Substrate binding in the processive cellulase Cel7A

Transition state of complexation and roles of conserved tryptophan residues

Røjel, Nanna Sandager; Kari, Jeppe; Sørensen, Trine Holst; Badino, Silke Flindt; Morth, J. Preben; Schaller, Kay; Cavaleiro, Ana Mafalda; Borch, Kim; Westh, Peter

Substrate binding in the processive cellulase Cel7A: Transition state of complexation and roles of conserved tryptophan residues.

Nanna Røjel,[1,#] Jeppe Kari,[2] Trine Holst Sørensen,[3] Silke F. Badino,[1,#] J. Preben Morth,[2] Kay Schaller,[2] Ana Mafalda Cavaleiro,[3] Kim Borch[3] and Peter Westh[2,*]

[1]Roskilde University, INM, 1 Universitetsvej, Building 28, DK-4000, Roskilde, Denmark
[2]Technical University of Denmark, Department of Biotechnology and Biomedicine. Building 224, DK-2800, Kgs. Lyngby, Denmark
[3]Novozymes A/S, 2 Biologiens Vej, DK-2800 Kgs. Lyngby Denmark

Running title: Cellulose binding and transition state of Cel7A

[#]Present address: Technical University of Denmark, Department of Biotechnology and Biomedicine. Building 224, DK-2800, Kgs. Lyngby, Denmark

[*]Corresponding author. Email: petwe@dtu.dk Tel. +45 3091 5488

ABSTRACT

Cellobiohydrolases effectively degrade cellulose and are of biotechnological interest because they can convert lignocellulosic biomass to fermentable sugars. Here, we implemented a fluorescence-based method for real-time measurements of complexation and decomplexation of the processive cellulase Cel7A and its insoluble substrate, cellulose. The method enabled detailed kinetic and thermodynamic analyses of ligand binding in a heterogeneous system. We studied wildtype Cel7A and several variants in which one or two of four highly conserved Trp residues in the binding tunnel had been replaced with Ala. Wildtype Cel7A had on/off rate constants of 1×10^5 $M^{-1}s^{-1}$ and $5 \times 10^{-3}s^{-1}$, respectively, reflecting the slow dynamics of a solid, polymeric ligand. Especially the off-rate constant was many orders of magnitude lower than typical values for small, soluble ligands. Binding rate and strength both were typically lower for the Trp variants, but effects of the substitutions were moderate and sometimes negligible. Hence, we propose that lowering the activation barrier for complexation is not a major driving force for the high conservation of the Trp residues. Using so-called Φ-factor analysis, we analyzed the kinetic and thermodynamic results for the variants. The results of this analysis suggested a transition state for complexation and decomplexation in which the reducing end of the ligand is close to the tunnel entrance (near Trp-40), whereas the rest of the binding tunnel is empty. We propose that this structure defines the highest free-energy barrier of the overall catalytic cycle and hence governs the turnover rate of this industrially important enzyme.

INTRODUCTION

Cellobiohydrolases (EC 3.2.1.21 and EC 3.2.1.91) are among the most effective enzymes for the breakdown of cellulose. They are dominant in the secretome of many cellulose degrading microorganisms and hence play a vital role in the natural carbon cycle. They are also of major technological interest because they make up the principle component in enzyme cocktails used for the conversion of lignocellulosic biomass to fermentable sugars. This industrial process (so-called saccharification), is critically important for the development of efficient biorefineries that produce fuels and chemicals from lignocellulosic feedstock. Reaching the desired goals for production of renewable fuels will involve an unprecedented consumption of industrial enzymes, and improved technical and scientific understanding of the saccharification process hence appears valuable. However, many mechanistic aspects of cellobiohydrolases (CBHs) remain poorly understood and this is at least in part because these enzymes operate by a complex, processive mechanism.

In the current work, we address the kinetics and thermodynamics of substrate complexation and decomplexation of the CBH, Cel7A, from the cellulolytic fungus *Trichoderma reesei*. These two steps appear to be particularly important, as they have alternately been proposed to be rate-limiting for the overall reaction (see (1) for a recent review). Focus on the on- and off-rates of Cel7A has been accentuated by the observation that covalent transitions associated with the actual bond cleavage proceed quite rapidly. Thus, the sequence of events (sometimes called the inner catalytic cycle; step 3 in Fig. 1A), which includes breaking of one β-1,4 glycosidic bond, expulsion of product (cellobiose) and one processive step forward occurs at a rate of about 5 s^{-1} at room temperature (2). Yet, the maximal turnover of Cel7A, V_{max}/E_0, is much lower at about 0.1 – 0.3 s^{-1} (3,4), and this clearly suggests the existence of slow (non-covalent) transitions outside the inner catalytic cycle. Slow complexation, for example, appears likely as the

process involves "threading" of at least nine pyranose moieties into the tunnel-shaped binding area of the enzyme (*c.f.* Fig. 1). This process could well be associated with a high activation energy because the piece cellulose strand must be pulled out of a bound state in the cellulose particle before it can engage in favorable interactions with the enzyme (5). Conversely, some studies have ascribed low turnover to a particularly slow rate of decomplexation, and a concomitant accumulation of inactive enzyme on the substrate surface (4,6). These interpretations of either on- or off-rate limitation of Cel7A may appear discordant, but they could also reflect that the rate-limiting step changes with experimental conditions such as temperature or substrate load (7,8). In any event, more insights into the binding and un-binding of a cellulose strand in Cel7A will be required to understand the mechanism of this enzyme. In particular, it would be rewarding to gain insights into the nature of the transition state for the processes of substrate complexation and decomplexation, as these structures may govern the overall reaction rate under most conditions.

The structural element that has received most interest regarding threading and substrate binding for Cel7A is four highly conserved Trp residues placed along the binding tunnel (9,10) as shown in Fig. 1B. They have all been shown to interact with the ligand (11) and two of them, W38 and W40, are located in "relaxed sub-sites" (11) (sub-sites number -7 and -4) quite far from the catalytic residues, where the ligand is in an extended conformation. The third, W367, is in a "twisting sub-site" (11) (subsite -2), where it helps to flip the chain and hence expose the scissile bond for nucleophilic attack, and the last, W376, is in a "product sub-site" (sub-site +1) where the product is transiently located after hydrolysis of the glycosidic bond. Different functional roles of these Trp-residues in Cel7A and other cellulases have been discussed extensively (reviewed in (9)), and in the current work, we address this through quantitative kinetic data. Specifically, we have implemented a continuous

fluorescence method by which complexation of Cel7A and insoluble cellulose can be monitored after a dead-time of a few hundred ms. We have used this method to study the dynamics of complex formation for both Cel7A and a number of enzyme variants in which one or two of the Trp residues had been replaced with Ala. The results elucidated both wild-type ligand binding and the roles of the four Trp residues for the rate and strength of complexation. Finally, the data allowed so-called Φ-factor analysis to elucidate the nature of the transition state in the complexation/decomplexation steps.

Figure. 1. Panel A shows a simplified illustration of the reaction mechanism of Cel7A. Panel B: The enlargement shows the structure of the bound cellononaose ligand (green/red sticks) and the location of the four conserved Trp residues. Numbers in the enlargement identify the subsites; the scissile bond is between subsites +1 and -1. PDB 4C4C.

RESULTS AND DATA ANALYSIS

Titration experiments. The fluorescence emission at 328 nm increased distinctly when wild type Cel7A bound its insoluble substrate (see Fig. S1 in the SI). For the W376A+W38A mutant, the change was much less pronounced (about 15 % of the change for Cel7A$_{WT}$), and the other enzymes in Tab. 1 showed changes between these two values. In general, the changes were large enough to be readily measurable, and probably reflect de-quenching as water molecules are removed from the indole side chains of Trp-residues (12) during ligand binding. It follows that the populations of free and bound enzyme can be estimated by a linear combination of the fluorescence signals from enzyme solutions with respectively no substrate and saturating substrate loads. Under this interpretation, the fluorescence method distinguishes between free enzyme (E) and enzyme in complex with substrate

(ES), while it is blind to activity and (un-complexed) adsorption (*i.e.* step 1 in Fig. 1A). In light of this, we will interpret the current data along the lines of a simple binding model (Scheme 1).

$$E + S \underset{k_{off}}{\overset{k_{on}}{\rightleftarrows}} ES \quad \text{(Scheme 1)}$$
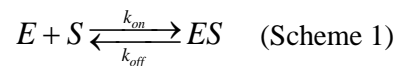
We note that Scheme 1 assumes binding equilibrium, and this criterion has previously been confirmed for short experiments with the current systems (13).

The fraction of enzyme molecules that are engaged in a complex, θ, may be determined from as

$$\theta = \frac{[ES]}{E_0} = \frac{F_x}{F_{max}} \quad (1)$$

where E_0 is the total enzyme concentration, F_{max} is the increment in the fluorescence signal at saturation (measured at high substrate loads where $E_0 \sim [ES]$), and F_x is the increment in signal in a titration experiment with a given cellulose load (c.f. Fig. S1 in the SI). As the signals in the titration experiments were allowed to become constant, we assume that θ represents the equilibrium distribution. However, the equilibrium constant for Scheme 1 is not straight forward to define because the substrate load, S, is in mass units (the accessible molar concentration of sites at the surface of the insoluble substrate is unknown). To address this, we introduce a parameter, Γ, which specifies the density of sites on the substrate surface (in mol/(g cellulose)), to which the enzyme can bind and form the ES complex (sometimes called an attack site). The total (molar) concentration of attack sites is then $S_0^* \Gamma$ where S_0^* is the (known) initial cellulose load in g/L. We may then write the following mass balance

$$[S] = S_0^* \Gamma - [ES] \quad (2)$$

where [S] is the (molar) concentration of unoccupied attack sites. The overall strength of cellulase binding

can be described by the so-called partition coefficient, K_P (14), which may be expressed

$$K_P = \frac{\Gamma}{K_d} \qquad (3)$$

In eq. (3), K_d is the usual dissociation constant ($K_d=[E][S]/[ES]$) for scheme 1). The partitioning coefficient takes into account both intrinsic binding strength and the availability of sites, and hence provides a weighted affinity parameter. This is an important property because distinctly different Γ-values among cellulases (15,16) tend to decouple K_d and the overall binding capacity. For this reason, we will use K_P in discussions of the relative binding strength of the investigated variants. We note, however, that according to eq. (3), K_P can be readily converted to a dissociation constant (in units of M^{-1}), $K_d=K_P/\Gamma$. This latter parameter carries larger uncertainty because it requires conversion to molar substrate concentration, and it is hence subject to error propagation through division of two parameters from the regression analysis (Γ and K_p). However, it is useful to use K_d in comparisons with other studies that reports affinity through the dissociation constant.

To derive experimental values of K_P and Γ from the titration measurements, we combine eqs. (1), (2) and (3). This yields

$$\frac{S_0^*\Gamma}{E_0} - \theta\left(1 + \frac{S_0^*\Gamma}{E_0} + \frac{\Gamma}{E_0 K_P}\right) + \theta^2 = 0 \quad (4)$$

Solving the quadratic eq. (4) for θ and combining with eq. (1) gives

$$F_x = F_{max} \frac{\beta - \sqrt{\beta^2 - \alpha}}{2} \qquad (5)$$

$$\alpha = 4\frac{S_0^*\Gamma}{E_0} \quad , \quad \beta = 1 + \frac{S_0^*\Gamma}{E_0} + \frac{\Gamma}{E_0 K_P}$$

Equation 5, which is a standard relationship for ligand binding (17), was implemented in a non-linear regression routine (Origin 2018b, OriginLab Wellesley Hills, MA), and used to derive maximum likelihood values for K_P, Γ and F_{max}. We consistently found that eq. (5) accounted well for the experimental data, and examples of measurements and regression results are shown in Fig. 2 (a schematic representation of how the raw data from the titratration experiments were converted into binding curves can be found in SI Fig. S2). The parameters (K_p, Γ and F_{max}) for all investigated enzymes are listed in Tab. 1. Note that F_{max} is reported as the ratio (in %) of variant and Cel7A_WT.

Figure 2. Examples of binding curves obtained from the titration experiments. Each point represents the apparent equilibrium fluorescence emission (in arbitrary units) at a given load of amorphous cellulose (RAC). Data is shown for the wild type Cel7A and the W40A mutant at starting concentrations of 250 nM enzyme. The lines show the best fits of eq. (5) to the experimental data.

Time-course experiments: Figure 3 shows representative examples of the real-time data for ligand binding; results for other enzymes may be found in the Supporting Information (SI, Fig. S3). We used 250 nM enzyme and at $t = 0$, we added enough RAC suspension (always 100 µL) to reach the total substrate load listed in the legend of Fig. 3. To analyze the data, we first assessed the instrument dead time in control experiments with a non-binding protein. As detailed in Fig. S4 of the SI, effects of mixing tapered off with a half-time of 150 ms, and became undetectable (compared to the experimental scatter) after about 0.5 s. Hence, we discarded data points for $t < 0.5$ s. for all runs. The results could be described by a single exponential function of the type $F_x(t) = F_{eq}(1 - e^{-kt})$, where $F_x(t)$ is the real-time fluorescence signal and F_{eq} is the equilibrium value specified by the plateau level for each RAC load as illustrated in Fig. 3.

4

In some cases, a slight improvement of the variance of the fits could be gained by using the sum of two exponential terms. We did not associate any theoretical meaning with the exponential fitting parameters but used them as a practical way to find the initial slope of the fluorescence traces as the mathematical derivative for t→0. Combining with eq. (1) and F_{max} from the titration experiments (Tab. 1), we converted slopes for t→0 to the initial rate of complex formation, v_{on}, in units M/s. We plotted v_{on} as a function of the RAC load as exemplified in Fig. 4. These plots consistently showed that v_{on} scaled proportionally to the RAC load, and hence confirmed that the reaction was first order with respect to the substrate at least in the concentration range studied here.

Figure 3. Real time fluorescence data for the complexation of RAC and respectively the Cel7A wild type and the W40A mutant (250 nM enzyme). The ordinate, θ, shows the fraction of enzyme molecules in an ES complex (*c.f.* eq. 1). Analogous data for the other enzymes are shown in Fig S3 of the SI. Black lines represent best fit to an exponential function.

According to Scheme 1, the initial on-rate may be expressed as $v_{on} = k_{on}S_0E_0$, and it follows that the slope, α, of the lines in Fig. 4 is $\alpha = k_{on}E_0$. We used α-values from Fig. 4 and (known) E_0 to calculate the k_{on} values listed in Tab. 1. On-rate constants determined in this way are in the somewhat unusual (mass-based) units of $(g/L)^{-1}s^{-1}$, and to emphasize this, we will henceforth use the symbol $^{mass}k_{on}$ for this parameter. However, as we know the density of attack sites on the substrate (Γ in μmol/g cellulose) from the titration experiments, we may convert $^{mass}k_{on}$ to an on-rate constant, $^{molar}k_{on}$, in the usual units of $M^{-1}s^{-1}$ ($^{molar}k_{on}=^{mass}k_{on}/\Gamma$). This value is also listed in Tab. 1. Finally, combination of eq. (3) and the general relation between K_d and rate constants, $K_d=k_{off}/^{molar}k_{on}$, allows us to determine the off-rate constant, $k_{off}=^{mass}k_{on}/K_P$. We note in passing that all parameters could have been derived from the time-course measurements without conducting titration measurements at all. However, we found that the experimental scatter was lessened by making separate titration experiments with repeated additions of RAC to the same enzyme sample and we hence preferred this procedure.

Cellulase kinetics is known to depend on the physical properties of the substrate, and to investigate this aspect we attempted to repeat the above experiments on bacterial microcrystalline cellulose (BMCC), which is mainly crystalline. Unfortunately, excessive light scattering from BMCC limited this work to fairly low loads of cellulose (*i.e.* far from saturation with superimposed curves at increasing loads as in Fig. 3). This prevented systematic analysis, but raw data for Cel7A WT at low BMCC loads suggested kinetics comparable to that observed on RAC (see Fig. S5 in the SI). In fact, the movement towards equilibrium was slightly faster (half-time of about 5 s) on this more crystalline substrate.

Figure 4. Initial rate of complexation, v_{on}, derived from Fig. 3 and plotted as a function of the RAC load for two enzymes, Cel7A_{WT} and W40A. The observed proportionality supports the interpretation that the complexation is first order with respect to the load of cellulose, and the slope signifies the parameter $^{mass}k_{on}E_0$ (see main text).

Table 1. Parameters from titration- and time-course experiments.

* $F_{max}/F_{max\_WT}\cdot 100$

DISCUSSION

Cellulases and other enzymes that attack solids, must break interactions in the substrate surface before a complex can be formed, and this "excavation" of the

ligand is likely to be important for the energy landscape of heterogeneous enzyme reactions. Heterogeneous reactions are widespread both *in vivo* and technical enzyme applications (18-22), and insights into the thermodynamics and kinetics of ligand binding appears particularly important for mechanistic understanding. Earlier works on this topic for cellulases have mostly reported adsorption data derived from assays that distinguishes respectively bulk- and surface-adsorbed enzymes. Results from this type of work have provided quite consistent conclusions regarding the strength of binding (see below), but the picture has been less clear with respect to dynamics. Some works have found that Cel7A and other cellulases build up quite slowly on the substrate surface, with apparent adsorption equilibrium established after tens of minutes (23-27). Others have reported a faster process, which leveled off towards a constant adsorbed population within seconds or tens of seconds (28,29). Recently, more detailed insights have emerged from single-molecule imaging methods (24,30-33). This work has elucidated key parameters such as the rate of movement and residence times of enzymes on the substrate, although on- and off-rate constants derived in different studies have been quite divergent (24,30,31,33). Other studies have used specialized biochemical assays, which enables distinction between cellulases that are either free in solution, complexed with substrate or surface adsorbed with empty binding tunnel (*i.e.* not complexed, see Fig. 1) (29,34-36). This type of work has elucidated the population of these three states and the dynamics of their interconversion, but the necessary assays are very labor-intensive and hence not apt for comparative investigations. In the current work we address the complexation and decomplexation of Cel7A through a new fluorescence-based method that monitors ligand binding in real-time. The observed changes in fluorescence signal occurs because the ligand causes de-quenching of Trp residues, when it enters the binding tunnel, and the method hence distinguishes threaded enzyme from apoenzyme, whereas it is insensitive to whether the enzyme is adsorbed with

empty binding tunnel or free in the bulk. We argue that this makes it particularly useful in studies of complexation dynamics. We first focus on the rate of threading of Cel7A$_{WT}$ and subsequently discuss the roles of the conserved Trp residues and the possible structure of the transition state for the complexation step.

Wild type Cel7A and variants with native Trp pattern. The partitioning coefficient, K_P, for Cel7A$_{WT}$ was 122 ± 33 L/g (Tab. 1), which corresponds to a dissociation constant, $K_d = K_P/\Gamma$ of about 52 nM. This binding strength falls in the same range as earlier reports on Cel7A (14,27,36-38). Affinity data from these and other references are listed in Table S5 of the SI. This table shows that earlier work with sub µM enzyme concentrations (like the current work) found K_d values in the low nM range in accordance with the results reported here. Experiments with tens or hundreds of µM enzyme, on the other hand, indicate much weaker binding (K_d in the µM range). This difference probably reflects heterogeneity of the binding sites, and thus a tendency to detect weaker, unproductive binding at higher concentrations (37,39). The value for the attack site density on amorphous cellulose ($\Gamma = 6.4 \pm 0.6$ µmol attack sites per gram of RAC, Tab. 1) is in excellent accordance with earlier adsorption studies and biochemical assessments of productive binding capacities (13,40). Thus, results from the titration experiments generally supported the validity of the approach and interpretation based on Scheme 1. This simplified interpretation is in line with the argument that the method only differentiates complexed and un-complexed states, and further supported by the observation that the wild type and E212Q-variant have almost the same on/off rate constants (Tab. 1). Thus, the E212Q variant is practically inactive on insoluble cellulose (41), and hence formally meets the premises of Scheme 1.

The off-rate constant for wild type Cel7A was 0.0048 ± 0.0013 s^{-1} (Tab. 1), and this corresponds to a mean ES life-time (1/k_{off}) of about 200 s. This number is in

6

good agreement with recent measurements of Cel7A residence times based on single-molecule fluorescence imaging, and it is also in line with biochemical data (4,33). The low k_{off}-value has the important corollary that slow dissociation governs the reaction rate at saturation (4,28), and as a result, the maximal turnover number may be approximated $V_{max}/E_0 \sim n k_{off}$ (42), where n is the average processivity number. Earlier work has suggested that n ~ 15-20 for Cel7A acting on RAC (42-45), and multiplying this value by k_{off} from Tab. 1 predicts a maximal turnover of about 0.1 s^{-1}. This matches V_{max}/E_0 –values measured directly in biochemical assays (3,4), and this prediction of the maximal turnover from the current ligand binding studies (without any activity measurements) further supports applicability of the method. The on-rate constant for Cel7A$_{WT}$ determined from the initial slope in Fig. 3 was 0.59 ± 0.01 (g/L)$^{-1}$s^{-1} (Tab. 1). As seen directly in Fig. 3, this corresponds to a rather swift complexation process, which moves towards equilibrium with a half-time around 10 s under the conditions studied here. As argued above, the mass-based on-rate constants ($^{mass}k_{on}$) in Tab. 1 can readily be converted to $^{molar}k_{on}$ in the conventional units of M^{-1}s^{-1}. Values of $^{molar}k_{on}$ calculated in this way are also listed in Tab. 1. We posit that this parameter scales with the molar concentrations of enzyme and attack sites like a normal bimolecular rate constant, and hence that it could be meaningful to compare it with on-rate constants for other enzyme reaction. Typical values for on-rate constants fall in the $(1 – 50) \times 10^6$ M^{-1}s^{-1} range for enzymes binding small, soluble ligands such as amino acids, nucleotides, small metabolites or hydrogen peroxide (46). The current $^{molar}k_{on}$ for Cel7A was 0.1×10^6 M^{-1}s^{-1} (Tab. 1) and thus 1-2 orders of magnitude slower than typical values for small-molecule ligands. This slower complexation probably reflects energy barriers associated with the polymeric nature of the ligand and its binding to the surface prior to complexation, but we note that the difference in on-rate constants for solid cellulose and soluble ligands is not exceptionally large. This obviously suggests that

Cel7A has developed an effective stabilization of the transition state for complexation and we return to this below. We may also compare the off-rate constant in Tab. 1 with typical values for small, soluble ligands. This comparison suggests that Cel7A is far more atypical with respect to decomplexation, as typical off-rate constants for small ligands are in the $10^3 – 10^4$ s^{-1} range (46), and hence, larger than k_{off} for Cel7A by some 6 orders of magnitude. Overall, these numbers illustrate the slow complexation/decomplexation dynamics of Cel7A, and the uncommonly low value of k_{off} underscores that Cel7A binds its substrate more tightly than average enzymes (as $K_d = k_{off}/^{molar}k_{on}$). This latter point is confirmed by comparing the standard free energy of binding for Cel7A$_{WT}$, $\Delta G° = RT\ln K_d \sim -42$ kJ/mol, with values in a recent meta-analysis of literature data for hydrolases acting on soluble substrates (47). Thus, only a few percent of the hydrolases bound their cognate substrate as strongly as Cel7A$_{WT}$, and the average $\Delta G°$ was about -20 kJ/mol.

Trp-variants and Φ-factor analysis. Tryptophan residues play a key role as driver of protein-carbohydrate interactions. This is reflected, for example, in a distinct overrepresentation of Trp in contact regions of non-covalently bound carbohydrates in protein crystal structures (48). This special role of the indole side chain of Trp for carbohydrate binding is also evident in Cel7A, which has four Trp residues along the substrate-binding tunnel (11,49) as illustrated in Fig. 1. Functional roles of these Trp residues in Cel7A and analogous Trp residues in other processive glycoside hydrolases have been widely discussed and several mutational studies have highlighted their importance for *e.g.* ligand binding strength, initial threading, processivity number and processive sliding (10,32,50-57). We first notice that the maximal change in fluorescence emission, F$_{max}$, differed markedly among the investigated enzymes (Tab. 1 and Fig. S1). The wild type and variants with a native Trp pattern (*i.e.* E212Q and CD) all showed about the same F$_{max}$, while the Trp-variants had smaller changes. This is in

line with the interpretation that the fluorescence changes represent de-quenching of these Trp residues. The most conspicuous effect was in the two variants that included the W38A mutation. In these cases, F_{max} was reduced to respectively 15% (W38A+W376A) and 26% (W38A) of the WT-value (Tab. 1), and this suggested that W38 makes a particularly large contribution to measured F_{max}-values in Cel7A$_{WT}$. This may reflect a particularly strong dehydration of W38 upon ligand binding, and this interpretation was confirmed by MD simulations as described in the SI S-7. Thus, in the complex, no water was found within 5 Å of the center of mass of W38, while the other three Trp residues showed some residual hydration with 1-3 water molecules within this cutoff (Fig. S7 in SI).

While the parameters in Tab. 1 generally revealed a role of the Trp residues for the thermodynamics and kinetics of ligand binding, the observed effects were moderate, and this is in contrast to some earlier suggestions on the roles of these residues. In particular, the on-rate constants $^{mass}k_{on}$, were quite similar and fell within 30% of the wild type for all Trp point-mutants. We note that $^{mass}k_{on}$ is a robust, model free, parameter, and that the low sensitivity of the on-rate to Trp point mutations appeared directly from the initial slopes of the raw data in Figs. 3 and S3. Variants with two Trp replacements showed some reduction in the rate of complexation, but again the changes were not dramatic and the drop in $^{mass}k_{on}$ for both W367A+W376A and W38A+W376A was less than three-fold compared to Cel7A$_{WT}$. We conclude that at least on amorphous cellulose, one Trp residue can be replaced with limited effect on the threading rate of Cel7A. Even two replacements resulted in on-rates of the same order of magnitude, and this suggests that maintaining a high on-rate is not the major evolutionary driving force for the conservation of Trp in the binding tunnel of Cel7A. Binding affinity of the Trp point mutants was similar or moderately lowered compared to the wild type. The relative ligand affinity was Cel7A$_{WT}$~W376A > W367A ~ W40A > W38A. The difference in standard

free energy of binding, ΔΔG°, between the boundaries in this sequence (i.e. between W38A and the wild type) was about 4 kJ/mol. This ΔΔG° is in line with an earlier kinetic investigation of these two enzymes (53). For the variants W367A and W40A the free energy of ligand binding was about 2 kJ/mol, higher than the WT thus signifying weak favorable interactions from the Trp residues in these positions.

The results in Tab. 1 lists both rate constants and ligand affinities for variants with mutations in different regions of the binding area, and this opens for the application of so-called Φ-factor analysis to elucidate the transition state (TS) of the complexation/decomplexation reactions. This approach has been successfully used in protein folding studies, and the basic idea is to compare changes in respectively binding- and activation energy. Principles and practices of the approach have been lucidly discussed elsewhere (58,59), and will not be reiterated here, except for some key aspects of the current application, which are sketched out in Fig. 5. For the present data, we may define Φ as:

$$\Phi \equiv \frac{\Delta\Delta G^{\ddagger}}{\Delta\Delta G^{\circ}} = \frac{\ln\left(\dfrac{k_{wt}}{k_{var}}\right)}{\ln\left(\dfrac{K_{P,WT}}{K_{P,var}}\right)} \qquad (6)$$

If we first insert on-rate constants for different mutants in eq. (6), we obtain Φ-factors for complexation, Φ_{on}. A Φ_{on}-value about 1, for a given point mutation suggests that ligand interactions at this position are established *before* the TS (the mutation affects the free energy of ES and ES$^{\neq}$ equally, see Fig 5B). Conversely, if ligand interactions are established in the downhill part of the energy diagram *after* the TS, the Φ_{on}-value will be around 0 (only ES is affected, see Fig. 5C). The analogous values for decomplexation (i.e. going from right to left in Figs. 5B and 5C) are $\Phi_{off} = 0$ and $\Phi_{off} = -1$, respectively. We plotted the Φ-factors derived by eq. (6) from the data in Tab. 1 in the histogram in Fig. 5A. In this figure, the columns are organized according to

the position of the mutation along the binding tunnel, and the horizontal axis hence represents the reaction coordinate of the essentially one-dimensional threading process (the ligand enters the tunnel opening near the CBM and W40, and gradually moves forward to fill the tunnel). Although Φ-factors are usually calculated for point mutations, we include the CD-variant in this analysis as the CD is a distinct domain separated by a linker (the CD retains its native structure without the CBM (11)). As other Φ-factor analyses (60), interpretation of Fig. 5A was challenged by experimental errors, but we see that Φ_{on} is about 1 for the CD-variant, attains an intermediate for W40A and is close to 0 for the next three Trp residues. This suggests that the transition state for ligand binding occurs early in the threading process at a stage where the CBM is interacting with the substrate and the cellulose reducing end makes contact with W40 at the tunnel entrance. This implies that upon further forward movement, favorable enzyme-ligand interactions in the tunnel compensates the work required to release more glucose moieties from the cellulose surface, and hence makes this part of the complexation process thermodynamically downhill. This interpretation is supported by analyzing Φ_{off}-values derived from the decomplexation rate constants (Tab. 1). Thus, Φ_{off} for residues near the middle or exit of the tunnel (W376, W367, E212 and W38) were around -1 (Fig. 5A), suggesting the TS of decomplexation occurs after ligand-interaction of these residues have been broken. W40A has an intermediate value whereas the CBM-less variant (CD) has Φ_{off} near 0, and this again points toward a TS with the reducing end near the tunnel entry and W40.

Figure 6 provides a structural interpretation of results from the Φ-factor analysis. We stress that this is a coarse representation and that other structures could be in accord with the data, but the figure serves to illustrate that the TS has several glucose moieties broken away from the cellulose surface. However, the TS does not involve ligand interactions at the middle of the tunnel or near the catalytic residues. This observation of a TS (in both directions) with limited ligand penetration was also proposed in an enzyme kinetic study (61), and it is interesting to note that ligand interactions with W40, which are established around the TS in Fig. 6, have been highlighted for their role in Cel7A efficacy in different earlier works (10,32,44,62). The unfavorable free energy of the structure in Fig. 6 relies on a balance between attractive ligand interactions in respectively the cellulose surface layer and the enzyme complex. Strong interactions in the complex are illustrated by the large binding free energy found here ($\Delta G° \sim$ -42 kJ/mol), and elsewhere (14,27,36-38), but in order for the complex to form, several pyranose moieties must be dislodged from a bound state in the solid surface.

Figure 5. Φ-factor analysis. Panel A shows experimental Φ-factors for respectively complexation (Φ_{on}) and decomplexation (Φ_{off}). The cartoon above the histogram indicates the location of the investigated mutations. Panels B and C illustrates the interpretation of Φ-factors on the basis of two hypothetical transition states for the complexation process. In panel B, the TS is late in the process just before the ligand fully fills the tunnel, while in C, it is early in the complexation process. If we mutate a residue located at the yellow star, which interacts with the ligand, the energy landscape changes differently in the two cases (dashed lines). Thus, in B the free energies of the bound state (ES) and the TS (ES$^{\neq}$) will shift in parallel because the mutated residue interacts with the ligand in both states, and this yield $\Phi_{on} = 1$. Conversely, in C, only ES (and not ES$^{\neq}$) will be affected by the mutation and Φ_{on} will be 0. This difference allowed us to assess whether a given mutation was located before or after the reaction coordinate of the TS. An analogous analysis may be conducted for decomplexation. In this case, the Φ_{off}-factors that indicates whether the mutation is before or after the TS are 0 and -1 respectively.

The current results suggest that before the reducing end reaches W40, it is favorable for the cellulose strand to

flick back to a bound state on the cellulose surface. However, when the reducing end has passed W40, it is thermodynamically downhill to proceed forward and establish enzyme-ligand interactions.

Figure 6. Structural interpretation of the Φ-factor analysis. The transition state of complexation and decomplexation of Cel7A is proposed to occur at a stage, where the reducing end is near W40 while the rest of the tunnel is empty. The unfavorable free energy of this structure arises from the (uncompensated) detachment of several glucose moieties from the cellulose surface. At a higher reaction coordinate of the complexation process (when the tunnel fills) compensating ligand interactions occur in the tunnel and hence lower the free energy. At a lower reaction coordinate, the free energy becomes more favorable as the ligand interacts with the cellulose surface.

To address this interpretation quantitatively, one may (as a first approximation) estimate the free energy of this TS by inserting $^{molar}k_{on}$ in the Eyring equation (although this approach may be questionable for a diffusive process like complexation). Doing so suggests an activation free energy of about 45 kJ/mol, and it is interesting to compare this to computational estimates of the work required to pull off cellooligosaccharides from the surface of crystalline cellulose. This work is in the range of 12-18 kJ/(mol glucose moieties) (5,63), and these numbers suggest that the activation free energy of complexation corresponds to pulling off 3-4 glucose moieties. While this is only a very coarse estimate, this stretch of cellulose strand seems to be in a realistic order of magnitude

Conclusions. The intrinsic fluorescence of Cel7A has previously been shown to change distinctively upon ligand binding (64,65). Based on this, we have implemented a novel, real-time method and a simple kinetic analysis, which provides on- and off rates for Cel7A on its insoluble substrate. The approach appears robust and derives on-rates directly from initial slopes of raw progress curves, independent of models or regression analyses. The setup used a stirred cuvette, hence obviating potential problems of handling insoluble particles in *e.g.* stopped-flow systems. The setup had sub-second dead-time, and that was sufficient to monitor essentially the full complexation process in real-time. We envision that the method could be broadly applicable to cellulases and other glycosyl hydrolases as changes in tryptophan fluorescence is quite common for these enzymes (66). We found that complexation occurred fairly rapidly with an on-rate constant of the wild type of 0.59 (g/L)$^{-1}$s^{-1}. Taking the density of attack sites on RAC into account, this on-rate constant corresponded to about 0.1×10^6 M^{-1}s^{-1}, and this is only moderately slower than typical on-rate constants for enzymes and small, soluble substrates. The off-rate constant was 0.0048 s^{-1} and this is similar to some values reported earlier for Cel7A, and many orders of magnitude slower than typical off-rates for soluble substrates. The roles of the four highly conserved Trp residues in the binding tunnel were elucidated through the characterization of several mutants. We found that these Trp residues promoted the strength and rate of complexation, but the contributions were low or moderate, and we propose that their high conservation must rely on other functional factors than rapid complexation. Finally, we used the kinetic and thermodynamic data for the mutants in a so-called Φ-factor analysis, which suggested that the transition state for the on/off step occurred when the reducing end of the ligand reached the tunnel entrance near Trp40, and the remainder of the tunnel was empty. As earlier work has suggested that complexation or decomplexation (rather than the chemical changes associated with bond breakage) govern the maximal turnover of Cel7A, we propose that a structure akin to the one in Fig. 6 underlies the apex of the complex free energy landscape of this enzyme.

EXPERIMENTAL PROCEDURES

Enzymes, substrate and buffer. The cellobiohydrolase Cel7A form *Trichoderma reesei* and a number of Cel7A variants specified below were expressed in *Aspergillus oryzae* as detailed elsewhere (67). The enzymes were purified from fermentation broth, validated to show a single band on NuPAGE 4-12% Bis-Tris SDS-PAGE and quantified by UV absorption as described previously (8). We expressed and purified nine enzymes for this work. In addition to the wild type (denoted Cel7A$_{WT}$) we made six enzymes with variations in the pattern of Trp residues shown in Fig. 1. Specifically, we made four mutants in which each of the Trp residues had been replaced with Ala (W38A, W40A, W367A and W376A) and two mutants with two replacements of Trp (W367A+W376A and W38A+W376A). Finally, we expressed the catalytic domain (CD) form Cel7A (*i.e.* a variant with truncated linker and carbohydrate binding module, CBM) and a point mutant in which the catalytic nucleophile, Glu212, had been replaced with Gln. This latter variant, E212Q, has previously been shown to retain WT-like structure while being essentially inactive on insoluble cellulose (41). The substrates were Regenerated Amorphous Cellulose (RAC) and Bacterial Microcrystalline Cellulose (BMCC) and both were prepared by standard methods (68). The load of cellulose in the stocks was determined as dry matter content. All experiments were conducted in 50 mM sodium acetate buffer pH 5, referred to as standard buffer.

Fluorescence spectroscopy. Fluorescence measurements were conducted in a Jasco FP-8200 spectrofluorometer equipped with a custom-made injection inlet directly to the cuvette, as well as a standard thermostat and magnetic stirrer (Jasco STR-811). The experimental temperature was 20 °C. All measurements relied on changes in the emission spectrum, when Cel7A binds the ligand, and to assess this, we initially recorded full spectra (310-500 nm) for all investigated enzymes at several (typically 10) different loads of RAC. We used an excitation wavelength of 295 nm, which primarily probes Trp residues (12). In accordance with earlier reports on Cel7A, (64,65), the emission spectra (see Supplementary Information, S1) showed blue-shift and increased intensity upon ligand binding. The changes were particularly pronounced around 328 nm (Fig. S1), and we used this wavelength in all subsequent measurements. The strength of ligand binding was assessed from *titration measurements*. In these trials, 2 mL enzyme dissolved in pure standard buffer was initially transferred to a stirred 1x1 cm quartz cuvette. Subsequently, the enzyme solution was successively titrated with 10 - 20 aliquots of RAC suspension until we did not observe further changes in the signal (typically at a total RAC load of 0.2 – 0.4 g/L). After each RAC addition, we waited for the fluorescence signal to reach a constant value (*i.e.* apparent binding equilibrium) and then measured the average emission over 20 s. These titration trials were conducted for all variants at two enzyme concentrations (0.25 µM and 0.50 µM) and used to derive binding constants and binding site densities as described below. The rate of ligand-enzyme binding was investigated in real-time in a separate series of *time-course experiments.* In this case, the fluorescence from 2 mL, 0.25 µM enzyme solution with vigorous stirring was monitored continuously. When the signal was stable, one 100 µL aliquot of cellulose suspension was added quickly from a syringe, and the changes in fluorescence recorded (10 data points per s.) over approximately 500 s. The experiment was repeated (with a fresh enzyme solution) for about 10 different RAC loads for each enzyme variant so that the kinetics of binding was monitored for RAC loads between 0.005 and 0.3 g/L (the concentration of the injected RAC-aliquot was adjusted to reach the desired final load upon addition of 100 µL sample in all cases). We made a number of control measurements with similar additions of RAC suspension to either pure standard buffer or a solution of a non-binding protein (Bovine Serum Albumin, BSA, Sigma Aldrich Product #05470). These results were used partly to estimate the instrument dead time

11

(*i.e.* the time required to mix the reactants in the cuvette), and partly to quantify the loss emission associated with light scattering from RAC particles.

ACKNOWLEDGMENTS

CONFLICT OF INTEREST

T.H.S, A.M.C and K.B. work at Novozymes, a major enzyme producing company.

REFERENCES

1. Jeoh, T., Cardona, M. J., Karuna, N., Mudinoor, A. R., and Nill, J. (2017) Mechanistic kinetic models of enzymatic cellulose hydrolysisA review. *Biotechnol. Bioeng.* **114**, 1369-1385
2. Nill, J., Karuna, N., and Jeoh, T. (2018) The impact of kinetic parameters on cellulose hydrolysis rates. *Process Biochem.* **74**, 108-117
3. Badino, S. F., Kari, J., Christensen, S. J., Borch, K., and Westh, P. (2017) Direct kinetic comparison of the two cellobiohydrolases Cel6A and Cel7A from Hypocrea jecorina. *Biochimica et biophysica acta. Proteins and proteomics* **1865**, 1739-1745
4. Kurasin, M., and Valjamae, P. (2011) Processivity of Cellobiohydrolases Is Limited by the Substrate. *J. Biol. Chem.* **286**, 169-177
5. Payne, C. M., Himmel, M. E., Crowley, M. F., and Beckham, G. T. (2011) Decrystallization of Oligosaccharides from the Cellulose I beta Surface with Molecular Simulation. *Journal of Physical Chemistry Letters* **2**, 1546-1550
6. Cruys-Bagger, N., Tatsumi, H., Ren, G. R., Borch, K., and Westh, P. (2013) Transient Kinetics and Rate-Limiting Steps for the Processive Cellobiohydrolase Cel7A: Effects of Substrate Structure and Carbohydrate Binding Domain. *Biochemistry* **52**, 8938-8948
7. Kari, J., Olsen, J. P., Jensen, K., Badino, S. F., Krogh, K. B. R. M., Borch, K., and Westh, P. (2018) Sabatier Principle for Interfacial (Heterogeneous) Enzyme Catalysis. *ACS Catalysis* **8**, 11966-11972
8. Sørensen, 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 cellobiohydrolases. *Journal of Biological Chemistry*, jbc. M115. 658930
9. Payne, C. M., Knott, B. C., Mayes, H. B., Hansson, H., Himmel, M. E., Sandgren, M., Stahlberg, J., and Beckham, G. T. (2015) Fungal Cellulases. *Chem. Rev.* **115**, 1308-1448
10. Nakamura, A., Tsukada, T., Auer, S., Furuta, T., Wada, M., Koivula, A., Igarashi, K., and 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**, 13503-13510
11. 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 angstrom long tunnel of cellobiohydrolase I from Trichoderma reesei. *J. Mol. Biol.* **275**, 309-325
12. Lakowicz, J. R. (2006) *Principles of Fluorescence Spectroscopy*, 3rd ed., Springer
13. Pellegrini, V. O., Lei, N., Kyasaram, M., Olsen, J. P., Badino, S. F., Windahl, M. S., Colussi, F., Cruys-Bagger, N., Borch, K., and Westh, P. (2014) Reversibility of substrate adsorption for the cellulases Cel7A, Cel6A, and Cel7B from *Hypocrea jecorina*. *Langmuir* **30**, 12602-12609
14. Palonen, H., Tenkanen, M., and Linder, M. (1999) Dynamic interaction of Trichoderma reesei cellobiohydrolases Cel6A and Cel7A and cellulose at equilibrium and during hydrolysis. *Appl. Environ. Microbiol.* **65**, 5229-5233
15. Schiano-di-Cola, C., Rojel, N., Jensen, K., Kari, J., Sorensen, T. H., Borch, K., and Westh, P. (2019) Systematic deletions in the cellobiohydrolase (CBH) Cel7A from the fungus Trichoderma reesei reveal flexible loops critical for CBH activity. *J. Biol. Chem.* **294**, 1807-1815
16. Kari, J., Andersen, M., Borch, K., and Westh, P. (2017) An Inverse Michaelis Menten Approach for General Description of Interfacial Enzyme Kinetics. *ACS Catalysis* **7**, 4904-4914
17. Klostermeier, D., and Rudolph, M. G. (2017) *Biophysical Chemistry*, CRC Press, Taylor & Francis Group
18. Kirk, O., Borchert, T. V., and Fuglsang, C. C. (2002) Industrial enzyme applications. *Curr. Opin. Biotechnol.* **13**, 345-351

19. Sinsabaugh, R. L., Belnap, J., Findlay, S. G., Shah, J. J. F., Hill, B. H., Kuehn, K. A., Kuske, C. R., Litvak, M. E., Martinez, N. G., Moorhead, D. L., and Warnock, D. D. (2014) Extracellular enzyme kinetics scale with resource availability. *Biogeochemistry* **121**, 287-304

20. Marangoni, A. G. (1994) Enzyme kinetics of lipolysis revisited: the role of lipase interfacial binding. *Biochem. Biophys. Res. Commun.* **200**, 1321-1328

21. Illanes, A., Fernández-Lafuente, R., Guisán, R., and Wilson , L. (2008) Heterogeneous Enzyme Kinetics. in *Enzyme biocatalysis:* (Illanes, A. ed.), Springer, Dordrecht pp

22. McLaren, A. D., and Packer, L. (1970) Some aspects of enzyme reactions in heterogenous systems. *Adv. Enzymol. Relat. Areas Mol. Biol.* **33**, 245-308

23. Liu, H., Zhu, J. Y., and Chai, X. S. (2011) In Situ, Rapid, and Temporally Resolved Measurements of Cellulase Adsorption onto Lignocellulosic Substrates by UV-vis Spectrophotometry. *Langmuir* **27**, 272-278

24. Zhu, P., Moran-Mirabal, J. M., Luterbacher, J. S., Walker, L. P., and Craighead, H. G. (2011) Observing Thermobifida fusca cellulase binding to pretreated wood particles using time-lapse confocal laser scanning microscopy. *Cellulose* **18**

25. Moran-Mirabal, J. N., Santhanam, N., Corgie, S. C., Craighead, H. G., and Walker, L. P. (2008) Immobilization of Cellulose Fibrils on Solid Substrates for Cellulase-Binding Studies Through Quantitative Fluorescence Microscopy. *Biotechnol. Bioeng.* **101**, 1129-1141

26. Steiner, W., Sattler, W., and Esterbauer, H. (1988) Adsorption of *Trichoderma reesei* cellulase on cellulose: experimental data and their analysis by different equations. *Biotechnology and bioengineering* **32**, 853-865

27. Nidetzky, B., Steiner, W., and Claeyssens, M. (1994) Cellulose hydrolysis by the cellulases from *Trichoderma reesei*: adsorptions of two cellobiohydrolases, two endocellulases and their core proteins on filter paper and their relation to hydrolysis. *Biochemical Journal* **303**, 817

28. Cruys-Bagger, N., Elmerdahl, J., Praestgaard, E., Tatsumi, H., Spodsberg, N., Borch, K., and Westh, P. (2012) Pre-steady state kinetics for the hydrolysis of insoluble cellulose by Trichoderma reesei Cel7A. *J. Biol. Chem.* **287**, 18451-18458

29. Jalak, J., Kurašhin, M., Teugjas, H., and Väljamäe, P. (2012) Endo-exo synergism in cellulose hydrolysis revisited. *Journal of Biological Chemistry*, jbc. M112. 381624

30. Shibafuji, Y., Nakamura, A., Uchihashi, T., Sugimoto, N., Fukuda, S., Watanabe, H., Samejima, M., Ando, T., Noji, H., Koivula, A., Igarashi, K., and Iino, R. (2014) Single-molecule Imaging Analysis of Elementary Reaction Steps of Trichoderma reesei Cellobiohydrolase I (Cel7A) Hydrolyzing Crystalline Cellulose I (alpha) and III (I). *J. Biol. Chem.* **289**, 14056-14065

31. Moran-Mirabal, J. M., Corgie, S. C., Bolewski, J. C., Smith, H. M., Cipriany, B. R., Craighead, H. G., and Walker, L. P. (2009) Labeling and Purification of Cellulose-Binding Proteins for High Resolution Fluorescence Applications. *Anal. Chem.* **81**, 7981-7987

32. Igarashi, K., Koivula, A., Wada, M., Kimura, S., Penttila, M., and Samejima, M. (2009) High Speed Atomic Force Microscopy Visualizes Processive Movement of Trichoderma reesei Cellobiohydrolase I on Crystalline Cellulose. *J. Biol. Chem.* **284**, 36186-36190

33. Jung, J., Sethi, A., Gaiotto, T., Han, J. J., Jeoh, T., Gnanakaran, S., and Goodwin, P. M. (2013) Binding and movement of individual Cel7A cellobiohydrolases on crystalline cellulose surfaces revealed by single-molecule fluorescence imaging. *Journal of Biological Chemistry* **288**, 24164-24172

34. Kostylev, M., Moran-Mirabal, J. M., Walker, L. P., and Wilson, D. B. (2012) Determination of the molecular states of the processive endocellulase Thermobifida fusca Cel9A during crystalline cellulose depolymerization. *Biotechnol. Bioeng.* **109**, 295-299

35. Kuusk, S., Sorlie, M., and Valjamae, P. (2015) The Predominant Molecular State of Bound Enzyme Determines the Strength and Type of Product Inhibition in the Hydrolysis of Recalcitrant Polysaccharides by Processive Enzymes. *J. Biol. Chem.* **290**, 11678-11691

36. Cruys-Bagger, N., Alasepp, K., Andersen, M., Ottesen, J., Borch, K., and Westh, P. (2016) Rate of threading a cellulose chain into the binding tunnel of a cellulase. *The Journal of Physical Chemistry B* **120**, 5591-5600

37. Jalak, J., and Valjamae, P. (2014) Multi-Mode Binding of Cellobiohydrolase Cel7A from Trichoderma reesei to Cellulose. *PLoS One* **9**, e108181

38. Ståhlberg, J., Johansson, G., and Pettersson, G. (1991) A new model for enzymatic hydrolysis of cellulose based on the two-domain structure of cellobiohydrolase I. *Bio/Technology* **9**, 286

39. Medve, J., Ståhlberg, J., and Tjerneld, F. (1997) Isotherms for adsorption of cellobiohydrolase I and II from *Trichoderma reesei* on microcrystalline cellulose. *Applied biochemistry and biotechnology* **66**, 39-56

40. Karuna, N., and Jeoh, T. (2017) The productive cellulase binding capacity of cellulosic substrates. *Biotechnol. Bioeng.* **114**, 533-542

41. Stahlberg, 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 cellobiohydrolase I from Trichoderma reesei. *J. Mol. Biol.* **264**, 337-349

42. Cruys-Bagger, N., Elmerdahl, J., Praestgaard, E., Borch, K., and Westh, P. (2013) A steady-state theory for processive cellulases. *FEBS J.* **280**, 3952-3961

43. Nakamura, A., Watanabe, H., Ishida, T., Uchihashi, T., Wada, M., Ando, T., Igarashi, K., and Samejima, M. (2014) Trade-off between Processivity and Hydrolytic Velocity of Cellobiohydrolases at the Surface of Crystalline Cellulose. *J. Am. Chem. Soc.* **136**, 4584-4592

44. von Ossowski, I., Stahlberg, J., Koivula, A., Piens, K., Becker, D., Boer, H., Harle, R., Harris, M., Divne, 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 cellobiohydrolase, Ce17A. A comparison with Phanerochaete chrysosporium Cel7D. *J. Mol. Biol.* **333**, 817-829

45. Kurasin, M. (2017) *Processivity of cellulases and chitinases*, Tartu University

46. Fersht, A. (1998) *Structure and Mechanism in Protein Science*, 2nd ed., W.H. Freeman and Company, New York

47. Sousa, S. F., Ramos, M. J., Lim, C., and Fernandes, P. A. (2015) Relationship between Enzyme/Substrate Properties and Enzyme Efficiency in Hydrolases. *ACS Catalysis* **5**, 5877-5887

48. Hudson, K. L., Bartlett, G. J., Diehl, R. C., Agirre, J., Gallagher, T., Kiessling, L. L., and Woolfson, D. N. (2015) Carbohydrate-Aromatic Interactions in Proteins. *J. Am. Chem. Soc.* **137**, 15152-15160

49. Knott, B. C., Momeni, M. H., Crowley, M. F., Mackenzie, L. F., Gotz, A. W., Sandgren, M., Withers, S. G., Stahlberg, J., and Beckham, G. T. (2014) The Mechanism of Cellulose Hydrolysis by a Two-Step, Retaining Cellobiohydrolase Elucidated by Structural and Transition Path Sampling Studies. *J. Am. Chem. Soc.* **136**, 321-329

50. Sorlie, M., Zakariassen, H., Norberg, A. L., and Eijsink, V. G. H. (2012) Processivity and substrate-binding in family 18 chitinases. *Biocatal. Biotransform.* **30**, 353-365

51. Zakariassen, H., Aam, B. B., Horn, S. J., Varum, K. M., Sorlie, M., and Eijsink, V. G. H. (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**, 10610-10617

52. Kont, R., Kari, J., Borch, K., Westh, P., and Valjamae, P. (2016) Inter-domain Synergism Is Required for Efficient Feeding of Cellulose Chain into Active Site of Cellobiohydrolase Cel7A. *J. Biol. Chem.* **291**, 26013-26023

53. Kari, J., Olsen, J., Borch, K., Cruys-Bagger, N., Jensen, K., and Westh, P. (2014) Kinetics of Cellobiohydrolase (Cel7A) Variants with Lowered Substrate Affinity. *J. Biol. Chem.* **289**, 32459-32468

54. Taylor, C. B., Payne, C. M., Himmel, M. E., Crowley, M. F., McCabe, C., and Beckham, G. T. (2013) Binding Site Dynamics and Aromatic-Carbohydrate Interactions in Processive and Non-Processive Family 7 Glycoside Hydrolases. *J. Phys. Chem. B* **117**, 4924-4933

55. Zhang, S., Irwin, D. C., and Wilson, D. B. (2000) Site-directed mutation of noncatalytic residues of Thermobifida fusca exocellulase Cel6B. *Eur. J. Biochem.* **267**, 3101-3115

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

57. Davies, G. J., Ducros, V., Lewis, R. J., Borchert, T. V., and Schülein, M. (1997) Oligosaccharide specificity of a family 7 endoglucanase: insertion of potential sugar-binding subsites. *J. Biotechnol.* **57**, 91-100

58. Fersht, A. R., and Sato, S. (2004) Phi-Value analysis and the nature of protein-folding transition states. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 7976-7981

59. Ferguson, N., and Fersht, A. (2009) Φ - Value Analysis of Protein Folding Transition States. in *Protein Engineering Handbook* (Lutz, S., and Bornscheuer, U. eds.), WILEY-VCH, Weinheim. pp 81-120

60. Naganathan, A. N., and Muñoz, V. (2010) Insights into protein folding mechanisms from large scale analysis of mutational effects. *Proceedings of the National Academy of Sciences* **107**, 8611-8616

61. Sørensen, T. H., Cruys-Bagger, N., Borch, K., and Westh, P. (2015) Free energy diagram for the heterogeneous enzymatic hydrolysis of glycosidic bonds in cellulose. *Journal of Biological Chemistry* **290**, 22203-22211

62. Ghattyvenkatakrishna, P. K., Alekozai, E. M., Beckham, G. T., Schulz, R., Crowley, M. F., Uberbacher, E. C., and Cheng, X. (2013) Initial recognition of a cellodextrin chain in the cellulose-binding tunnel may affect cellobiohydrolase directional specificity. *Biophys. J.* **104**, 904-912

63. Bergenstrahle, M., Thormann, E., Nordgren, N., and Berglund, L. A. (2009) Force Pulling of Single Cellulose Chains at the Crystalline Cellulose-Liquid Interface: A Molecular Dynamics Study. *Langmuir* **25**, 4635-4642

64. Yan, B. X., Sun, Y. Q., and Gao, P. (1997) Intrinsic fluorescence in endoglucanase and cellobiohydrolase from Trichoderma pseudokiningii S-38: effects of pH, quenching agents, and ligand binding. *J Protein Chem* **16**, 681-688

65. Boer, H., and Koivula, A. (2003) The relationship between thermal stability and pH optimum studied with wild-type and mutant Trichoderma reesei cellobiohydrolase Cel7A. *European Journal of Biochemistry* **270**, 841-848

66. Zolotnitsky, G., Cogan, U., Adir, N., Solomon, V., Shoham, G., and Shoham, Y. (2004) Mapping glycoside hydrolase substrate subsites by isothermal titration calorimetry. *Proceedings of the National Academy of Sciences* **101**, 11275-11280

67. Borch, K., Jensen, K., Krogh, K., Mcbrayer, B., Westh, P., Kari, J., Olsen, J. P., Sørensen, T. H., Windahl, M. S., and Xu, H. (2014) New cellobiohydrolase variant having cellobiohydrolase activity useful for making transgenic plant, and for fermenting a cellulosic material for producing fermentation product e.g. ethanol, n-butanol, isobutanol, acids, alcohols.

68. Zhang, Y. H., Cui, J., Lynd, L. R., and Kuang, L. R. (2006) A transition from cellulose swelling to cellulose dissolution by o-phosphoric acid: evidence from enzymatic hydrolysis and supramolecular structure. *Biomacromolecules* **7**

**Table 1.** Parameters from titration- and time-course experiments.

| Enzyme	Titration exp.			Real-time exp.	Derived parameter	
	Signal* (%)	K_P (L/g)	Γ_{max} (μmol/g)	$^{mass}k_{on}$ (L/g s^{-1})	$^{molar}k_{on} \cdot 10^6$ (M^{-1} s^{-1})	$k_{off} \cdot 10^{-3}$ (s^{-1})
WT	100 ± 2	122 ± 33	6.4 ± 0.6	0.59 ± 0.01	0.093±0.009	4.8 ± 1.3
E212Q	109 ± 3	63 ± 12	6.6 ± 0.7	0.63 ± 0.03	0.094±0.011	9.9 ± 2.0
CD	92 ± 5	55 ± 25	2.0 ± 0.2	0.33 ± 0.02	0.163±0.016	5.9 ± 2.8
W38A	26 ± 2	25 ± 12	2.6 ± 0.7	0.41 ± 0.06	0.157±0.046	16.6 ± 8.1
W40A	72 ± 3	42 ± 13	4.2 ± 0.6	0.39 ± 0.01	0.094±0.013	9.4 ± 3.0
W367A	80 ± 2	53 ± 8	7.0 ± 0.6	0.59 ± 0.05	0.085±0.010	11.1 ± 1.8
W376A	53 ± 1	193 ± 57	9.0 ± 0.9	0.43 ± 0.06	0.048±0.008	2.2 ± 0.7
W367A+W376A	46 ± 1	99 ± 22	9.5 ± 1.4	0.26 ± 0.01	0.027±0.004	2.6 ± 0.6
W38A+W376A	15 ± 1	62 ± 28	7.2 ± 1.9	0.23 ± 0.02	0.032±0.009	3.7 ± 1.7

* $F_{max}/F_{max\_WT} \cdot 100$
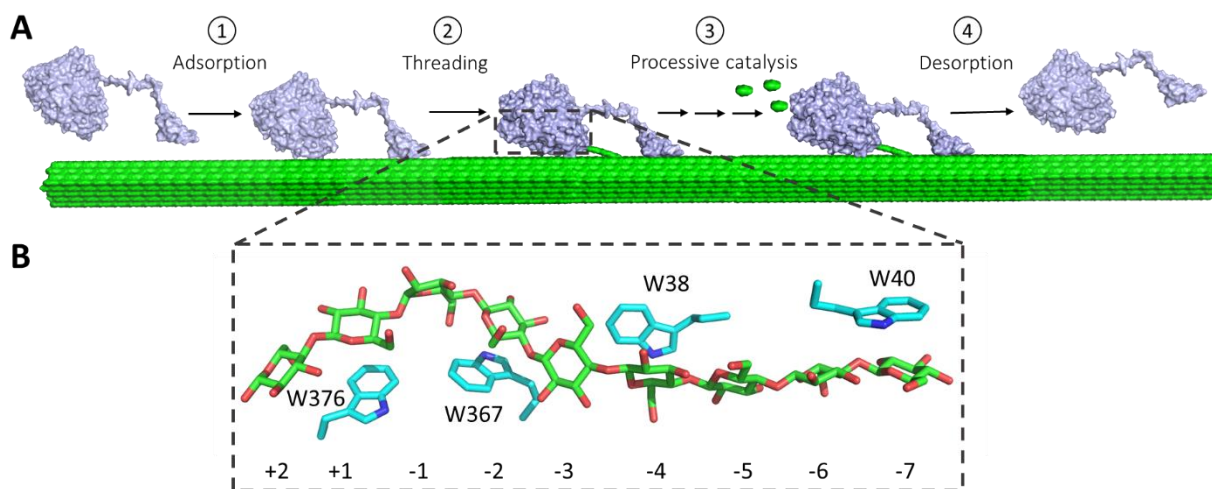
17

Figure. 1. Panel A shows a simplified illustration of the reaction mechanism of Cel7A. Panel B: The enlargement shows the structure of the bound cellononaose ligand (green/red sticks) and the location of the four conserved Trp residues. Numbers in the enlargement identify the subsites; the scissile bond is between subsites +1 and -1. PDB 4C4C.
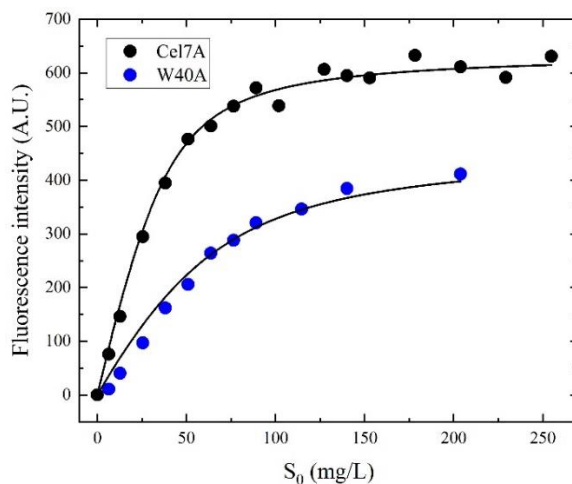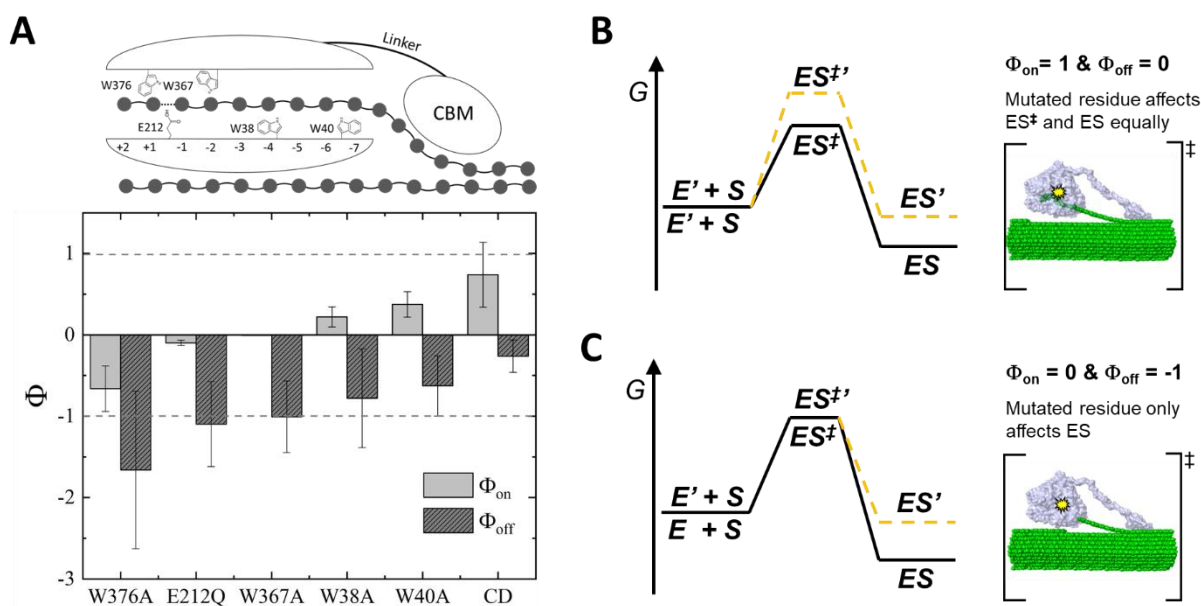

Figure 2. Examples of binding curves obtained from the titration experiments. Each point represents the apparent equilibrium fluorescence emission (in arbitrary units) at a given load of amorphous cellulose (RAC). Data is shown for the wild type Cel7A and the W40A mutant at starting concentrations of 250 nM enzyme. The lines show the best fits of eq. (5) to the experimental data.
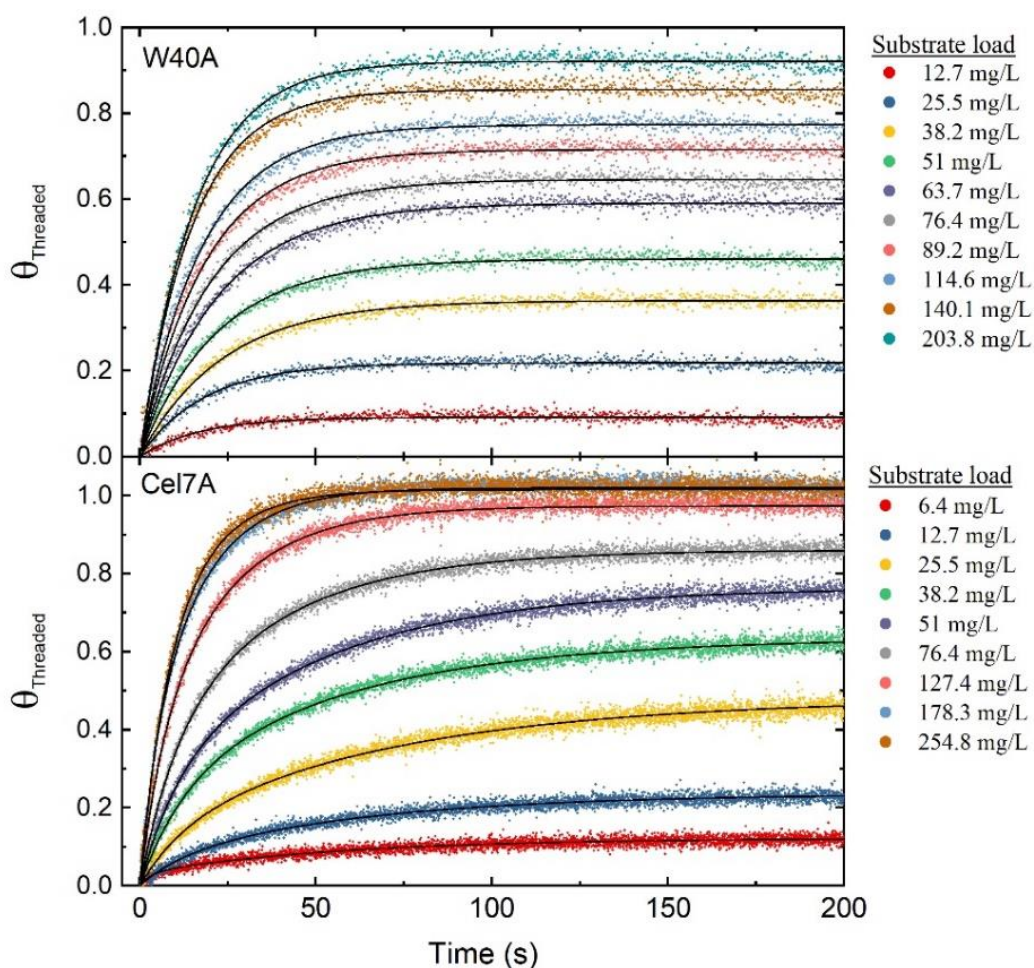
Figure 3. Real time fluorescence data for the complexation of RAC and respectively the Cel7A wild type and the W40A mutant (250 nM enzyme). The ordinate, θ, shows the fraction of enzyme molecules in an ES complex (*c.f.* eq. 1). Analogous data for the other enzymes are shown in Fig S3 of the SI. Black lines represent best fit to an exponential function.
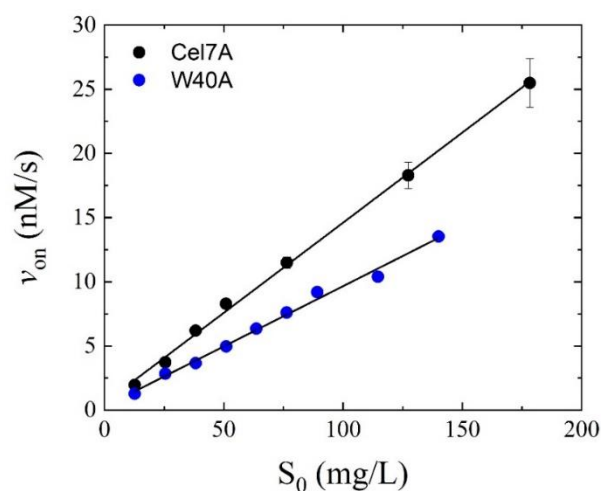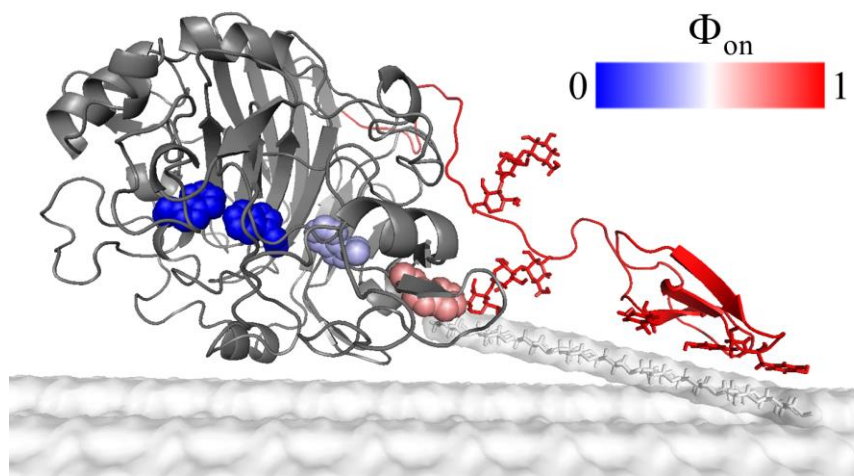
Figure 4. Initial rate of complexation, v_{on}, derived from Fig. 3 and plotted as a function of the RAC load for two enzymes, Cel7A$_{WT}$ and W40A. The observed proportionality supports the interpretation that the complexation is first order with respect to the load of cellulose, and the slope signifies the parameter $^{mass}k_{on}E_0$ (see main text).

Figure 5. Φ-factor analysis. Panel A shows experimental Φ-factors for respectively complexation (Φ_{on}) and decomplexation (Φ_{off}). The cartoon above the histogram indicates the location of the investigated mutations. Panels B and C illustrates the interpretation of Φ-factors on the basis of two hypothetical transition states for the complexation process. In panel B, the TS is late in the process just before the ligand fully fills the tunnel, while in C, it is early in the complexation process. If we mutate a residue located at the yellow star, which interacts with the ligand, the energy landscape changes differently in the two cases (dashed lines). Thus, in B the free energies of the bound state (ES) and the TS (ES$^{\neq}$) will shift in parallel because the mutated residue interacts with the ligand

in both states, and this yield $\Phi_{on} = 1$. Conversely, in C, only ES (and not ES$^{\neq}$) will be affected by the mutation and Φ_{on} will be 0. This difference allowed us to assess whether a given mutation was located before or after the reaction coordinate of the TS. An analogous analysis may be conducted for decomplexation. In this case, the Φ_{off}-factors that indicates whether the mutation is before or after the TS are 0 and -1 respectively.

Figure 6. Structural interpretation of the Φ-factor analysis. The transition state of complexation and decomplexation of Cel7A is proposed to occur at a stage, where the reducing end is near W40 while the rest of the tunnel is empty. The unfavorable free energy of this structure arises from the (uncompensated) detachment of several glucose moieties from the cellulose surface. At a higher reaction coordinate of the complexation process (when the tunnel fills) compensating ligand interactions occur in the tunnel and hence lower the free energy. At a lower reaction coordinate, the free energy becomes more favorable as the ligand interacts with the cellulose surface.

Substrate binding in the processive cellulase Cel7A: Transition state of complexation and roles of conserved tryptophan residues.

Nanna Røjel, Jeppe Kari, Trine Holst Sørensen, Silke F. Badino, J. Preben Morth, Kay Schaller, Ana Mafalda Cavaleiro, Kim Borch and Peter Westh

J. Biol. Chem. published online December 17, 2019

Access the most updated version of this article at doi: 10.1074/jbc.RA119.011420

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts