Molecular recognition in the product site of cellobiohydrolase Cel7A regulates processive step length

Johan Pelck Olsen¹, Jeppe Kari², Michael Skovbo Windahl³, Kim Borch¹ and Peter Westh²*

¹Novozymes A/S, Denmark, ²Department of Biotechnology and Biomedicine, Technical University of Denmark, ³Bioneer A/S, Denmark

*Corresponding author: Email: petwe@dtu.dk; Telephone: +45 45 25 26 41

Keywords: Enzyme kinetics, inhibition, proofreading, cellobiohydrolase, processivity, cellulose.

Abstract
Cellobiohydrolase Cel7A is an industrial important enzyme that breaks down cellulose by a complex processive mechanism. The enzyme threads the reducing end of a cellulose strand into its tunnel-shaped catalytic domain and progresses along the strand while sequentially releasing the disaccharide cellobiose. While some molecular details of this intricate process have emerged, general structure-function relationships for Cel7A remain poorly elucidated. One interesting aspect is the occurrence of particularly strong ligand interactions in the product binding site. In this work, we analyze these interactions in Cel7A from *Trichoderma reesei* with special emphasis on the Arg251 and Arg394 residues. We made extensive biochemical characterization of enzymes that were mutated in these two positions and showed that the arginine residues contributed strongly to product binding. Specifically, about 50% of the total standard free energy of product binding could be ascribed to four hydrogen bonds to Arg251 and Arg394, which had previously been identified in crystal structures. Mutation of either Arg251 or Arg394 lowered production inhibition of Cel7A, but at the same time altered the enzyme product profile and resulted in about 50% reduction in both processivity and hydrolytic activity. The position of the two arginine residues closely matches the two-fold screw axis symmetry of the substrate, and this energetically favors the productive enzyme-substrate complex. Our results indicate that the strong and specific ligand interactions of Arg251 and Arg394 provide a simple proofreading system that controls the step-length during consecutive hydrolysis and minimizes dead time associated with transient, non-productive complexes.

Introduction
Cellobiohydrolases (CBH’s) are an industrially important group of glycoside hydrolase (GH) enzymes, which catalyze the hydrolysis of cellulose with cellobiose as the primary product. They work predominantly in a processive manner, meaning that they slide along one individual cellulose chain successively releasing the disaccharide cellobiose without dissociating from the substrate. The structural basis of this behavior is a remarkable substrate binding region involving up to 10 pyranose sub-sites located in a groove or tunnel (1,2). The sub-sites are sequentially numbered according to the conventional GH nomenclature with +1/-1 on either site on the scissile bond, where positive and negative integers denote the reducing and non-
reducing end of the substrate, respectively (3). The most studied CBH is Cel7A from *Trichoderma reesei*. This enzyme catalyzes cellulose hydrolysis in the direction from the reducing to the non-reducing end, which means that the cellobiose product is transiently bound in the +1/+2 site following hydrolysis (see Fig. 1). For this reason, the +1/+2 site is also denoted the product binding site or sometimes – perhaps to emphasize the pivotal role for the catalytic cycle – the product expulsion site.

In 1998 Divne and co-workers reported the first high-resolution crystal structure of Cel7A with bound substrate, and noted that the intensity of enzyme-substrate interactions increased towards the product site (2). The authors suggested the abundance of hydrogen bonds in the product-binding site might be important both for the advancement of the substrate strand during processive movement and for the pronounced product inhibition, which had been reported earlier (2,4). These interpretations by Divne et al. have promoted subsequent interest in ligand interactions in the product-binding site, and motivated both computational- (5-7) and experimental (8,9) studies. In accordance with the suggestion by Divne et al., a dual role of product site interactions has been found experimentally in variants with both lower product inhibition and reduced catalytic efficacy on insoluble cellulose (8,9). Other experimental studies have shown that in addition to strong ligand interactions, (10) the product site also exhibits an anomeric selectivity, which influences the product profile (11). Motivated by these results, we have studied interrelationships of product site interactions and function of *T. reesei* Cel7A. Based on the crystal structure, we identified the arginine residues at positions 251 and 394 as especially relevant for this analysis. These two residues are situated on either side of the product in the +1 and +2 pyranose sub-site, respectively (see Fig. 1). This positioning has a conspicuous correspondence to the two-fold screw axis symmetry of the substrate, resulting in the two residues interacting with the same functional groups of the glucopyranose moieties in subsite +1 and +2 (2). Moreover, the two residues have been identified *in silico* as the most important for interactions in the product-binding site (2,5-7). To elucidate the functional role of these residues we made variants of Cel7A, where Arg251 and Arg394 were substituted by alanine. By measuring the binding of α- and β-anomers of glucose and cellobiose to wild-type and variants, we found experimental evidence for the strength of individual interactions. Combining these results with measurements of processivity, maximal turnover and product profiles revealed that these residues are important for productive positioning of the cellulose strand during processive movement. This gives new insights into the roles of product binding site interactions as a proofreading system for the step length in the processive cycle.
Figure 1. Illustration of Cel7A during processive hydrolysis of a cellulose chain in a microfibril. The enlargement shows the positions of the two arginine residues studied here, as well as the catalytic nucleophile Glu212. Hydrolytic cleavage occurs between pyranose subsites -1 and +1.

Experimental

Enzymes and reagents: Chemicals and reagents were purchased from Sigma-Aldrich. All reactions were carried out at 25°C in 50 mM acetate buffer (pH 5) with 2 mM CaCl$_2$. Mutagenesis, heterologous expression in Aspergillus oryzae, and purification of TrCel7A were carried out as described previously (11). Beta-glycosidase (BG) from Aspergillus fumigatus and pyranose dehydrogenase (PDH) from Agaricus meleagris were expressed and purified as described elsewhere (12,13). Enzyme concentrations were determined by absorbance at 280 nm measured on a Nanodrop spectrophotometer (Thermo scientific) using a theoretical extinction coefficient of 86.8 mM$^{-1}$ cm$^{-1}$ for TrCel7A (wild type and variants), 185.6 mM$^{-1}$ cm$^{-1}$ for AfBG, and 67.8 mM$^{-1}$ cm$^{-1}$ for AmPDH.

Inhibition – soluble substrate: Inhibition of hydrolytic activity against para-nitrophenyl lactopyranoside (pNPL) by both the α and β anomers of glucose and cellobiose was determined by measuring the release of para-nitrophenol by absorbance at 405 nm (Spectramax I3 plate reader, Molecular Devices) as described elsewhere (14), with the only difference that hydrolysis time was reduced to 5 minutes. Inhibitor concentrations were 0, 100, 250 and 500 mM for glucose and 0, 100, 250 and 500 μM for cellobiose. To ensure activities were within the linear range of the spectrophotometer enzyme concentrations were adjusted to 1.60 μM, 0.32 μM and 1.28 μM for wild-type TrCel7A, R251A and R394A, respectively. In all cases, the substrate concentration was in excess with regards to the enzyme by at least two orders of magnitude.
magnitude. Inhibition by α- and β–glucose as well as β–cellobiose was measured by the same procedure as above, except with freshly dissolved ligands. The supplier states the enantiomeric purity to be 96% and 97% for α-glucose and β-glucose respectively, and we have previously determined the cellobiose preparation to be almost pure β–cellobiose (15). We were unable to obtain pure α–cellobiose and relied on an equilibrated solution with 64% β–cellobiose (16) for estimation of the enantiomeric effect for α–cellbiose (see Results). The dissolution of ligands was meticulously timed so that the reaction was initiated exactly 120 seconds after addition of buffer to the dry ligand for glucose and 90 seconds for cellobiose. Using a previously determined mutarotation rate of 5·10⁻⁴ s⁻¹ for both glucose and cellobiose (15) we estimated the enantiomeric composition halfway through the reaction to be 88%, 7%, 36% and 4% α-configuration for the preparations designated α-glucose, β–glucose, equilibrated cellobiose and β–cellobiose, respectively. These values were used to calculate actual concentrations of the respective enantiomers in the reaction mixtures, which allowed determination of inhibition constants for all four enantiomers as described below. Through the experiments, we used different stock preparations of enzyme, substrate and inhibitor and this resulted in slight variations of $k_{cat}$. Hence, a reference (uninhibited enzyme) was included in all inhibition experiments, and $k_{cat}$ reported in Tab.1 is an average. The Michaelis-Menten constants ($K_M$) for the uninhibited enzymes were indistinguishable within the standard error for independent experimental series. 

Product profile: Production of glucose (G₁), cellobiose (G₂) and cellotriose (G₃) from the hydrolysis of insoluble cellulose was measured in a standard endpoint assay. The substrate was 70 g/l washed Avicel PH 101 prepared as previously reported (17). Reactions were carried out in 2.0 ml microcentrifuge tubes with a final reaction volume of 1.0 ml and a final enzyme concentration of 1.0 µM enzyme. After 30 minutes of incubation in an Eppendorf Thermomixer at 25 °C, 750 rpm, the reaction was stopped by centrifugation for 3 minutes at 1.4·10⁴ relative centrifugal force. The supernatant was diluted 1:10 in milliQ water, transferred to glass vials and analyzed by HPAEC-PAD (Dionex ICS-5000, Thermo Fischer Scientific) for glucose, cellobiose and cellotriose against a 6-point combined external standard curve.

Biosensor activity measurements: We used a pyranose dehydrogenase (PDH) modified carbon paste electrode to determine the release of soluble cello-oligosaccharides in the early phase of the reaction in real-time for the wild type and the two variants. The substrate was the same, 70 g/l washed Avicel, as used in the end-point measurement described above and the enzyme concentration was 100 nM. The preparation and use of PDH biosensors were in accordance with our previously published protocol (13,15). The sensors were covered by a 100 nm pore size polycarbonate membrane as described in the mentioned protocol, but we used 1,4-benzoquinone as a mediator as described elsewhere (14). We used averages of duplicate measurements, and for all three enzymes, reproducibility was excellent, making distinction between repeated measurements practically impossible.
**Structural- and sequence analysis.** All structures (see fig. 4) were prepared in PyMol® (Schrödinger LLC, V. 2.3.0) using a crystal structure of the catalytic domain of TrCel7A in complex with a cellononanose (PDB ID: 4C4C) (18). Each structure was energy minimized in a steepest descent in GROMACS (2018.6) (19-23) while using CHARMM36 (24-27) force field to describe the systems.

To test the conservation of the Arg251 and Arg394 in the GH7 family a database with 1296 putative GH7 sequences was created using sequences from the Pfam database (28). The database was cleaned with the online tool MaxAlign 1.1 (29) and redundancy was reduced using CD-HIT (30) with a 70% sequence identity cut-off. Sequences that did not contain all catalytic amino acids (according to Payne et al. (31)) were removed manually. The resulting 281 sequences were aligned using Fast Fourier Transform (MAFFT) (32).

**Results**

**Ligand binding:** We utilized the suppression of hydrolytic activity on the soluble substrate analog para-nitrophenyl lactopyranoside (pNPL) to assess the binding of different ligands. The inhibition mechanism for the investigated systems has previously been found to be competitive (14,33,34), and this means that the inhibition constant, $K_i$, derived from kinetic measurements is equivalent to the equilibrium dissociation constant $K_d$ for the inhibitor (35). The experiments were designed to single out $K_d$ for both $\alpha$- and $\beta$-anomers of cellobiose and glucose. For glucose, this was done by determining enzymatic activity with systematically varied amounts of two different glucose preparations containing primarily the $\alpha$- or $\beta$-configuration. For cellobiose, we were unable to obtain preparations dominated by the $\alpha$-anomer, and instead we applied respectively $\beta$-cellobiose and an equilibrated sample as inhibitors in the kinetic experiments. In the analysis of results from the equilibrated sample, we assumed an $\alpha$-cellobiose content of 36% in the equilibrated samples (16). We analyzed all results with respect to a two-inhibitor competitive inhibition model, using the estimated concentrations of $\alpha$- and $\beta$-anomers ($I_\alpha$ and $I_\beta$) as inhibitor concentrations. The rate equation in eq. (1) is analogous to the well-known description of single-inhibitor competitive inhibition, but introduces one inhibition constant for each of the two ligands, $K_\alpha$ and $K_\beta$ (36).

\[
v = \frac{V_{\text{max}} \cdot S_0}{S_0 + K_m \left(1 + \frac{I_\alpha}{K_\alpha} + \frac{I_\beta}{K_\beta}\right)}
\]  

(Eq. 1)

Results from the inhibition studies are shown in Fig. 2. To derive the inhibition constants from this data, we analyzed all data in each panel by global, non-linear regression using Eq. 1. The best fits of the global analysis are illustrated by the lines in Fig. 2, and it appears that Eq. 1 accounted well for all measurements. The resulting parameters are listed in Tab. 1. We note that kinetic measurements are an unusual approach.
to obtain dissociation constants, however the ligand interactions with the variants were too week to be measured with sufficient accuracy by ITC (data not shown). To confirm the validity of the kinetically derived dissociation constants we measured the binding of equilibrated cellobiose to TrCel7A by ITC. Results in the Supplementary material show that equilibrated cellobiose bound to TrCel7A with a \(K_d\) of 19 \(\mu\)M, in excellent agreement with the inhibition measurements (see table 1).

Figure 2. Michaelis Menten curves showing activity on the soluble substrate analogue pNPL in presence of inhibitor (glucose and cellobiose) against substrate concentration for the three investigated enzymes. Left panel shows MM-curves in presence of different concentrations of glucose while right panel shows MM-curves in presence of different concentrations of cellobiose. The three enzymes were: TrCel7A (Top), R251A (middle) and R394A (bottom). In all cases except for cellobiose inhibition of R394A there was a clear tendency that the binding strength of the \(\beta\)-anomer was higher than that of the \(\alpha\)-anomer. In all experiments the temperature was 25 °C and the enzyme concentration was 1.6 \(\mu\)M, 0.32 \(\mu\)M and 1.28 \(\mu\)M for TrCel7A, R251A and R394A respectively. Error bars indicate the standard deviation of duplicate measurements.
Table 1. Michaelis-Menten parameters ($K_M$ and $k_{cat}$) and apparent competitive inhibition constants for the hydrolysis of the soluble substrate analog pNPL at 25°C. Inhibition was determined by fitting a two-inhibitor competitive inhibition model (eq. 1) to data obtained with either glucose or cellobiose in different anomeric forms.

|        | $k_{cat} \cdot 10^3$ (s$^{-1}$) | $K_M$ (mM) | $K_i$ (mM) | $K_i$ (μM) |
|--------|---------------------------------|------------|------------|------------|
|        |                                  | $\alpha$-glucose | $\beta$-glucose | $K_i(\beta)/K_i(\alpha)$ | $K_i(\beta)/K_i(\alpha)$ |
| Cel7A  | $142 \pm 2$                     | 0.52 ± 0.04 | 254 ± 23   | 126 ± 9    | 0.5 ± 0.1   | 18.8 ± 0.5 | 150 ± 60 | 12.7 ± 0.3 | 0.1 ± 0.1 |
| R251A  | $2259 \pm 60$                   | 3.1 ± 0.6   | 748 ± 28   | 440 ± 12   | 0.6 ± 0.1   | 171 ± 6   | 607 ± 180 | 119 ± 4   | 0.2 ± 0.1 |
| R394A  | $192 \pm 4$                     | 1.7 ± 0.2   | 453 ± 20   | 171 ± 5    | 0.4 ± 0.04  | 519 ± 31  | 524 ± 73  | 507 ± 29  | 1.0 ± 0.3 |

Activity and processivity on cellulose: Fig. 3 shows progress curves measured by the PDH functionalized electrode. As observed before (37) Cel7A shows a conspicuous burst phase followed by a quasi-steady-state regime, where the progress curve is essentially linear. The current measurements were made at high substrate loads (70 g/l), which is 30-50 times higher than reported $K_M$ values (38). It follows that the enzyme is essentially saturated and that the steady-state rate (the slope of the linear part in Fig. 3) is close to $V_{max}$. Under these conditions, the processivity number, $N$, can be estimated from the amplitude of the burst, $\pi$, defined as the intersect of the (extrapolated) tangent to the progress curve at steady state (see Fig. 3). Rigorous analysis has shown that $\pi = E_0 N$ (39,40), and we note in passing that this is analogous to the relationship used in so-called active site titration, which is commonly used for non-processive enzymes (i.e. for $N=1$) (41). In practice, we fitted a straight line to the progress curve between 120-300 seconds (dashed lines in Fig. 3), and extrapolating it to $t = 0$ sec to obtain $\pi$. This is the amount of soluble sugars released during the first processive run by the total enzyme population. Normalizing this value to the enzyme concentrations results in the product released from each enzyme during its first processive run, which is a measure of $N$. Compared to the wild type, the value of $N$ determined by this approach was reduced by approximately 50% and 70% for R251A and R394A, respectively. The maximal steady-state rates on Avicel (slope of linear parts in Fig.3) of the two variants were also reduced compared to the wild type. Specifically, we found maximal turnover numbers of 0.24 s$^{-1}$, 0.13 s$^{-1}$ and 0.078 s$^{-1}$ for TrCel7A, R251A and R394A, respectively (Tab. 2). While possibly fortuitous, the relative losses with respect to the wild type (about 50% for R251A and 70% for R394A) were the same as the reductions in processivity given above.
Figure 3. Time-resolved data for the release of soluble saccharides during hydrolysis of 70 g/l avicel by 100 nM TrCel7A (black), R251A (red) and R394A (blue). Dashed lines indicate linear regression to hydrolysis curves from 120-300 seconds used to determine the steady-state rate and amplitude of the burst ($\eta$). The processivity number $N$ was estimated from $\eta$ as detailed in the text.

**Product profile:** To evaluate the role of product site interactions for the positioning of substrate prior to hydrolysis, we measured the distribution of products from the hydrolysis of Avicel (70g/l). Specifically, we measured the concentrations of glucose, cellobiose and cellotriose by HPAEC-PAD. Results in Tab. 2, reiterate the result from the biosensor measurements that the variants are less active on Avicel compared to the wild type. More importantly, we found that the two variants produced glucose less frequently than the wild type. Thus, glucose made up 20% of total products for the wild type, and only about 10% for R251A. The lower output of glucose was associated with a higher relative production of cellobiose for this variant. R394A produced intermediate levels of glucose (15% of total products), but twice as much cellotriose as the wild type.

Table 2. Activity measurements for the three enzymes on 70 g/l Avicel and 100 nM enzyme. Soluble products were quantified by HPAEC-PAD. From real-time biosensor data, $k_{obs}$ and $N$ were determined by normalizing to the enzyme concentration respectively the slope and intercept of linear regression to the hydrolysis curve from 120-300 seconds (see Fig. 2)
Sequence analysis: The conservation of the two investigated arginines in the GH7 family was analyzed by sequence alignment of 281 putative GH7 cellulases (data not shown). Of these, 103 had substantial truncation in the peripheral loops enclosing the catalytic tunnel (mainly B2, B3 and A4, using the nomenclature of Momeni et al. (42)). These truncations strongly suggest that these enzymes are endoglucanases (31). Interestingly, none of these sequences contained arginine residues on positions corresponding to R251 and R394. The remaining 178 sequences were putative cellobiohydrolases (with all loops intact). Of these sequences, R251 was ~90% conserved while R394 was ~99% conserved.

Discussion
The purpose of this study was to elucidate ligand interactions between the +1 and +2 subsites and their effect on catalysis for TrCel7A. Based on structural evidence and other earlier work (5-8), we identified the residues Arg251 and Arg394 for mutational analysis. We were particularly interested in the almost perfect positioning of these residues along the two-fold screw axis symmetry of the substrate; i.e. their match with the 180° flip of consecutive pyranose rings in cellulose as illustrated in Fig. 1. We emphasize that computational- (5-7) and experimental (8,9) studies have pinpointed at least 8 residues (including the two investigated arginines) which are involved in product binding. However, in the current study have focused on the functional properties of the two conserved arginines.

In the following, we analyze ligand interactions of these residues and argue that during processive hydrolysis, Arg251 and Arg394 promote the formation of an enzyme-substrate complex with the scissile glycosidic bond in the right position near the nucleophile Glu212 and cellobiose in the product site. Conversely, complexes in which the polymeric substrate is shifted one pyranose unit forward or backward have higher energies and are hence less populated. This ability of Cel7A to preferentially form productive complexes during processive hydrolysis arises not only from strong interactions in the expulsion site, but relies on the enzyme’s ability to recognize substrate symmetry.
Figure 4. Hypothetical structure of the product binding site with α- or β-cellobiose (cyan) bound in the product site of Cel7A (left) and the two variants R251A (middle) and R394A (right). The catalytic nucleophile Glu212 is shown in gray sticks to indicate the position of the scissile bond. Hydrogen bonds to Arg251 and Arg394 as identified by Divne et al (2) are indicated in yellow, dashed lines. The images were made in Pymol (Schrödinger, LLC, New York) based on PDB-entry 4C4C. All structures was prepared in Pymol and energy minimized as stated in the method section.

Binding of cellobiose and glucose. In accordance with previous reports (9,10,43), we found a dissociation constant, $K_d$, for (anomer-equilibrated) cellobiose of 18.8 µM. This corresponds to a standard free energy of binding, $\Delta G^\circ$, of -27 kJ/mol. The role of the investigated arginine residues appeared clearly, as the dissociation constant for equilibrated cellobiose increased about an order of magnitude in both of the variants. The changes in $K_d$ (anomer-equilibrated) corresponded to $\Delta \Delta G^\circ$ values of 5 kJ/mol for R251A and 8 kJ/mol for R394A, or a relative reduction in binding strength of respectively 20% and 30% in the variants.

In an earlier computational study, Bu et al. reported a reduction in $\Delta G^\circ$ from -60 kJ/mol in TrCel7A to -55 kJ/mol in R251A (5). Possible origins of the difference in absolute binding strengths between simulation and experiment have been discussed elsewhere (10), and will not be repeated here, but we note that the $\Delta \Delta G^\circ$ value for R251A shows good agreement for the two approaches. The dissociation constants for α and β-cellobiose (Tab. 1) confirmed an earlier suggestion (44) that the TrCel7A wild type has a distinct preference for the β-anomer. Thus, $K_d$ was about ten-fold lower for the β-form. Interestingly, this anomicselectivity was completely lost in the R394A variant but mostly conserved in R251A (Tab. 1). Obviously, these results suggest that Arg394 is more involved in interactions with the reducing end of cellobiose in the expulsion site than Arg251, and this conclusion is in line with crystallographic evidence of the hydrogen bonding
pattern (2,4) illustrated in Fig. 4. To obtain quantitative insights into this hydrogen bonding, we made a combined analysis of the binding data in Tab. 1 and the structure of the enzyme-ligand complex, c.f. Fig. 4. If, for example, we consider the two panels to the left in Fig. 4 it appears that the hydrogen bond between Arg394 and the O1,2 oxygen atom only occurs for β-cellobiose (O1 follows the usual nomenclature for carbohydrates, and subscript +2 identifies the pyranose sub-site of the enzyme, see Fig. 1). Hence, we tentatively ascribe the strength of the R394-O1,2 interaction to the difference $\Delta G^\circ$($\beta$-cellobiose) - $\Delta G^\circ$($\alpha$-cellobiose) for the wild type enzyme. This difference is -6 kJ/mol as shown in Tab. 3. Analogous comparisons of binding data and proposed structures in Fig. 4 were used to estimate the contribution from the other interactions. Details of this analysis can be found in the Supplementary Material, and the results are listed in Tab. 3. While these values are only approximate as we neglected other (unknown) changes arising from local rearrangements in the product site of the mutants, they signify strong interactions, particularly between Arg394 and the reducing end of the cellulose strand.

Table 3 Estimated interaction strength of the hydrogen bonds formed by Arg251 and Arg394.

| Interaction | $\Delta G^\circ$ (kJ/mol) |
|-------------|--------------------------|
| R251-O6,1   | -6                       |
| R251-O3,2   | -3                       |
| R394-O6,2   | -3                       |
| R394-O1,2   | -6                       |

The binding of the α- and β-anomer of glucose to wild-type Cel7A gave $K_d$-values of respectively 254 mM and 126 mM (Tab. 1). This is in the same order of magnitude as the concentration of glucose that inhibits TrCel7A by 50% on insoluble cellulose, i.e. the so-called IC$_{50}$-value (45,46), but we are not aware of earlier work that directly measured the binding of glucose to Cel7A. The R251A variant, having significantly decreased hydrogen-bonding capabilities in the +1 site (Fig. 4), exhibited a clear reduction of the glucose binding strength (~ 3 fold increase in $K_d$-values). For R394A we observed only a moderate reduction in glucose binding strength (~ 1.5 fold increase in $K_d$-values), and all three enzymes in Tab. 1 showed a weak preference for the β-anomer of glucose. We suggest that this reflects some affinity of glucose for both sites, but that a more stable complex can be made in the +1 site. This is the first experimental evidence for the preferential binding of glucose to the +1 site, but the same conclusion was reached in a computational study (5). If indeed glucose has higher affinity for the +1 site, this could have important implications for the design of industrial cellulases since glucose is the dominant product in industrial breakdown of lignocellulosic biomass. We hasten to say, however, that engineering enzymes with lower product inhibition may be a difficult task considering the importance of this site for processive hydrolysis. This connection between inhibition and activity was elegantly demonstrated by Atreya et al, who showed that a
number of product site variants in a homologous Cel7A from *T. emersonii* were less inhibited, but also less catalytically efficient compared to the wild type (8). We observed a similar trade-off in this study as the activity on Avicel (Tab. 2) decreased commensurately to the inhibitor binding strength (Tab. 1).

**Proofreading of processive step length.** The energy barriers for the motion of a cellulose strand inside the binding tunnel of Cel7A are quite small, and translocation (in both directions) probably occurs much faster than the chemical processes in the catalytic cycle (6). This underscores the need for some proofreading mechanism that can help the enzyme stabilize the productive complex during continuous processive reactions. Due to the 180° rotation of consecutive glucose units in cellulose, the glucosidic bond can be positioned productively with the bond facing the catalytic nucleophile E212 (see Fig. 1) or unproductively where the bond points away from E212. We suggest that the productive complex with celllobiose in the product site is favored by hydrogen bonding to the two Arg residues, which recognizes the two-fold symmetry of the cellulose strand. This molecular recognition is achieved, at least in part through interactions with the O6 oxygens in both subsites and the O1 oxygen in subsite +2, and these interactions are all likely to be weaker or absent in other shifted complexes (c.f. Figs. 1 and 4). These other complexes include structures, where the product site is occupied with odd-numbered saccharide (e.g. glucose or cellotriose) or even-numbered ligands that are longer than celllobiose (e.g. cellotetraose). We emphasize that this proofreading of the processive step length results not simply from strong interactions in the product site, but relies on the match of the arginine residues and the two-fold symmetry of the substrate. From sequence alignment we found that Arg251 and Arg394 are highly conserved (≥90%) among processive GH7 cellobiohydrolases, while the non-processive structural homolog GH7 endoglucanases do not have these residues. This observation was also reported by Knott et al. (6) from sequence alignment of 7 GH7 CBHs and 3 EGs with solved crystal structure. Hence, the primitive proofreading mechanism provided by Arg251 and Arg394 may be essential for an efficient processive mechanism. This idea was supported in a recent work by Wang et al. (47), which introduced Arg394 at the corresponding position in endoglucanase Cel7B from *Trichoderma reesei*, and found increased processivity in the variant.

**Initial cut.** The previous section discussed consecutive steps during processive hydrolysis. There is, however, one more aspect of product site interactions that is worth considering, and that is their role for the initial cut in a processive run. This first cut is intricate and may produce a range of different products including both anomers of glucose, celllobiose, cellotriose and cellotetraose (44). This complexity results partially from the two-fold symmetry of a cellulose chain, and partly from the α/β-distribution of the reducing end (44). It is generally deduced that odd-length products (in practice glucose and cellotriose) in hydrolysates of Cel7A mainly stem from initial-cuts (48), and recently it was found that α-glucose dominated the among the odd-length products of Cel7A (44). In light of this, we may use the
concentrations of odd-length products in Tab.2 to elucidate the roles of the two arginine residues for the first cut of Cel7A. The most conspicuous result in Tab. 2 was that the R251A variant produces much less glucose (10% of total products) compared to the wild type (20%). We suggest that lowered glucose production for R251A reflected lower stability of the complex with filled +1, but empty +2 subsite. This complex is favored by the Arg251-O6\textsubscript{+1} hydrogen bond, which cannot be formed in R251A. This interaction seems less dependent on the anomeric configuration (see Tab. 1), and the glucose production of R251A will probably be low regardless of the configuration of the reducing end of the cellulose strand. Hence, the results suggest that strong and anomer-independent interactions to Arg251 are at least part of the reason for the dominance of glucose as first-cut product in the wild type.

**Conclusion**

We have shown that replacement with alanine of either Arg251 or Arg394 in the product site of Cel7A severely affects both activity, processivity and the binding of glucose and cellobiose. The maximal hydrolytic rate for R251A measured by biosensors was about half the value of the wild type and lower still for R394A. The processivity of the variants was reduced by approximately the same fractions, and we suggest that these kinetic changes occur as a direct consequence of weaker interactions in the product site. Hence, strong interactions here are necessary both for the forward movement of substrate in the binding tunnel and for keeping the enzyme bound to a cellulose strand during processive hydrolysis. Studies of ligand binding showed that Arg251 had a major effect on glucose binding while Arg394 was particularly important for the binding of cellobiose. We assessed binding to TrCel7A and variants of both α- and β-anomers of glucose and cellobiose, and combining this binding data with previously published enzyme-ligand structures, we were able estimate contributions of individual interactions. We found that interactions with the O6 oxygen of pyranose rings in subsites +1 and +2 contributed equally to the standard free energy of ligand binding by approximately 3 kJ/mol each, and that the interaction of Arg394 and the reducing end was particularly strong (6 kJ/mol). All four hydrogen bonds discussed here rely on the two-fold symmetry of the cellulose strand (the 180° flip of consecutive pyranose rings). As a result, these interactions strongly favor productive complexes with cellobiose in the product site and the scissile bond in the correct position. This minimizes the population of unproductive complexes and hence, in essence, provides a proofreading mechanism for the length of the processive step during consecutive hydrolysis of a cellulose strand. We conclude that the product site of Cel7A is designed not only to provide strong attraction, which is required to drive the forward motion of the cellulose strand. It is also finely tuned to recognize the substrate symmetry in order to increase the likelihood of productive enzyme-substrate complexes, thereby avoiding the dead time that would arise if complexes of the wrong conformation were
significantly populated. This gives new insight into function of the product-binding site of celllobiohydrolases as the structural architecture seems to provide a proofreading mechanism for processive step length, which may be essential for an efficient processive breakdown of cellulose.

Abbreviations
CBH, celllobiohydrolase; GH, glycoside hydrolase; MM, Michaelis-Menten; BG, Beta-glycosidase; pNPL, para-nitrophenyl lactopyranoside; PDH, pyranose dehydrogenase; HPAEC-PAD, High-Performance Anion-Exchange Chromatography Coupled with Pulsed Electrochemical Detection.

Author contribution
J.P.O., J.K., K.B. and P.W. conceived and designed the study. J.P.O performed the kinetic measurements. M.S.W. mutated, expressed and purified the enzymes. J.P.O, J.K. and P.W. wrote the manuscript. All authors reviewed and approved the manuscript.

Funding
This work was financially supported by grants from Innovation Fund Denmark (Grant no. 0603-00496B).

Competing interests
J.P.O and K.B. work for Novozymes A/S, a major manufacturer of industrial enzymes.

Supplementary material
This article contains supplementary material.

Acknowledgement
We thanks S. J. Christensen and K. Schaller for performing the sequence and structural analysis.

References
1. Breyer, W. A., and Matthews, B. W. (2001) A structural basis for processivity. Protein Sci 10, 1699-1711
2. Divne, C., Stahlberg, J., Teeri, T. T., and Jones, T. A. (1998) High-resolution crystal structures reveal how a cellulose chain is bound in the 50 Å long tunnel of celllobiohydrolase I from Trichoderma reesei. J Mol Biol 275, 309-325
3. Davies, G. J., Wilson, K. S., and Henrissat, B. (1997) Nomenclature for sugar-binding subsites in glycosyl hydrolases. Biochem J 321 ( Pt 2), 557-559
4. Ståhlberg, J., Divne, C., Koivula, A., Piens, K., Claeyssens, M., Teeri, T. T., and Jones, T. A. (1996) Activity studies and crystal structures of catalytically deficient mutants of celllobiohydrolase I from Trichoderma reesei. J. Mol. Biol. 264, 337-349
5. Bu, L., Beckham, G. T., Shirts, M. R., Nimlos, M. R., Adney, W. S., Himmel, M. E., and Crowley, M. F. (2011) Probing carbohydrate product expulsion from a processive cellulase with multiple absolute binding free energy methods. The Journal of biological chemistry 286, 18161-18169
6. Knott, B. C., Crowley, M. F., Himmel, M. E., Stahlberg, J., and Beckham, G. T. (2014) Carbohydrate-protein interactions that drive processive polysaccharide translocation in enzymes revealed from a computational study of cellulobiohydrolase processivity. J Am Chem Soc 136, 8810-8819

7. Silveira, R. L., and Skaf, M. S. (2015) Molecular Dynamics Simulations of Family 7 Cellulobiohydrolase Mutants Aimed at Reducing Product Inhibition. Journal of Physical Chemistry B 119, 9295-9303

8. Atreya, M. E., Strobel, K. L., and Clark, D. S. (2016) Alleviating product inhibition in cellulase enzyme Cel7A. Biotechnology and bioengineering 113, 330-338

9. von Ossowski, I., Stahlberg, J., Koivula, A., Piens, K., Becker, D., Boer, H., Harle, R., Harris, M., Divine, C., Mahdi, S., Zhao, Y. X., Driguez, H., Claeyssens, M., Sinnott, M. L., and Teeri, T. T. (2003) Engineering the exo-loop of Trichoderma reesei cellulobiohydrolase, Cel7A. A comparison with Phanerochaete chrysosporium Cel7D. J Mol Biol 333, 817-829

10. Colussi, F., Sorensen, T. H., Alasepp, K., Kari, J., Cruys-Bagger, N., Windahl, M. S., Olsen, J. P., Borch, K., and Westh, P. (2015) Probing substrate interactions in the active tunnel of a catalytically deficient cellulobiohydrolase (Cel7). The Journal of biological chemistry 290, 2444-2454

11. Westh, P., Kari, J., Olsen, J., Borch, K., Jensen, K., and Krogh, K. B. R. M. (May 1, 2014) PCT International Patent Appl. WO/2014/064115. C12N 9/00 Ed.

12. Bohlin, C., Olsen, S. N., Morant, M. D., Patkar, S., Borch, K., and Westh, P. (2010) A comparative study of activity and apparent inhibition of fungal beta-glucosidases. Biotechnology and bioengineering 107, 943-952

13. Cruys-Bagger, N., Badino, S. F., Tokin, R., Gontsariik, M., Fathalinejad, S., Jensen, K., Toscano, M. D., Sorensen, T. H., Borch, K., Tatsumi, H., Valjamae, P., and Westh, P. (2014) A pyranose dehydrogenase-based biosensor for kinetic analysis of enzymatic hydrolysis of cellulose by cellulases. Enzyme Microb Technol 58-59, 68-74

14. Olsen, J. P., Alasepp, K., Kari, J., Cruys-Bagger, N., Borch, K., and Westh, P. (2016) Mechanism of product inhibition for cellulobiohydrolase Cel7A during hydrolysis of insoluble cellulose. Biotechnology and bioengineering 113, 1178-1186

15. Cruys-Bagger, N., Ren, G., Tatsumi, H., Baumann, M. J., Spodsberg, N., Andersen, H. D., Gorton, L., Borch, K., and Westh, P. (2012) An amperometric enzyme biosensor for real-time measurements of cellulobiohydrolase activity on insoluble cellulose. Biotechnology and bioengineering 109, 3199-3204

16. Kabayama, M. A., Patterson, D., and Piche, L. (1958) The thermodynamics of mutarotation of some sugars: I. Measurement of the heat of mutarotation by micrololorimetry. Canadian Journal of Chemistry 36, 557-562

17. Sorensen, T. H., Cruys-Bagger, N., Windahl, M. S., Badino, S. F., Borch, K., and Westh, P. (2015) Temperature Effects on Kinetic Parameters and Substrate Affinity of Cel7A Cellulbiohydrolases. The Journal of biological chemistry 290, 22193-22202

18. Knott, B. C., Haddad Momeni, M., Crowley, M. F., Mackenzie, L. F., Götz, A. W., Sandgren, M., Withers, S. G., Ståhlberg, J., and Beckham, G. T. (2014) The Mechanism of Cellulose Hydrolysis by a Two-Step, Retaining Cellulbiohydrolase Elucidated by Structural and Transition Path Sampling Studies. Journal of the American Chemical Society 136, 321-329

19. Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., and Lindahl, E. (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1-2, 19-25

20. Berendsen, H. J. C., van der Spoel, D., and van Drunen, R. (1995) GROMACS: A message-passing parallel molecular dynamics implementation. Computer Physics Communications 91, 43-56

21. Lindahl, E., Hess, B., and van der Spoel, D. J. M. m. a. (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis. 7, 306-317

22. Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., van der Spoel, D., Hess, B., and Lindahl, E. (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 29, 845-854
41. Fersht, A. (1998) Structure and Mechanism in Protein Science, 2nd ed., W.H. Freeman and Company, New York
42. Momeni, M. H., Payne, C. M., Hansson, H., Mikkelsen, N. E., Svedberg, J., Engstrom, A., Sandgren, M., Beckham, G. T., and Stahlberg, J. (2013) Structural, biochemical, and computational characterization of the glycoside hydrolase family 7 cellobiohydrolase of the tree-killing fungus Heterobasidion irregularare. The Journal of biological chemistry 288, 5861-5872
43. Claeyssens, M., Van Tilbeurgh, H., Tomme, P., Wood, T. M., and McRae, S. I. (1989) Fungal cellulase systems. Comparison of the specificities of the cellobiohydrolases isolated from Penicillium pinophilum and Trichoderma reesei. Biochem J 261, 819-825
44. Kari, J., Kont, R., Borch, K., Buskov, S., Olsen, J. P., Cruyz-Bagger, N., Valjamae, P., and Westh, P. (2017) Anomeric Selectivity and Product Profile of a Processive Cellulase. Biochemistry 56, 167-178
45. Teugjas, H., and Väljamäe, P. (2013) Product inhibition of cellulases studied with 14C-labeled cellulose substrates. Biotechnology for biofuels 6, 1-14
46. Murphy, L., Bohlin, C., Baumann, M. J., Olsen, S. N., Sørensen, T. H., Anderson, L., Borch, K., and Westh, P. (2013) Product inhibition of five Hypocrea jecorina cellulases. Enzyme and Microbial Technology 52, 163-169
47. Wang, Y., Zhang, S., Song, X., and Yao, L. (2016) Cellulose chain binding free energy drives the processive move of cellulases on the cellulose surface. Biotechnology and bioengineering 113, 1873-1880
48. Horn, S. J., Sorlie, M., Varum, K. M., Valjamae, P., and Eijsink, V. G. (2012) Measuring processivity. Methods in enzymology 510, 69-95