asparagine 29 (N29), and tyrosine 32 (Y32). Figure 7 shows the residue-specific PES for Y5, which is located at the distal end of the CBM and in the potential energy wells forms both a favorable ring-stacking interaction and a hydrogen bond with a primary alcohol group of a glucose residue. This is highlighted in yellow in Figure 7b with the hydrogen bond between Y5 and the primary alcohol group on cellulose shown as a dotted red line. An interesting feature regarding the Y5 stacking and hydrogen bonding interaction is that favorable interactions are formed every other glucose

molecule. This is because, as shown in Figure 7b, the primary alcohol group on the surface of cellulose points out of the plane and into the solution. Since cellobiose units are β -glucose dimers with a rotation of approximately 180° about the glycosidic linkage between monomers, there are unfavorable steric interactions with the Y5 residue every other glucose unit, shown as Y5 highlighted in green in Figure 7b.

As with Y5, the amino acids Q7, N29, and Y32 form favorable hydrogen bonds with the primary alcohol groups on the surface of cellulose, again every cellobiose unit. As shown in Figure 8a–c, hydrogen bonds (highlighted in red in the case of O–H hydrogen bonds and blue in the case of N–H hydrogen bonds) are formed between the side chains and the primary alcohol hydrogen pointing out of the cellulose slab plane. Again as with Y5, the geometry of the CBM and the cellulose prevent these interactions from forming every glucose unit. For Q7, little to no net interactions (in red) form with the cellulose surface except in the energy wells every cellobiose unit. N29 also forms strong hydrogen bonds to the primary alcohol groups on the cellobiose length scale. The Y32 PES, although not as dramatic as Q7 or N29, shows the steric hindrance of the tyrosine group every other glucose and lack of favorable hydrogen bonds, as shown in Figure 7d in red between the stable wells.

The results we obtained for the four specific residues, Y5, Q7, N29, and Y32, may be extendable to other Family 1 CBMs. Similar to Nimlos and co-workers,¹⁹ we conducted a sequence comparison of Family 1 CBMs with the Basic Local Alignment Search Tool (BLAST).³¹ From the top 250 sequences, the average identity is 60% and the average positive similarity is 73%, which is expected as Family 1 CBMs are known to exhibit high homology.² Table 1 shows several of the results from the BLAST search with the species name and enzyme type (e.g., Cel7A, Cel6A, etc.). The full table with the accession codes can be found in the Supporting Information. As Table 1 illustrates, the residues at sites 5, 7, 29, and 32 possess high functional homology. In the cases in which differences exist (e.g., tryptophan instead of tyrosine), the substituted residue will be able to accomplish the same function (i.e., hydrogen bonding and/or ring stacking) as the respective residue in the wild-type Cel7A CBM (i.e., positive similarity). We show a quantitative summary of our results from the BLAST search in Table 2, in which the Cel7A CBM residue is listed along with the number of identical residues, the number of functionally similar residues, and the number of functionally dissimilar residues. As shown in all four cases, the percent functional similarity is over 98%. It should be noted that several of the BLAST hits that are found

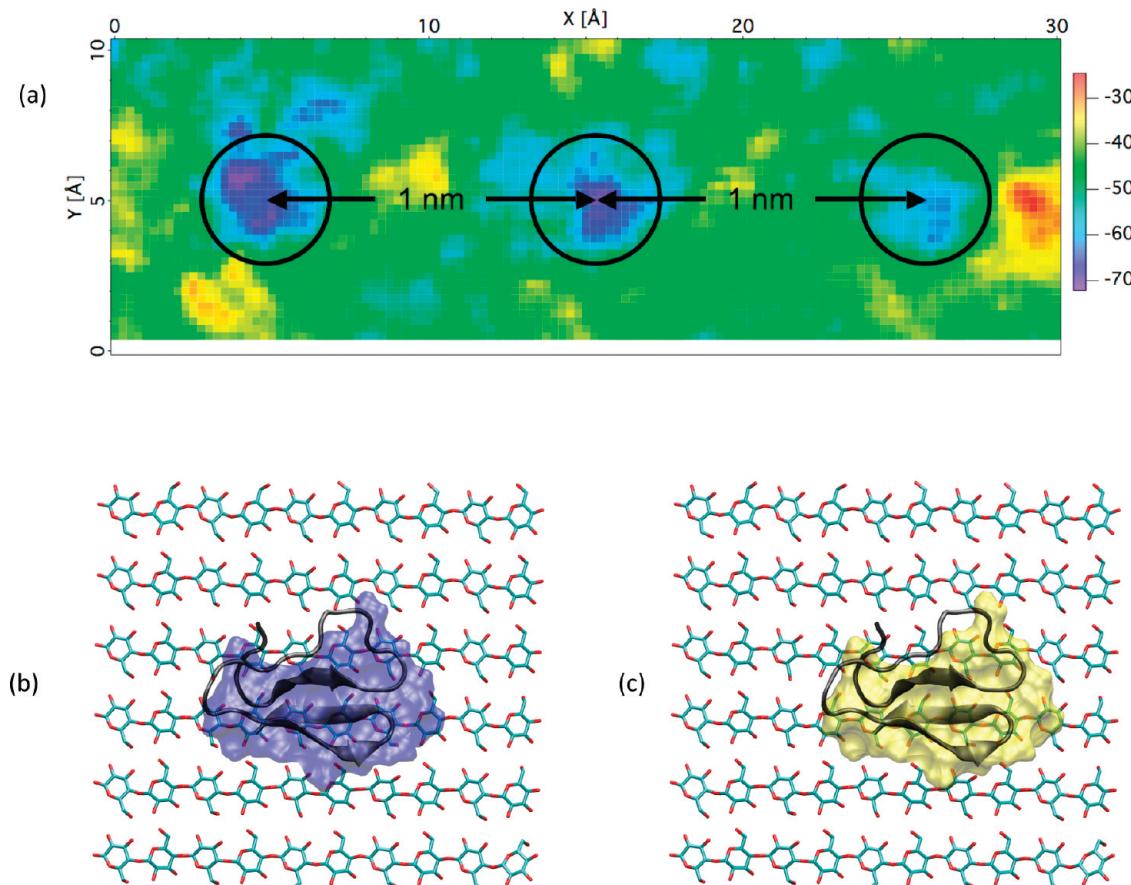


Figure 5. (a) Potential energy surface for the CBM–cellulose interaction. The circles highlight the energy minima approximately every 1 nm, or every 1 cellobiose unit. (b) Snapshot of the CBM in the equilibrium position at approximately $x = 15 \text{ \AA}$ and $y = 5 \text{ \AA}$. The blue shading outlines the CBM footprint on cellulose. (c) Snapshot of the CBM at $x = 20 \text{ \AA}$ and $y = 5 \text{ \AA}$. The yellow shading outlines the CBM footprint on cellulose.

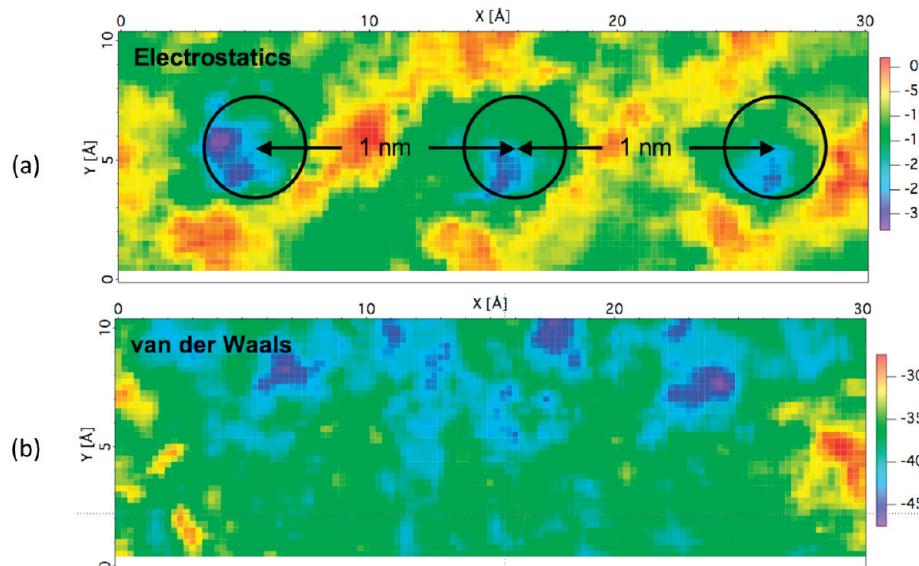


Figure 6. Separation of the CBM–cellulose interaction into energetic components. (a) The electrostatic interactions exhibit similar behavior as that shown in Figure 5. (b) The van der Waals interactions are low in energy and the range is significantly smaller than the electrostatic interactions.

are known mutations of the Cel7A CBM,^{13,15} which were excluded in our analysis shown in Table 2. Overall, the homology comparison results shown here strongly suggest that Family 1 CBMs may generally exhibit energy minima on crystalline cellulose every cellobiose unit.

In this study, we have used a fully atomistic model to demonstrate that the Cel7A CBM exhibits thermodynamically stable, periodic minima on the surface of cellulose I β on the length scale of a cellobiose molecule. This result is both enabled by and complementary to the work of Lehtiö and co-workers,¹⁸

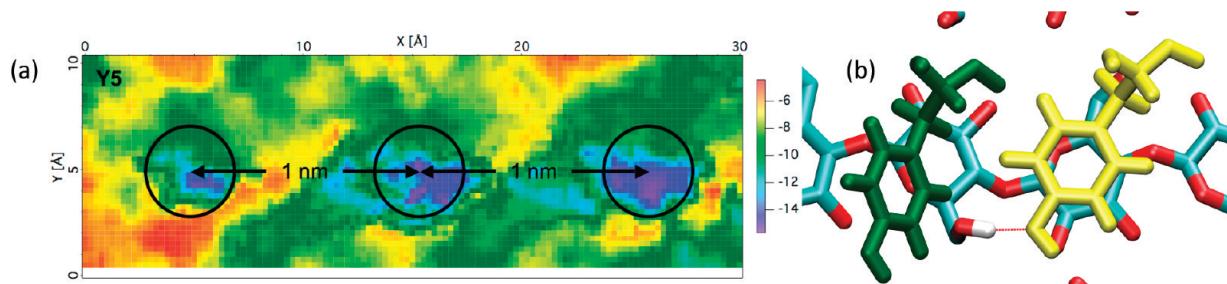


Figure 7. (a) The residue-cellulose PES for Y5. (b) A corresponding snapshot of two configurations are shown with Y5 highlighted in yellow in an energy minimum and highlighted in green in a location where the interactions are unfavorable. The grid points from where the configurations were taken are $X = 10 \text{ \AA}$, $Y = 5 \text{ \AA}$ for the green configuration and $X = 15 \text{ \AA}$, $Y = 5 \text{ \AA}$ for the yellow configuration. The CBM is not shown for clarity.

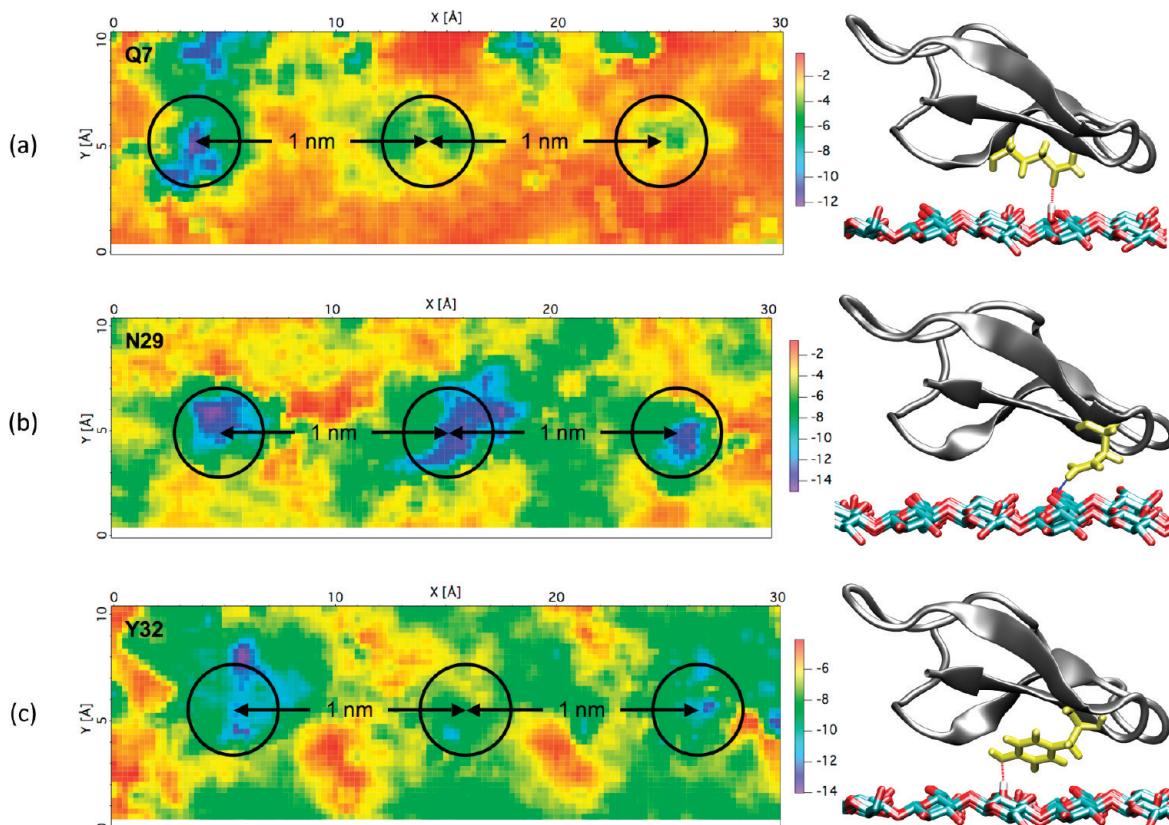


Figure 8. Residue-cellulose PESs for (a) Q7, (b) N29, and (c) Y32. Corresponding snapshots of configurations in an energy minimum are shown with each PES, with the relevant amino acid highlighted in yellow.

in which the Cel7A CBM was experimentally shown to bind to the hydrophobic face of cellulose. We made use of their observations of the binding specificity of the Cel7A CBM to probe the CBM behavior on the appropriate crystal face. The primary significance of this work is that it demonstrates that the Cel7A CBM has evolved to aid Cel7A in substrate targeting of specific locations on the hydrophobic face. The ability of the CBM to be stabilized in given wells may aid in recognition and processivity of a cellobextrin chain for processive hydrolysis because it will anchor the enzyme in a given location for time scales commensurate with the barriers to translate to new sites along the same chain. If the glycosylated linker is sufficiently stiff or explores relatively few conformations on the time scale of CBM translation along a chain, this implies that the Cel7A CD will be anchored in a relatively firm manner, which may help the CD “recognize” free chain ends, as it is hypothesized to do with an entrance tryptophan residue.⁹ Experimental studies with small-angle X-ray scattering have demonstrated that the linker conformations in solution with the intact Cel7A

enzyme and other fungal cellulases are extended and relatively stiff.^{32–34}

Additionally, since these types of extracellular enzymes act under equilibrium conditions, this implies that the CBM will undergo random, undirected diffusion between the thermodynamically stable wells along a given, intact cellulose chain. This result, obtained with a fully atomistic model, is a molecular-level validation of a previously published result using a multiscale model from our group with the same CBM.²⁰ In our previous study, we showed that the Cel7A CBM has global, periodic minima along a cellulose chain every cellobiose unit, but that there were also metastable minima every glucose unit. The previous study applied a coarse-grained model for cellulose in which a glucose molecule was represented by three beads. Therefore, the primary alcohol group in the coarse-grained model did not protrude from the surface, thus both the steric hindrance and the explicit hydrogen bonding observed on the fully atomistic model were not present in the coarse-grained model.

TABLE 1: Sequence Alignment for Several Family 1 CBMs from Commonly Studied Cellulose-Degrading Enzymes^a

Organism	Enzyme	5	7	29	32
<i>Trichoderma reesei</i>	Cel7A	T Q S H	Y G Q	C G G I G Y S G P T V C A S G T T C Q V L N P Y Y S Q C L	
<i>Trichoderma reesei</i>	Cel6A	C S S V W G Q	C G G Q N W S G P T C C A S G S T C V Y S N D Y Y S Q C L		
<i>Trichoderma reesei</i>	Cel7B	T Q T H W G Q	C G G I G Y S G C K T C T S G T T C Q Y S N D Y Y S Q C L		
<i>Trichoderma reesei</i>	Cel5A	Q Q T V W G Q	C G G I G W S G P T N C A P G S A C S T L N P Y Y A Q C I		
<i>Trichoderma reesei</i>	Cel61A	T Q T L Y G Q	C G G S G Y S G P T R C A P P A T C S T L N P Y Y A Q C L		
<i>Agaricus bisporus</i>	Cel6A	Q S P V W G Q	C G G N G W T G P T T C A S G S T C V K Q N D F Y S Q C L		
<i>Aspergillus aculeatus</i>	Cel7A	V A Q L Y G Q	C G G Q G W T G P T T C A S G - T C T K Q N D Y Y S Q C L		
<i>Fusarium oxysporum</i>	Cel7A	S V D Q W G Q	C G G Q N Y S G P T T C K S P F T C K K I N D F Y S Q C Q		
<i>Humicola grisea</i>	Cel7A	K A G R W Q Q	C G G I G F T G P T Q C E E P Y T C T K L N D W Y S Q C L		
<i>Phanerochaete chrysosporium</i>	Cel7A	T V P Q W G Q	C G G I G Y T G S T T C A S P Y T C H V L N P Y Y S Q C Y		
<i>Talaromyces emersonii</i>	Cel6A	Q Q S L W G Q	C G G S S W T G A T S C A A G A T C S T I N P Y Y A Q C V		
<i>Thielavia australiensis</i>	Cel7A	T A K H W Q Q	C G G N G W T G P T V C E S P Y K C T K Q N D W Y S Q C L		
<i>Trichoderma harzianum</i>	Cel7A	T Q T H Y G Q	C G G T G W T G P T R C A S G Y T C Q V L N P F Y S Q C L		
<i>Trichoderma longibrachiatum</i>	Cel7B	T Q T H W G Q	C G G I G Y T G C K T C T S G T T C Q Y G N D Y Y S Q C L		
<i>Trichoderma viride</i>	Cel7A	T Q T H Y G Q	C G G I G Y S G P T V C A S G S T C Q V L N P Y Y S Q C L		

^a Starting from the *T. reesei* Cel7A CBM, we highlight the Y5, Q7, N29, and Y32 amino acids. As shown, there is high positive similarity (or high “functional homology”) in these four sites among Family 1 CBMs from various cellulase enzymes.

TABLE 2: Summary of the Sequence Comparison from a Set of 250 Aligned Sequences of Family 1 CBMs from a BLAST Search

residue	identical	functionally similar	functionally dissimilar	percent functional similarity
Y5	130 Y	114 W	6 F	98%
Q7	244 Q	4 R	2 T	100%
N29	233 N	12 S	3 D	99%
Y32	245 Y	2 W	3 F	99%

Moreover, the mutation of residues Y5, Q7, N29, and Y32, to small, nonpolar amino acids, such as alanine, is known to reduce the binding affinity and activity of Cel7A to crystalline cellulose.^{14–17} However, from the results presented here, it is expected that mutation of Y5, Q7, N29, and Y32 will affect the activity not only by decreasing the binding affinity of Cel7A, but also by disrupting the structure–function relationship that yields the energy minima observed every cellobiose unit along a cellulose chain, as predicted. Because activity improvements can be made to these types of enzymes via increased binding affinity,^{11,12} the results presented here highlight the importance of an additional, predicted function of CBMs bound to crystalline cellulose, that of potentially assisting the cellobiose based processive action of Cel7A. This result will further aid in the construction and validation of both molecular-level descriptions of individual enzyme action as well as lower-resolution, mechanistic models of enzymatic cellulose degradation.³⁵

Additionally, as highlighted in Figure 2, there is an additional tyrosine residue, Y31 on the flat face of the Cel7A CBM, which has also received significant attention in biochemical studies of the Cel7A CBM.^{13,15} A mutational biochemistry study demonstrated that mutation of the Y31 residue, which is located at the proximal end of the CBM, decreases both the binding and the specific activity of Cel7A to nearly that of the isolated catalytic domain without a CBM.¹³ Although our results indicate that Y31 is likely to not be functionally important for the thermodynamics of the CBM processivity on the cellulose surface, the previous experimental results and our results show that it does form favorable interactions with cellulose (see the Supporting Information). Additionally, we hypothesize that it may be relevant for initially aligning the CBM along a cellobextrin chain on the hydrophobic face, and maintaining that alignment during processive cellulose hydrolysis, although this question is outside of the scope of our present study.

From a computational perspective, we note that the ranges in potential energies obtained in these PESs are quite high.

Because we are only considering potential energy in this study, the absolute magnitudes of the energies will likely be damped in reality. Additionally, the most appropriate quantity to probe in a study of this type would be free energy; however, as with many biological systems with explicit solvation and fully atomistic models, the convergence of free energy calculations such as umbrella sampling along a chosen reaction coordinate is incredibly computationally prohibitive. Therefore, we have chosen to limit our study to a subset of probable conformational states, obtained from a converged MD simulation, in the direction that Cel7A likely takes on the known binding face of crystalline cellulose.¹⁸ Because enthalpic contributions to CBM–cellulose interaction (post-binding) are substantial, it is unlikely that this limitation will significantly affect the computational predictions.

Conclusions

In this study, we have demonstrated that the CBM from *Trichoderma reesei* Cel7A alone exhibits stable energy minima corresponding to a cellobiose unit along a chain on the hydrophobic face of cellulose. This result is significant because it enhances our molecular-level understanding of the ability of CBMs from cellulose-degrading enzymes to recognize and interact with carbohydrates. Our results suggest that activity changes in cellulases and other carbohydrate-active enzymes via CBM mutations may not only arise from changes in binding affinity, but also in the ability of the CBM to preferentially bind to locations along a cellulose chain commensurate with the catalytic length scale(s) of a given carbohydrate-degrading enzyme. This implies that artificial evolutionary strategies for specific activity improvements via CBM binding affinity should be designed to not only increase binding affinity, but also to enhance or at least retain the molecular-level function of CBMs on carbohydrates. Our results also will aid in the construction of models for carbohydrate-active enzyme function at multiple length and time scales.

Acknowledgment. We thank the Department of Energy Office of the Biomass Program for funding. Computational time for this research was supported in part by the Golden Energy Computing Organization at the Colorado School of Mines using resources acquired with financial assistance from the National Science Foundation and the National Renewable Energy Laboratory. Simulations were also performed in part using the Texas Advanced Computing Center Ranger cluster under the National Science Foundation Teragrid grant number TG-MCA08X015.

Supporting Information Available: Y31 potential energy surface. Potential energy as a function of the CBM rotation angle. Sequence alignments for 250 proteins from a BLAST search. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References and Notes

- (1) Lynd, L. R.; Weimer, P. J.; van Zyl, W.; Pretorius, I. S. *Microbiol. Molec. Biol. Rev.* **2002**, *66*, 506–577.
- (2) Boraston, A. B.; Bolam, D. N.; Gilbert, H. J.; Davies, G. J. *Biochem. J.* **2004**, *382*, 769–781.
- (3) Ragauskas, A. J.; Williams, C. K.; Davison, B. H.; Britovsek, G.; Cairney, J.; Eckert, C. A.; Frederick, W. J.; Hallett, J. P.; Leak, D. J.; Liotta, C. L. *Science* **2006**, *311*, 484–489.
- (4) Himmel, M. E.; Ding, S. Y.; Johnson, D. K.; Adney, W. S.; Nimlos, M. R.; Brady, J. W.; Foust, T. D. *Science* **2007**, *315*, 804–807.
- (5) Eijsink, V. G. H.; Vaaje-Kolstad, G.; Varum, K. M.; Horn, S. J. *Trends Biotechnol.* **2008**, *26*, 228–235.
- (6) Martinez, D.; Berka, R. M.; Henrissat, B.; Saloheimo, M.; Arvas, M.; Baker, S. E.; Chapman, J.; Chertkov, O.; Coutinho, P. M.; Cullen, D. *Nat. Biotechnol.* **2008**, *26*, 553–560.
- (7) Divine, C.; Stahlberg, J.; Reinikainen, T.; Ruohonen, L.; Pettersson, G.; Knowles, J. K. C.; Teeri, T. T.; Jones, T. A. *Science* **1994**, *265*, 524–528.
- (8) Barr, B. K.; Hsieh, Y. L.; Ganem, B.; Wilson, D. B. *Biochemistry* **1996**, *35*, 586–592.
- (9) Divine, C.; Stahlberg, J.; Teeri, T. T.; Jones, T. A. *J. Mol. Biol.* **1998**, *275*, 309–325.
- (10) Kraulis, P. J.; Clore, G. M.; Nilges, M.; Jones, T. A.; Pettersson, G.; Knowles, J.; Gronenborn, A. M. *Biochemistry* **1989**, *28*, 7241–7257.
- (11) Stahlberg, J.; Johannsson, G.; Pettersson, G. *BioTechnology* **1991**, *9*, 286–290.
- (12) Srisodsuk, M.; Lehtiö, J.; Linder, M.; Margolles Clark, E.; Reinikainen, T.; Teeri, T. T. *J. Biotechnol.* **1997**, *57*, 49–57.
- (13) Reinikainen, T.; Ruohonen, L.; Nevanen, T.; Laaksonen, L.; Kraulis, P.; Jones, T. A.; Knowles, J. K. C.; Teeri, T. T. *Proteins-Struct. Funct. Genet.* **1992**, *14*, 475–482.
- (14) Linder, M.; Lindeberg, G.; Reinikainen, T.; Teeri, T. T.; Pettersson, G. *FEBS Lett.* **1995**, *372*, 96–98.
- (15) Linder, M.; Mattinen, M. L.; Kontteli, M.; Lindeberg, G.; Stahlberg, J.; Drakenberg, T.; Reinikainen, T.; Pettersson, G.; Annila, A. *Protein Sci.* **1995**, *4*, 1056–1064.
- (16) Srisodsuk, M.; Lehtiö, J.; Linder, M.; Margolles Clark, E.; Reinikainen, T.; Teeri, T. T. *J. Biotechnol.* **1997**, *57*, 49–57.
- (17) Takashima, S.; Ohno, M.; Hidaka, M.; Nakamura, A.; Masaki, H. *FEBS Lett.* **2007**, *581*, 5891–5896.
- (18) Lehtiö, J.; Sugiyama, J.; Gustavsson, M.; Fransson, L.; Linder, M.; Teeri, T. T. *Proc. Natl. Acad. Sci.* **2003**, *100*, 484–489.
- (19) Nimlos, M. R.; Matthews, J. F.; Crowley, M. F.; Walker, R. C.; Chukkapalli, G.; Brady, J. W.; Adney, W. S.; Cleary, J. M.; Zhong, L. H.; Himmel, M. E. *Prot. Eng. Des. Select.* **2007**, *20*, 179–187.
- (20) Bu, L.; Beckham, G. T.; Crowley, M. F.; Chang, C. H.; Matthews, J. F.; Bomble, Y. J.; Adney, W. S.; Himmel, M. E.; Nimlos, M. R. *J. Phys. Chem. B* **2009**, *113*, 10994–11002.
- (21) Nishiyama, Y.; Langan, P.; Chanzy, H. *J. Am. Chem. Soc.* **2002**, *124*, 9074–9082.
- (22) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (23) MacKerell, A. D.; Feig, M.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1400–1415.
- (24) Brooks, B. R.; Brooks, C. L.; MacKerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S. *J. Comput. Chem.* **2009**, *30*, 1545–1614.
- (25) Kuttel, M.; Brady, J. W.; Naidoo, K. *J. J. Comput. Chem.* **2002**, *23*, 1236–1243.
- (26) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (27) Durell, S. R.; Brooks, B. R.; Ben-Naim, A. *J. Phys. Chem.* **1994**, *98*, 2198–2202.
- (28) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8857.
- (29) Ryckaert, J.; Ciccotti, G.; Berendsen, H. *J. Comput. Phys.* **1977**, *23*.
- (30) Creagh, A. L.; Ong, E.; Jervis, E.; Kilburn, D. G.; Haynes, C. A. *Proc. Natl. Acad. Sci.* **1996**, *93*, 12229–12234.
- (31) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. *J. Mol. Biol.* **1990**, *215*, 403–410.
- (32) Abuja, P. M.; Pilz, I.; Claeysse, M.; Tomme, P. *Biochem. Biophys. Res. Commun.* **1988**, *156*, 180–185.
- (33) Abuja, P. M.; Schmuck, M.; Pilz, I.; Tomme, P.; Claeysse, M.; Esterbauer, H. *Eur. Biophys. J.* **1988**, *15*, 339–342.
- (34) Receveur, V.; Czjzek, M.; Schulein, M.; Panine, P.; Henrissat, B. *J. Biol. Chem.* **2002**, *277*, 40887–40892.
- (35) Ting, C. L.; Makarov, D. E.; Wang, Z. G. *J. Phys. Chem. B* **2009**, *113*, 4970–4977.

JP908810A