ARTICLE

BIOTECHNOLOGY and BIOENGINEERING

Cel48A From *Thermobifida fusca*: Structure and Site Directed Mutagenesis of Key Residues

Maxim Kostylev,[1] Markus Alahuhta,[2] Mo Chen,[3] Roman Brunecky,[2] Michael E. Himmel,[2] Vladimir V. Lunin,[2] John Brady,[3] David B. Wilson[1]

[1]Department of Molecular Biology and Genetics, Cornell University, 458 Biotechnology Building, Ithaca, New York 14853; telephone: +1-607-255-6476; fax: 607-255-6249; e-mail: mk377@cornell.edu

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

[3]Department of Food Science, Cornell University, Ithaca, New York

ABSTRACT: Lignocellulosic biomass is a potential source of sustainable transportation fuels, but efficient enzymatic saccharification of cellulose is a key challenge in its utilization. Cellulases from the glycoside hydrolase (GH) family 48 constitute an important component of bacterial biomass degrading systems and structures of three enzymes from this family have been previously published. We report a new crystal structure of TfCel48A, a reducing end directed exocellulase from *Thermobifida fusca*, which shows that this enzyme shares important structural features with the other members of the GH48 family. The active site tunnel entrance of the known GH48 exocellulases is enriched in aromatic residues, which are known to interact well with anhydroglucose units of cellulose. We carried out site-directed mutagenesis studies of these aromatic residues (Y97, F195, Y213, and W313) along with two non-aromatic residues (N212 and S311) also located around the tunnel entrance and a W315 residue inside the active site tunnel. Only the aromatic residues located around the tunnel entrance appear to be important for the ability of TfCel48A to access individual cellulose chains on bacterial cellulose (BC), a crystalline substrate. Both Trp residues were found to be important for the processivity of TfCel48A on BC and phosphoric acid swollen cellulose (PASC), but only W313 appears to play a role in the ability of the enzyme to access individual cellulose chains in BC. When acting on BC, reduced processivity was found to correlate with reduced enzyme activity. The reverse, however, is true when PASC is the substrate. Presumably, higher density of available cellulose chain ends and the amorphous nature of PASC explain the increased initial activity of mutants with lower processivity.

Correspondence to: M. Kostylev

Received 29 July 2013; Revision received 5 October 2013; Accepted 21 October 2013
Accepted manuscript online 28 October 2013;
Article first published online 21 November 2013 in Wiley Online Library (http://onlinelibrary.wiley.com/doi/10.1002/bit.25139/abstract).
DOI 10.1002/bit.25139

Biotechnol. Bioeng. 2014;111: 664–673.

KEYWORDS: Cel48A; *Thermobifida fusca*; cellulose; cellulase

Introduction

Lignocellulosic biomass is a potential source of sustainable transportation fuels with a low carbon footprint. Among the key challenges for its utilization is the ability to release sugars stored in cellulose and hemicellulose in an efficient and cost-effective manner. In nature, most biomass is digested by cellulolytic microorganisms using complex mixtures of enzymes, which have evolved complementary properties in order to overcome the recalcitrance of their substrates. Understanding the mechanisms by which such enzymes degrade biomass will aid the development of lignocellulosic biofuel industry and the consequential displacement of petroleum-based transportation fuels.

Cellulose is the main component of lignocellulosic biomass. Individual cellulose chains consist of anhydroglucose units covalently linked via β-1,4 glycosidic bonds. Hydrogen bonding and van der Waals forces between individual chains result in the formation of a composite insoluble substrate that typically contains crystalline, semi-crystalline, and amorphous fractions (Larsson et al., 1997; Park et al., 2010). Cellulose is predominantly hydrolyzed by cellulases, which belong to various glycoside hydrolase (GH) families and historically have been categorized into two main classes, although experimental evidence suggests that this may be an oversimplification. Exocellulases usually contain the active site inside a tunnel or a deep cleft, which in some cases may be enclosed by additional loops. Due to the physical constraints of their active site, exocellulases appear to preferentially initiate hydrolysis either at the reducing or the nonreducing end of the cellulose chain (Barr et al., 1996; Fagerstam and Pettersson, 1980). Endocellulases typically

contain the active site in a relatively shallow cleft, which should enable them to initiate hydrolysis anywhere along a cellulose chain. The processivity of cellulases is defined as the average number of cleavage events a cellulase can carry out before dissociation from the cellulose chain in its active site. Exocellulases are believed to have higher processivity than endocellulases, although direct experimental evidence for this is limited (Kurasin and Valjamae, 2011). The rate of crystalline cellulose hydrolysis is much lower than that of soluble cellooligosaccharides and it is widely believed that access to substrate—the individual cellulose chains—is the rate-limiting step in the enzymatic digestion of cellulose. Recent experimental evidence further supports this hypothesis (Fox et al., 2012; Kostylev et al., 2011; Zakariassen et al., 2010). It is likely, therefore, that at least some cellulases have evolved residues whose primary function is to enhance the ability of the enzyme to separate individual cellulose chains away from the bulk substrate and thread them to the active site.

Reducing-end-directed exocellulases from the GH48 family form an important component of bacterial cellulose-degrading systems. It has been shown that their expression is up-regulated when cellulolytic bacteria are grown on cellulose (Gold and Martin, 2007; Raman et al., 2009; Spiridonov and Wilson, 1998). Furthermore, the knockout of Cel48S from *Clostridium thermocellum* was shown to reduce that strain's rate of Avicel® hydrolysis by 60% (Olson et al., 2010). Even though GH48 exocellulases have relatively low specific activity on crystalline cellulose, they appear to be particularly important in synergistic mixtures (Berger et al., 2007; Irwin et al., 1993, 2000). Mechanistic studies of these enzymes will provide insight into bacterial cellulose degradation and may have important implications for enzymatic hydrolysis of cellulose in the industrial production of lignocellulosic biofuels.

Several structures of three unique GH48 enzymes—Cel48F from *Clostridium cellulolyticum* (Parsiegla et al., 1998), Cel48S from *C. thermocellum* (Guimaraes et al., 2002) and GH48A from *Hahella chejuensis* (Sukharnikov et al., 2012)have been published to date. All three enzymes share the same (alpha/alpha)$_6$ barrel fold and an almost identical active site tunnel that can accommodate seven glucose units at subsites −7 to −1, followed by an open product binding site, best adapted for cellobiose. Here we report the crystal structure of another GH48 exocellulase, Cel48A from *Thermobifida fusca*, a model cellulolytic bacterium. The structure was obtained with X-ray diffraction using the previously published TfCel48A catalytic base mutant D224N (Kostylev and Wilson, 2011) complexed with cellobiose and cellohexaose. In order to better understand the mechanisms by which cellulases are able to hydrolyze crystalline substrates, we generated and tested point mutants of several aromatic and two non-aromatic residues near the entrance of the active site tunnel. Our results suggest that aromatic residues near the binding site of cellulases play an important role in the ability of these enzymes to access individual cellulose chains when acting on microcrystalline cellulose.

Materials and Methods

Substrates

Bacterial cellulose (BC) was a gift from Monsanto. BC cake was washed three times with deionized (DI) water by centrifugation and resuspended in DI water with 0.04% sodium azide (Sigma–Aldrich, St. Louis, MO). Concentration was determined as dry weight per volume. Phosphoric acid swollen cellulose (PASC) was prepared from Avicel® powder (PH-105; FMC Corporation, Philadelphia, PA) using procedures described in (Zhang et al., 2006) and was stored in DI water with 0.04% sodium azide.

Enzymes

The CD of TfCel48A-D224N for structure analysis was expressed and purified as previously described in (Irwin et al., 2000). WT and all other mutant versions of TfCel48A were expressed and purified as previously described in (Kostylev and Wilson, 2013). TfCel48A mutants were generated using Agilent Quickchange® II XL Site Directed Mutagenesis Kit following the manufacturer's instructions. All mutations and sequences were verified by Sanger sequencing. All enzyme concentrations were determined by spectroscopy using NanoDrop® 1000 spectrophotometer.

Crystallization

TfCel48A-D224N crystals were obtained with sitting drop vapor diffusion using a 96-well plate with Crystal Screen HT from Hampton Research (Aliso Viejo, CA). Fifty microliter of well solution was added to the reservoirs and drops were made with 0.2 μL of well solution and 0.2 μL of protein solution using a Phoenix crystallization robot (Art Robbins Instruments, Sunnyvale, CA). The crystals were grown at 20°C with 0.1 M sodium cacodylate trihydrate pH 6.5, 18% (w/v) polyethylene glycol 8000 and 0.2 M zinc acetate dihydrate as the well solution. The protein solution contained 1.2 mg/mL of protein, 20 mM acetate pH 5.0, 100 mM NaCl, 5 mM $CaCl_2$, 5 mM cellohexaose and 1 mM cellobiose.

Data Collection and Processing

The TfCel48A-D224N crystal was flash frozen in a nitrogen gas stream at 100 K before data collection. The crystallization solution with 10% (v/v) ethylene glycol and 10% (v/v) glycerol was used for freezing the crystal. Data was collected in an in-house Bruker X8 MicroStar X-Ray generator with Helios mirrors and Bruker Platinum 135 CCD detector. Data were indexed and processed with the Bruker Suite of programs version 2011.2-0 (Bruker AXS, Madison, WI).

Structure Solution and Refinement

Intensities were converted into structure factors and 5% of the reflections were flagged for Rfree calculations using

programs SCALEPACK2MTZ, ctruncate, MTZDUMP, Unique, CAD, FREERFLAG and MTZUTILS from the CCP4 package of programs (Winn et al., 2011). The program MOLREP version 11.0.05 (Sukharnikov et al., 2012) was used for the molecular replacement. The best solution was achieved using *H. chejuensis* GH48 (PDB code 4FUS) as the search model. The structure was refined and manually build using REFMAC5 (Murshudov et al., 2011) version 5.7.0032 and Coot (Emsley et al., 2010) version 0.6.2. The MOLPROBITY method (Chen et al., 2010) was used to analyze the Ramachandran plot and root mean square deviations (rmsd) of bond lengths and angles were calculated from ideal values of Engh and Huber stereochemical parameters (Engh and Huber, 1991). Wilson B-factor was calculated using ctruncate version 1.5.1. Average B-factors, were calculated using program ICM version 3.7-2a (Molsoft LLC, La Jolla, CA). The data collection and refinement statistics are shown in Table I.

Table I. X-ray data collection and refinement statistics.

Data collection	
Space group	P21
Unit cell, Å, °	$a = 56.458$, $b = 88.494$, $c = 66.753$ $\beta = 113.7$
Wavelength, Å	1.54178
Temperature (K)	100
Resolution, Å	25–1.6 (1.7–1.6)
Unique reflections	78,727 (12,722)
$R_{int}^{\dagger}$	0.089 (0.251)
Average redundancy	7.93 (3.32)
$<I>/<\sigma(I)>$	15.10 (3.81)
Completeness, %	99.4 (96.9)
Refinement	
Resolution, Å	25–1.6 (1.64–1.60)
R/R_{free}	0.121 (0.190)/0.164 (0.244)
Protein atoms	5,415
Water molecules	1,175
Other atoms	160
RMSD from ideal bond length, Å#	0.023
RMSD from ideal bond angles, ° #	2.152
Wilson B-factor	7.25
Average B-factor for protein atoms, Å^2	10.2
Average B-factor for water molecules, Å^2	23.6
Ramachandran plot statistics, %*	
Allowed	99.9
Favored	97.5
Outliers	1 (Ala41)

Statistics for the highest resolution bin are in parenthesis.

Structure Analysis

Programs Coot, PyMOL (http://www.pymol.org) and ICM (http://www.molsoft.com) were used for comparing and analyzing structures. This structure has been deposited to the protein data bank (PDB; www.rcsb.org) with entry code 4JJJ.

Time Course and Endpoint Activity Assays

All reactions were conducted in triplicate in Eppendorf 2 mL Protein LoBind plastic tubes. 1.5 mg BC or PASC was combined with 0.1 μM enzyme in 0.6 mL 50 mM sodium acetate buffer, pH 5.5. Only buffer and substrate were combined for negative controls. Upon mixing reactions were immediately placed in a 50°C (unless otherwise indicated) water bath. Samples in triplicate were removed at the given time points and placed on dry ice to stop the reaction. Frozen samples were later placed in a boiling bath for 10 min in order to denature the enzyme. It was verified experimentally that boiling does not alter soluble sugar profiles detected by high performance liquid chromatography (HPLC). G5 endpoint hydrolysis assays were prepared similarly to BC and PASC assays, mixing 25 μM substrate with 0.01 μM enzyme. Reactions were incubated for 45 min at 50°C. All samples were filtered using Corning® Spin-X® Centrifuge tube filters and the soluble sugar concentrations were measured using a Shimadzu HPLC system fitted with a Bio-Rad (Hercules, CA) Aminex® HPX-87P analytical column and a refractive index detector. The mobile phase was Milli-Q water at a flow rate of 0.6 mL/min. Sample injection (25–50 μL volume) was performed by an autosampler installed on the instrument.

Apparent K_M Assays

0.1 μM TfCel48A was combined with the indicated amount of PASC in 0.6 mL of 50 mM sodium acetate, pH 5.5 and reactions were incubated in a 50°C water bath for 1 h. Samples were boiled for 5 min to denature the enzyme and were processed and analyzed by HPLC in the same manner as described for the time course assays.

Data Analysis

HPLC data were processed with OriginPro 8. Product identities and concentrations were determined by Gaussian peak fitting, using standard solutions with known concentrations of soluble cellooligosaccharides for reference. Soluble sugar concentrations at time zero were subtracted from all of the subsequently obtained concentrations. Soluble sugar produced upon initial mixing of the enzyme and the substrate is primarily due to the burst activity, as described in (Cruys-Bagger et al., 2012; Praestgaard et al., 2011), whereas the model used here is concerned with the digestion of the more recalcitrant portions of cellulose. Apparent K_M and the A and b parameter values of time course profiles were determined using the nonlinear least squares fit of the Michaelis–Menten expression ($V_{max}[S]/(K_M + [S])$) and Equation (1), respectively.

Results and Discussion

The Crystal Structure of *T. fusca* Cel48A

The structure of TfCel48A-D224N was refined to a resolution of 1.6 Å with R and R_{free} of 0.121 and 0.164, respectively. There is one molecule in the asymmetric unit in complex with a cellobiose and a cellohexaose molecule (Fig. 1). It has

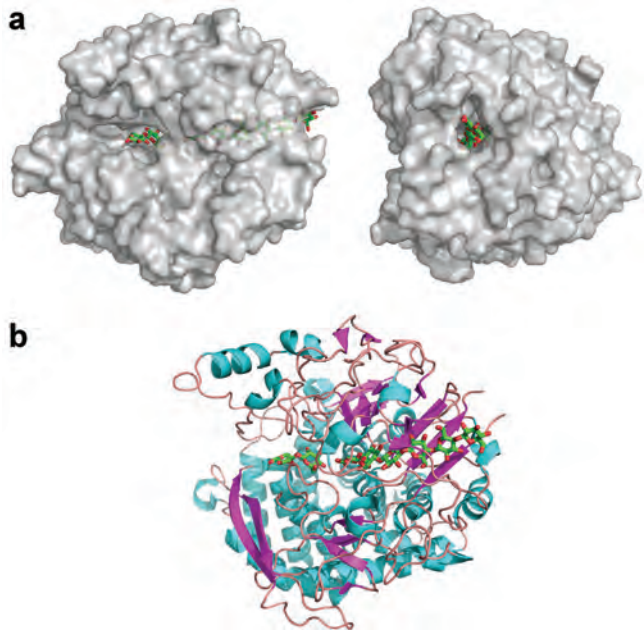

Figure 1. Overall structure of *T. fusca* Cel48A. Surface (**a**) and ribbon (**b**) diagrams are shown to illustrate the active site tunnel (with cellohexaose) and the product binding site (which contains cellobiose). The structures in (a) are rotated 90 degrees with respect to each other.

an (alpha/alpha)$_6$ barrel fold with several calcium, sodium, and zinc atoms and one ethylene glycol and one acetate molecule bound on the surface. The mutated 224 residue is well defined in the electron density.

Structural Comparison

Pair-wise secondary-structure matching of structures with at least 70% secondary structure similarity by PDB fold (Krissinel and Henrick, 2004) found 22 unique structural matches for GH48 from the protein data bank. Out of these only three were unique. Most were CcCel48F and CtCel48S structures with one GH48 from *H. chejuensis*. TfCel48A has 54% sequence identity with CtCel48S (Guimaraes et al., 2002) and 56% sequence identity with CcCel48F (Parsiegla et al., 1998) and *H. chejuensis* GH48. The Cα root mean square deviations of all similar structures varied between 0.85 and 1.01 Å showing that the overall backbone of all of the known GH48 structures is similar.

Closer inspection of the active site tunnels of the other three unique GH48 structures (PDB entry codes 1FCE, 1L2A and 4FUS) with TfCel48A shows it to be highly conserved. The catalytic glutamate and the cellobiose and cellohexaose molecules superimpose almost perfectly with those of the other structures. When residues located within 5 Å from cellobiose and cellohexaose are compared, only two differences stand out. Ile106 is alanine and a tryptophan at position 406 is tyrosine in the other GH48 structures.

The asparagine residue of the D224N mutation adopts a slightly different conformation compared to the aspartate residues of the other GH48 structures. The mutation makes it impossible for this residue to function as the catalytic base and causes it to lose its hydrogen bond with the main chain oxygen of A222. N224 instead forms a hydrogen bond with R424, consequently changing its orientation.

Cel48A Activity and Processivity on Crystalline and Amorphous Cellulose

We recently developed a kinetic model of enzymatic cellulose hydrolysis in order to aid mechanistic studies of cellulases (Kostylev and Wilson, 2013). The model relies on two parameters to fit experimentally obtained time course profiles of individual cellulases and their mixtures acting on insoluble cellulose. The equation used in the model is based on a pseudo zero order Michaelis–Menten scheme (Michaelis and Menten, 1913), but contains a time (digestion) dependent activity coefficient rather than a constant to account for the continuous decrease in activity typically observed for cellulose hydrolysis. It can be expressed either as product formation over time, or as percent digestion over time. The latter is used in this manuscript:

$$X = At^b \quad 0 \leq b \leq 1 \tag{1}$$

where X is % digestion of the substrate, A is the net activity of the added enzyme, and b is an intrinsic constant, which quantifies the curvature of the time course profile and reflects the ability of a given enzyme to degrade a given substrate. Parameter values are obtained by fitting experimental time course data to Equation (1) and the model is applicable under conditions of low to medium-low enzyme loads (i.e., when the substrate concentration is well above the enzyme concentration). The A parameter is a product of specific activity and the concentration of the productively bound enzyme. It is strongly dependent on the total added enzyme concentration and is exponentially dependent on temperature, following the Arrhenius equation. As an intrinsic constant, b parameter of individual cellulases is not dependent on enzyme concentration and shows only a slight dependence on temperature. The theoretical limits of b are 0, in which case no products would be generated over time and 1, in which case the rate coefficient is constant over time, as is the case in classical kinetics. Consequently, cellulases with higher b values on a given substrate are more effective at overcoming the recalcitrance of that substrate.

By itself, TfCel48A is the least effective of *T. fusca* cellulases in the digestion of insoluble cellulose (Irwin et al., 2000). This is reflected by its low A and b parameter values obtained from the time course profiles on BC and PASC, shown in Figure 2a. On BC, which has high crystallinity (Park et al., 2010), its A and b values are 0.45 and 0.34 respectively. PASC is a mostly amorphous form of cellulose, obtained by treatment of crystalline cellulose with phosphoric acid (Walseth, 1952; Zhang et al., 2006). It is more readily degraded than

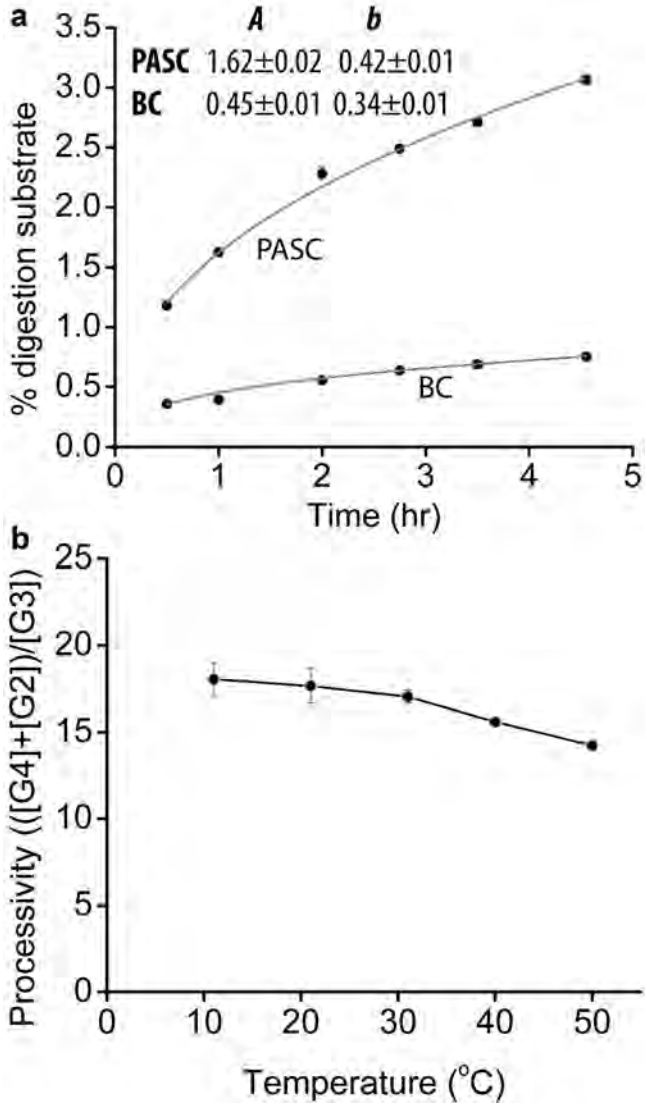

Figure 2. Activity and processivity of wild type TfCel48A. (**a**) Time course profiles of 0.1 μM TfCel48A on 1.5 mg BC and PASC with corresponding *A* and *b* parameter values from Equation (1). (**b**) Relative processivity (Equation (2)) of TfCel48A on BC as a function of temperature. Processivity values were calculated after an 18 h incubation period of 0.1 μM enzyme with 1.5 mg BC. All points were obtained in triplicate and error bars represent one standard deviation.

crystalline cellulose and this is reflected by the higher TfCel48A *A* and *b* values of 1.51 and 0.41, respectively. The higher *b* value on PASC indicates that this substrate is more accessible for TfCel48A, which results in a slower decline of the digestion rate over time when compared to BC hydrolysis. The higher *A* value observed on PASC may be due to an increase in either or both the productively bound fraction of the total added enzyme and the specific activity of TfCel48A acting on less tightly packed cellulose chains.

Processivity is an important consideration in understanding the mechanisms of cellulose hydrolysis. If access to substrate is indeed the rate-limiting step on crystalline cellulose, it should be advantageous for a cellulase to carry out multiple cleavage steps on the same chain that has been threaded in the active site tunnel. This is not necessarily the case when acting on soluble and amorphous substrates (Horn et al., 2006). Most commonly, relative rather than absolute processivity values have often been reported in literature. For exocellulases, processivity is typically determined by measuring the relative concentrations of glucose (G1), cellobiose (G2), and cellotriose (G3) (Medve et al., 1998; von Ossowski et al., 2003; Vuong and Wilson, 2009). Depending on the mode of initial binding of the cellulose chain in the active site, the first cleavage product may be G1, G2, or G3, but only G2, the repeating unit of a cellulose chain, is produced for all subsequent processive cleavages. In the case of TfCel48A acting on BC or PASC, the initial cleavage appears to be only either G3 or G2, as no G1 is observed. In addition, at low enzyme concentration, a trace amount of cellotetraose (G4) is also produced (Fig. S1), which is most likely the result of a skipping event during procession of the enzyme. G4 is relatively slowly cleaved by the enzyme to produce two G2 units. As a result, relative processivity (*P*) of TfCel48A is given here as:

$$P = \frac{[G_4] + [G_2]}{[G_3]} \qquad (2)$$

where brackets indicate product concentration. Under the reaction conditions employed here, the calculated processivity values reach equilibrium within a 90 min incubation period and remain unchanged through at least 18 h incubation.

The effect of temperature on the activity of TfCel48A was previously studied by our group, with optimal temperature determined to be around 50°C (Irwin et al., 2000). The effect on processivity, however, was not determined at that time. As Figure 2b illustrates, when the temperature of BC digestion by TfCel48A is increased from 10 to 50°C, the ratio of the sum of [G4] and [G3] to [G2] is decreased from ca. 18 to 14, indicating a decrease in processivity. This result is most likely due to an increase with temperature in the value of k_{off}, which governs the binding affinity of TfCel48A to the cellulose chain inside the active site tunnel.

The Role of Aromatic Residues in Binding Chain Ends

Aromatic residues play an important role in the ability of carbohydrate-active enzymes to bind via stacking interactions with pyranose rings (Spurlino et al., 1991) and their importance for cellulose binding by cellulases has been demonstrated in literature (Horn et al., 2006; Payne et al., 2011; Poole et al., 1993; Vuong and Wilson, 2009). The tunnel entrance of TfCel48A and other GH48 cellulases is enriched in aromatic residues and it is possible that these residues are important for the ability of these enzymes to separate and thread individual cellulose chains from the bulk substrates. In order to test this hypothesis, we prepared

alanine point mutations of the four aromatic residues in TfCel48A, as shown in Figure 3a: Y97, F195, Y213, and W313. Another aromatic residue located inside the tunnel, W315, was mutated in order to compare its role with that of the residues immediately outside the tunnel. As shown in Figure 3b, both Trp residues are in line with each other and form stacking interactions with the cellulose chain located in the active site tunnel. Two non-aromatic residues, N212 and S311, which are located next to Y211 and W313, respectively, were also mutated. Finally, double mutations—Y213A-S311A and Y213A-W313A—were generated to see whether the effect of the tested mutations on the properties of TfCel48A is additive. All mutants were analyzed by native polyacrylamide gel electrophoresis to verify that the mutations did not compromise protein folding or stability (Fig. S2).

Soluble cellooligosaccharides, such as cellopentaose (G5) can serve as a proxy for loose chain ends on bulk cellulose. Endpoint activity of TfCel48A mutants on G5 is shown in Figure 4. Moderately reduced activity is observed for Y97A, F195A, and Y213A single mutants; as well as for both double mutants, which contain the Y213A mutation. This result suggests that these residues may be important for the initial binding of TfCel48A to soluble chain ends.

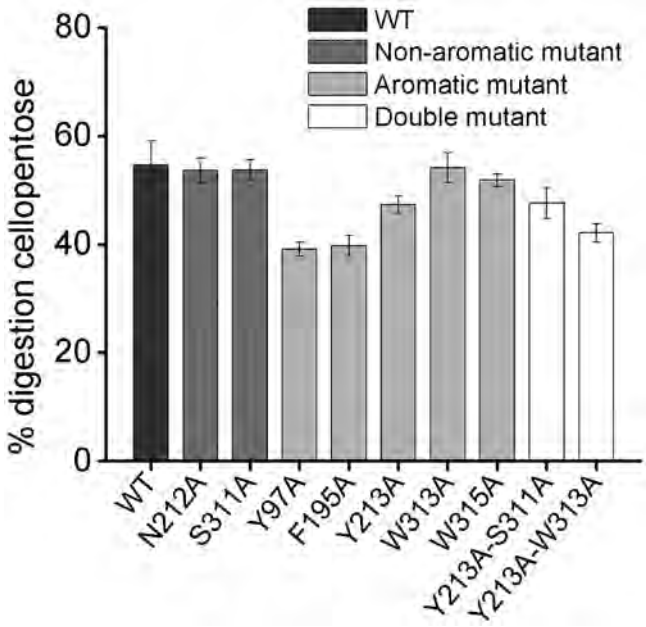

Figure 4. Endpoint activity of TfCel48A wild type and mutants on cellopentaose (G5), measured as % digestion. All points were obtained in triplicate and error bars represent one standard deviation.

The Role of Aromatic Residues in the Hydrolysis of BC

Time course profiles of BC hydrolysis were fit with Equation (1) and the obtained parameter values are plotted in Figure 5. A relatively small, but consistent, reduction of the *b* value (0.26–0.29) is apparent for all point mutants of the aromatic residues located around the active site tunnel entrance. WT value (0.34) is retained for both non-aromatic residue mutations N212A (0.36) and S311A (0.34). The effect of the aromatic residue mutations is more pronounced in the double mutant, Y213A-W313A, as evidenced by its significantly lower *b* value (0.14). On the other hand, the value of the double mutant, Y213A-S311A (0.30) is similar to that observed for the point mutant, Y213A (0.29). Of particular importance is the observation that the mutation of W315, which is aromatic, but is located inside the tunnel, does not seem to affect the *b* value (0.33) of the protein,

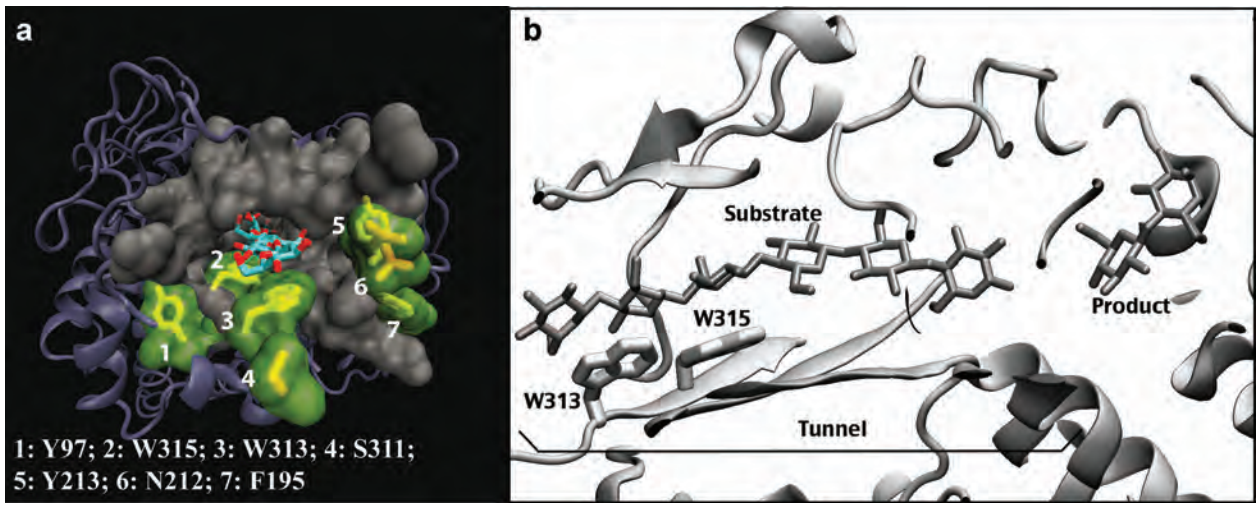

Figure 3. TfCel48A active site tunnel with substrate. Highlighted residues were mutated individually or in pairs to alanines. (**a**) Residues around the tunnel entrance are fully or partially exposed on the surface of the enzyme. (**b**) W313 and W315 are located at the mouth and inside of the tunnel respectively and both interact with the cellulose chain bound in the tunnel.

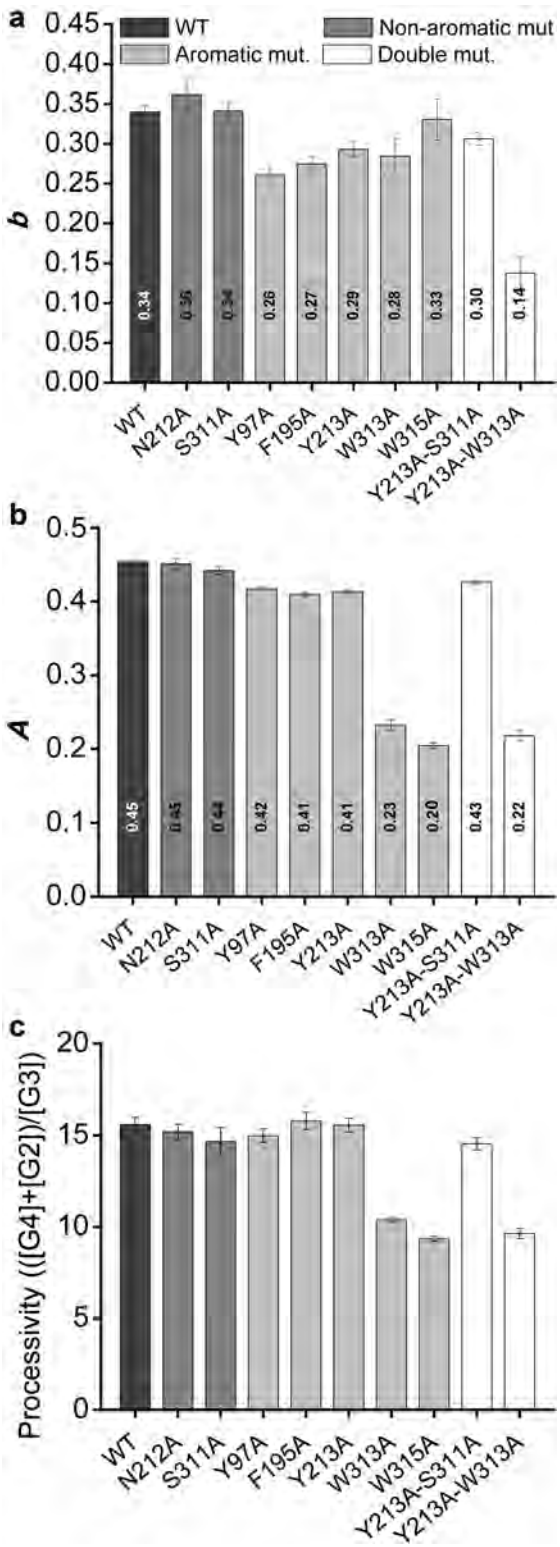

Figure 5. TfCel48A wild type and mutants on BC. *b*, the hydrolysis power factor (**a**) and *A*, the net activity (**b**) were obtained by fitting the time course of each protein variant with Equation (1). All data points were obtained in triplicate and error bars represent standard error of the fit. (**c**) Relative processivity values (calculated with Equation (2)) are the average for time points 1.5–4 h.

as we previously showed for slightly different enzyme and substrate concentrations (Kostylev and Wilson, 2013). Lower *b* values reflect a flattened time course profile, which is indicative of a faster digestion rate decline by a given mutant. Hence, these results suggest that of the tested residues, the aromatic amino acids located around the active site tunnel entrance are particularly important for the ability of TfCel48A to access the individual cellulose chains packed in BC and their removal compromises the ability of the enzyme to find productive binding sites. This finding is consistent with the demonstrated ability of aromatic residues to interact with anhydroglucose units of cellulose (Spurlino et al., 1991) and with the location of these residues near the active site tunnel (Fig. 3a). Similar conclusions have been reached for aromatic residue function in a number of cellulase and chitinase studies (Igarashi et al., 2009; Koivula et al., 1998; Nakamura et al., 2013; Uchiyama et al., 2001; Zakariassen et al., 2009, 2010).

The *A* parameter in Equation (1) is the net activity of the added enzyme, and changes in its value may be indicative of the change in the specific activity of the enzyme or the productively bound enzyme fraction or both. Compared to WT (0.45), the *A* value of the mutants, (Fig. 5b) is slightly reduced for Y97A (0.42), F195A (0.41), and Y213A (0.41), all of which are located near the tunnel entrance, but according to structural data do not interact with the cellulose chain inside the tunnel (Fig. 3). The slight reduction in the *A* value of these mutants may be indicative of a smaller than WT fraction of productively bound enzyme, which would be consistent with their reduced *b* parameter values. The *A* value of W313A and W315A is significantly lower than that of WT, and correlates well with the lower processivity of these mutants, as quantified with Equation (2) (Fig. 5c). Lower processivity for these mutants is expected from the structural data, which indicates that both Trp residues form stacking interactions with the cellulose chain inside the active site tunnel (Fig. 3b). The lower *A* values are likely the result of a smaller than WT fraction of productively bound enzyme. On a crystalline substrate, such as BC, access to individual cellulose chains is presumably the rate limiting step of digestion and the premature release of the chain from the active site tunnel would thus be disadvantageous for the overall rate of hydrolysis. The fact that none of the other mutations result in lower processivity of TfCel48A suggests that the other tested residues do not interact with the cellulose chain once it is inside the active site tunnel. These results are consistent with other cellulase and chitinase studies showing that the removal of aromatic residues that interact with the substrate inside the active site tunnel or cleft result in reduced processivity and activity of the enzymes on crystalline substrates (Horn et al., 2006; Nakamura et al., 2013; Uchiyama et al., 2001; Zakariassen et al., 2009, 2010; Zhang et al., 2000).

Effect of the Mutated Residues on the Digestion of PASC

PASC is a mostly amorphous form of cellulose with different properties compared to highly crystalline substrates such as

BC. It is more digestible by cellulases due to its smaller particle size, lack of supramolecular structure, and lower degree of polymerization (Zhang et al., 2006). WT TfCel48A is about five times more active on PASC than on BC (Fig. 3). In order to further understand the function of the mutated residues, similar time course experiments were carried out on PASC to determine the effect of the mutations on the parameter values of Equation (1), shown in Figure 6. Of the tested mutants, only W313A and W315A have a significant effect on the value of the *b* parameter (0.34 and 0.33, respectively) in comparison to WT (0.41). This finding is interesting because W315A did not affect the *b* value for the enzyme acting on BC and it suggests that the basis of rate decline is different on BC and PASC. As a regenerated form of cellulose with no supramolecular structure (Zhang et al., 2006), it is likely that PASC contains many obstacles due to crisscrossing of individual cellulose chains and a lack of defined fibers. The reduced *b* value of both Trp mutants correlates with reduced processivity. Published modeling and experimental data show that the removal of aromatic residues that interact with the cellulose chain in the tunnel of exocellulases can reduce the binding affinity of the enzyme (Horn et al., 2006; Payne et al., 2011; Poole et al., 1993). This appears to be the case when considering the W313A and W315A mutations, as further supported by their higher apparent K_M values on PASC (1.7 and 2.4 mg/mL PASC, respectively, compared to 0.7 for WT; Fig. S3). Inherent processivity of exocellulases is governed by the k_{off} constant, which in turn is related to their binding affinity for the cellulose chain undergoing hydrolysis (Zakariassen et al., 2010). Thus, affinity for the substrate inside the active site tunnel may play a dominant role in the ability of an exocellulase to effectively digest PASC and for this reason the *b* value of TfCel48A correlates with processivity. In turn, this suggests that unlike BC, digestibility of PASC by an exocellulase is related to the ability of the enzyme to overcome surface obstacles (formed when the substrate is rapidly regenerated to the insoluble form by the removal of phosphoric acid (Zhang et al., 2006)) before dissociating from the chain. Given the non crystalline, gel-like nature of PASC (Zhang et al., 2006), the compromised ability to retain the chain inside the active site tunnel would be expected to lead to a more rapid digestion rate decline, as reflected by the lower *b* value of the W313A and W315A mutants.

The *A* value of the aromatic residue mutants (1.58–2.19) is slightly to moderately higher than that of WT (1.51). This result is somewhat surprising for the Y97A, F195A, Y213A, and Y213A-S311A mutants, which display WT processivity and slightly decreased *b* values. One possible explanation for this result is that the removal of these residues leads to a small increase in the specific activity of TfCel48A (i.e., due to a weaker interaction with the cellulose chain end upon initial binding), without compromising the ability of the enzyme to bind productively to the substrate. In the case of both the W313A and W315A mutants, the increased *A* value (2.06 and 2.19) correlates with, and can be explained by their lower processivity. Since lower processivity indicates weaker

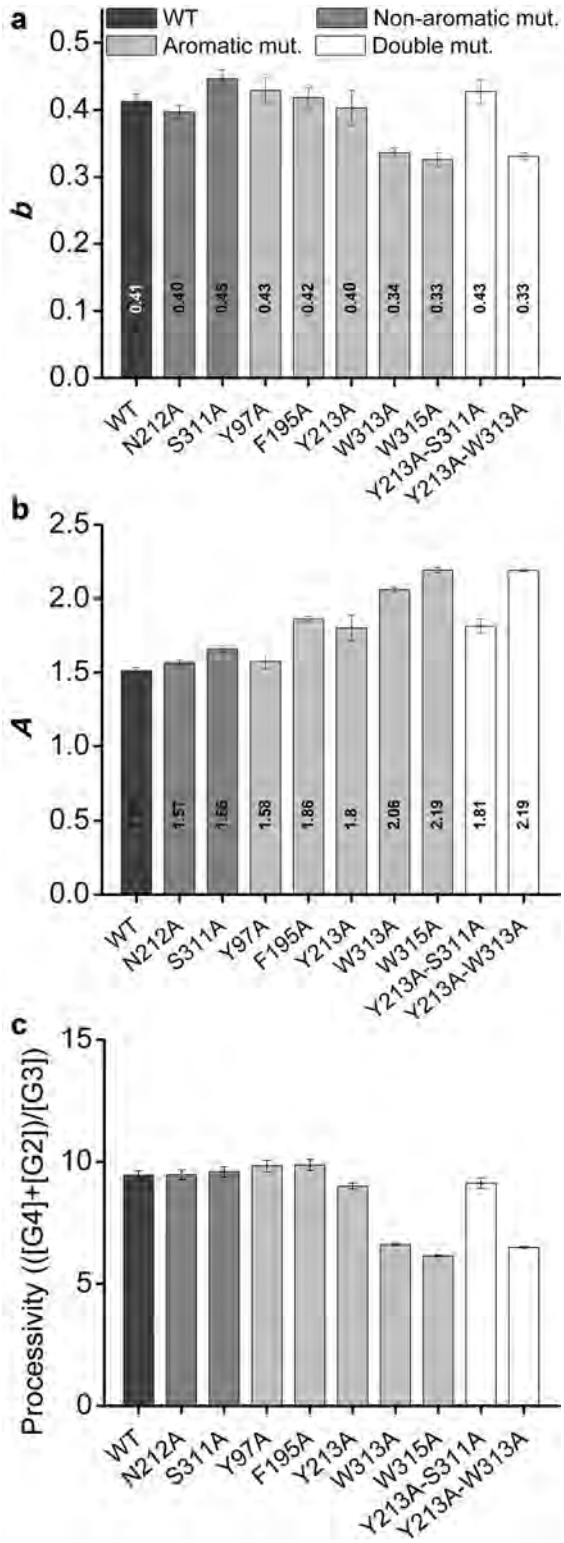

Figure 6. TfCel48A wild type and mutants on PASC. *b*, the hydrolysis power factor (**a**) and *A*, the net activity (**b**) were obtained by fitting the time course of each protein variant on PASC with Equation (1). All data points were obtained in triplicate and error bars represent standard error of the fit. (**c**) Relative processivity values (calculated with Equation (2)) are the average for time points 1.5–4 h.

binding of the enzyme to the cellulose chain undergoing hydrolysis, it is possible that these mutants are able to process faster along that chain until they dissociate. Given the lower degree of polymerization and higher accessibility of PASC (Zhang et al., 2006), higher frequency of dissociation from the chains is not necessarily detrimental to the overall rate of digestion, as a dissociated enzyme should be able to rapidly find another productive binding site to continue hydrolysis. This observation is consistent with other published data on the effect of aromatic residue mutations in processive cellulases and chitinases acting on amorphous substrates (Horn et al., 2006; Nakamura et al., 2013; Zakariassen et al., 2009; Zhang et al., 2000). The apparent discrepancy between the effect of the removal of either Trp residue on *A* and *b* parameter values suggests that lower processivity may be beneficial only for the initial digestion of amorphous substrates such as PASC (as reflected by the increase in the value of *A*). The flattened time course profile of the Trp mutants indicates that despite having higher initial rates, their ability to continue to hydrolyze PASC declines faster than WT. This may also explain why in some studies lower processivity correlates with lower measured activity on amorphous substrates under certain conditions (Nakamura et al., 2013; Vuong and Wilson, 2009; Zhang et al., 2000).

Conclusions

We have determined the X-ray crystal structure of the *T. fusca* TfCel48A D224N mutant in complex with cellobiose and cellohexaose, and structural comparisons show high similarity with the active site tunnels of other known GH48 enzymes. Four aromatic residues around the TfCel48A active site tunnel entrance and one inside the tunnel were mutated and the mutated enzymes were tested for activity on cellopentaose, BC, and PASC. According to the data obtained for cellopentaose, Y97, F195, and Y213, all of which are located around the tunnel entrance but are not bound to the cellulose chain undergoing hydrolysis, may play a role in helping thread cellulose chain ends into the active site tunnel. Both Trp residues—W313 and W315—which are located at the tunnel entrance and inside the active site tunnel, respectively, and interact with the cellulose chain undergoing hydrolysis, are important for processivity of TfCel48A on BC, PASC, and probably other substrates. All aromatic residues located around the tunnel entrance seem to be important for the ability of TfCel48A to digest BC, presumably due to the affinity of aromatic residues for the anhydroglucose units of cellulose. On the other hand, the ability of TfCel48A to digest PASC extensively appears to be more dependent on its processivity, which is governed by the strength of its interaction with the cellulose chain inside its active site tunnel. Both Trp mutants with reduced processivity demonstrate flattened time course profiles on PASC in comparison to WT despite having higher initial activity on this substrate (higher *A* parameter). The lack of conservation of most of the residues around the active site tunnel, including three of the aromatic residues that are required for crystalline cellulose activity, is in stark contrast to the conservation of the active site residues in the tunnel. One possible explanation for this observation is that the tunnel entrance residues have been selected for the types of plant cell walls utilized by *T. fusca*, which suggests that each organism has evolved a set of such residues that is optimal for its commonly encountered substrates. If this is true, then it should be possible to produce engineered variants that display higher than WT activity on specific, pretreated biomass substrates.

References

Barr BK, Hsieh YL, Ganem B, Wilson DB. 1996. Identification of two functionally different classes of exocellulases. Biochemistry 35(2):586–592.

Berger E, Zhang D, Zverlov VV, Schwarz WH. 2007. Two noncellulosomal cellulases of *Clostridium thermocellum*, Cel9I and Cel48Y, hydrolyse crystalline cellulose synergistically. FEMS Microbiol Lett 268(2):194–201.

Chen VB, Arendall WB III, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. 2010. MolProbity: All-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 66(Pt 1):12–21.

Cruys-Bagger N, Elmerdahl J, Praestgaard E, Tatsumi H, Spodsberg N, Borch K, Westh P. 2012. Pre-steady-state kinetics for hydrolysis of insoluble cellulose by cellobiohydrolase Cel7A. J Biol Chem 287(22):18451–18458.

Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66(Pt 4):486–501.

Engh RA, Huber R. 1991. Accurate bond and angle parameters for X-ray protein-structure refinement. Acta Crystallogr A 47:392–400.

Fagerstam L, Pettersson L. 1980. The 1.4-β-glucan cellobiohydrolases of *Trichoderma reesei* QM 9414A new type of cellulolytic synergism. FEBS Lett 119(1):97–100.

Fox JM, Levine SE, Clark DS, Blanch HW. 2012. Initial- and processive-cut products reveal cellobiohydrolase rate limitations and the role of companion enzymes. Biochemistry 51(1):442–452.

Gold ND, Martin VJ. 2007. Global view of the *Clostridium thermocellum* cellulosome revealed by quantitative proteomic analysis. J Bacteriol 189 (19):6787–6795.

Guimaraes BG, Souchon H, Lytle BL, David Wu JH, 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(3):587–596.

Horn SJ, Sikorski P, Cederkvist JB, Vaaje-Kolstad G, Sorlie M, Synstad B, Vriend G, Varum KM, Eijsink VG. 2006. Costs and benefits of processivity in enzymatic degradation of recalcitrant polysaccharides. Proc Natl Acad Sci USA 103(48):18089–18094.

Igarashi K, Koivula A, Wada M, Kimura S, Penttila M, Samejima M. 2009. High speed atomic force microscopy visualizes processive movement of *Trichoderma reesei* Cellobiohydrolase I on crystalline cellulose. J Biol Chem 284(52):36186–36190.

Irwin DC, Spezio M, Walker LP, Wilson DB. 1993. Activity studies of eight purified cellulases: Specificity, synergism, and binding domain effects. Biotechnol Bioeng 42(8):1002–1013.

Irwin DC, Zhang S, Wilson DB. 2000. Cloning, expression and characterization of a family 48 exocellulase, Cel48A, from *Thermobifida fusca*. Eur J Biochem/FEBS 267(16):4988–4997.

Koivula A, Kinnari T, Harjunpaa V, Ruohonen L, Teleman A, Drakenberg T, Rouvinen J, Jones TA, Teeri TT. 1998. Tryptophan 272: An essential determinant of crystalline cellulose degradation by *Trichoderma reesei* cellobiohydrolase Cel6A. FEBS Lett 429(3):341–346.

Kostylev M, Moran-Mirabal JM, Walker LP, Wilson DB. 2012. Determination of the molecular states of the processive endocellulase *Thermobifida fusca* Cel9A during crystalline cellulose depolymerization. Biotechnol Bioeng 109:295–299.

Kostylev M, Wilson D. 2013. Two-parameter kinetic model based on a time-dependent activity coefficient accurately describes enzymatic cellulose digestion. Biochemistry 52(33):5656–5664.

Kostylev M, Wilson DB. 2011. Determination of the catalytic base in family 48 glycosyl hydrolases. Appl Environ Microbiol 77(17):6274–6276.
Krissinel E, Henrick K. 2004. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr D Biol Crystallogr 60:2256–2268.
Kurasin M, Valjamae P. 2011. Processivity of cellobiohydrolases is limited by the substrate. J Biol Chem 286(1):169–177.
Larsson P, Wickholm K, Iversen T. 1997. A CP/MAS13C NMR investigation of molecular ordering in celluloses. Carbohydr Res 302(1–2):19–25.
Medve J, Karlsson J, Lee D, Tjerneld F. 1998. Hydrolysis of microcrystalline cellulose by cellobiohydrolase I and endoglucanase II from *Trichoderma reesei*: Adsorption, sugar production pattern, and synergism of the enzymes. Biotechnol Bioeng 59(5):621–634.
Michaelis L, Menten ML. 1913. Die kinetik der invertinwirkung. Biochem Z 49:333–369.
Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA. 2011. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr 67(Pt 4):355–367.
Nakamura A, Tsukada T, Auer S, Furuta T, Wada M, Koivula A, Igarashi K, Samejima M. 2013. The tryptophan residue at the active site tunnel entrance of *Trichoderma reesei* cellobiohydrolase Cel7A is important for initiation of degradation of crystalline cellulose. J Biol Chem 288 (19):13503–13510.
Olson DG, Tripathi SA, Giannone RJ, Lo J, Caiazza NC, Hogsett DA, Hettich RL, Guss AM, Dubrovsky G, Lynd LR. 2010. Deletion of the Cel48S cellulase from *Clostridium thermocellum*. Proc Natl Acad Sci USA 107 (41):17727–17732.
Park S, Baker JO, Himmel ME, Parilla PA, Johnson DK. 2010. Cellulose crystallinity index: Measurement techniques and their impact on interpreting cellulase performance. Biotechnol Biofuels 3:10–20.
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 A resolution. EMBO J 17(19):5551–5562.
Payne CM, Bomble YJ, 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(47):41028–41035.
Poole DM, Hazlewood GP, Huskisson NS, Virden R, Gilbert HJ. 1993. The role of conserved tryptophan residues in the interaction of a bacterial cellulose binding domain with its ligand. FEMS Microbiol Lett 106 (1):77–83.
Praestgaard E, Elmerdahl J, Murphy L, Nymand S, McFarland KC, Borch K, Westh P. 2011. A kinetic model for the burst phase of processive cellulases. FEBS J 278(9):1547–1560.
Raman B, Pan C, Hurst GB, Rodriguez M Jr, McKeown CK, Lankford PK, Samatova NF, Mielenz JR. 2009. Impact of pretreated Switchgrass and biomass carbohydrates on *Clostridium thermocellum* ATCC 27405 cellulosome composition: A quantitative proteomic analysis. PloS ONE 4(4):e5271.
Spiridonov NA, Wilson DB. 1998. Regulation of biosynthesis of individual cellulases in Thermomonospora fusca. J Bacteriol 180(14):3529–3532.
Spurlino JC, Lu GY, Quiocho FA. 1991. The 2.3-A resolution structure of the maltose- or maltodextrin-binding protein, a primary receptor of bacterial active transport and chemotaxis. J Biol Chem 266(8):5202–5219.
Sukharnikov LO, Alahuhta M, Brunecky R, Upadhyay A, Himmel ME, Lunin VV, Zhulin IB. 2012. Sequence, structure, and evolution of cellulases in glycoside hydrolase family 48. J Biol Chem 287(49):41068–41077.
Uchiyama T, Katouno F, Nikaidou N, Nonaka T, Sugiyama J, Watanabe T. 2001. Roles of the exposed aromatic residues in crystalline chitin hydrolysis by chitinase A from Serratia marcescens 2170. J Biol Chem 276(44):41343–41349.
von Ossowski I, Stahlberg J, Koivula A, Piens K, Becker D, Boer H, Harle R, Harris M, Divne C, Mahdi S, Zhao Y, Driguez, H, Claeyssens M, Sinnot ML, Teeri TT. 2003. Engineering the exo-loop of *Trichoderma reesei* cellobiohydrolase, Ce17A. A comparison with Phanerochaete chrysosporium Cel7D. J Mol Biol 333(4):817–829.
Vuong TV, Wilson DB. 2009. Processivity, synergism, and substrate specificity of *Thermobifida fusca* Cel6B. Appl Environ Microbiol 75 (21):6655–6661.
Walseth CS. 1952. The influence of the fine structure of cellulose on the action of cellulases. Tappi 35(5):233–238.
Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS Potterton EA, Power HR, Read RJ, Vagin A, Wilson KS. 2011. Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr 67(Pt 4):235–242.
Zakariassen H, Aam BB, Horn SJ, Varum KM, Sorlie M, Eijsink VG. 2009. Aromatic residues in the catalytic center of chitinase A from Serratia marcescens affect processivity, enzyme activity, and biomass converting efficiency. J Biol Chem 284(16):10610–10617.
Zakariassen H, Eijsink VGH, Sørlie M. 2010. Signatures of activation parameters reveal substrate-dependent rate determining steps in polysaccharide turnover by a family 18 chitinase. Carbohydr Polym 81(1):14–20.
Zhang S, Irwin DC, Wilson DB. 2000. Site-directed mutation of noncatalytic residues of *Thermobifida fusca* exocellulase Cel6B. Eur J Biochem/FEBS 267(11):3101–3115.
Zhang YH, Cui J, Lynd LR, Kuang LR. 2006. A transition from cellulose swelling to cellulose dissolution by o-phosphoric acid: Evidence from enzymatic hydrolysis and supramolecular structure. Biomacromolecules 7(2):644–648.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web-site.