ARTICLE

BIOTECHNOLOGY and BIOENGINEERING

Simulation Studies of Substrate Recognition by the Exocellulase CelF From *Clostridium cellulolyticum*

Mo Chen,[1] Michael E. Himmel,[2] David B. Wilson,[3] John W. Brady[1]

[1]Department of Food Science, Cornell University, Ithaca, New York 14853; telephone: +1 (607) 255-2897; fax: 607-254-4686; e-mail: jwb7@cornell.edu

[2]Biosciences Center, National Renewable Energy Laboratory, Golden, Colorado

[3]Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York

ABSTRACT: Molecular dynamics (MD) simulations were used to study substrate recognition by the family 48 exocellulase CelF from *Clostridium cellulolyticum*. It was hypothesized that residues around the entrance of the active site tunnel of this enzyme might serve to recognize and bind the substrate through an affinity for the cellulose monomer repeat unit, β-D-glucopyranose. Simulations were conducted of the catalytic domain of this enzyme surrounded by a concentrated solution of β-D-glucopyranose, and the full three-dimensional probability distribution for finding sugar molecules adjacent to the enzyme was calculated from the trajectory. A significant probability of finding the sugar stacked against the planar faces of Trp 310 and Trp 312 at the entrance of the active site tunnel was observed.

Biotechnol. Bioeng. 2016;113: 1433–1440.

KEYWORDS: enzyme mechanisms; enzymatic hydrolysis; MD simulations; substrate recognition; substrate binding; cellulases

Introduction

Cellulose, the most abundant biomolecule by weight, due to its structural role in plant cell walls, is very resistant to enzymatic deconstruction for several reasons, including its partial crystallinity, insolubility in water, and its interactions with other cell wall polymers such as lignin and hemicelluloses. Various microorganisms have evolved a wide array of enzymes to degrade cellulose, some of which are quite complex, multi-domain "molecular machines." In all of these enzymes, however, the recognition and binding of the cellulose substrate is an essential initial step in hydrolysis. Many cellulases possess carbohydrate-binding modules (CBMs) in addition to the catalytic domains (CDs), promoting the association of the enzymes with their insoluble cellulose substrates (Boraston et al., 2004). In several cases, the catalytic domains retain some activity even in the absence of their binding domains, implying that these modules are independently able to make favorable substrate interactions (Pakarinen et al., 2014). In some of the most interesting of these cellulases, the catalytic cleavage site is located inside a long tunnel that spans the globular domain and requires that the substrate cellulose chain find and then be threaded into this tunnel, with the cellobiose cleavage product exiting from the opposite end of the tunnel (Beckham et al., 2014).

Numerous studies have found evidence that the planar aromatic amino acid side chains, and particularly the indole group of tryptophan, are important for substrate or ligand recognition in proteins that bind glucose or glucose-like molecules. Molecular dynamics (MD) simulations have provided an explanation for the affinity of glucose for such nonpolar surfaces in terms of hydrophobic association between the non-polar H1-H3-H5 triad of the glucose ring and the nonpolar planes (Brady et al., 2012). Although glucose is an osmolyte and tends to exhibit a negative surface excess concentration with respect to protein surfaces, being drawn away into the aqueous bulk phase (Mason et al., 2011), some proteins have been designed by evolution to specifically bind to sugars, including of course enzymes like cellulases that act upon carbohydrate substrates. The substrate binding sites in these proteins usually contain planar side chains (Quiocho, 1989; Quiocho et al., 1991), and quite commonly these sites feature at least one tryptophan residue. Exploiting this fact, MD simulations have been used not only to calculate the approximate binding energy of such interactions (Mason et al., 2011), but also as a general means of identifying sugar binding sites on protein surfaces (Tavagnacco et al., 2011).

The crystal structures of many CDs, particularly those of processive exocellulases, such as *Trichoderma reesei* Cel6A (Zou et al., 1999) and Cel7A (Becker et al., 2001), and family

Current address of Mo Chen is Department of Biomedical Engineering and OCSSB, Oregon Health and Science University, Portland, Oregon.
Correspondence to: J.W. Brady
Contract grant sponsor: U.S. Department of Energy (DOE)
Contract grant sponsor: National Science Foundation
Contract grant number: ACI-1053575
Received 18 August 2015; Revision received 7 December 2015; Accepted 18 December 2015
Accepted manuscript online 22 December 2015;
Article first published online 28 January 2016 in Wiley Online Library (http://onlinelibrary.wiley.com/doi/10.1002/bit.25909/abstract).
DOI 10.1002/bit.25909

48 exocellulases from various microorganisms (Guimarães et al., 2002; Parsiegla et al., 2008), have been reported and illustrate that Trp residues are present at the entrances and exits of the active site tunnels (Fig. 1). It has been suggested that the Trp residues at the tunnel entrances might function in cellulose recognition and acquisition, and that those at the tunnel exits might function in stabilizing the product side of the substrate prior to its hydrolysis (Chen et al., 2015; Kostylev et al., 2014; Payne et al., 2011). Free energy calculations using molecular dynamics simulations have provided support for such mechanisms at the molecular level. For example, Payne and co-workers have calculated the relative ligand binding free energies between the wild type and Ala mutants substituted at each of the four tunnel Trp residues in *T. reesei* Cel6A, and revealed that removing only the Trp residues at the tunnel entrance and exit could dramatically impact the binding affinity to the substrate (Payne et al., 2011). Site-directed mutagenesis experiments have confirmed this theoretical prediction by demonstrating that replacement of these residues reduces this enzyme's activity (Kostylev et al., 2014). Wohlert and co-workers have calculated the free energy landscape for the interaction between indole (the sidechain of Trp) and β-D-glucopyranose (the monomer repeat unit of cellulose) in aqueous solution, showing that these two molecules favor stacking interactions, with a binding energy of ∼1.2 kJ/mol (Wohlert et al., 2010).

The family 48 cellulases have a long active site tunnel (∼43 Å in length), which contains four conserved Trp residues that make stacking interactions with the monomer units of the cellooligomer substrate chains in the reported crystal structures. Two of these Trp residues are located at the entrance (residues 310 and 312 in CelF; Fig. 1), one is at the exit, and another is near the active site (Guimarães et al., 2002; Parsiegla et al., 2008). This substrate tunnel guides the cellulose chain to the active site. In the present study, MD simulations of a solution of free β-D-glucopyranose monomers interacting with the CD of CelF, a family 48 exocellulase from *Clostridium cellulolyticum*, were used to identify any possible additional binding sites on the surface of the CD that might play a role in directing the substrate chain into the active site tunnel, and thus might be useful candidates for substitution in site-directed mutagenesis experiments aimed at improving turnover rates. Because glucose is an osmolyte that favors aqueous solution rather than tending to bind on protein surfaces, a weak, but sufficiently significant binding affinity between glucose and surface residues in the protein would suggest a possible role in the binding mechanism of the processive exocellulases.

Methods

The crystal structure of CelF was used as the starting structure in the simulation. Several CelF crystal structures have been reported

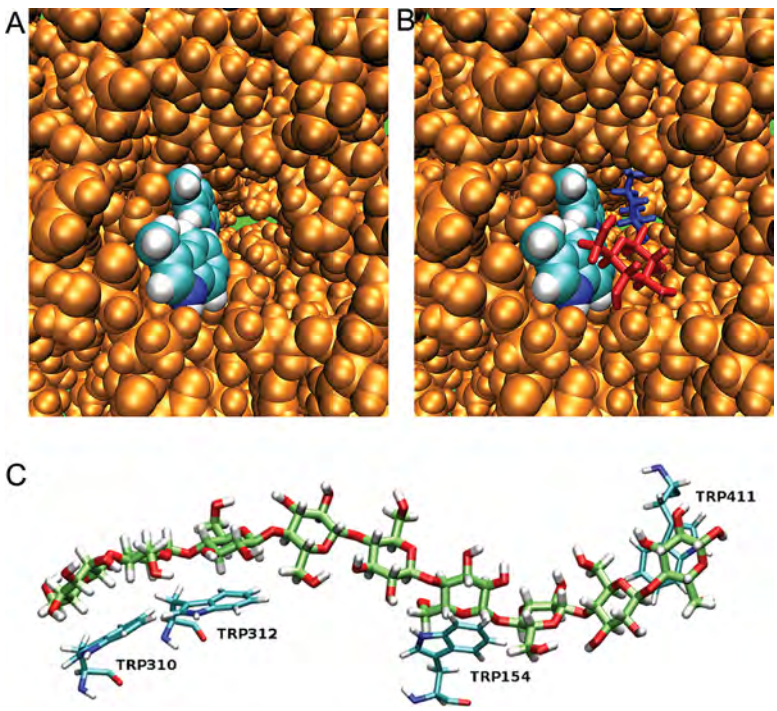

Figure 1. (**a**) The active site tunnel in CelF, with the side chains of Trp 310 and 312 shown using atomic coloring to illustrate their placement at the entrance in a fashion that facilitates the stacking of the cellulose substrate and its entrance into the tunnel. A green background highlights the rear exit of the tunnel. (**b**) The same image with typical instantaneous positions for β-D-glucopyranose molecules, in red and blue, found stacking against these indole groups in the present MD simulations (see text). (**c**) The "lower" path of the substrate chain in the CelF point mutant crystal structure, 2QNO, showing the positions of the four tunnel Trp residues relative to the monomers in the chain (Parsiegla et al., 2008).

with various ligands occupying the substrate and product sites in the protein (Parsiegla et al., 1998, 2000, 2008). The most recently reported CelF structures correspond to two CelF single mutants with different substrate pathway configurations (Parsiegla et al., 2008). In the present simulations, the structure with the so-called "lower" substrate pathway (E55Q) was used to construct a starting structure for the wild type CelF CD. However, no cellooligomer substrate was placed in the enzyme active site tunnel in these simulations.

The model system was constructed by positioning a CelF molecule in the center of a large cubic box with the dimension of ~96 Å under fluctuating volume, constant pressure simulation conditions, containing a concentrated solution of β-D-glucopyranose molecules. This box initially contained 864 β-D-glucopyranose molecules, and was built by assembling together 27 smaller cubes, each of which contained 32 randomly positioned β-D-glucopyranose molecules. Those glucose molecules that overlapped the CelF protein were then removed. Next, water molecules were added to fill in the vacant space in the large cubic system, and sodium counter ions were placed randomly to neutralize the system. Overall, the system contained 1 CelF protein, 707 sugar molecules, 13 sodium cations, and 20,787 water molecules, for a total of 89,010 atoms in the primary simulation box. This box thus represented a glucose concentration of ~1.89 molal. None of the retained sugar molecules were inside the active site tunnel of the enzyme.

MD simulations were used to model the evolution of the protein/sugar solution. The CHARMM22/CMAP force field (MacKerell et al., 1998, 2004a,b), the all-atom CHARMM carbohydrate force field (Guvench et al., 2008), and the TIP3P water model (Jorgensen et al., 1983; Neria et al., 1996) were used to describe the protein, the glucose, and the water molecules, respectively. The general molecular mechanics program CHARMM (Brooks et al., 1983, 2009) was used to build the molecular system. The CHAMBER program (Crowley et al., 2009) was used to convert the CHARMM starting files into AMBER format, and the PMEMD engine of AMBER (Case et al., 2012) was used to carry out the MD simulations. The SHAKE algorithm (Ryckaert et al., 1977; van Gunsteren and Berendsen, 1977) was applied in the simulation to constrain the bond distances involving hydrogen atoms. The non-bond cutoff distance used was 8 Å.

The solvent of the system was subjected to minimization to eliminate any possible atomic overlaps resulting from the setup procedure. Then, keeping the protein solute restrained with a force constant of 10 kcal/molÅ^2, the system was thermalized at constant volume from 0 to 300 K over an interval of 20 ps. The production run of the molecular dynamics simulation was carried out under a constant temperature of 300 K and at a constant pressure of 1 atm for 250 ns using a timestep of 2 fs. Constant temperature coupling was regulated by a Langevin thermostat (Schneider and Stoll, 1978a,b; van Gunsteren and Berendsen, 1988), and constant pressure coupling was controlled using the Berendsen weak coupling algorithm (Berendsen et al., 1984).

Complete atomic coordinate sets were extracted every 10 ps of the production run trajectory and saved as a new more sparsely sampled trajectory containing 25,000 frames representing the entire 250 ns of the production run. This new trajectory was mapped onto a cubic box centered on the CelF with an RMS fitting algorithm, which superimposed the CelF in each frame onto that in the initial frame, while moving all the other molecules inside the cubic box. All frames of this re-centered trajectory were then used to calculate the average volume density map of glucose (Schmidt et al., 1996; Tavagnacco et al., 2011) with respect to the protein molecule.

The volume density map was calculated using the VolMap tool of VMD (Humphrey et al., 1996). A selected set of atoms that included all of the glucose ring heavy atoms (C1, C2, C3, C4, C5, and O5) was used to represent the movement of the glucose molecules. The resolution of the density map was set to be 1 Å, and the weight was set to be the "occupancy," with 1 referring to the grid occupied and 0 referring to the grid being unoccupied.

Results and Discussion

As would be expected for an osmolyte such as glucose, there was no general tendency observed for the glucose to associate with the protein surface, consistent with results of a similar previous study of glucose interacting with the small bee venom peptide melittin (Mason et al., 2011). However, certain specific locations on the enzyme surface did exhibit an affinity for the sugar solute. The volume density map of β-D-glucopyranose at the density isovalue of 3.5 times bulk density found that multiple small regions on the CelF protein surface had a high glucose occupancy during the simulation. These higher densities can be interpreted as an increased affinity of the surface for glucose (Fig. 2). Two of the most significant of these binding sites are adjacent to the flat surfaces of Trp 310, at the outer edge of the entrance to the active site tunnel (site "1"), and Trp 312, which is further along this tunnel, roughly one cellobiose unit further in toward the active site (site "2"; Fig. 3). A third site of significant glucose association was at the exit opening of the active site tunnel, on the opposite side of the protein (site "3"; Fig. 4). Clouds 1 and 2 are located at the active site tunnel entrance subsites −7 and −5, respectively, stacking onto a Trp residue at each site, suggesting that Trp residues facilitate cellulose chain binding to the tunnel. This result is consistent with the theoretical study by Wohlert and co-workers, which demonstrated weak binding affinity between indole and β-D-glucopyranose via a stacking interaction (Wohlert et al., 2010). The glucose molecules did not exhibit any increased residence densities adjacent to any of the other residues constituting the surface region immediately around the tunnel entrance.

Insufficient sampling can potentially be an issue for MD simulations on large diffusion-controlled molecular systems like this. In the present case, however, many exchanges were observed for each occupancy site (Table I). In addition, qualitatively similar results were obtained for both the first and second halves of the simulation. Cloud 2 is located in the inner region of the tunnel entrance, and the presence of a glucose molecule at the site 1 region could easily prevent a glucose at site 2 from diffusing back to the cloud 1 region or escaping from the tunnel altogether, reducing the number of glucose exchanges at this site. However, even for this position, 16 exchanges were observed during the simulation period. Only one jump was observed where the glucose occupying site 2 diffused further along the tunnel toward the catalytic active site. On the other side of the enzyme, cloud 3 is surrounded by a funnel-shaped tunnel exit (Fig. 4). Glucose

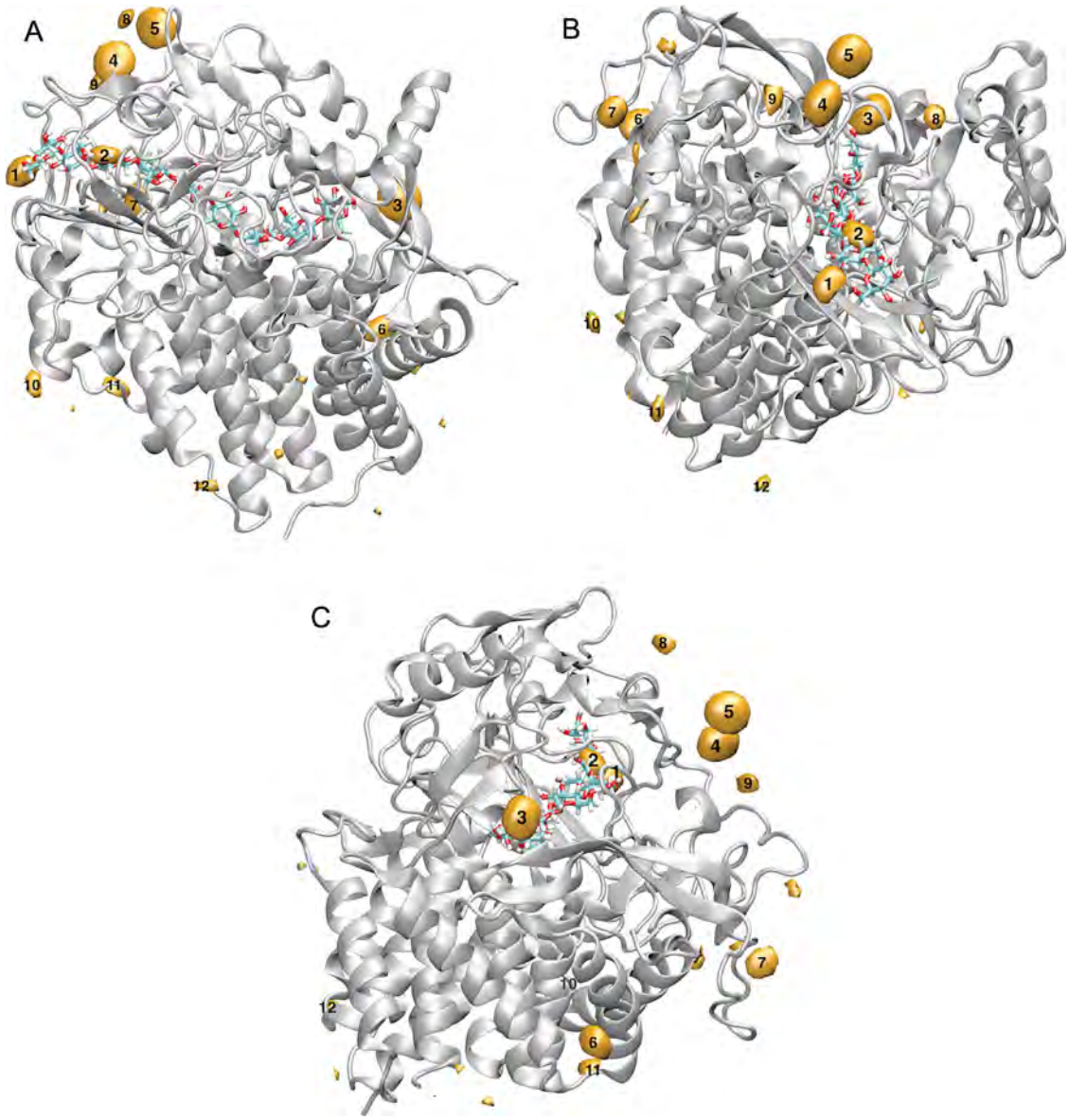

Figure 2. The volume density map of β-D-glucopyranose ring heavy atoms (C1, C2, C3, C4, C5, and O5) at the isovalue of ~3.5 times bulk density. The CelF backbone is shown as a ribbon cartoon. The density clouds of the selected glucose ring atoms are shown in orange. The substrate cellooligomer (DP = 9) in the active site tunnel is superimposed to highlight the tunnel, but was not present in the MD simulation. The substrate monomer subsites at the entrance of the tunnel are labeled as "1" and "2," and the comparable subsite at the tunnel exit is labeled as "3"; the first two subsites are defined by tryptophan side chain indole groups.

molecules in cloud 3 had a residence time at this site that was relatively longer, on the order of ~11 ns on average, with a very large standard deviation, and with the lowest number of exchanges of any of the major sites, including even the sterically constrained site 2. This strong glucose binding at the tunnel exit implies significant product inhibition, as was also observed in another study (Chen et al., 2015). The localized glucose molecules in clouds 1, 2, and 3 do not exhibit preferences in terms of binding direction within this short simulation. As family 48 cellulases are known to acquire cellulose chains from their reducing end, other mechanisms might exist for progressing of the correct chain end into the tunnel.

Hydrogen bonding between glucose molecules and the local amino acid residues certainly played an important role in the glucose localization. Analysis of the temporal evolution of the intermolecular hydrogen bonds between the localized glucose molecules in each density cloud and the local protein residues was carried out. Geometric criteria were used to define the hydrogen bonds; these were: a donor–acceptor distance less than 3.4 Å and an angle cutoff greater than 145°. Additionally, the occurrence of glucose localization in each density cloud was characterized; the glucose was considered localized if any atom of a glucose molecule was within 1 Å of the geometric center of the density cloud. The results demonstrated that formation of multiple hydrogen

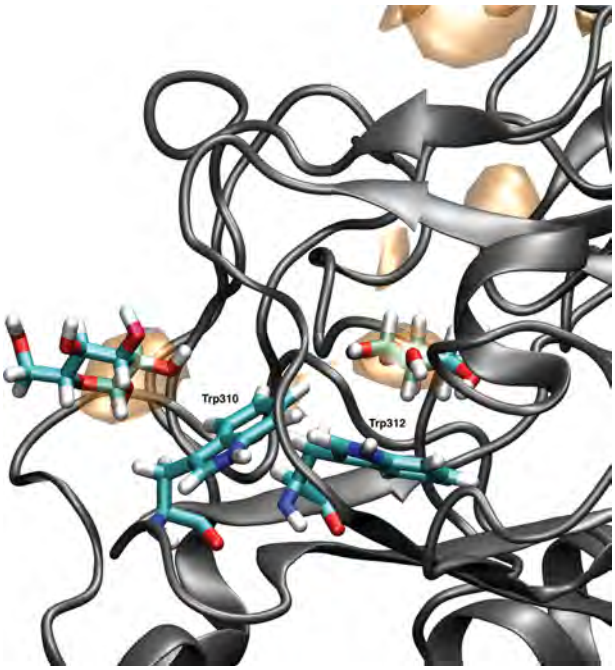

Figure 3. The volume density map of β-D-glucopyranose ring atoms around the entrance to the active site tunnel, displaying the Trp 310 (left, site 1) and Trp 312 (right, site 2) side chains, and with a typical instantaneous orientation of a glucose ligand indicated at each site.

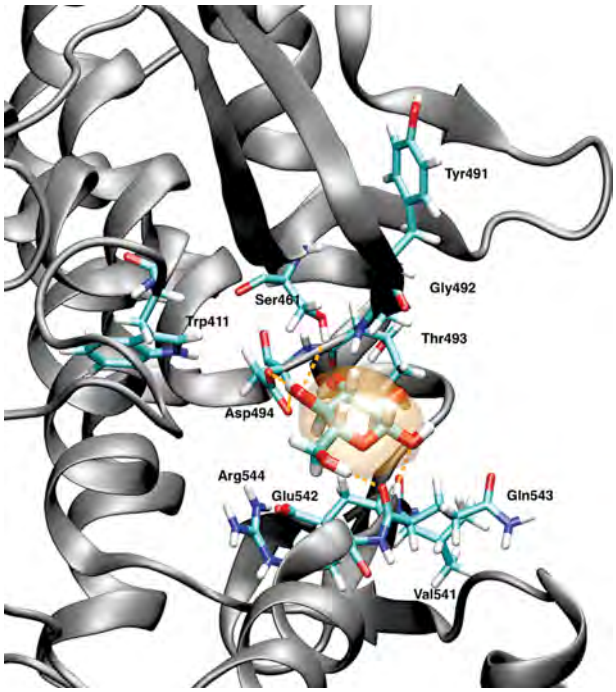

Figure 4. The exit opening of the active site tunnel of CelF, illustrating selected residues and the density cloud associated with binding site 3, with a representative instantaneous orientation of a glucose ligand superimposed. Hydrogen bonds are indicated in orange.

bonds was a requirement for the occurrence of the localized glucose molecules in each density cloud, even those stacked against a Trp residue, and an average of 3–5 hydrogen bonds were involved in glucose binding to the local protein surface (see Figs. 4–8).

Two additional significant binding sites, numbered 4 and 5 in Figure 2, were also observed. These sites are quite close to one another on the surface, and the sugar molecules that occupied these sites generally made two hydrogen bonds to one another, as well as three additional hydrogen bonds each to protein residues (Figs. 6–8). Neither of these sites featured a tryptophan indole side chain for the sugar rings to stack against, but in the case of site 4, the aromatic ring of Phe 115 was adjacent to the site (Fig. 6). In site 5, the planar side chain guanidinium group of Arg 381 was also adjacent to the binding site and could serve as a stacking surface for the sugar molecules in site 5 (Fig. 7). Although the guanidinium group is formally charged, previous work has shown that the planar faces of this molecule do behave like a non-polar surface, leading to guanidinium association in aqueous solution and explaining how guanidinium functions as a denaturant (Mason et al., 2004). The two glucose molecules occupying sites 4 and 5 are somewhat further apart than successive residues in cellobiose or a cellulose chain would be (Fig. 8), and it is unclear from the present results whether this collective site would have any enhanced affinity for a cellulose chain, but the possibility would make an interesting subject for future study in a larger simulation with cellobiose solutes.

The binding free energy for glucose molecules in the density cloud at the various binding sites was estimated from the volume density map using the same procedures previously used for other systems (Chen et al., 2012; Mason et al., 2011). Here, the bulk density of the selected atoms is estimated to be the averaged density of all the grids that are at least double the glucose length (~11 Å) away from the CelF protein surface (treating the protein as an approximate sphere with $d = 65$ Å), giving a value of 0.00523 atoms per Å^3. The binding free energy of glucose in density cloud 1 is ~1.0 kcal/mol (Table II), which is within the same magnitude as the calculated Trp-glucose binding energy in melittin (Mason et al., 2011). A comparable binding energy was estimated for site 2, but with significantly fewer exchanges. The energy of site 3 was estimated to be somewhat stronger, ~1.3 kcal/mol,

Table I. The number of glucose exchanges for each density cloud.

Cloud label	Number of glucose exchanges[a]	Residence time (ns) (mean ± std)
1	48	0.94 ± 4.05
2	16	3.46 ± 13.25
3	14	10.95 ± 36.88
4	24	5.22 ± 23.69
5	30	3.92 ± 20.06
6	34	2.01 ± 11.19
7	74	0.66 ± 3.41
8	93	0.41 ± 1.26
9	51	0.63 ± 1.54
10	61	0.55 ± 1.37
11	87	0.37 ± 0.49
12	129	0.27 ± 0.71

[a]This is the count of localized β-D-glucopyranose molecules with different residue IDs.

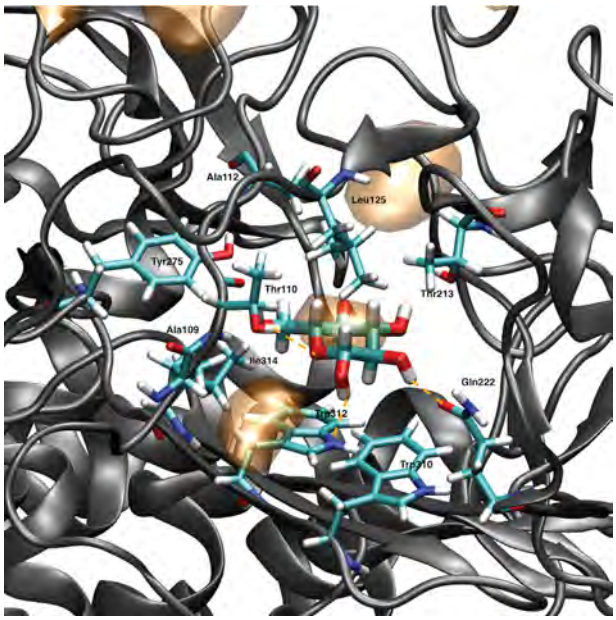

Figure 5. The entrance of the active site tunnel of CelF, illustrating selected residues in atomic detail, including Trp 310 and 312, and the glucose ring density adjacent to Trp 312, site 2, calculated as in Figure 1. Three stabilizing hydrogen bonds are shown in orange.

suggesting significant product inhibition. As can be seen from the table, the estimated energies were not strongly sensitive to the arbitrary choice of the energy contour selected as the site boundaries.

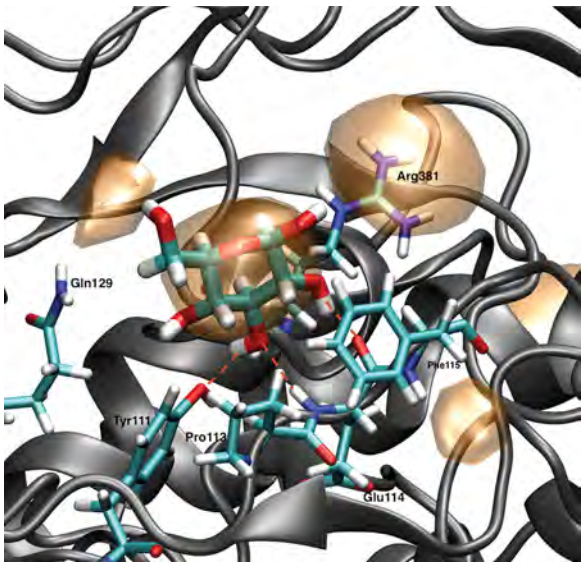

Figure 6. An illustration of a typical instantaneous configuration of a glucose molecule occupying binding site 4 in the simulations. Three stabilizing hydrogen bonds to protein surface residues are shown in orange. The density cloud for site 5 is to the right of this molecule in the figure, and the glucose molecules occupying that site typically also made two hydrogen bonds to the sugar in site 4.

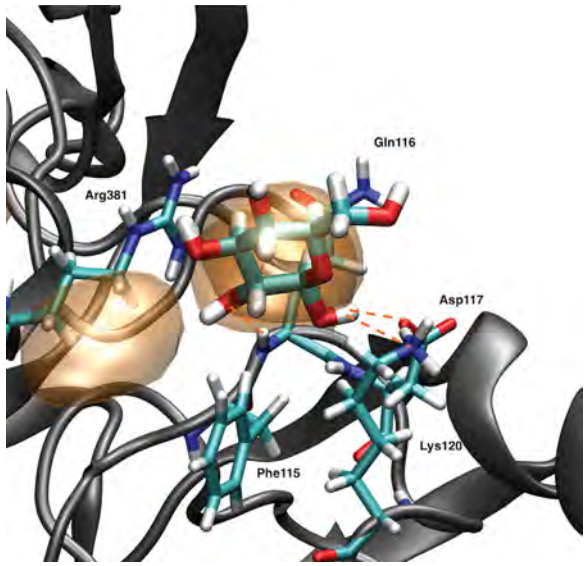

Figure 7. An illustration of a typical instantaneous configuration of a glucose molecule occupying binding site 5 in the simulations. Three stabilizing hydrogen bonds to protein surface residues are shown in orange.

Glucose binding was observed at several other sites on the surface of CelF, labeled as sites 6–12 in Figure 2, but all of these locations are far from either the active site tunnel entrance or exit, and thus are unlikely to play a role in directing substrate chains into the tunnel. As already stated, no additional binding sites were identified at the tunnel entrance; however, the proximity of other residues to sites 1 and 2 besides Trp 310 and 312 might suggest that

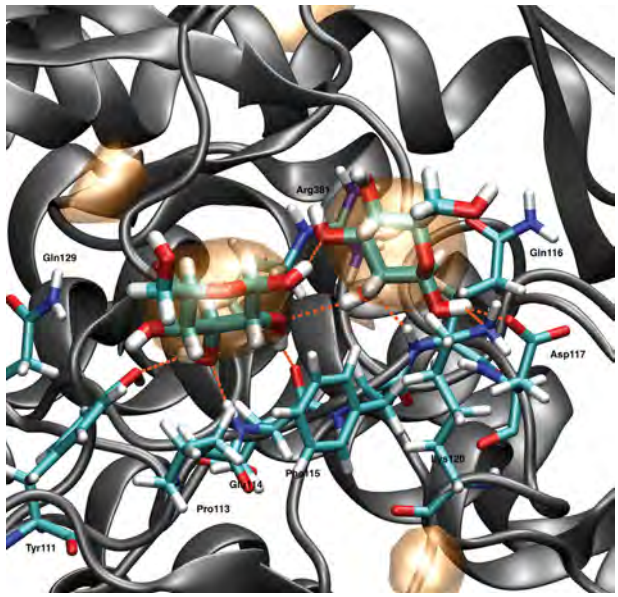

Figure 8. An illustration of a typical instantaneous configuration of glucose molecules simultaneously occupying binding sites 4 and 5 in the simulations. Two stabilizing hydrogen bonds between the sugar molecules are indicated, as well as hydrogen bonds to protein surface residues.

Table II. The binding energy of β-D-glucopyranose at sites 1 and 2 at the tunnel entrance and at site 3 at the tunnel exit.

Site 1	Contour level (×bulk density)	K_{eq}	Calculated binding energy (kcal/mol)
	3.5	4.56	−0.90
	4.0	4.97	−0.95
	4.5	5.43	−1.00
	5.0	5.62	−1.02
Site 2	Contour level (×bulk density)	K_{eq}	Calculated binding energy (kcal/mol)
	3.5	4.96	−0.95
	4.0	5.49	−1.01
	4.5	5.72	−1.03
	5.0	6.34	−1.09
Site 3	Contour level (×bulk density)	K_{eq}	Calculated binding energy (kcal/mol)
	3.5	8.51	−1.27
	4.0	9.31	−1.32
	4.5	9.87	−1.35
	5.0	10.3	−1.38

they play an additional role in the substrate affinity, perhaps through hydrogen bonding to the sugar hydroxyl groups, and might be candidates for site-directed mutagenesis experiments.

Conclusions

Although glucose is an osmolyte, β-D-glucopyranose, the monomer repeat unit of a cellulose chain, has a weak binding affinity to certain regions on the surface of CelF, which is a family 48 exocellulase, as is shown by the calculated volume probability density map of glucose on the CelF surface. The localized glucose molecules make multiple hydrogen bonds with the local protein residues. In particular, glucose tends to bind at the active site tunnel entrance and exit, where it forms a stacking interaction with Trp residues, or hydrogen bonds with the surrounding protein residues, in the case of site 3. From the 250 ns simulation, a sufficient number of glucose exchanges at the tunnel entrance −7 subsite (site "1" of Fig. 1) were observed to allow the binding affinity to be calculated. The binding energy between glucose and this specific site was estimated to be ∼1.0 kcal/mol. This result, to some degree, supports the general assumption that family 48 cellulases acquire cellulose chains directed into the entrance of the active site tunnel via favorable interactions with residues that surround the entrance. Additionally, glucose binding at the tunnel exit (labeled as site "3" in Fig. 1) was observed to last for a longer period of time (on the order of >11 ns on average). This binding phenomenon indicates significant product inhibition at this site (Chen et al., 2015).

This project was supported by the U.S. Department of Energy (DOE) Office of Science, Office of Biological and Environmental Research, through the BioEnergy Science Center (BESC), a DOE Bioenergy Research Center. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1053575.

References

Becker D, Braet C, Brumer IH, Claeyssens M, Divne C, Fagerstrom B, Harris M, Jones T, Kleywegt G, Koivula A, Mahdi S. 2001. Engineering of a glycosidase Family 7 cellobiohydrolase to more alkaline pH optimum: The pH behaviour of Trichoderma reesei Cel7A and its E223S/A224H/L225V/T226A/D262G mutant. Biochem J 356:19–30.

Beckham GT, Ståhlberg J, Knott BC, Himmel ME, Crowley MF, Sandgren M, Sørlie M, Payne CM. 2014. Towards a molecular-level theory of carbohydrate processivity. Curr Opin Biotechnol 27:96–106.

Berendsen HJC, Postma JPM, van Gunsteren WF, Di Nola A, Haak JR. 1984. Molecular dynamics with coupling to an external bath. J Chem Phys 81:3684–3690.

Boraston AB, Bolam DN, Gilbert HJ, Davies GJ. 2004. Carbohydrate-binding modules: Fine-tuning polysaccharide recognition. Biochem J 382:769–781.

Brady JW, Tavagnacco L, Ehrlich L, Chen M, Schnupf U, Himmel ME, Saboungi M-L, Cesàro A. 2012. Weakly-hydrated surfaces and the binding interactions of small biological solutes. Eur Biophys J 41:369–377.

Brooks BR, Brooks CL, MacKerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A. 2009. CHARMM: The biomolecular simulation program. J Comput Chem 30(10):1545–1614.

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.

Case DA, Darden TA, Cheatham III, TE, Simmerling CL, Wang J, Duke RE, Luo R, Walker RC, Zhang W, Merz KM, Roberts B. 2012. AMBER 12. San Francisco: University of California, *1*(3).

Chen M, Bomble YJ, Himmel ME, Brady JW. 2012. Molecular dynamics simulations of the interaction of glucose with imidazole in aqueous solution. Carbohydr Res 349:73–77.

Chen M, Bu L, Alahuhta M, Brunecky R, Xu Q, Lunin VV, Brady JW, Crowley MF, Himmel ME, Bomble YJ. 2015. Strategies to reduce end-product inhibition in family 48 glycoside hydrolases. Proteins Struct Funct Bioinf in press.

Crowley MF, Williamson MJ, Walker RC. 2009. CHAMBER: Comprehensive support for CHARMM force fields within the AMBER software. Int J Quantum Chem 109:3767–3772.

Guimarães BG, Souchon H, Lytle BL, Wu JHD, Alzari PM. 2002. The crystal structure and catalytic mechanism of cellobiohydrolase CelS, the major enzymatic component of the *Clostridium thermocellum* cellulosome. J Mol Biol 320: 587–596.

Guvench O, Greene SN, Kamath G, Brady JW, Venable RM, Pastor RW, Mackerell A D. 2008. Additive empirical force field for hexopyranose monosaccharides. J Comput Chem 29:2543–2564.

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.

Kostylev M, Alahuhta M, Chen M, Brunecky R, Himmel ME, Lunin VV, Brady J, Wilson DB. 2014. Cel48A from thermobifida fusca: Structure and site directed mutagenesis of key residues. Biotechnol Bioeng 111:664–673.

MacKerell AD, Bashford D, Bellott MLDR, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha SA, Joseph-McCarthy D. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102(18):3586–3616.

Mackerell AD, Feig M, Brooks CL. 2004a. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J Comput Chem 25:1400–1415.

MacKerell AD, Feig M, Brooks CL. 2004b. Improved treatment of the protein backbone in empirical force fields. J Am Chem Soc 126:698–699.

Mason PE, Lerbret A, Saboungi M-L, Neilson GW, Dempsey CE, Brady JW. 2011. The interactions of glucose with a model peptide. Proteins Struct Funct Bioinf 79:2224–2232.

Mason PE, Neilson GW, Enderby JE, Saboungi M-L, Dempsey CE, MacKerell AD, Brady JW. 2004. The structure of aqueous guanidinium chloride solutions. J Am Chem Soc 126:11462–11470.

Neria E, Fischer S, Karplus M. 1996. Simulation of activation free energies in molecular systems. J Chem Phys 105:1902–1919.

Pakarinen A, Haven MØ, Djajadi DT, Várnai A, Puranen T, Viikari L. 2014. Cellulases without carbohydrate-binding modules in high consistency ethanol production processes. Biotechnol Biofuels 7:27.

Parsiegla G, Juy M, Reverbel-Leroy C, Tardif C, Belaich JP, Driguez H, Haser R. 1998. The crystal structure of the processive endocellulase CelF of Clostridium cellulolyticum in complex with a thiooligosaccharide inhibitor at 2.0 Ångstrom resolution. EMBO J 17:5551–5562.

Parsiegla G, Reverbel C, Tardif C, Driguez H, Haser R. 2008. Structures of mutants of cellulase Cel48F of Clostridium cellulolyticum in complex with long hemithiocellooligosaccharides give rise to a new view of the substrate pathway during processive action. J Microbiol 375:499–510.

Parsiegla G, Reverbel-Leroy C, Tardif C, Belaich JP, Driguez H, Haser R. 2000. Crystal structures of the cellulase Cel148F in complex with inhibitors and substrates give insights into its processive action. Biochemistry 39:11238–11246.

Payne CM, Bomble Y, Taylor CB, McCabe C, Himmel ME, Crowley MF, Beckham GT. 2011. Multiple functions of aromatic-carbohydrate interactions in a processive cellulase examined with molecular simulation. J Biol Chem 286:41028–41035.

Quiocho FA. 1989. Protein-carbohydrate interactions: Basic molecular features. Pure Appl Chem 61:1293–1306.

Quiocho FA, Vyas NK, Spurlino JC. 1991. Atomic interactions between proteins and carbohydrates. Trans Am Crystallogr Assoc 25:23–35.

Ryckaert JP, Ciccotti G, Berendsen HJ. 1977. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. J Comput Phys 23(3):327–341.

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.

Schneider T, Stoll E. 1978a. Molecular-dyamics study of a 3-dimensional one-component model for distortive phase-transitions. Phys Rev B 17:1302–1322.

Schneider T, Stoll E. 1978b. Molecular-dynamics study of 2nd sound. Phys Rev B 18:6468–6482.

Tavagnacco L, Mason PE, Schnupf U, Pitici F, Zhong L, Himmel ME, Crowley M, Cesàro A, Brady JW. 2011. Sugar binding sites on the surface of the carbohydrate binding module of CBH I from *Trichoderma reesei*. Carbohydr Res 346:839–846.

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

van Gunsteren WF, Berendsen HJC. 1988. A leap-frog algorithm for stochastic dynamics. Mol Sim 1:173–185.

Wohlert J, Schnupf U, Brady JW. 2010. Free energy surfaces for the interaction of glucose with planar aromatic groups in aqueous solution. J Chem Phys 133:155103.

Zou J-Y, Kleywegt GJ, Ståhlberg J, Driguez H, Nerinckx W, Claeyssens M, Koivula A, Teeri TT, Jones TA. 1999. Crystallographic evidence for substrate ring distortion and protein conformational changes during catalysis in cellobiohydrolase Cel6A from *Trichoderma reesei*. structure 7:1035–1045.