Published in final edited form as:

J Mol Biol. 2013 November 15; 425(22): 4267–4285. doi:10.1016/j.jmb.2013.05.030.

Structure, Dynamics, and Specificity of Endoglucanase D from *Clostridium cellulovorans*

Christopher M. Bianchetti^{1,2}, Phillip Brumm^{2,3}, Robert W. Smith^{1,2}, Kevin Dyer⁴, Greg L. Hura⁴, Thomas J. Rutkoski^{1,2}, and George N. Phillips Jr.^{1,2,5}

¹Department of Biochemistry, University of Wisconsin, Madison, WI 53706, USA

²Great Lakes Bioenergy Research Center, University of Wisconsin, Madison, WI 53706, USA

³Lucigen Corporation and C5-6 Technologies, Madison WI 53562, USA

⁴Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, 94720, USA

⁵Department of Biochemistry and Cell Biology Rice University, Houston, TX 77005

Abstract

The enzymatic degradation of cellulose is a critical step in the biological conversion of plant biomass into an abundant renewable energy source. An understanding of the structural and dynamic features that cellulases utilize to bind a single strand of crystalline cellulose and hydrolyze the β -1,4-glycosidic bonds of cellulose to produce fermentable sugars would greatly facilitate the engineering of improved cellulases for the large-scale conversion of plant biomass. Endoglucanase D (EngD) from *Clostridium cellulovorans* is a modular enzyme comprising a N-terminal catalytic domain and a C-terminal carbohydrate-binding module (CBM), which is attached *via* a flexible linker. Here, we present the 2.1 Å resolution crystal structures of full-length EngD with and without cellotriose bound, solution small angle X-ray scattering (SAXS) studies of the full-length enzyme, the characterization of the active cleft glucose binding subsites, and substrate specificity of EngD on soluble and insoluble polymeric carbohydrates. SAXS data support a model in which the linker is flexible, allowing EngD to adopt an extended conformation in solution. The cellotriose-bound EngD structure revealed an extended active site cleft that contains seven glucose-binding subsites, but unlike the majority of structurally determined endocellulases the active site cleft of EngD is partially enclosed by Trp162 and Tyr232. EngD variants, which lack Trp162, showed a significant reduction in activity and an alteration in the distribution of cellohexaose degradation products suggesting that Trp162 plays a direct role in substrate binding.

Address correspondence to: George N. Phillips, Jr., Department of Biochemistry and Cell Biology, Rice University, MS140, Houston, TX 77005 USA; Telephone: (713) 348-6951, georgep@rice.edu.

Accession numbers

Coordinates and structure factors for both the EngD and cellotriose-bound EngD structures have been deposited in the Protein Data Bank with accession numbers 3NDY and 3NDZ, respectively.

Keywords

cellulase; endoglucanase; cellulose degradation; small angle X-ray scattering; X-ray crystallography

Introduction

Cellulolytic microorganisms produce a diverse array of glycosyl hydrolases (GH) that depolymerize plant cell wall polysaccharides. To date, 131 families of GHs and 66 families of carbohydrate-binding modules (CBM) have been identified and organized by sequence similarity (Carbohydrate Active Enzyme database, <http://www.cazy.org/>)¹. In particular, cellulases catalyze the hydrolysis of the β -1,4-glycosidic bonds of cellulose. These microorganisms utilize the resulting degradation products as both a carbon and energy source and play an essential role in biomass degradation. The potential utilization of plant biomass to provide an abundant renewable “green” energy source has reinvigorated both the search for novel cellulases from extremophiles as well as efforts to engineer more physically robust cellulases for use in industrial biomass conversion^{2,3}. The structural characterization of model cellulases, such as Endoglucanase D (EngD), will guide these design efforts⁴ as well as contribute to the basic understanding of how these enzymes accomplish their formidable task.

EngD is produced by the anaerobic, mesophilic, cellulolytic bacterium *Clostridium cellulovorans*⁵. The catalytic domain of EngD belongs to the GH family 5 and exhibits both cellulase and xylanase activities^{6,7}. Members of the GH5 family adopt a $(\beta/\alpha)_8$ fold with the catalytic carboxylate residues located on the fourth and seventh strands. A large degree of sequence diversity exists among the GH5 family members with only eight strictly conserved residues observed across the entire family⁸. EngD does not associate with the cellulosome, but instead interacts with insoluble substrate *via* a CBM. The CBM of EngD, designated here as CBM_{EngD}, belongs to the CBM family 2 and is attached to the catalytic domain *via* a flexible proline/threonine-rich (PT) linker. Foong *et al.* demonstrated that CBM_{EngD} was capable of binding cellulose and that when this domain was removed the activity of EngD on insoluble substrate was diminished dramatically⁶.

One of the structural hallmarks of endoglucanases in general, and members GH5 in particular, is an open active site cleft that runs along the face of the catalytic domain. The cleft is large enough to accommodate a single strand of cellulose and contain several distinct glucose binding sites⁹⁻¹³. The EngD structures presented here reveal two residues on the exterior of the cleft, Trp162 and Tyr232, which form a “clamp” that encloses the cleft. Trp162 and Tyr232 would have to be transiently displaced in order for a cellulose strand to bind, suggesting that Trp162 and Tyr232 may play an important role in the interaction of cellulose. Previous studies have demonstrated that the aromatic residues lining the active site clefts of cellulases and chitinases are important in modulating substrate binding and reactivity. In particular, the mutation of aromatic residues on the exterior of the cleft reduces activity toward crystalline substrates, while not diminishing the reactivity towards soluble or amorphous substrates, and in some cases actually increasing rates¹⁴⁻¹⁹. In order to

understand the role Trp162 and Tyr232 play in substrate binding and reactivity we constructed and biochemically characterized the product distribution of three EngD variants which lack one (EngD_{W162A}, EngD_{Y232A}) or both (EngD_{DM}) of these aromatic residues.

Due to the inherent flexibility of the linker between the catalytic domain and CBM domains and the negative effects of disorder on the production of diffraction-quality protein crystals, the majority of known cellulase structures consist solely of the catalytic domain. Here, we present the structures of both the catalytic and CBM domains of full-length EngD, which was crystallized with and without bound cellotriose and solved to a resolution of 2.1 Å. This is the first family 2 CBM to be solved by X-ray crystallography. In addition to the X-ray structure, SAXS data from intact EngD allowed the modeling of the interactions between the catalytic and CBM domains. Biochemical characterization of the EngD variants along with the wild type EngD crystal structure with and without bound cellotriose provides insight into how endocellulases interact with a bound cellulose chain. Understanding of these critical cellulase-cellulose interactions will improve our ability to design synthetic cellulases.

Results

Overall Structure and Structure Quality

The EngD construct utilized for crystallization contained a catalytic domain, a PT linker, a CBM, and a C-terminal non-cleavable His₆ affinity tag. Both the unliganded and cellotriose-bound EngD structures had well-defined electron density corresponding to the catalytic domain (residues 4-348) and the CBM (residues 381-487) domains. Electron density was not observed in either of the EngD structures for either the PT linker (residues 349-380), presumably resulting from the intrinsic disorder of the flexible linker region, or the C-terminal non-cleavable His₆ affinity tag (residues 488-515). In both structures there are four molecules per asymmetric unit. The four monomers in the asymmetric unit are essentially identical with an average C_α root mean square deviation of 0.30 Å between monomers in the cellotriose-bound structure. All pertinent information on data collection, refinement, and model statistics are summarized in Table 1.

EngD Catalytic Domain

The catalytic domain of EngD (residues 4-348) adopts a slightly modified (β/α)₈ barrel fold which is similar to the fold observed in several GH families including the GH5 family^{11, 20-22} (Figure 1A). One departure from the canonical (β/α)₈ fold observed in the EngD structure is the presence of an additional α-helix (α0) at the N-terminus enclosing one end of the β-barrel. A second departure from the prototypical (β/α)₈ fold is the truncation of α-helix 5 relative to the other helices surrounding the internal β-barrel. Finally, there are two single-turn helices located on extended loops connecting β-strand 4 to α-helix 4 and β-strand 6 to α-helix 6. The two extended loops and the truncated α-helix 5 form a cleft that runs along the face of the β-barrel. Two residues, Trp162 and Tyr232, located on the extended loops, form a “clamp” that encloses the cleft. The positions of Trp162 and Tyr232 near the active site suggest a role in substrate binding and EngD variants, which lack one (EngD_{W162A} and EngD_{Y232A}) or both (EngD_{DM}) residues were constructed to study their function.

EngD Carbohydrate-Binding Module

The CBM of EngD, here designated CBM_{EngD}, adopts a β -sandwich fold composed of two four-stranded antiparallel β -sheets, which stack on top of each other forming a wedge-like structure. One face of CBM_{EngD} is flat and displays three aromatic residues, Trp393, Trp430, and Tyr448, which likely define a cellulose-binding site (Figure 1D). Trp393, Trp430, and Tyr448 are structurally homologous to three tryptophan residues observed in the CBM of C_{ex} from *Cellulomonas fimi* (PDB ID: 1EXG)²³, which have been shown to be involved in cellohexaose binding²⁴, suggesting that Trp393, Trp430, and Tyr448 are involved in cellulose binding. Several polar residues, Ser395, Asn391, Ser398, Thr427, Asn428, Ser431, Asn449, Asn464, and Asn466, flanking Trp393, Trp430, and Tyr448 could potentially form hydrogen bonds with the hydroxyl groups of a bound cellulose chain (Figure 1D).

Interactions Between the Catalytic and Cellulose Binding Domain

Due to the inherent flexibility of the linker region that connects the catalytic domain to the CBM, nearly all the structures of GH5 family members have been solved without their CBM domains. In the asymmetric unit there are four CBM_{EngD} monomers and each directly interacts with one catalytic domain. The interface between a CBM_{EngD} and the corresponding catalytic domain has an average surface area of 591 Å² as calculated by PISA²⁵. Without definitive electron density for the PT linker, it is not possible to assign a particular CBM_{EngD} to a corresponding catalytic domain. Consequently, each of the CBM_{EngD} and catalytic domains were modeled and refined as separate chains.

To better understand the interactions between the catalytic domain and the CBM domain we performed SAXS experiments on the same full-length EngD construct that was used for crystallization. Monomers created from the crystal structure by modeling realistic linkers between the catalytic and the CBM_{EngD} domains fit the SAXS data poorly with a best Chi agreement of 6.2. The calculated distance distribution function, P(r), from SAXS data had significantly larger dimensions than was observed in the crystal structure and had two separated distributions of mass (Fig 2A). Utilizing a standard of known molecular weight and concentration, the mass calculated from the extrapolated X-ray intensity at zero scattering angle from EngD was 70kDa with and expected monomer molecular weight of 58kDa. The single peak from size exclusion chromatography was symmetric and characteristic of a single species (Data not shown). Thus SAXS data analysis was conducted assuming the dominant species in solution was monomeric. The over-estimate of molecular weight is likely due to errors concentration measuring upon which the molecular weight estimate relies. The Radius of gyration (R_g) was significantly larger than crystallographic models and that expected of a globular proteins of the same size. Larger than expected R_g have been observed in other full-length cellulases²⁶⁻²⁹. The large R_g and bimodal P(r) suggested EngD adopts an extended conformation in solution where the catalytic domain and the CBM_{EngD} do not interact and are separated by the flexible linker.

Due to the flexible nature of full-length EngD a library of conformations was generated to fit the SAXS data (Figure 2C). BILBOMD was utilized to create a large and comprehensive ensemble of possible conformations of EngD where the catalytic domain and CBM_{EngD}

independently retain their structures but are flexibly linked (Fig 2C). This large ensemble was then further refined to identify minimal ensembles. The single best conformer fit with a Chi of 2.5 and a center-to-center distance from catalytic domain to CBM_{EngD} of 60 Å. Three conformers fit with a Chi of 1.1. A further improved fit (χ of 0.95, where a χ = 1 indicates agreement within error bars) was obtained when five conformers of EngD with near even fractional contributions were utilized (Figure 2D). The compact conformations of the catalytic domain and CBM_{EngD} observed in the crystal structure were not components of the five conformers used to fit the SAXS data. Suggesting that the tight association in the crystal structure is driven by the solution conditions leading to crystallization and does not reflect the flexible nature of full-length EngD in solution. The average center-to-center distance for the five member ensemble was 73 Å but included members with a low of 53 and high of 111 Å. Thus the position and orientation of the catalytic relative to the CBM_{EngD} domain was diverse among the five conformations with similar weightings. This results supports flexibility between the domains as no single dominating conformations could be found with a similar quality of fit to the experiment.

Active Site and Cellulose-Binding Cleft

There are eight strictly conserved residues across GH5 family members ^{8,30,31}. In EngD these residues correspond to Arg63, His107, Asn151, Glu152, His227, Tyr229, Glu275, and Trp308. The conserved residues are clustered together on one face of the β -barrel corresponding to the active site cleft. EngD hydrolyzes a glycosidic bond with retention of the anomeric configuration through a double replacement mechanism ^{32,33}. Based on structural alignments with other GH5 family members, Glu152 has been identified as the general acid/base and Glu275 as the nucleophile ^{11,13,34}, which is consistent with previous biochemical studies ^{8,35,36}.

Aromatic residues are known to be involved in the binding of carbohydrates through stacking ^{37,38,39}. Trp41, Trp162, Tyr229, Tyr232, and Tyr308 are located along the active site cleft and provide a favorable platform for cellulose binding (Figure 1B). Trp308 and Trp41 in the cellotriose-bound structure are observed to interact directly with cellotriose through a hydrogen bond with the glucose moiety occupying the -2 subsite (Trp308) and by stacking against the glucose moiety located at subsite -3 (Trp41). The binding cleft of EngD is sufficiently long to bind six or possibly seven glycosyl units. The cellotriose-bound structure clearly identifies four of the glucose-binding subsites (-3, -2, -1, and +3) and a possible fifth subsite (+4) (Figure 1C). The residues that interact directly with the cellotriose molecules are shown in Figure 3.

EngD•Cellotriose Complex

In the asymmetric unit of the cellotriose-bound EngD structure there are four EngD monomers and six cellotriose molecules. Each catalytic domain directly interacts with two cellotriose molecules. One cellotriose molecule occupies glucose binding subsites -3, -2, and -1 and appears to bind with high affinity. The second cellotriose molecule bridges two adjacent catalytic monomers (subsites +3 and +4) and could represent a low-affinity binding site that is occupied due to crystal packing. The electron density was well defined for both the active site and bridging cellotriose molecules and all cellotriose molecules were modeled

at 100% occupancy (Figure 4). Attempts were made to solve the crystal structure of EngD in complex with xylooligosaccharides and mixed β -1-3, 1-4 linked glucans but these attempts resulted in partial binding and ambiguous electron density (data not shown) when compared to the cellotriose-bound EngD structure. Between the two cellotriose molecules in the active site cleft, sufficient room remains to accommodate a cellobiose unit, which would occupy putative glucose binding subsites +1 and +2 (Figure 1C). Trp162, which is positioned on a long flexible loop, could interact with a glucose moiety located at the +2 subsite. In addition to Trp162, Tyr229 and Tyr232 could provide a binding platform for glucose moieties located at subsites +1 and +2.

The glucose moiety in the -1 site is anchored by several hydrogen bonds from four of the eight conserved residues in GH5 family members (His107, Asn151, Glu152, and Glu275) (Figure 3). The C-1 hydroxyl of the glucose at the -1 position participates in a hydrogen bond with the catalytic acid/base Glu152. Asn151 and the nucleophilic Glu275 hydrogen bond with the hydroxyl at the C-2 position. One of the conserved histidine residues, His107, forms a hydrogen bond with the C-3 hydroxyl of the glucose moiety at the -1 subsite. Glu319 interacts with both the C-6 hydroxyl of the glucose in subsite -1 and the C-2 hydroxyl of the glucose in the -2 subsite. The strictly conserved Trp308, along with Asn30 and Asn310, forms a series of hydrogen bonds with the glucose in the -2 subsite. The C-2 hydroxyl of the glucose moiety at the -2 subsite hydrogen bonds with both Trp308 and Asn310. Finally, Asn30 forms a hydrogen bond with the C-3 hydroxyl of the glucose molecule at the -2 position. The glucose molecule at the -3 subsite is mostly solvent-exposed and other than a stacking interaction with Trp41, there are no direct interactions with the protein.

Compared to the cellotriose bound near the active site, the bridging cellotriose molecule makes fewer interactions with the surrounding polypeptide (Figure 3). In the structure of EngD without cellotriose, the cavity where the bridging cellotriose would bind is completely solvated. Even though a non-crystallographic 2-fold axis passes through the bridging cellotriose molecule, the electron density can be satisfactorily fit with a single conformer of cellotriose (Figure 4B). This results in two possible binding conformations of the +3 glucose. In two of the four catalytic monomers (chain A and chain B), the +3 glucose is hydrogen bonded with the carbonyl oxygen of Ala205, the backbone nitrogen of Met206, and Ser204. In chains C and D, the +3 glucose interacts directly with Glu256 and indirectly with Ser230 through a water molecule. Unlike the glucose bound at the +3 subsite, which is coordinated by a single catalytic domain, two catalytic domains coordinate the glucose located at the possible +4 glucose subsite. The +4 glucose forms a hydrogen bond with Ser256 from one monomer (chain A or chain B) and Ser207 from the adjacent monomer (chain C or chain D). There are relatively few interactions with the glucose moiety bound in the +4 subsite when compared to the other subsites. It could be that the +4 subsite is a crystallographic artifact or represents a low-affinity binding site.

Sequence Homologues

In an attempt to clarify relationships amongst sequence-similar enzymes, a BLAST analysis of the entire EngD sequence was performed against the Uniprot database. The results show

that EngD is closely related to a number of enzymes from other cellulolytic organisms. The homology resides primarily in the GH5 domain of the enzymes. However, the organization of the enzymes shows three distinct structural motifs (Table 2): 1. Analogous to EngD, two of the homologues possess CBM 2 domain connected to the GH5 via a linker region. These enzymes may all be involved in degradation of cellulose or cellulosic fragments as described above; 2. Cellulosomal variants contain no CBM, rather a dockerin domain separating the GH5 from a lipase/esterase domain. This combination of several varied catalytic domains on one scaffold suggests these enzymes do not attack just cellulose, as it contains no ester linkages. A different cell-wall substrate is suggested, possibly xylan, where the combination of esterase and GH5 could act synergistically on a substituted xylan; 3. Variants containing only a GH5 domain would not be expected to be highly active on cellulosic substrates. These enzymes may act on soluble substrates such as *beta*-glucans xyloglucans, and related carbohydrates.

Substrate specificity

To better understand the function of EngD, the substrate specificity of EngD was determined using soluble and insoluble polymeric carbohydrates as described in Methods. The specific activity values (Table 3) demonstrate that the active site of the enzyme effectively binds and cleaves a range of substrates containing bonds other than β -(1-4) (β -glucan), substrates with side chain carbohydrates (xyloglucan), and substrates with β -(1-4) sugars other than glucose (xylan and glucomannan). EngD is unable to hydrolyze galactomannan, a β -(1-4)-linked mannan backbone with β -(1-6)-linked galactose residues.

To further study cellulose hydrolysis by EngD, extent of hydrolysis reactions were carried out comparing EngD to *Clostridium thermocellum* cellulases using the florescent reporter MUC, and 0.2% Avicel microcrystalline cellulose as substrate as described in Methods. Incubations with MUC showed EngD and all *C. thermocellum* GH5 enzymes hydrolyzed MUC (Table 4), an indicator of exo-cellulase activity. Avicel hydrolysis experiments with EngD were carried out for 113 hr at 40°C and with *C. thermocellum* enzymes at 60°C for 67 hr. The results (Table 4) show that the presence or absence a CBM module does not increase the amount of glucose produced.

Because EngD normally acts in the presence of cellulosomal enzymes, it is possible that either cellulosomal enzymes convert cellulose into a substrate for EngD, or EngD produces an intermediate product from cellulose that is a substrate for cellulosomal cellulases. To evaluate the possibility that additional cellulases are needed to demonstrate cellulolytic activity for EngD, cellulose hydrolysis was performed at 40°C using purified soluble *C. thermocellum* cellulases either alone or in combination with EngD. The results (Table 5) show that EngD acts synergistically with combinations of either 3 or 4 *C. thermocellum* cellulases. Synergy factors do not have a significant dependence on the combination of *C. thermocellum* cellulases used; synergy factors range from 1.3 to 1.6 for the three combinations tested.

Degradation of Crystalline Cellulose

To investigate the role of Trp162 and Tyr232, two single point EngD variants (EngD_{W162A}, EngD_{Y232A}) and a variant that replaced both Trp162 and Tyr232 with alanine (EngD_{DM}) were generated and the soluble products formed from the hydrolysis of Sigmacell 50 catalyzed by these variants were determined (Figure 5). The formation of glucose and cellobiose can be fit to a previously described logarithmic function⁴⁰. Cellobiose and glucose were the major products formed during the 120 hour hydrolysis. There was a small amount of cellotriose formed during hydrolysis, but only after 96 hours (data not shown). Based on the amount of glucose and cellobiose produced, EngD_{Y232A} displayed a slightly lower activity than the wild type. On the other hand, the EngD_{W162A} and EngD_{DM} variants showed a significant decrease in activity on Sigmacell 50. There was a 2 fold and 8 fold reduction in the amount of glucose produced by EngD_{W162A} and EngD_{DM}, respectively, compared to wild type (Figure 5A). Unlike EngD_{Y232A}, which showed a similar decrease in the production of both glucose and cellobiose, EngD_{W162A} and EngD_{DM} showed a larger decrease in production of cellobiose than glucose. EngD_{W162A} and EngD_{DM} produced 3 fold and 10 fold less cellobiose than wild type (Figure 5B). The binding of the EngD constructs towards Sigmacell 50 was measured to determine if the reduction in glucose and cellobiose was the result of decreased cellulose binding. The three EngD constructs showed similar Sigmacell 50 binding when compared to wild type and it appears that Trp162 and Tyr232 mutations do not affect cellulose binding (Figure 5C). Removal of CBM_{EngD} results in an insoluble protein product, and this construct was not tested.

Hydrolysis of Soluble Oligosaccharides

Previous studies have shown that mutation of aromatic residues near the active site cleft of cellulases and chitinases can cause an increase in activity towards soluble substrates¹⁴⁻¹⁹. To determine if the loss of Trp162 and/or Tyr232 would affect activity towards soluble substrates the release of products from the hydrolysis of cellobiose, cellotriose, cellotetraose, cellopentaose, and cellohexaose were monitored for wild type EngD and the three aromatic mutants. All EngD constructs were unable to hydrolyze cellobiose and cellotriose within 60 minutes and cellobiose and cellotriose appeared to inhibit activity similar to previous studies⁴¹. In addition to cellobiose and cellotriose, EngD_{W162A} and EngD_{DM} were unable to degrade cellotetraose and only produced measurable amounts of products on cellopentaose and cellohexaose (Figure 6). In a similar manner to the hydrolysis of Sigmacell 50, EngD_{Y232A} showed a small decrease in the amount of products compared to wild type EngD. Both wild type EngD and EngD_{Y232A} produced a similar ratio of cellobiose and cellotriose when reacted with cellotetraose. With the exception of EngD_{DM}, which produced marginal amounts of breakdown products, all of the EngD constructs primarily produced cellobiose and cellotriose from cellopentaose (Figure 6). A major shift in product distribution was seen in the hydrolysis of cellohexaose. EngD_{W162A} produces a similar amount of cellotriose as wild type, but only a fraction of the cellobiose (Figure 6). Given the small amount of products produced by EngD_{DM} it is not clear if EngD_{DM} follows a similar pattern. The reduction in the formation of cellobiose by EngD_{W162A} from cellohexaose is similar to the reduction in cellobiose formation seen in the hydrolysis of Sigmacell 50.

DISCUSSION

There has been increased interest in cellulases and their ability to degrade lignocellulosic biomass into fermentable sugars for the production of bioethanol. Crystalline cellulose poses several challenges to hydrolytic enzymes. The cellulase must first bind to an insoluble substrate, then disrupt the crystalline packing in order to extract a single cellulose strand, and finally position the newly freed strand in the catalytic site. It is known that the CBM of EngD is the driving force that localizes the catalytic domain to cellulose, but what is less well characterized is the role of the active site aromatic residues play in binding and positioning a single cellulose strand. The cellotriose bound structure and active site cleft aromatic mutants presented here elucidate the role of two key aromatic residues (Trp162 and Tyr232) and provide insight into the role these non-catalytic residues play in substrate binding.

The Role of Aromatic Residues Near the Active Site

One of the most prominent features of GH5 cellulases is the large cleft that runs across the catalytic face of the (β/α)₈ barrel to the active site. Several aromatic (Trp41, Trp162, Tyr229, Tyr232, and Tyr308) are located along the active site cleft and provide a favorable platform for cellulose binding (Figure 1B and Figure 3). Unlike the other aromatic residues that line the active site cleft of EngD, Trp162 and Tyr232 are located on flexible, solvent-exposed loops and form a “clamp” that encloses the active site cleft (Figure 1A). Trp162 and Tyr232 have higher than average B factors (26.5) compared to the average B factors of the cellotriose-bound EngD structure (16.3). Thus, the extended loops, which encompass Trp162 and Tyr232, are two of the more flexible portions of the structure. The flexibility of these loop regions should allow Trp162 and Tyr232 to interact with a bound cellulose strand and would explain the reduced active of EngD_{W162A}, EngD_{Y232A}, and EngD_{DM} and the altered product formation of EngD_{W162A}.

EngD_{W162A}, when compared to wild type EngD, showed significant reduction in the amount of glucose and cellobiose produced from the hydrolysis of Sigmacell 50. Although both glucose and cellobiose production was reduced, the ratio of glucose to cellobiose was altered in EngD_{W162A}. Wild type EngD produced approximately 2 mol of cellobiose for every 1 mol of glucose from Sigmacell 50, while EngD_{W162A} produced glucose and cellobiose in approximately equal molar amounts (Figure 5). A similar trend was observed in the breakdown products of cellohexaose. Wild type EngD produced an equal amount cellotriose and cellobiose in the degradation of cellohexaose. On the other hand, EngD_{W162A} produced approximately 5 times more cellotriose than cellobiose (Figure 6). These results and the cellotriose-bound structure suggest that Trp162 interacts directly with the +2 subsite and without this interaction the formation of cellobiose is less favorable and the overall activity is decreased. Similar reductions in activity were observed when a solvent exposed Trp residue (Trp272) was mutated in Cel6A from *Trichoderma reesei*¹⁹. The authors speculated that Trp272 of Cel6A is involved in the rate limiting step of crystalline substrate hydrolysis and that it may play a role in positing the cellulose chain within the active site cleft¹⁹. The Cel6A results are similar to what is observed in the Trp162 EngD variant and suggest that

Trp162 may play a similar role in the initial contact with cellulose in addition to stabilizing a bound cellulose strand at subsite +2.

Unlike Trp162, Tyr232 seems to have a limited influence on EngD activity. The EngD_{Y232A} variant showed only modest reductions in glucose and cellobiose formation from Sigmacell 50 when compared to wild type. Despite reduced activity, EngD_{Y232A} produced similar amounts of glucose from cellotetraose, cellopentaose, and cellohexaose when compared to wild type EngD (Figure 6). These assays suggest that Tyr232 may play a secondary role in stabilizing the bound substrate at the +2 subsite. Given the position of Tyr232, it is possible that Tyr232 acts to block bound substrate from exiting the active site cleft. Preventing bound substrate from dissociating would increase the amount of productive binding leading to an increase in activity. However, mutations of Tyr245 of Cel5A from *Acidothermus cellulolyticus*, which is located on an extended loop between β -strand 6 and α -helix 6 in a similar environment to Tyr232 of EngD, resulted in a substantial increase in the rate of hydrolysis of soluble substrates⁴². When the two structures are aligned, Tyr245 of Cel5A occupies a similar position as Tyr232 of EngD. Baker *et al.* suggested that the increased reaction rate was due to a reduction in product inhibition⁴². Based on the data presented here it is not possible to determine if EngD_{Y232A} has reduced cellobiose inhibition when compared to wild type EngD but it is clear that the loss of Tyr232 affects activity. It is possible that even though Tyr245 of Cel5A and Tyr232 of EngD are structurally similar they may play a different role in catalysis, but this remains to be determined.

EngD_{DM}, which lacks both Trp162 and Tyr232, showed a dramatic reduction in activity compared to wild type and EngD_{Y232A}. Only after 24 hours of Sigmacell 50 degradation did EngD_{DM} produce enough glucose or cellobiose to be measured by HPLC (Figure 5). EngD_{DM} was also ineffective in degrading cellotetraose, cellopentaose, and only produced measureable amount of cellotriose and cellobiose from cellohexaose (Figure 6). EngD_{W162A} showed that Tyr232 could marginally compensate for the loss of Trp162, but with the loss of both Trp162 and Tyr232 EngD is essentially inactive, which suggests that Tyr232 plays a supporting role in catalysis.

Oligosaccharide Binding

The cellotriose-bound crystal structure and degradation assays provide insight into the relative affinities of the subsites of EngD. None of the EngD constructs were able to degrade cellobiose or cellotriose, which suggests that cellobiose and cellotriose bind in a similar manner to what is observed in the cellotriose-bound crystal structure. If EngD bound cellobiose and cellotriose in the -3 to -1 subsites then there would be no degradation products (Figure 7). Unlike cellopentaose and cellohexaose, cellotetraose was only partially degraded by wild type EngD and EngD_{Y232A}. If cellotetraose bound to the high affinity site (-3 to -1 subsites) then glucose and cellotriose would be the only products formed, but there was a significant amount of cellobiose from the degradation of cellotetraose, demonstrating that there is substantial binding to the -2 to +2 subsites (Figure 7). The binding of cellotetraose to the -2 to +2 subsites further strengthens the hypothesis that Trp162 stabilizes the +2 subsite and would explain why EngD_{W162A} and EngD_{DM} were not able to degrade cellotetraose. With the exception of EngD_{DM}, which shows insignificant activity,

all of the EngD constructs produced cellotriose and cellobiose from cellopentaose. Based on the crystal structure it is likely that cellopentaose binds in the -3 to +2 subsite and not the -2 to +3 subsite, but the definitive position of cellopentaose is impossible to determine with this assay (Figure 7). The most pronounced shift in product distribution was observed in the degradation of cellohexaose. EngD_{W162A} favors the formation of cellotriose, suggesting that cellohexaose binds in the -3 to +3 subsites, while wild type and EngD_{Y232A} produce equal amounts of cellotriose and cellobiose, which suggest an alternative binding mode for cellohexaose (Figure 7). The binding in the -3 – +3 subsite of EngD_{W126A} may be due to the loss of Trp162 and the destabilization of the +2 subsite. The reduction in EngD_{W162A} and EngD_{DM} activity can be explained by a decrease in binding affinity at the +2 subsite for substrates longer than cellotriose. Taken as a whole, these results suggest that EngD is preferentially active on substrates longer than cellotetraose, bind substrates primarily to the -3 to -1 subsites, and the +2 subsite is important in catalysis.

Dynamics of EngD in solution

Many cellulases have a common architecture, a catalytic domain tethered by a PT-rich linker to a CBM domain. We have shown here both the detailed structures of both domains by X-ray crystallography and also shown by SAXS that the linker is indeed flexible and that, free in solution the relative positions of the CBM and the catalytic domain are quite variable (Figure 2D). This property is consistent with the idea that the CBM fixes the enzyme to the surface of the cellulose, allows the catalytic subunit to achieve a high local concentration and probe the nearby surface of the cellulose for productive binding and catalysis.

Physiological function of EngD

The physiological function of EngD is difficult to elucidate from *in vitro* observations alone. The presence of EngD homologues in many cellulolytic organisms argues strongly for a function related to biomass degradation. The enzyme is a true cellulase, possessing both endocellulase and exocellulase activities. Like many other GH5 cellulases EngD can hydrolyze synthetic small molecule substrates, crystalline cellulose, amorphous cellulose, and cellulose derivatives. The extent of cellulose hydrolysis is unexceptional, but within the range seen with *C. thermocellum* GH5 cellulases. Given the presence of a strong cellulose-binding domain, one would expect EngD to perform as well or better than GH5 cellulases not containing a cellulase-binding domain. Yet, EngD performs worse than several *C. thermocellum* cellulases with no cellulose-binding domain.

EngD appears to act synergistically with other normally cellulosomal cellulases, a factor that may be more physiologically relevant to its activity. The results of the experiments combining EngD with *C. thermocellum* cellulases suggests that the activity of EngD may be limited to initiating hydrolysis at a limited number of sites on the crystalline surface of cellulose. The open, but clamp-enabled cleft appears posed to accept a strand directly from the surface of cellulose, an event that may well be the limiting factor for initial strand breakage in crystalline cellulose⁴³. Other *C. thermocellum* cellulases may also partially degrade the cellulose crystal, exposing more sites for EngD hydrolysis to initiate cleavage. In the presence of *C. thermocellum* cellulases EngD may also function to further degrade the cellooligosaccharides generated by the *C. thermocellum* cellulases to cellobiose.

A final option for EngD's activity does not involve action on cellulose. EngD's strong hydrolytic activity on substrates other than cellulose or *beta*-glucan is rare, but not unheard of with GH5 cellulases. A number of GH5 cellulases have been reported to be active on mannan, glucomannan, galactomannan, xyloglucan and xylan⁴⁴. The ability of EngD to degrade xyloglucan, and xylan may have physiological relevance on real biomass, where cellulosomal enzymes are unable to penetrate into the cellular matrix of cellulose, hemicellulose, pectins, and other polysaccharides. Being a soluble, rather than cellulosomal enzyme, EngD may also function to open up the biomass, allowing access to the cell wall for the cellulosomes.

Materials and Methods

Protein Expression and Purification

The gene encoding wild type EngD (NCBI database P28623) was commercially synthesized after its sequence was codon optimized for expression in *Escherichia coli* by GeneArt (Regensburg, Germany). W162A, Y232A, and W162A Y232A variants were generated with a QuikChange II Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA). The EngD genes were amplified by PCR using Easy-A High-Fidelity Cloning enzyme (Stratagene, La Jolla, CA) and subsequently subcloned into the pBAD/Thio TOPO vector (Invitrogen, Carlsbad, CA). After transformation into TOP10 cells, the clone harboring the desired thioredoxin-EngD construct was identified by DNA sequencing. Expression and purification of wild type, W162A, Y232A, and W162A Y232A EngD were completed as described previously with several modifications⁴⁵. *E. coli* harboring the thioredoxin-EngD constructs were grown in LB at 37° C to an OD₆₀₀ of 0.5. Following induction with 0.002% L-arabinose, cells were grown for an additional 20 hours at 25° C before harvesting. Harvested cells were resuspended in 40 mL of lysis buffer (50 mM NaH₂PO₄ buffer, pH 8.0 containing NaCl (300 mM), imidazole (10 mM), lysozyme (0.5 mg•mL⁻¹), E-64 (1.0 μM), benzamidine (0.5 μM), and EDTA (1.0 mM)) and incubated on ice for 30 min. The cells were lysed by sonication, and the C-terminally His-tagged thioredoxin-EngD fusion protein was purified from the supernatant by immobilized nickel affinity chromatography. Fractions containing EngD, as determined by SDS-PAGE, were pooled and dialyzed overnight against 20 mM Tris buffer, pH 8.0 containing NaCl (100 mM). Cleavage of the thioredoxin from the fusion protein was achieved using EKMax recombinant enterokinase (0.5 U/mg of EngD; Invitrogen, Carlsbad, CA). Following cleavage, the thioredoxin and free EngD were separated using a HiTrap Q HP anion exchange column (GE Healthcare, Piscataway, NJ). Pooled fractions exhibiting activity against the substrate 4-methylumbelliferyl β-D-cellobioside were dialyzed against 5 mM HEPES buffer, pH 7.0 containing NaCl (50 mM).

4-Methylumbelliferyl β-D-cellobioside Activity Assay

To confirm the enzymatic activity of the four EngD constructs, purified protein preparations were assayed for their ability to hydrolyze the fluorogenic substrate 4-methylumbelliferyl β-D-cellobioside (MUC Sigma, St. Louis, MO). Specifically, enzyme was diluted in assay buffer (50 mM sodium citrate buffer, pH 6.0) to a final volume of 50 μL/well in a black, round-bottom 96-well microtiter plate. The reaction was initiated upon addition of 50 μL/well assay buffer containing 1 mM MUC and was allowed to progress for 10 min at room

temperature with continuous shaking. The reaction was terminated by addition of 25 μL /well of 1 M sodium carbonate. End-point fluorescence (λ_{ex} : 360 nm, λ_{em} : 455 nm) was recorded using a Spectramax Gemini XS plate reader (Molecular Devices, Sunnyvale, CA). Relative fluorescence units were compared to those of controls.

Activity on Insoluble Crystalline Cellulose

Enzymatic activity on Sigmacell 50 was determined by mixing 1 $\text{mg}\cdot\text{mL}^{-1}$ of enzyme at 37°C with 1% (w/v) of substrate in 100 mM sodium citric buffer (pH 5.0). 250 μL samples were taken at specific time points and the reactions were quenched by boiling the samples for 5 minutes. The quenched samples were filtered through a 0.22 μm Millex-GV syringe filter (Millipore, Ireland) and loaded onto a Shimadzu Liquid Chromatograph HPLC (Shimadzu Scientific Instruments, Columbia, MD) equipped with a refractive index detector. A Phenomenex Rezex RPM-Monosaccharide column heated to 85°C was utilized to separate the soluble oligosaccharides.

Activity on Soluble Substrates

Enzyme specific activity was measured using a micro version of the Modified Somogyi Method for reducing sugars as previously described⁴⁶. The reaction mixtures containing 200 μL of substrate (1% β -glucan or other carbohydrate in 50 mM acetate buffer, pH 5.8) and 5 μL enzyme sample were incubated at 70°C for 10 minutes. Micromoles of sugars formed were determined using a glucose standard curve, and unit activity calculated as micromoles of reducing sugar per minute per milligram of protein.

Comparison of EngD to *C. thermocellum* cellulases

The exo-glucanase specificity of cellulases was determined by spotting 1.0 μg of enzyme directly on agar plates containing 10 mM 4-methylumbelliferyl β -D-cellobioside. Plates were incubated in a 40°C incubator for 60 minutes; after incubation, the plates were examined using a hand-held UV lamp and compared to negative and positive controls.

Digestion of Cellulose

Cellulose digestion experiments were conducted in a final volume of 1.0 ml of 50 mM acetate buffer, pH 5.8, containing 5 mM CaCl_2 and either 2.0 mg Avicel microcrystalline cellulose or 3.4 mg of Whatman 1 filter paper. Thermostable *beta*-glucosidase (*beta*-glucosidase 1, Lucigen, Middleton, WI) was added at a level of 0.1mg/reaction to all reactions to convert cellodextrin products to glucose. *C. thermocellum* cellulases (Lucigen, Middleton, WI) and EngD were added at 0.1mg/reaction. Assays were performed at 40°C, with shaking at 1000 rpm, in a Thermomixer R (Eppendorf, Hamburg, Germany). Glucose formed was determined using the Megazyme D-Glucose (GOPOD Format) Assay Kit according to the manufacturer's directions.

Product Distribution Assays

Initial product distribution was determined for cellobiose (G_2), cellotriose (G_3), cellotetraose (G_4), cellopentaose (G_5), and cellohexaose (G_6) for EngD, EngD_{W162A}, EngD_{Y232A}, and EngD_{DM}. 4 $\mu\text{g}/\text{mL}$ of enzyme was mixed with 2 mM substrate ($\text{Glc}_2 - \text{Glc}_6$) in 100 mM

sodium citrate buffer (pH 5.0). Each 500 μ l reaction was incubated for 30 minutes at 37° C. The samples were boiled to halt the reaction, filtered through an Amicon® Ultra-0.5 3K concentrator (Millipore, Ireland), and loaded onto an HPLC setup describe above.

Cellulose Binding Assay

Each 500 μ l binding reaction contained 200 μ l of 0.5% (w/v) cellulose, 25 μ l 1M citrate buffer (pH 5.0) and 275 μ l of enzymes and water. After a 2 hour incubation at 4° C the samples were filtered through a 0.22 μ m Millex-GV syringe filter. The concentration of unbound enzyme in the supernatant was determined by measuring the UV 280 absorbance. Bound enzyme was calculated by subtracting the unbound enzymes present in supernatant from total amount of added enzyme.

Small Angle X-ray Scattering

SAXS data were collected on the SIBYLS beamline using the high throughput data collection mode as described previously ⁴⁷. Glucose Isomerase at 1mg/ml was utilized as a molecular weight standard. Three concentrations of EngD were measured at 2.3mg/ml, 3.5mg/ml and 7.0mg/ml. Concentration dependence was observed and an extrapolation of the low q data to zero concentration was implemented. Each concentration was exposed 4 times to X-rays with 0.5, 1, 2 and 5 second exposure. No radiation damage was observed through comparing the 0.5 and 1 second exposure. Two buffer blanks were collected; one prior to the concentration series and one following. Both SAXS from buffer blanks were subtracted from the SAXS data collected from samples and the resulting subtracted files were compared for agreement. The 2 and 5 second exposures were merged with the 1 second exposure to produce a final scattering curve for further analysis using the program PRIMUS ⁴⁸.

The crystal structure is missing residues 349 –380 which were filled in by the program MODELLER ⁴⁹. The experimental scattering curve was then compared to that calculated from the crystal structure using the program FOXS ⁵⁰. Given that the missing residues in the crystal structure link domains and the significantly smaller radius of gyration relative to that measured from the experimental profiles, we tested whether the SAXS data was better fit by flexibly linked domains. Utilizing the program BilboMD ⁵¹ we created a large ensemble of over 16,000 conformations where the two domains were treated as rigid but were allowed to translate and rotate relative to one another within the biochemically realistic constraint of the linker. Scattering profiles were calculated from each conformation using FOXS. The best fitting structure was identified along with a minimal ensemble using the program MES ⁵¹. A five member ensemble was required to fit the data to a χ less than 1.

Crystallization and Cellotriose Soak

EngD crystals, approximately 200 μ m $\times$ 40 μ m $\times$ 40 μ m, were observed after one week in the IndexHT screen (Hampton research, Aliso Viejo, CA). Wild type EngD crystals utilized for data collection were grown at 4° C by hanging drop vapor diffusion by mixing 2 μ L of an 8.8 mg•ml⁻¹ protein solution, described above, with 2 μ L of well solution (100 mM Bis-Tris buffer, pH 5.5 containing polyethylene glycol 3350 (27%)). For the cellotriose-bound structure, wild type EngD crystals were soaked for up to 4 h in 100 mM Bis-Tris buffer, pH

5.5 containing polyethylene glycol 3350 (27%) and a saturating amount of cellotriose. Both the EngD and cellotriose-bound EngD crystals were cryoprotected by the addition of 15% ethylene glycol to the final well solutions described above and then frozen directly in liquid N₂.

Diffraction Data Collection and Structure Determination

X-ray diffraction data for EngD and cellotriose-bound EngD crystals were collected at the Life Sciences Collaborative Access Team (LS-CAT) 21-ID-G beamline at the Advance Photon Source, Argonne National Laboratory. Diffraction images from both EngD and cellotriose-bound EngD crystals were indexed, integrated, and scaled using HKL2000⁵². Molecular replacement was carried out with Phenix AutoMR using the catalytic domain of endoglucanase A from *Clostridium cellulolyticum* (PDB ID: 1EDG) as the initial model⁵³. The structure was completed with alternating cycles of manual model building in Coot and refinement in Phenix^{53,54}. The refined EngD structure was used as the molecular replacement model for the cellotriose-bound EngD structure. The cellotriose-bound EngD structure was completed as described for the initial EngD structure. All refinement steps for both structures were monitored using an R_{free} value based on selection of 5.0% of the independent reflections. Model quality was assessed using MolProbity⁵⁵. Molecular figures were generated using PyMOL (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.).

Bioinformatics

InterProScan Family analysis (<http://www.ebi.ac.uk/Tools/InterProScan/>), and BLASTP (Basic Local Alignment Search Tool⁵⁶) (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) analysis tools were used to compare CelA with other proteins in the UNIPROT database. Phylogeny analysis was performed using software at <http://www.phylogeny.fr/version2.cgi/index.cgi>. Multiple alignments were run using ClustalW⁵⁷ alignment with curation to remove positions with gaps⁵⁸. Construction of the phylogenetic tree was obtained using PhyML⁵⁹, and graphically displayed using TreeDyn <http://www.treedyn.org/>.

Acknowledgments

We thank Dr. Brian G. Fox for helpful discussions, and the NIH-funded Center for Eukaryotic Structural Genomics for general access to equipment and computers for the structural work. This work was funded in part by the DOE Great Lakes Bioenergy Research Center (DOE Office of Science BER DE-FC02-07ER64494). Use of the Advanced Photon Source was supported by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor for the support of this research program (Grant 085P1000817). The contribution of Dr. Greg L. Hura and Kevin N. Dyer was supported by DOE program Integrated Diffraction Analysis Technologies (IDAT) (IDAT-DE-AC02-05CH11231), and the National Institute of Health (NIH) grant GM105404.

REFERENCES

1. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res.* 2009; 37:D233–8. [PubMed: 18838391]
2. Maki M, Leung KT, Qin W. The prospects of cellulase-producing bacteria for the bioconversion of lignocellulosic biomass. *Int J Biol Sci.* 2009; 5:500–16. [PubMed: 19680472]

3. Percival Zhang YH, Himmel ME, Mielenz JR. Outlook for cellulase improvement: screening and selection strategies. *Biotechnol Adv.* 2006; 24:452–81. [PubMed: 16690241]
4. Mitchell, JC.; Rutkoski, TJ.; Bannen, RM.; Phillips, GN, Jr.. *Thermostable Proteins: Structural Stability and Design.* CRC Press; Computational and experimental approaches to sequence-based design of protein thermal stability
5. Sleat R, Mah RA, Robinson R. Isolation and Characterization of an Anaerobic, Cellulolytic Bacterium, *Clostridium-Cellulovorans* Sp-Nov. *Applied and Environmental Microbiology.* 1984; 48:88–93. [PubMed: 16346602]
6. Foong FC, Doi RH. Characterization and comparison of *Clostridium cellulovorans* endoglucanases-xylanases EngB and EngD hyperexpressed in *Escherichia coli*. *J Bacteriol.* 1992; 174:1403–9. [PubMed: 1735727]
7. Hamamoto T, Shoseyov O, Foong F, Doi RH. A *Clostridium-Cellulovorans* Gene, Engd, Codes for Both Endo-Beta-1,4-Glucanase and Cellobiosidase Activities. *Fems Microbiology Letters.* 1990; 72:285–288.
8. Wang Q, Tull D, Meinke A, Gilkes NR, Warren RA, Aebersold R, Withers SG. Glu280 is the nucleophile in the active site of *Clostridium thermocellum* CelC, a family A endo-beta-1,4-glucanase. *J Biol Chem.* 1993; 268:14096–102. [PubMed: 8100226]
9. Caines ME, Vaughan MD, Tarling CA, Hancock SM, Warren RA, Withers SG, Strynadka NC. Structural and mechanistic analyses of endo-glycoceramidase II, a membrane-associated family 5 glycosidase in the Apo and GM3 ganglioside-bound forms. *J Biol Chem.* 2007; 282:14300–8. [PubMed: 17329247]
10. Domínguez R, Souchon H, Lascombe M, Alzari PM. The crystal structure of a family 5 endoglucanase mutant in complexed and uncomplexed forms reveals an induced fit activation mechanism. *Journal of Molecular Biology.* 1996; 257:1042–51. [PubMed: 8632467]
11. Ducros V, Czjzek M, Belaich A, Gaudin C, Fierobe HP, Belaich JP, Davies GJ, Haser R. Crystal structure of the catalytic domain of a bacterial cellulase belonging to family 5. *Structure.* 1995; 3:939–49. [PubMed: 8535787]
12. Gloster TM, Ibatullin FM, Macauley K, Eklöf JM, Roberts S, Turkenburg JP, Bjørnvad ME, Jørgensen PL, Danielsen S, Johansen KS, Borchert TV, Wilson KS, Brumer H, Davies GJ. Characterization and three-dimensional structures of two distinct bacterial xyloglucanases from families GH5 and GH12. *J Biol Chem.* 2007; 282:19177–89. [PubMed: 17376777]
13. Lo Leggio L, Larsen S. The 1.62 Å structure of *Thermoascus aurantiacus* endoglucanase: completing the structural picture of subfamilies in glycoside hydrolase family 5. *FEBS Lett.* 2002; 523:103–8. [PubMed: 12123813]
14. Zakariassen H, Aam BB, Horn SJ, Varum KM, Sorlie M, Eijsink VG. Aromatic residues in the catalytic center of chitinase A from *Serratia marcescens* affect processivity, enzyme activity, and biomass converting efficiency. *J Biol Chem.* 2009; 284:10610–7. [PubMed: 19244232]
15. Watanabe T, Ariga Y, Sato U, Toratani T, Hashimoto M, Nikaidou N, Kezuka Y, Nonaka T, Sugiyama J. Aromatic residues within the substrate-binding cleft of *Bacillus circulans* chitinase A1 are essential for hydrolysis of crystalline chitin. *Biochem J.* 2003; 376:237–44. [PubMed: 12930197]
16. Horn SJ, Sikorski P, Cederkvist JB, Vaaje-Kolstad G, Sorlie M, Synstad B, Vriend G, Varum KM, Eijsink VG. Costs and benefits of processivity in enzymatic degradation of recalcitrant polysaccharides. *Proc Natl Acad Sci U S A.* 2006; 103:18089–94. [PubMed: 17116887]
17. Katouno F, Taguchi M, Sakurai K, Uchiyama T, Nikaidou N, Nonaka T, Sugiyama J, Watanabe T. Importance of exposed aromatic residues in chitinase B from *Serratia marcescens* 2170 for crystalline chitin hydrolysis. *J Biochem.* 2004; 136:163–8. [PubMed: 15496586]
18. Zhang S, Irwin DC, Wilson DB. Site-directed mutation of noncatalytic residues of *Thermobifida fusca* exocellulase Cel6B. *Eur J Biochem.* 2000; 267:3101–15. [PubMed: 10824094]
19. Koivula A, Kinnari T, Harjunpaa V, Ruohonen L, Teleman A, Drakenberg T, Rouvinen J, Jones TA, Teeri TT. Tryptophan 272: an essential determinant of crystalline cellulose degradation by *Trichoderma reesei* cellobiohydrolase Cel6A. *FEBS Lett.* 1998; 429:341–6. [PubMed: 9662445]

20. Davies GJ, Dauter M, Brzozowski AM, Bjørnvad ME, Andersen KV, Schülein M. Structure of the *Bacillus agaradherans* family 5 endoglucanase at 1.6 Å and its cellobiose complex at 2.0 Å resolution. *Biochemistry*. 1998; 37:1926–32. [PubMed: 9485319]
21. Hilge M, Gloor SM, Rypniewski W, Sauer O, Heightman TD, Zimmermann W, Winterhalter K, Piontek K. High-resolution native and complex structures of thermostable beta-mannanase from *Thermomonospora fusca* – substrate specificity in glycosyl hydrolase family 5. *Structure*. 1998; 6:1433–44. [PubMed: 9817845]
22. Larson SB, Day J, Barba de la Rosa AP, Keen NT, McPherson A. First crystallographic structure of a xylanase from glycoside hydrolase family 5: implications for catalysis. *Biochemistry*. 2003; 42:8411–22. [PubMed: 12859186]
23. Xu GY, Ong E, Gilkes NR, Kilburn DG, Muhandiram DR, Harris-Brandts M, Carver JP, Kay LE, Harvey TS. Solution structure of a cellulose-binding domain from *Cellulomonas fimi* by nuclear magnetic resonance spectroscopy. *Biochemistry*. 1995; 34:6993–7009. [PubMed: 7766609]
24. McLean BW, Bray MR, Boraston AB, Gilkes NR, Haynes CA, Kilburn DG. Analysis of binding of the family 2a carbohydrate-binding module from *Cellulomonas fimi* xylanase 10A to cellulose: specificity and identification of functionally important amino acid residues. *Protein Eng*. 2000; 13:801–9. [PubMed: 11161112]
25. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol*. 2007; 372:774–97. [PubMed: 17681537]
26. Receveur V, Czjzek M, Schülein M, Panine P, Henrissat B. Dimension, shape, and conformational flexibility of a two domain fungal cellulase in solution probed by small angle X-ray scattering. *J Biol Chem*. 2002; 277:40887–92. [PubMed: 12186865]
27. Violot S, Aghajari N, Czjzek M, Feller G, Sonan GK, Gouet P, Gerday C, Haser R, Receveur-Brechot V. Structure of a full length psychrophilic cellulase from *Pseudoalteromonas haloplanktis* revealed by X-ray diffraction and small angle X-ray scattering. *J Mol Biol*. 2005; 348:1211–24. [PubMed: 15854656]
28. Abuja PM, Pilz I, Claeysens M, Tomme P. Domain-Structure of Cellobiohydrolase-Ii as Studied by Small-Angle X-Ray-Scattering – Close Resemblance to Cellobiohydrolase-I. *Biochemical and Biophysical Research Communications*. 1988; 156:180–185. [PubMed: 3178831]
29. Pilz I, Schwarz E, Kilburn DG, Miller RC, Warren RAJ, Gilkes NR. The Tertiary Structure of a Bacterial Cellulase Determined by Small-Angle X-Ray-Scattering Analysis. *Biochemical Journal*. 1990; 271:277–280. [PubMed: 2121133]
30. Bortoli-German I, Haiech J, Chippaux M, Barras F. Informational suppression to investigate structural functional and evolutionary aspects of the *Erwinia chrysanthemi* cellulase EGZ. *J Mol Biol*. 1995; 246:82–94. [PubMed: 7853408]
31. Henrissat B, Claeysens M, Tomme P, Lemesle L, Mornon JP. Cellulase families revealed by hydrophobic cluster analysis. *Gene*. 1989; 81:83–95. [PubMed: 2806912]
32. Vasella A, Davies GJ, Böhm M. Glycosidase mechanisms. *Current opinion in chemical biology*. 2002; 6:619–29. [PubMed: 12413546]
33. White A, Tull D, Johns K, Withers SG, Rose DR. Crystallographic observation of a covalent catalytic intermediate in a beta-glycosidase. *Nat Struct Biol*. 1996; 3:149–54. [PubMed: 8564541]
34. Sakon J, Adney WS, Himmel ME, Thomas SR, Karplus PA. Crystal structure of thermostable family 5 endocellulase E1 from *Acidothermus cellulolyticus* in complex with cellotetraose. *Biochemistry*. 1996; 35:10648–60. [PubMed: 8718854]
35. Baird SD, Hefford MA, Johnson DA, Sung WL, Yaguchi M, Seligy VL. The Glu residue in the conserved Asn-Glu-Pro sequence of two highly divergent endo-beta-1,4-glucanases is essential for enzymatic activity. *Biochem Biophys Res Commun*. 1990; 169:1035–9. [PubMed: 2363713]
36. Py B, Bortoli-German I, Haiech J, Chippaux M, Barras F. Cellulase EGZ of *Erwinia chrysanthemi*: structural organization and importance of His98 and Glu133 residues for catalysis. *Protein Eng*. 1991; 4:325–33. [PubMed: 1677466]
37. Ramirez-Gualito K, Alonso-Rios R, Quiroz-Garcia B, Rojas-Aguilar A, Diaz D, Jimenez-Barbero J, Cuevas G. Enthalpic nature of the CH/pi interaction involved in the recognition of carbohydrates by aromatic compounds, confirmed by a novel interplay of NMR, calorimetry, and theoretical calculations. *J Am Chem Soc*. 2009; 131:18129–38. [PubMed: 19928848]

38. Brandl M, Weiss MS, Jabs A, Suhnel J, Hilgenfeld R. C-H... π -interactions in proteins. *J Mol Biol.* 2001; 307:357–77. [PubMed: 11243825]
39. Vyas NK, Vyas MN, Quioco FA. Comparison of the periplasmic receptors for L-arabinose, D-glucose/D-galactose, and D-ribose. Structural and Functional Similarity. *J Biol Chem.* 1991; 266:5226–37. [PubMed: 1848243]
40. Bansal P, Hall M, Realff MJ, Lee JH, Bommarius AS. Modeling cellulase kinetics on lignocellulosic substrates. *Biotechnology Advances.* 2009; 27:833–848. [PubMed: 19577626]
41. Ishi A, Sheweita S, Doi RH. Characterization of EngF from *Clostridium cellulovorans* and identification of a novel cellulose binding domain. *Applied and Environmental Microbiology.* 1998; 64:1086–90. [PubMed: 9501449]
42. Baker JO, McCarley JR, Lovett R, Yu CH, Adney WS, Rignall TR, Vinzant TB, Decker SR, Sakon J, Himmel ME. Catalytically enhanced endocellulase Cel5A from *Acidothermus cellulolyticus*. *Appl Biochem Biotech.* 2005; 121:129–148.
43. Chundawat SP, Bellesia G, Uppugundla N, da Costa Sousa L, Gao D, Cheh AM, Agarwal UP, Bianchetti CM, Phillips GN, Langan P, Balan V, Gnanakaran S, Dale BE. Restructuring the Crystalline Cellulose Hydrogen Bond Network Enhances Its Depolymerization Rate. *J Am Chem Soc.* 2011
44. Chen Z, Friedland GD, Pereira JH, Reveco SA, Chan R, Park JI, Thelen MP, Adams PD, Arkin AP, Keasling JD, Blanch HW, Simmons BA, Sale KL, Chivian D, Chhabra SR. Tracing determinants of dual substrate specificity in glycoside hydrolase family 5. *J Biol Chem.* 2012; 287:25335–43. [PubMed: 22645145]
45. Murashima K, Kosugi A, Doi RH. Thermostabilization of cellulosomal endoglucanase EngB from *Clostridium cellulovorans* by in vitro DNA recombination with non-cellulosomal endoglucanase EngD. *Mol Microbiol.* 2002; 45:617–26. [PubMed: 12139610]
46. Brumm P, Mead D, Boyum J, Drinkwater C, Deneke J, Gowda K, Stevenson D, Weimer P. Functional annotation of *Fibrobacter succinogenes* S85 carbohydrate active enzymes. *Applied biochemistry and biotechnology.* 2011; 163:649–57. [PubMed: 20803100]
47. Hura GL, Menon AL, Hammel M, Rambo RP, Poole FL 2nd, Tsutakawa SE, Jenney FE Jr. Classen S, Frankel KA, Hopkins RC, Yang SJ, Scott JW, Dillard BD, Adams MW, Tainer JA. Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nat Methods.* 2009; 6:606–12. [PubMed: 19620974]
48. Konarev PV, Petoukhov MV, Volkov VV, Svergun DI. ATSAS 2.1, a program package for small-angle scattering data analysis. *Journal of Applied Crystallography.* 2006; 39:277–286.
49. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol.* 1993; 234:779–815. [PubMed: 8254673]
50. Schneidman-Duhovny D, Hammel M, Sali A. FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res.* 2010; 38:W540–4. [PubMed: 20507903]
51. Pelikan M, Hura GL, Hammel M. Structure and flexibility within proteins as identified through small angle X-ray scattering. *Gen Physiol Biophys.* 2009; 28:174–89. [PubMed: 19592714]
52. Otwinowski Z, Minor W. Processing of x-ray diffraction data collected in oscillation mode. *Methods in Enzymology.* 1997; 276:307–326.
53. Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr.* 2002; 58:1948–54. [PubMed: 12393927]
54. Emsley P, Cowtan K. Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallographica Section D – Biological Crystallography.* 2004; 60:2126–2132.
55. Lovell SC, Davis IW, Arendall WB, de Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC. Structure validation by C-alpha geometry: phi, psi, and C-beta deviation. *Proteins: Structure, Function, and Genetics.* 2003; 50:437–450.
56. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997; 25:3389–402. [PubMed: 9254694]

57. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994; 22:4673–80. [PubMed: 7984417]
58. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic biology.* 2003; 52:696–704. [PubMed: 14530136]
59. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* 2008; 36:W465–9. [PubMed: 18424797]

Figure 1.

The overall structure of full-length EngD. (A) Cartoon representation of the overall structure of the catalytic domain of EngD is shown going from blue at the N-terminus to red at the C-terminus. The eight α -helices that flank the internal β -barrel of EngD are labeled $\alpha 1 - \alpha 8$. The two single turn helices located on extended loops and the N-terminal helix that encloses one side of the β -barrel are labeled $\alpha'1$, $\alpha'2$, and $\alpha 0$, respectively. A 125° turn from the view seen in the left panel shows the active site binding cleft highlighted in red and the residues that enclose the cleft, Trp162 and Tyr232, are shown explicitly as sticks. (B) View of the

aromatic residues lining the active site binding cleft. **(C)** Location of the glucose binding subsites. Crystallography observed sugars are shown as white and red and a modeled in cellobiose is shown in purple. The glucose binding subsites are labeled. **(D)** Cartoon representation of the overall structure of the carbohydrate-binding module of EngD (CBM_{EngD}). The left view depicts the planar strip of aromatic amino acids (Trp393, Trp430, and Tyr448) proposed to be involved in cellulose binding. The right view is rotated 90° with respect to the left view. Hydrogen bonding side chains that flank Trp393, Trp430, and Tyr448 are shown.

Figure 2.

Analysis of EngD flexibility by SAXS. (A) The pair distribution function, $P(r)$, from the SAXS experiment (black) compared to that calculated from the crystal structure (blue). (B) SAXS scattering curve from EngD (black) with fit (red) by an ensemble of five conformations. (C) The Chi agreement to the experiment plotted as a function of Radius of Gyration (R_g) for each conformation created from BILBOMD (grey circles), the best fitting EngD monomer in the crystallographic lattice (cyan circle, $\text{Chi} = 6.2$) and the five conformations (triangles) from the BILBOMD ensemble that together fit the data to better than a Chi of 1 (red line, plotted in (A)). (D) The best fitting EngD monomer from the crystallographic lattice (black) along with the best fitting five conformation ensemble (catalytic domain (red), flexible linker (green) and CBM_{EngD} (blue)). The percent contribution is noted and used to scale models.

Figure 3.

A schematic representation of the interactions involved in the binding of the two cellotriose molecules. Water molecules are represented as gray spheres. The top half of the figure shows subsites -3, -2, and -1 while the bottom half show subsites +3 and +4. All interactions in the -3, -2, and -1 subsite are from a single EngD monomer. The cellotriose molecule that occupies subsite +3 and +4 is coordinated by two EngD monomers and the residues are labeled with the corresponding chain id.

Figure 4.

(A) Stereo view of the $2F_o - F_c$ map contoured at 1.3σ of the cellotriose bound in the active site cleft. EngD is shown as red sticks and the bound cellotriose molecule is shown as sticks with carbon in white and oxygen in red. Solvent molecules have been removed for clarity.

(B) Stereo view of the $2F_o - F_c$ map contoured at 1.3σ of the cellotriose bound between EngD monomers. One monomer of EngD is shown as red sticks while the adjacent monomer is shown as green sticks. Solvent molecules have been removed for clarity.

Figure 5.

The formation of glucose (A) and cellobiose (B) due to the hydrolysis of Sigmacell 50 by wild type EngD (◆) EngD_{W162A} (■), EngD_{Y232A} (▲), and EngD_{DM} (✱) at 37° C. (C) Equilibrium binding isotherms of wild type EngD, (◆) EngD_{W162A} (■), EngD_{Y232A} (▲), and EngD_{DM} (✱) on Sigmacell 50. The formation of glucose and cellobiose were fit with a log function as described previously⁴⁰.

Figure 6. HPLC traces of the hydrolysis products of wild type EngD (blue), EngD_{W162A} (red), EngD_{Y232A} (green), and EngD_{DM} (purple) from cellotriose (G₃), cellotetraose (G₄), cellopentaose (G₅), and cellohexaose (G₆).

Figure 7.

The possible interactions between the subsites of EngD and cellobiose, cellotriose, cellotetraose, cellopentaose, and cellohexaose are shown. Subsites that were observed to interact with cellotriose in the cellotriose-bound crystal structure are marked (*). Bond cleavage occurs between subsite -1 and +1.

TABLE 1

Summary of crystal parameters, data collection, and refinement statistics. Values in parentheses are for the highest resolution shell.

	EngD	Cellotriose-bound EngD
Crystal parameters		
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Unit-cell parameters (Å)	a = 85.3, b = 119.1, c = 198.5	a = 85.3, b = 119.1, c = 198.5
Data collection statistics		
Wavelength (Å)	0.97857	0.97857
Resolution range (Å)	37.22–2.10 (2.15–2.1)	41.70–2.08 (2.13–2.08)
No. of reflections (measured / unique)	592041 / 118099	470557 / 113804
Completeness (%)	99.7 (100)	93.0 (83.0)
R _{merge} [*]	0.117 (0.456)	0.091 (0.373)
Redundancy	5.0 (5.0)	4.1 (3.1)
Mean I / sigma (I)	13.3 (4.0)	11.6 (2.7)
Refinement and model statistics		
Resolution range (Å)	37.22–2.10	41.70–2.08
No. of reflections (work / test)	112183 / 5916	108107 / 5697
R _{cryst} [§]	0.147 (0.171)	0.144 (0.195)
R _{free} [¶]	0.195 (0.242)	0.193 (0.260)
RMSD bonds (Å)	0.007	0.007
RMSD angles (°)	1.01	1.11
B factor (protein / solvent / ligands) (Å ²)	16.4 / 27.4 / 28.9	16.3 / 27.4 / 31.2
No. of protein atoms	14024	13990
No. of waters	1970	1927
No. of auxiliary molecules	4 Bis-Tris	6 Cellotriose
Ramachandran plot (%)		
Favorable region	96.81	97.0
Additional allowed region	3.19	3.0
Disallowed region	0.0	0.0
PDB ID	3NDY	3NDZ

^{*} $R_{\text{merge}} = \sum_h \sum_i |I_i(h) - \langle I(h) \rangle| / \sum_h \sum_i I_i(h)$, where $I_i(h)$ is the intensity of an individual measurement of the reflection and $\langle I(h) \rangle$ is the mean intensity of the reflection.

[§] $R_{\text{cryst}} = \sum_h ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum_h |F_{\text{obs}}|$, where F_{obs} and F_{calc} are the observed and calculated structure-factor amplitudes, respectively.

[¶] R_{free} was calculated as R_{cryst} using 5.0 % of randomly selected unique reflections that were omitted from the structure refinement.

TABLE 2

Organization of EngD homologues

Gene	Accession	Structure
<i>C. cellulovorans</i> EngD	P28623	GH5 - CBM2
<i>C. longisporum</i> Ce1A	P54937	GH5 - CBM2
<i>B. fibrisolvens</i> end1	P20847	GH5 - CBM2
<i>R. albus</i> Ce1A	P23660	GH5
<i>C. acetobutylicum</i> CA_C0825	Q97KU1	GH5
<i>C. lentocellum</i> Clole_2863	F2JL20	GH5
<i>C. cellulovorans</i> Cloce1_3359	D9SV64	Unknown - GH5 – Dockerin
<i>Ac. cellulolyticus</i> AceceDRAFT_0142	E1K831	GH5 - Dockerin – Lipase/esterase
<i>C. thermocellum</i> Ce1E	P10477	GH5 - Dockerin - Lipase/esterase

TABLE 3

Substrate Specificity of EngD

Soluble Substrate	Specific Activity (u/mg)
1.0% β -glucan	42
0.25% xyloglucan	36
1.0% glucomannan	31
1.0% CMC	15
1.0% birchwood xylan	0.5
1.0% galactomannan	n.d.

TABLE 4

Hydrolysis of Avicel by GH5 Cellulases

Enzyme name	Gene locus	GH Family	CBM Family	MUC Activity	% Avicel Degraded
EngD	Clocel_3242	GH5	CBM2	Positive	1.7
CelC	Cthe_2807	GH5		Positive	0.6
CelE	Cthe_0797	GH5		Positive	5.8
CelG	Cthe_2872	GH5		Positive	2.9
CelL	Cthe_0405	GH5		Positive	4.5
CelH	Cthe_1472	GH26 GH5	CBM11	Positive	0.5
CelO	Cthe_2147	GH5	CBM3	Positive	2.1

TABLE 5Hydrolysis of Whatman filter paper by EngD and *C. thermocellum* Cellulases

<i>C. thermocellum</i> Enzymes	EngD	21 hr	46 hr	164 hr
-	+	0.8%	1.6%	1.8%
CbhA, CelG, CelK,	–	5.0%	10.0%	16.5%
CbhA, CelG, CelK	+	7.7%	15.3%	23.9%
Cell, CelG, CelK	–	7.3%	14.6%	30.5%
Cell, CelG, CelK	+	11.2%	23.5%	39.6%
CbhA, Cell, CelG, CelK	–	6.4%	12.7%	24.3%
CbhA, Cell, CelG, CelK	+	10.9%	21.6%	38.7%