Thermodynamic signatures of substrate binding for three *Thermobifida fusca* cellulases with different modes of action

Anne Grethe Hamre[1], Anita Kaupang[1], Christina M. Payne[2], Priit Väljamäe[3], and Morten Sørlie[1*]

[1] *Department of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, PO 5003, N-1432 Ås, Norway*.

[2] *Department of Chemical and Materials Engineering, University of Kentucky, 177 F. Paul Anderson Tower, Lexington, KY, 40506, USA.*

[3] *Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia*

[*] To whom correspondence should be addressed: Morten Sørlie (morten.sorlie@nmbu.no), Tel.: +47-67232562 and Fax: +47-64965901

ABSTRACT

The enzymatic breakdown of recalcitrant polysaccharides is achieved by synergistic enzyme cocktails of glycoside hydrolases (GHs) and accessory enzymes. Many GHs are processive meaning that they stay bound to the substrate in between subsequent catalytic interactions. Cellulases are GHs that catalyze the hydrolysis of cellulose (β-1,4-linked glucose (Glc)). Here, we have determined the relative subsite binding affinity for a glucose moiety, as well as the thermodynamic signatures for $(Glc)_6$ binding to three of the seven cellulases produced by the bacteria *Thermobifida fusca. Tf*Cel48A is exo-processive, *Tf*Cel9A endo-processive, while *Tf*Cel5A is endo-nonprocessive. Initial hydrolysis of $(Glc)_5$ and $(Glc)_6$ was performed in $H_2{}^{18}O$ enabling the incorporation of an ^{18}O atom at the new reducing end anomeric carbon. An MALDI-TOF-MS analysis of the products reveal the intensity-ratios of otherwise identical ^{18}O- and ^{16}O-containing products to provide insight into how the substrate is placed during productive binding. The two processive cellulases have significant binding affinity in subsites where products dissociate during processive hydrolysis, aligned with a need to have a pushing potential to remove obstacles on the substrate. Moreover, we observed a correlation between processive ability and favorable binding free energy, as previously postulated. Upon ligand binding, the largest contribution to the binding free energy is desolvation for all three cellulases as determined by isothermal titration calorimetry. The two endo-active cellulases show a more favorable solvation entropy change than the exo-active cellulase, while the two processive cellulases have less favorable changes in binding enthalpy compared to the nonprocessive *Tf*Cel5A.

INTRODUCTION

Cellulose, a β-1,4-linked insoluble, linear polymer consisting of glucose (Glc) units, is the most abundant polysaccharide in nature. The glucose units are rotated 180° relative to each other; thus, the structural unit of cellulose is a dimer of Glc (cellobiose).[1] Native cellulose is cellulose made by living organisms and is found in two polymorphic variants, type Iα and Iβ. By alteration with strong alkali treatments, cellulose polymorphs of type II, III_I, III_{II}, IV_I, and IV_{II} can also be obtained.[2, 3] The Iβ variant is the most abundant form in nature and is also the most thermodynamically stable.[2, 4] The glucose units in cellulose form strong hydrogen bonds between adjacent chains in a cellulose sheet, preventing efficient hydrolysis of cellulose into fermentable glucose residues.[5-7]

The enzymatic hydrolysis of *O*-glycosidic linkages, the bonds joining two or more carbohydrates or a carbohydrate and a non-carbohydrate moiety, is generally acid-catalyzed by glycoside hydrolases (GHs). GHs use one of two different mechanisms where one yield retention of the anomeric carbon configuration and the other inversion.[8] Each enzyme has a customized mode of action to complement each other in the degradation of the recalcitrant polysaccharide architecture. Endo-acting GHs catalyze hydrolysis arbitrarily on the polymer chains. Exo-acting GHs catalyze hydrolysis starting from the reducing or the non-reducing end of the polymer. Both types of mechanisms can be performed in a processive manner. Then, the GH performs a series of catalysis before dissociating off the polymer.[8] Processive action by is not limited to carbohydrate active enzymes. Several other biopolymers such as DNA, RNA, and polypeptides are synthesized, modified, or degraded in a processive manner.[9] Examples include nucleic-acid polymerases, reverse transcriptases, and the 20S proteasome. All processive enzymes have in common that they stay bound to their substrates and perform multiple rounds of catalysis before dissociating. They achieve this by completely enclose their substrates or by using a large interaction surface. In GH catalysis, processivity has been linked

to the topology of the active site and the ligand binding free energy ($\Delta G_r°$).[8, 10] The topology of the active site is generally divided into three classes: pocket, cleft or tunnel.[8] Enzymes with the pocket topology are described by Davies and Henrissat to being optimal for saccharide non-reducing extremity and is found in monosaccharideases (i.e. β-galactosidase) and exopolysaccharidases (i.e. β-amylase), which are adapted for substrates with many available chain ends such as native starch granules.[8] Consequently, they are not efficient on fibrous substrates such as cellulose, which has almost no free chain ends. The cleft topology has a rather open structure allowing random binding of several units in polymeric substrates. This topology is mainly associated with endo-acting enzymes. The tunnel topology is characterized by long loops covering the active site being lined with aromatic amino acids providing strong binding interactions to the carbohydrate substrate. Thus, they are able to release products while remaining firmly bound to the substrate, increasing the probability of multiple catalytic events without dissociation from the substrate.[8]

Cellulases are GHs that catalyze the hydrolysis of cellulose into soluble sugars. They form a diverse collection of modular enzymes and are classified into families based on sequence similarities (www.cazy.org).[11] The Gram-positive, filamentous soil bacteria *Thermobifida fusca* (*Tf)* expresses a cellulolytic machinery responsible for the hydrolysis of cellulose consisting of 7 multi-modular cellulases from families 5, 6, 9, and 48, comprising *Tf*Cel5A, *Tf*Cel5B, *Tf*Cel6A, *Tf*Cel6B, *Tf*Cel9A, *Tf*Cel9B, and *Tf*Cel48A. All enzymes, except *Tf*Cel5B, possess a family 2 carbohydrate binding module (CBM), which aids in binding with crystalline substrates such as filter paper or bacterial microcrystalline cellulose (BMCC).[12] With regard to function, there are four endo-nonprocessive GHs (*Tf*Cel5A, *Tf*Cel5B, *Tf*Cel6A, and *Tf*Cel9B), two exo-processive GHs (*Tf*Cel6B and *Tf*Cel48A), and one endo-processive GH (*Tf*Cel9A).[13-15]

To better understand how GHs with different mode of action interact with the substrate, we have investigated the preferred positioning of cello-oligosaccharides during initial (less than 20 % substrate conversion) productive binding as analyzed by mass spectrometry and determined the thermodynamic signatures of binding binding to cellohexaose $(Glc)_6$ using isothermal titration calorimetry (ITC) for three *T. fusca* cellulases representing each possible mode: *Tf*Cel5A, *Tf*Cel9B, and *Tf*Cel48A. The initial hydrolysis of cello-oligosaccharides allows the determination of which subsites in the active site that provides strong interactions with substrate during the catalysis of hydrolysis. Thermodynamics signatures of binding are in turn a function of the geometry, dynamics, and chemical composition of the substrate tunnels or clefts. *Tf*Cel5A is a 46 kDa endoglucanase with some cellobiosidase activity.[15, 16] In the N-terminus it possesses a type B CBM,[17] and in line with its mode of action, the active site is an open, shallow cleft (PDB ID: 2cks).[15] According to the CAZy-database, family 5 GHs perform catalysis with retention of the anomeric configuration.[11, 18] *Tf*Cel9A is a unique cellulase in that it has characteristics of both exo- and endocellulases.[13, 19] Another striking feature is that it cleaves off cellotetraose units when it acts in the exo-mode.[20 21] *Tf*Cel9A is a 90.4 kDa secreted protein which contains four domains: an N-terminal family 9 catalytic domain, a CBM3c, a fibronectin III (FnIII) like domain, and a C-terminal CBM2. The catalytic domain has a shallow, open active site cleft with six glucose binding sites, subsites –4 to +2.[20] Here, the subsite terminology follows that of Davies *et al*. where subsites are labelled from $-n$ to $+n$ (where n is an integer). $-n$ represents the non-reducing end and $+n$ the reducing end, with cleavage taking place between the –1 and +1 subsites.[22] As in all family 9 members, *Tf*Cel9A hydrolyzes the *O*-glycosidic linkage with inversion of the anomeric configuration.[11, 18] *Tf*Cel48A is a 104 kDa exo-processive cellulase, attacking from the reducing end.[19] From the N-terminus, it possesses a type B CBM, followed by a FnIII-like domain, and ending with a family 48 catalytic domain having an active site with tunnel topology.[23, 24] Family 48 cellulases are quite similar, having

almost identical active site tunnels that can accommodate seven glucose units from subsite –7 to –1, followed by an open product binding site best adapted for cellobiose.[24] Family 48 cellulases also perform catalysis with inversion of the anomeric configuration.[11, 18]

By combining ITC and (MS) to determine the thermodynamic signatures and subsite preference, we show that there are significant dissimilarities between the three enzymes that to correlate to their mode of actions. This shows that enzymes with the same substrate are designed to act differently on this substrate and have subtle variances in their active site architectures that include differences in subsite affinity for substrate, solvation characteristics, and conformational dynamics.

MATERIALS AND METHODS

Chemicals. Cello-oligosaccharides were obtained from Megazyme (Wicklow, Ireland). All other chemicals were of analytical grade and purchased from standard manufacturers.

Enzymes. Initial hydrolysis of cellotetraose $(Glc)_4$, cellopentaose $(Glc)_5$, and $(Glc)_6$ was conducted using *Tf*Cel5A-WT, *Tf*Cel9A-WT, and *Tf*Cel48A-WT. In the isothermal titration calorimetry (ITC) experiments, variants where the catalytic acid is mutated to alanine, a mutation that inactivates the enzyme, were used.

Cloning. Three gene constructs were ordered from Genscript; *Tf*Cel5A-WT (Uniprot ID: Q47RH8), *Tf*Cel9A-WT (Uniprot ID: Q47MW0), and *Tf*Cel48A-E359A. For *Tf*Cel48A-E359A, the glutamic acid residue in position 359 in the wild type gene (Uniprot ID: Q47NH7) was exchanged with an alanine. The signal peptides were removed from the genes encoding the enzymes before the genes were codon optimized and cloned into the pET-22b(+)-vector by GenScript (Piscataway, NJ, USA). The received plasmids were transformed into *Escherichia coli* BL21Star (DE3) cells as described by the manufacturer (Life Technologies, Carlsbad, CA, USA).

Site directed mutagenesis. *Tf*Cel5A-E356A, *Tf*Cel9A-E425A, and *Tf*Cel48A-E359A (WT) were prepared using the QuikChange™ site directed mutagenesis kit from Stratagene (La Jolla, CA, USA), as described by the manufacturer. The Pellet Paint Coprecipitant kit ((Novagen, Madison, Wisconsin, USA) was utilized following the product protocol. The DNA templates and primers used for the mutagenesis (Table 1) were purchased from Life Technologies. To confirm that the respective genes contained the desired mutations and to check for the occurrence of non-desirable mutations, the mutated genes were sequenced using GATC Biotech's (Constance, Germany) LIGHTrun sequencing service before they were transformed into *E. coli* BL21Star (DE3) cells (Life Technologies).

Table 1. DNA-templates and primers used for site directed mutagenesis

Mutant	DNA template	Primer	Sequence
*Tf*Cel5A-E356A	*Tf*Cel5A-WT	Cel5A-E356A-fw	5'-CCGGTGTTTGTGACCGCATTTGGCACCGAAACC-3'
		Cel5A-E356A-rev	5'-GGTTTCGGTGCCAAATGCGGTCACAAACACCGG-3'
*Tf*Cel9A-E425A	*Tf*Cel9A-WT	Cel9A-E425A-fw	5'-GGATTATGTGGCGAACGCAGTGGCGACCGATTATA-3'
		Cel9A-E425A-rev	5'-TATAATCGGTCGCCACTGCGTTCGCCACATAATCC-3'
*Tf*Cel48A-A359E (WT)	*Tf*Cel48A-E359A	Cel48A-A359E-fw	5'-GATAATAGCTAAACGCTTCGCTGGTGGTCTGATGG-3'
		Cel48A-A359E-rev	5'-CCATCAGACCACCAGCGAAGCGTTTAGCTATTATC-3'

Protein production. For protein production, *E. coli* BL21(DE3) cells containing the appropriate plasmid were inoculated into 50 mL Luria-Bertani (LB) medium containing 100 µg/mL ampicillin. Then, the cells were allowed to grow at a temperature of 37 °C with shaking 200 rpm for a period of 18-20 hours. After inoculation of the cell culture into 500 mL Terrific Broth (TB) medium containing 100 µg/mL ampicillin, an OD_{600} of 0.1 was reached. Then, the culture was allowed to grow until and of OD_{600} 0.6-0.8 was attained followed by a change in temperature to 20 °C. Gene expression was induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) over an period of 20 hours. Then, the cells were harvested by

centrifugation (8000 rpm, 20 min at 4 °C). Cell pellets were resuspended in lysis buffer (50 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.3 mg/mL lysozyme chloride, and 1 protease inhibitor cocktail tablet per 60 mL solution (Roche, Basel, Switzerland)) before 1 h incubation at 30 °C. The cells were then lysed by sonication for 4 min with 5 s pulses, using a VibraCell Ultrasonic Liquid Processor VCX500/VCX750 (Sonics, Newtown, Connecticut, USA). Cell debris was removed by centrifugation (8000 rpm, 30 min at 4 °C). The supernatant was collected, and the volume measured (normally approximately 150 mL). Then, normally 15 mL (10 % of total of the supernatant volume) of a streptomycin sulphate solution (adjusted to pH 7 with 2.5 % NH_3) was added dropwise to the supernatant with careful stirring over a period of 5 minutes before the solution was incubated for 10 minutes at room temperature. The solution was then centrifuged for 20 minutes at 4 °C and 8000 rpm. The supernatant was sterilized by filtration (0.2 μm) and stored at 4 °C prior to purification.

Protein purification. *Tf*Cel5A was purified using a three-step protocol starting with heat treatment at 50 °C for 1 h. The *T. fusca* cellulases are heat stable; thus, this step effectively removes less stable *E. coli* proteins. Cell debris was removed with centrifugation (8000 rpm, 10 min at 4 °C), and the supernatant was sterilized by filtration (0.2 μm) and stored at 4 °C prior to the next purification step – ion exchange chromatography. The supernatant was adjusted to pH 8.0 and loaded onto a 5 mL HiTrap Q HP column (GE Healthcare, Chicago, Illinois, USA) connected to a BioLogic low-pressure protein purification system (Bio-Rad, Hercules, California, USA). Cellulases were eluted by applying a linear salt gradient from 0 % buffer A (50 mM Tris-HCl pH 8.0) to 60 % buffer B (50 mM Tris-HCl pH 8.0, 1 M NaCl) over 20 column volumes at a flow rate of 4 mL/min. *Tf*Cel5A, *Tf*Cel9A, and *Tf*Cel48A elute at approximately 30 %, 45 %, and 45 %, respectively. The cellulase-containing fractions were pooled before the third purification step – hydrophobic interaction chromatography. The pooled solution was altered to buffer A (50 mM Tris-HCl pH 8.0, 1 M $(NH_4)_2SO_4$) with 3M $(NH_4)_2SO_4$

and applied to a 5 mL HiTrap Phenyl HP column (GE Healthcare) in connection with a BioLogic low-pressure protein purification system from Bio-Rad. Cellulases were eluted by a two-step gradient. The first step applied was a linear salt gradient from 0 % buffer A to 70 % buffer B (50 mM Tris-HCl pH 8.0) for a total of 10 column volumes with a flow rate of 4 mL/min. The second step applied consisted of a linear salt gradient ranging from 70 to 100 % buffer B with a total of 15 column volumes and a flow rate of 4 mL/min. The, 5 column volumes of buffer B was added. All three cellulases eluted at 100 % buffer B. For *Tf*Cel9A and *Tf*Cel48A, a fourth step, size exclusion chromatography, in addition to the three previous steps, was necessary to get pure protein. The cellulase containing fractions were joined and concentrated to a final volume of 1 milliliter by the use of a Macrosep® Advances Centrifugal Devices (PALL Corporation, New York, USA) holding a molecular mass cutoff of 30000 Da. Next, the samples were applied to a HiLoad 16/600 Superdex 75 Prepgrade column (GE Healthcare). Here, a solution of 20 mM Tris-HCl pH 8.0 and 0.2 M NaCl was used as a running buffer with a flow rate of 1 mL/min. The cellulases eluted circa 70 and 50 min after injection for *Tf*Cel9A and *Tf*Cel48A, respectively.

Protein purity was analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis after each purification step. After the last step, fractions containing pure protein were pooled and concentrated using Macrosep® Advances Centrifugal Devices, followed by a buffer change to 20 mM potassium phosphate pH 6.0 by dialysis (SnakeSkin® Pleated Dialysis, 10000 MVCO from Thermo-Scientific). Protein concentrations were determined by absorbance at 280 nm, using the theoretical extinction coefficients (94225 $M^{-1}cm^{-1}$, 214435 $M^{-1}cm^{-1}$, and 259925 $M^{-1}cm^{-1}$ for *Tf*Cel5A, *Tf*Cel9A, and *Tf*Cel48A, respectively) (http://web.expasy.org/protparam/).

Initial hydrolysis of $(Glc)_4$, $(Glc)_5$, and $(Glc)_6$. To determine the preferred positioning of the substrate in the active site, initial hydrolysis of $(Glc)_5$ and $(Glc)_6$ in $H_2{}^{18}O$ (Larodan Fine

Chemicals, Malmö, Sweden) was performed, as described previously.[25-28] Initial hydrolysis of $(Glc)_4$ was performed in $H_2{}^{16}O$. The hydrolysis was performed at 37 °C and 600 rpm in $H_2{}^{16}O/H_2{}^{18}O$ containing 20 mM ammonium acetate (pH 6.1), 100 µm $(Glc)_x$, and various amounts of enzyme, depending on the oligosaccharide and cellulase in question (Table 2). Reactions were set up with such enzyme concentrations as to reach approximately 20 % substrate conversion within a short time interval in order to avoid non-enzymatic incorporation of ^{18}O, as described by Eide *et. al.* 2013.[25] To prevent incorporation of $H_2{}^{16}O$, the stock solutions containing GHs were very concentrated to ensure that added volume were under 2 % of the total assay volume.[27] Aliquots of 1 µL were withdrawn from the assay solution at different time intervals, immediately quenched by mixing with 1 µL of a 2,5 dihydroxybenzoic acid (DHB) solution (15 mg/mL DHB in 30 % ethanol), spotted directly on the target plate for an matrix assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF-MS) experiment and dried. The hydrolysis products were analyzed as described by Eide *et al*.[25]

Table 2. Enzyme concentrations used in the initial hydrolysis of $(Glc)_4$ in $H_2{}^{16}O$ and $(Glc)_5$ and $(Glc)_6$ in $H_2{}^{18}O$.

Enzyme	$[(Glc)_4]$	$[(Glc)_5]$	$[(Glc)_6]$
*Tf*Cel5A-WT	50 nM	25 nM	30 nM
*Tf*Cel9A-WT	5 µM	25 nM	25 nM
*Tf*Cel48A-WT	2 µM	2 µM	2 µM

Isothermal Titration Calorimetry Experiments. ITC experiments were performed with a VP-ITC system from Microcal, Inc. (Northampton, MA, USA).[29] Solutions were thoroughly degassed prior to experiments to avoid air bubbles in the calorimeter. Standard ITC

conditions were 500 μM of cellohexaose in the syringe and 15 μM of enzyme in the reaction cell in 20 mM potassium phosphate buffer pH 6.0. For *Tf*Cel5A-E356A, 1 mM $(Glc)_6$ and 8 μM of enzyme were used. During the experiments, 8 μL of the titrant were added into the reaction cell at 180 s intervals. To obtain a temperature dependence of the reaction enthalpy change ($\Delta H_r°$) of the reaction, this was determined at *t* of 20, 25, 30, and 37 °C. The stirring speed was set to be 260 rpm. 60 injections were performed in total. A minimum of two independent titrations was undertaken for each binding reaction. The buffer ionization heat is 1.22 kcal/mol.[30]

Analysis of calorimetric data. ITC data were collected by the Microcal Origin v.7.0 software associated with the VP-ITC system.[29] All data were corrected for heat of dilution by subtracting the heat remaining after saturation of binding sites on the enzyme prior to further data analysis. For the fit of ITC-data, a non-linear least-squares algorithm and a single-site binding model in the Origin software was employed. The binding reaction data followed a single-site binding model. From the fits, the stoichiometry (*n*) of the reaction, equilibrium binding association constant (K_a), and the $\Delta H_r°$ of the reaction were derived directly. Typically, the determined value of *n* was between 0.9 and 1.1 for each independent reaction. The equilibrium binding dissociation constant (K_d), reaction free energy change ($\Delta G_r°$) and the reaction entropy change ($\Delta S_r°$) were then calculated from the relations depicted in Equation 1.

$$\Delta G_r° = -RT\ln K_a = RT\ln K_d = \Delta H_r° - T\Delta S_r° \quad (1)$$

Errors are reported as standard deviations of at least two experiments at each temperature. A description of how the entropic term is parameterized has been described in detail previously.[31, 32]

RESULTS

Initial hydrolysis of $(Glc)_4$, $(Glc)_5$, and $(Glc)_6$. Initial hydrolysis of $(Glc)_4$ with all three cellulases was performed in $H_2{}^{16}O$ and yielded $(Glc)_2$ as the clear main product. This suggests that the simultaneous binding of a sugar moiety in subsites –2 and +2 are more important than binding to either the –3 or the +3 subsites.

To further investigate the contributions of other subsites involved in substrate binding, initial hydrolysis of $(Glc)_5$ and $(Glc)_6$ was performed in $H_2{}^{18}O$. Then the products were analyzed by MALDI-TOF-MS (Figure 1).[25, 27] Using this approach, an ^{18}O atom instead of an ^{16}O are bound to the anomeric carbon after hydrolysis. The sugar moiety that has the ^{18}O will be the sugar that interacted with subsite ─1 during productive binding. So, hydrolysis of a $(Glc)_5$ molecule bound in subsites –3 to +2 will generate a $(Glc)_3$ molecule with an ^{18}OH and a $(Glc)_2$ molecule with ^{16}OH at the reducing end, respectively. When a $(Glc)_5$ molecule bound to subsites –2 to +3 will generate a $(Glc)_3$ moelcule with a ^{16}OH and a $(Glc)_2$ moelcule with an ^{18}OH at the reducing end, respectively. To make sure initial turnover, the analysis of products analysis took place at circa 20 % substrate conversion. Furthermore, the handling of samples were swift to avoid non-enzymatic incorporation of ^{18}O because of the natural occurring anomeric equilibrium.[27] Therefore the intensity-ratios of otherwise identical ^{18}O-(heavy) and ^{16}O-containing (light) products in the MALDI-TOF MS spectra is equal to the ratio of the individual subsite occupancies and thus provide insight into how the substrate was bound while being cleaved. Using this heavy-oxygen water approach, $(Glc)_5$ was incubated with exo-processive *Tf*Cel48A. The results suggested that it has a $(Glc)_5$ occupancy of 70 ± 2 % in the –3 to +2 subsites (Figure 2). A reasonable conclusion to draw from this result is that the binding affinity is higher in subsite –3 compared to subsite +3. Still, there is affinity in the latter subsite (+3), as seen by the formation of $(Glc)_3$ as major products from productive binding of $(Glc)_6$. Here, some tetramers containing ^{18}O were detected as a minor product

originating from –4 to +2 binding. A direct comparison of the various binding modes (i.e. ─3 to +3 vs. ─4 to +2) is not possible since the MS is not fully quantitative. However, the data show a clear dominance for –3 to +3 binding compared to –4 to +2 binding, indicating that the +3 subsite has a higher affinity for a sugar moiety than the –4 subsite.

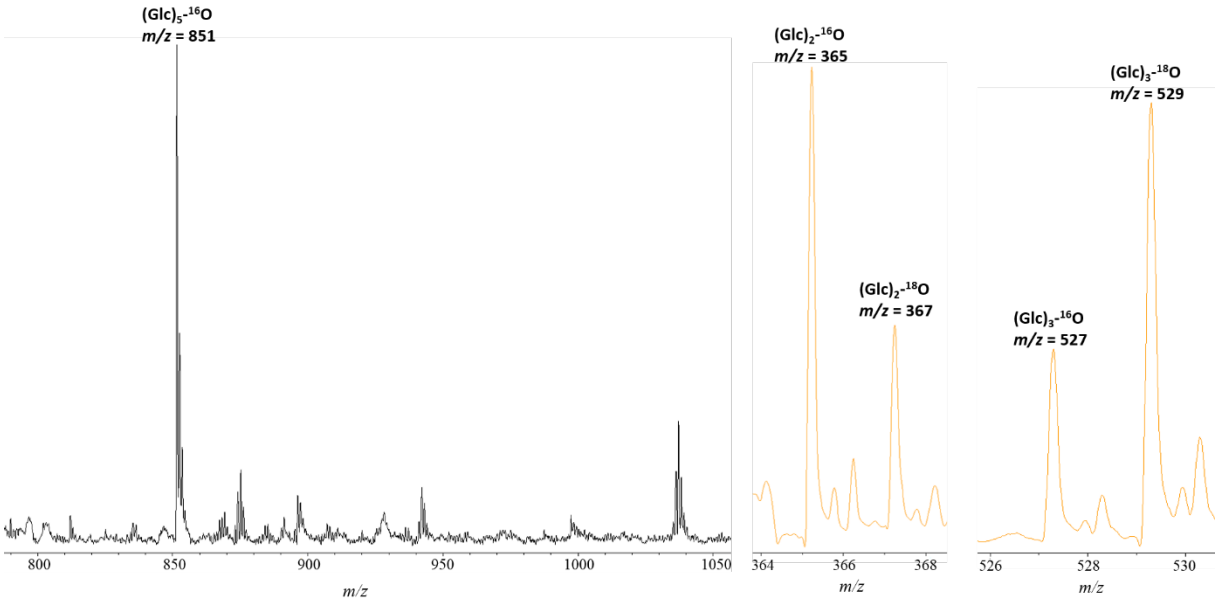

Figure 1. An example of a spectra from the MALDI-TOF-MS analyses of $(Glc)_5$ degradation with *Tf*Cel48A. The left panel (black) depicts the substrate before hydrolysis and the right panel depicts the products included with ^{16}O and ^{18}O (orange). The peaks labeled are sodium adducts of the respective oligomers.

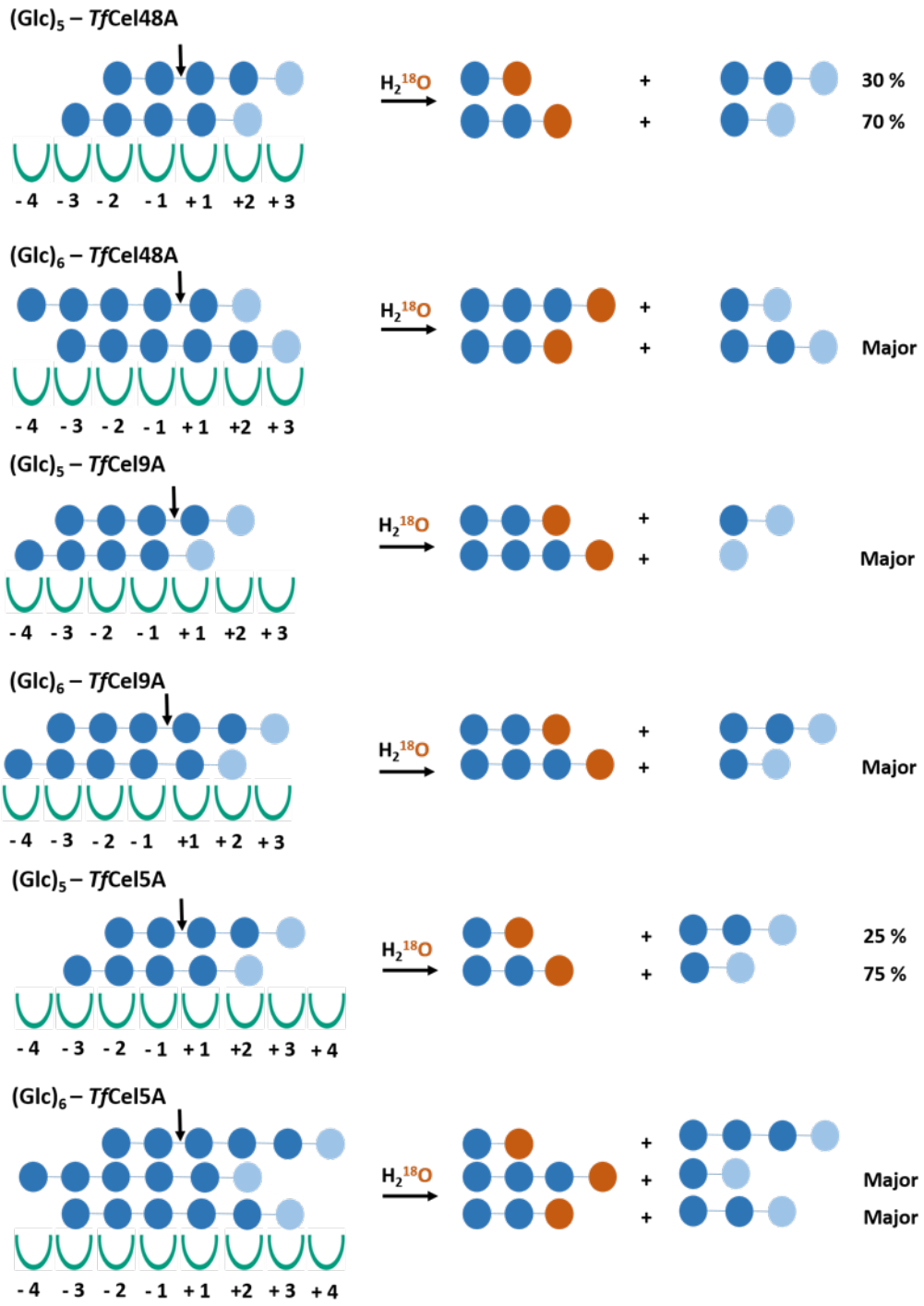

Figure 2. Productive binding modes of *Tf*Cel48A, *Tf*Cel9A, and *Tf*Cel5A. Hydrolysis of $(Glc)_5$ and $(Glc)_6$ resuling from experiments where $H_2{}^{18}O$ are present are shown. Suggested productive binding modes and their occurrences were obtained from the intensities of the signals (peak hights) in the MALDI-TOF-MS spectra. Light blue and orange circles represent the reducing end and the new reducing end with the ^{18}O atom bound to the anomeric carbon.

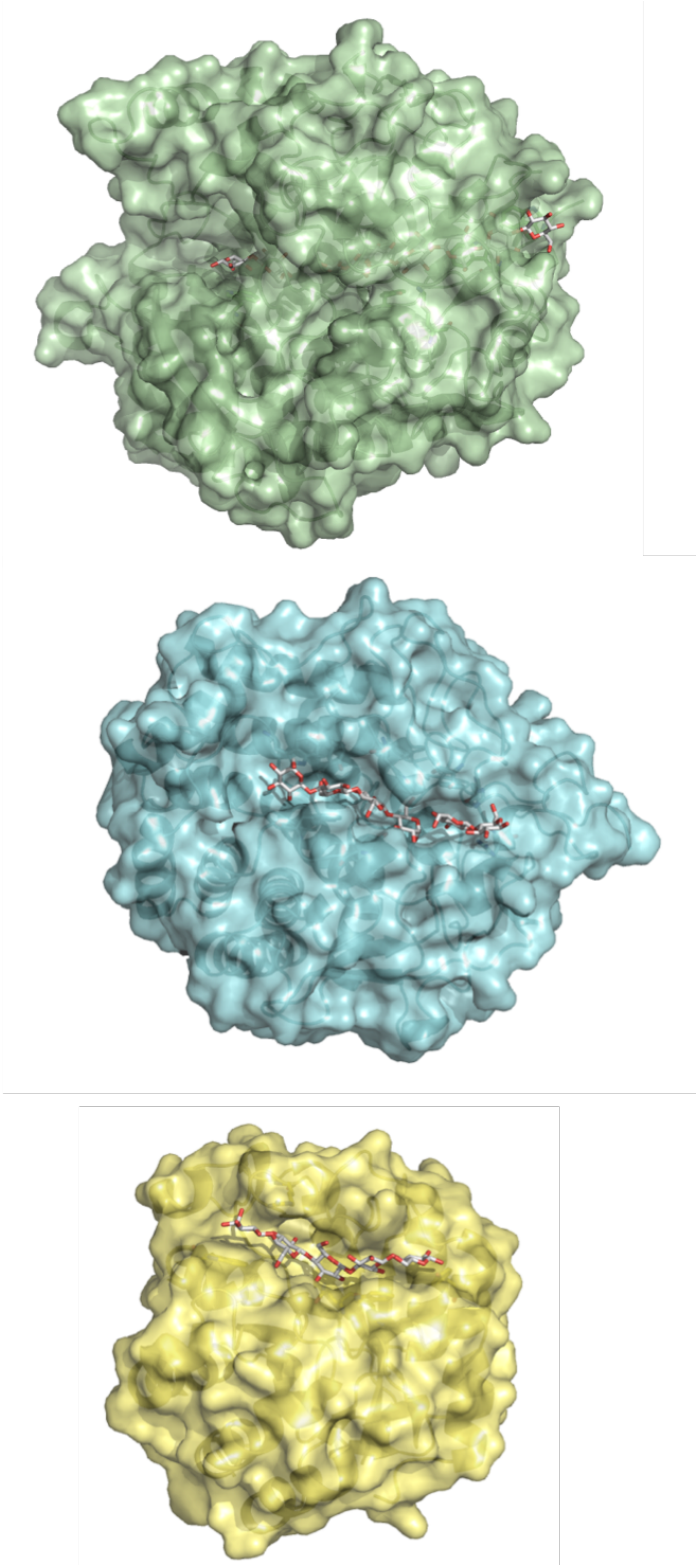

Figure 3. Crystal structures of the catalytic domains of *Tf*Cel48A[24] (pdb 4jjj, top), *Tf*Cel9A[20] (pdb 4tf4, middle), and *Tf*Cel5A (pdb 2ckr, bottom) in the presence of substrates (a $(Glc)_6$ and $(Glc)_2$ for *Tf*Cel48A, $(Glc)_4$ and $(Glc)_2$ for *Tf*Cel9A, and $(Glc)_5$ for *Tf*Cel5A).

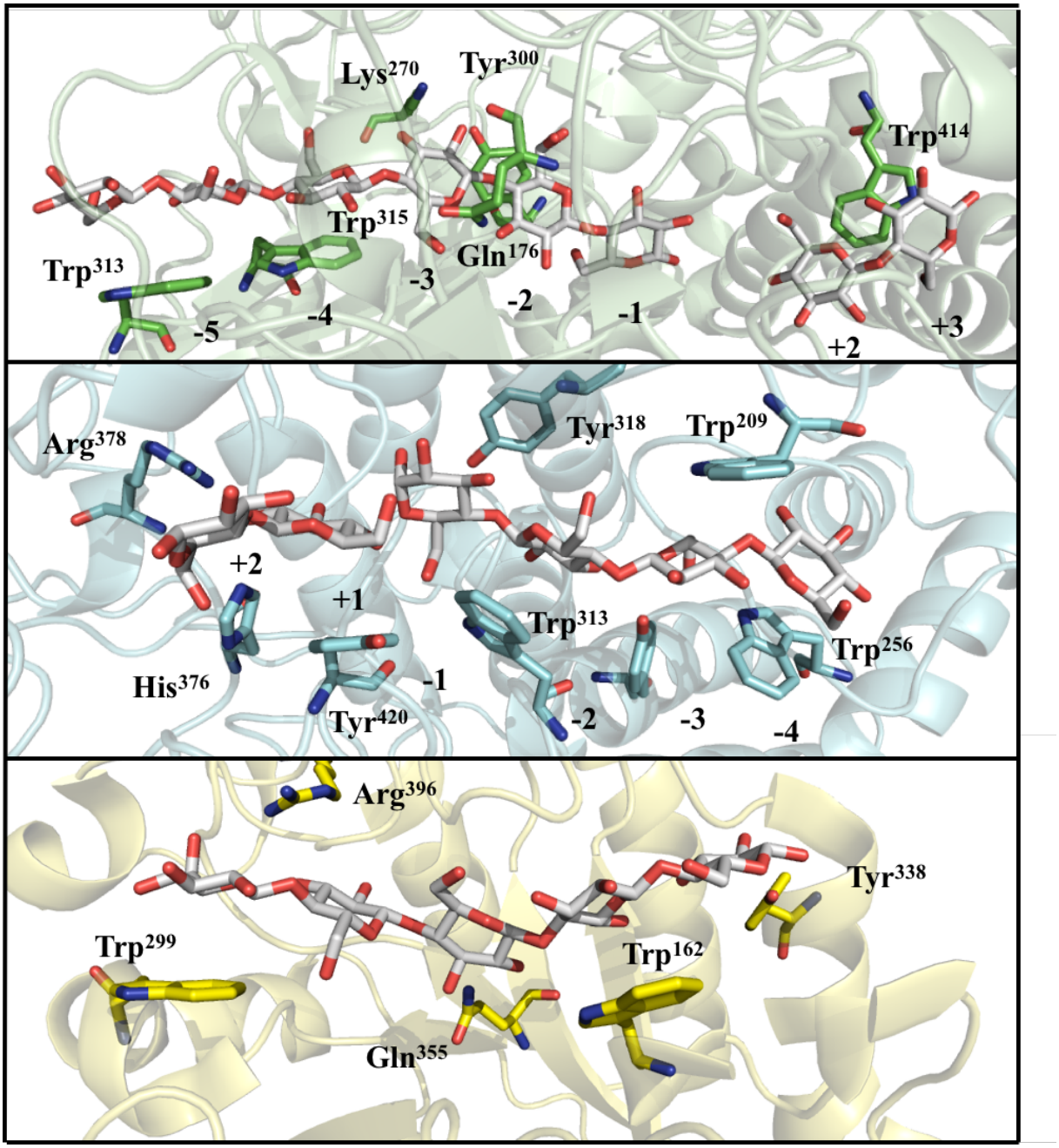

Figure 4. Crystal structures of the ligand-bound active sites of *Tf*Cel48A[24] (pdb 4jjj, green cartoon, top), *Tf*Cel9A[20] (pdb 4tf4, blue cartoon, middle), and *Tf*Cel5A (pdb 2ckr, yellow cartoon, bottom) in the presence of substrates. Key residues thought to participate in substrate binding are shown in stick representation and are labeled with sequence position numbers. The cello-oligomers are shown in stick representation wit gray carbon and red oxygen atoms. Binding subsites along the active site are labeled as positive and negative values (except for *Tf*Cel5A as detailed in the Discussion).

For endo-processive *Tf*Cel9A, initial hydrolysis of $(Glc)_5$ yielded a dominating peak in the mass spectrum equivalent to binding from subsite –4 to +1 (Figure 2). Some binding from subsite –3 to +2 was also observed. The main product after hydrolysis of $(Glc)_6$ was also a tetramer containing ^{18}O, being equivalent to binding from subsite –4 to +2.

Initial productive binding of $(Glc)_5$ for the endo-nonprocessive *Tf*Cel5A yielded $(Glc)_2$ and $(Glc)_3$ as the only products detected with a dominant occupancy from the –3 to +2 subsites (Figure 2). Moreover, $(Glc)_6$ hydrolysis, resulted in equal occupancies of productive binding from –4 to +2 and –3 to +3.

Binding of $(Glc)_6$ to *Tf*Cel48A, *Tf*Cel9A, and *Tf*Cel5A. The binding of cellohexaose to *Tf*Cel48A-E359A, *Tf*Cel9A-E425A, and *Tf*Cel5A-E356A at pH 6.0 (20 mM potassium phosphate buffer) at temperatures of 20, 25, 30, and – 37 °C was investigated using ITC. In Figure 5, a representative ITC thermogram and its theoretical fit to the data obtained in the experiment is depicted at t =30 °C. For simplicity, we will refer to the enzymes by their wild type name in the following text. At 30 °C, *Tf*Cel48A, binds $(Glc)_6$ with a K_d = 0.67 ± 0.2 μM ($\Delta G_r°$ = – 8.6 ± 0.2 kcal/mol). The enthalpic change of the reaction ($\Delta H_r°$) is –2.7 ± 0.3 kcal/mol while the entropic change of the reaction ($\Delta S_r°$) is 19.5 ± 1.3 cal/K·mol ($-T\Delta S_r°$ = –5.9 ± 0.4 kcal/mol). The change in the heat of the reaction, as determined by Equation 2, was determined to be –176 ± 25 cal/K·mol.

$$\Delta C_{p,r}^{o} = \left(\frac{\partial \Delta H_r^{o}}{\partial T} \right) \qquad (2)$$

For *Tf*Cel9A, the binding to $(Glc)_6$ has a K_d = 0.56 ± 0.16 μM which corresponds to a $\Delta G_r°$ = – 8.7 ± 0.2 kcal/mol. $\Delta H_r°$of the reaction is –1.1 ± 0.1 kcal/mol and $\Delta S_r°$ is 25.1 ± 0.7 cal/K·mol, corresponding to a $-T\Delta S_r°$ = –7.6 ± 0.2 kcal/mol. $\Delta C_{p,r}°$ was found to be –239 ± 21 cal/K·mol. For *Tf*Cel5A, the binding to $(Glc)_6$ has a K_d = 26 ± 2 μM ($\Delta G_r°$ = – 6.4 ± 0.0 kcal/mol). $\Delta H_r°$ was determined to be – 6.4 ± 0.2 kcal/mol and an $\Delta S_r°$ of 0.0 ± 0.6 cal/K·mol ($-T\Delta S_r°$ = 0.0 ±

0.2 kcal/mol). $\Delta C_{p,r}°$ was found to be –209 ± 17 cal/K mol. The results are summarized in Table 3.

Parameterization of the entropic term. Having access to $\Delta C_{p,r}$ for the binding reactions allows for a parameterization of the entropic term, which provides valuable information of the binding mechanism. The entropic term, $\Delta S_r°$, can be viewed as the sum of translational ($\Delta S°_{mix}$), solvation ($\Delta S°_{solv}$), and conformational ($\Delta S°_{conf}$) entropic changes as seen in Equation 6.[33]

$$\Delta S_r° = \Delta S°_{mix} + \Delta S°_{solv} + \Delta S°_{conf} \quad (6)$$

It has been observed that the change in entropy due to solvation is close to zero for proteins near T = 385 K. Thus, having access to an experimentally derived $\Delta C_{p,r}$ allows for the estimation solvation entropy change ($\Delta S°_{solv}$) of the binding reaction at t = 30 °C as depicted by Equation 7.[33-35]

$$\Delta S^{o}_{solv} = \Delta C^{o}_{p,r} \ln\left(\frac{303\,\mathrm{K}}{385\,\mathrm{K}}\right) \quad (7)$$

Using this relationship, a $\Delta S_{solv}°$ of 42.2 ± 5.9, 57.1 ± 5.3, and 50.6 ± 4.3 cal/K·mol can be calculated for the binding reaction between *Tf*Cel48A, *Tf*Cel9A, *Tf*Cel5A and $(Glc)_6$, respectively. These numbers represent –12.8 ± 1.8 kcal/mol, –17.3 ± 1.6, and –15.2 ± 1.3 ($-T\Delta S_{solv}°$) of the total free energy change, respectively. Moreover, the translational entropy change ($\Delta S°_{mix}$) of a binding reaction of a ligand to a protein has been proposed to be calculated as a "cratic" term, a statistical correction taking into account the mixing of solute and solvent molecules and the changes in entropy associated with translational/rotational degrees of freedom (Equation 8).[33]

$$\Delta S^{o}_{mix} = R \ln\left(\frac{1}{55.5}\right) \quad (8)$$

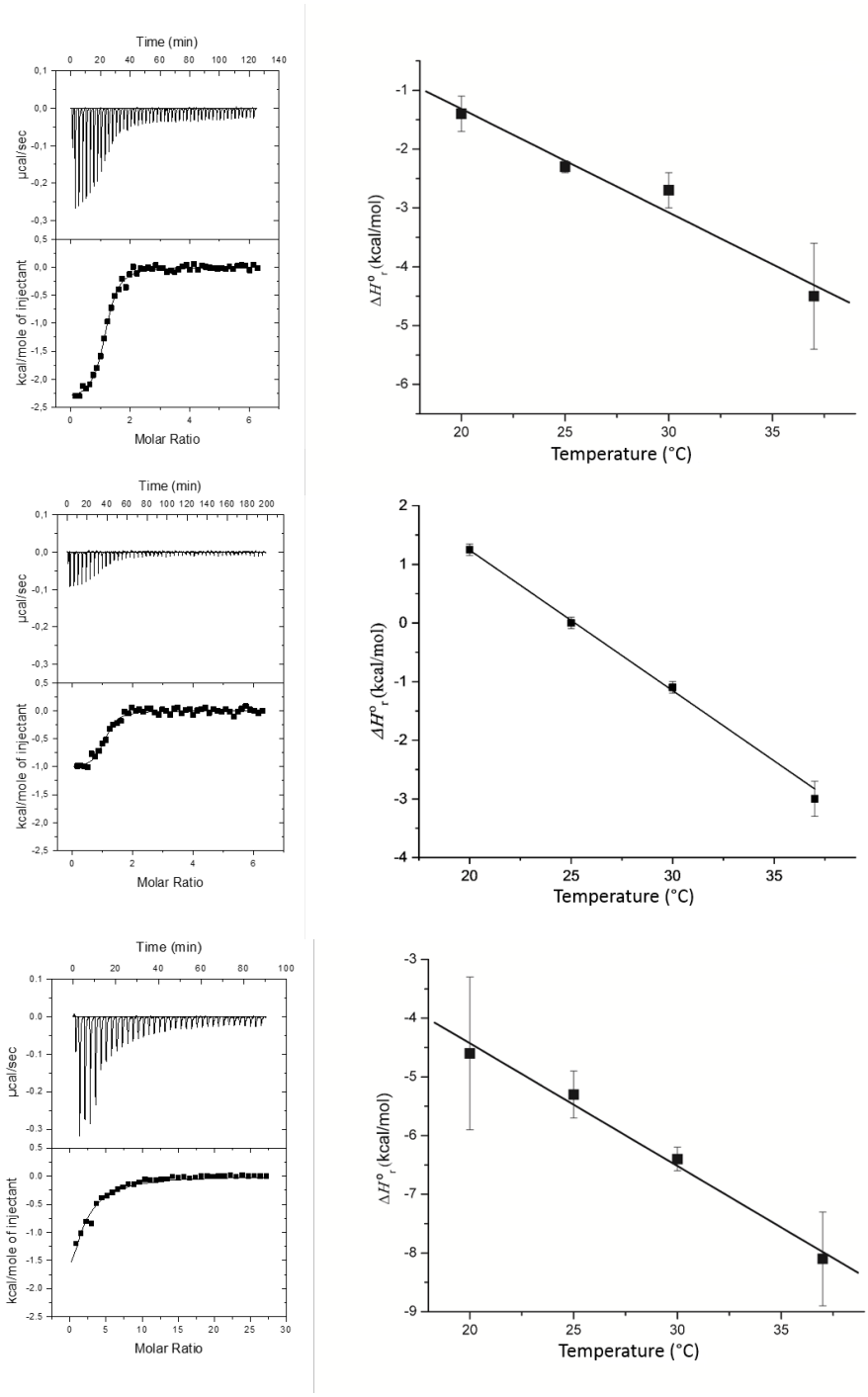

Figure 5. Left panel: Representative thermograms (top) and fitted theoretical data (bottom) for $(Glc)_6$ binding to *Tf*Cel48A (upper), *Tf*Cel9A (middle), and *Tf*Cel5A (bottom) at t = 30 °C in 20 mM potassium phosphate at pH 6.0. Right panel: Temperature dependence of $(Glc)_6$ to *Tf*Cel48A, *Tf*Cel9A, and *Tf*Cel5A. The value of $\Delta C_{p,r}$ is –176 cal/K·mol, –239 cal/K·mol, and –209 cal/K·mol for *Tf*Cel48A, *Tf*Cel9A, and *Tf*Cel5A, respectively.

Table 3. P^{app} and thermodynamic parameters for $(Glc)_6$ binding to *Tf*Cel5A, *Tf*Cel9A, T*f*Cel48A, and *Tr*Cel7A at t = 30 °C, pH = 6.0.

Enzyme	P^{app} [a,b]	K_d [c]	ΔG_r° [d]	ΔH_r° [d]	$-T\Delta S_r^{\circ}$ [d]	$-T\Delta S_{solv}^{\circ}$ [d,f]	$-T\Delta S_{conf}^{\circ}$ [d]	$-T\Delta S_{mix}^{\circ}$ [d]	$\Delta C_{p,r}^{\circ}$ [e, g]
*Tf*Cel5A[i]	2.2	26 ± 2	–6.4 ± 0.1	–6.4 ± 0.2	−0.0 ± 0.2	−15.2 ± 1.3	12.8 ± 1.3	2.4	–209 ± 17
*Tf*Cel9A[i]	7.0	0.56 ± 0.16	–8.7 ± 0.2	–1.1 ± 0.1	–7.6 ± 0.2	–17.3 ± 1.6	7.3 ± 1.6	2.4	–239 ± 21
*Tf*Cel48A[i]	23.4	0.67 ± 0.20	−8.6 ± 0.2	−2.7 ± 0.3	−5.9 ± 0.4	−12.8 ± 2.2	4.5 ± 2.2	2.4	−176 ± 25
*Tr*Cel7A	22.0	0.87[h]	−8.4[h]						

[a] soluble/non soluble reducing ends, [b] from *13*, [c] μM, [d] kcal/mol, [e] cal/K·mol, [f] $\Delta S_{solv}^{o} = \Delta C_p \ln(T_{303\ K}/T_{385\ K})$ derived using $\Delta S_r^{\circ} = \Delta S_{solv}^{o} + \Delta S_{mix}^{o} + \Delta S_{conf}^{o}$ where $\Delta S_{mix}^{o} = R\ln(1/55.5) = -8$ cal/K·mol ("cratic" term), [g] derived from the temperature dependence of ΔH_r°, [h] calculated from data published in *36*. [i] ITC measurements performed on mutants where the catalytic acid (Glu) has been exchanged to Ala as described in MATERIALS AND METHODS.

By means of this method, a ΔS_{mix}° of –8 cal/K·mol can be calculated, which corresponds to a $-T\Delta S_{mix}^{\circ}$ of 2.4 kcal/mol for all enzymes. Then, a value for the conformational entropy change can then be derived from Equation 6, yielding ΔS_{conf}° of –42.2 ± 4.3 cal/K·mol, –24.1 ± 5.3 cal/K·mol, and –14.9 ± 5.9 cal/K·mol. These numbers correspond to a $-T\Delta S_{conf}^{\circ}$ of 12.8 ± 1.3 kcal/mol, 7.3 ± 1.6 kcal/mol, and 4.5 ± 1.8 kcal/mol for *Tf*Cel5A, *Tf*Cel9A, and *Tf*Cel48A, respectively. All values are reported in tabular form (Table 3) for ease of comparison.

DISCUSSION

Preferred Binding Mode of Oligomeric Substrates to *Tf*Cel48A, *Tf*Cel9A, and *Tf*Cel5A. In this work, we have observed that the three studied cellulases have different subsite preferences and thermodynamic signatures upon substrate binding. From initial hydrolysis of $(Glc)_5$ and $(Glc)_6$, it is clear that *Tf*Cel48 has significant binding affinity in both subsites –3 and +3, and that the affinity is stronger in the –3 subsite than the +3 subsite (Figure 2). The exo-

processive *Tf*Cel48A acts from the reducing end of the cellulose chain (Figure 6).[24] This means that negative subsites are substrate binding sites, in other words, they remain bound to the

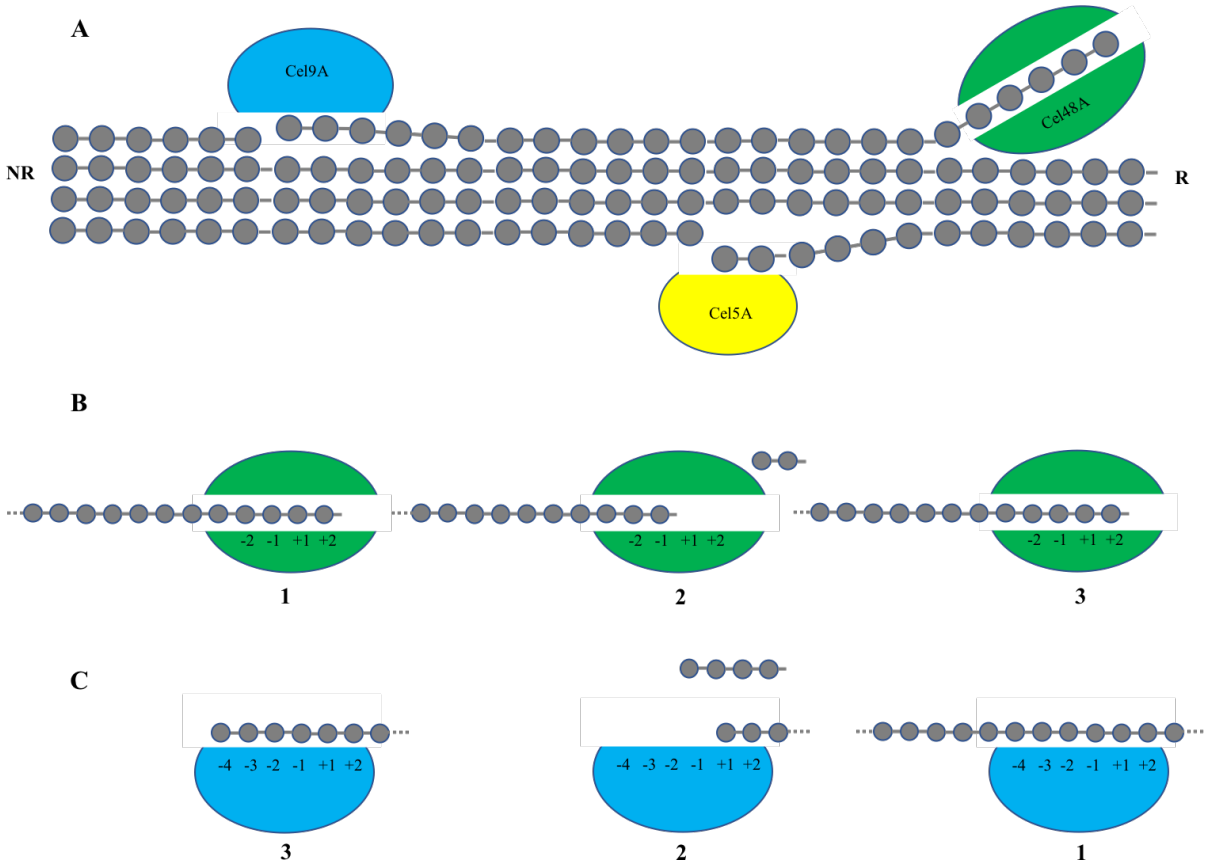

Figure 6. Schematic view of the mode of action on cellulose the studied *T. fusca* cellulases. (A) Cellulose is shown as tightly packed polymer chains of Glc-moieties (filled circles). *Tf*Cel48A (green), with its closed active site, degrades cellulose from the reducing end (labeled "R"). *Tf*Cel9A and *Tf*Cel5A with their relative open active sites bind to the more amorphous regions of the substrate for their catalyzes of hydrolysis. (B) After initial binding (1), *Tf*Cel48A catalyzes the hydrolysis and the product (2), here depicted as $(Glc)_2$, is displaced. Negative subsites, which are what we call substrate binding (only –2 and –1 are depicted for simplicity), are responsible for keeping the enzyme bound to the substrate. Subsites +1 and +2 (binding sites) are now available for the binding of new sugar moieties facilitating the moving of the polymer chain ("pushing potential" as describe below). Finally, the polymer chain is moved by two units (3) to set up for a new catalysis of hydrolysis. (C) After initial binding (1), *Tf*Cel9A

catalyzes the hydrolysis and the product (2), here depicted as $(Glc)_4$, is displaced. Positive subsites, which again are what we call substrate binding (only +2 and +1 are depicted for simplicity), are responsible for keeping the enzyme bound to the substrate. Subsites –1 to –4 (product binding sites) are now available for the binding of new sugar moieties facilitating the moving of the polymer chain. Finally, the polymer chain is moved by four units (3) to set up for a new catalysis of hydrolysis. *Tf*Cel5A dissociates of the substrate immediately after catalysis.

polymeric part of the cellulose molecule after hydrolysis has been performed. Having strong affinity in substrate binding subsites is central to the processive mechanism.[37-39] In *Tf*Cel48A, there are two aromatic amino acid residues in subsites –5 (Trp^{313}) and –4 (Trp^{315}) (Figure 4), respectively, that have been shown to be important for the processive ability.[24] Several studies show that a Trp-sugar interaction typically yields between 2 and 4 kcal/mol in binding free energy.[39-42] Still, such a hydrophobic surface also provides an inherent flexibility facilitating the sliding of the ligand in the active site.[43] Interestingly, the strong binding –3 subsite appears to lack a Trp-residue (Figure 4). Instead, there are several hydrogen bonding interactions (i.e., Lys^{270} and Gln^{176}) that seems to facilitate binding to the ligand. A similar observation is made for the exo-processive chitinase *Sm*ChiA from *Serratia marcescens*, which is also active from the reducing end of the polymer.[44, 45] Here, a threonine residue (Thr^{276}) contributes both to ligand binding free energy as well as the processive ability.[26] The strong interaction in the +2 subsite in *Tf*Cel48A is likely due to stacking with Tyr^{300}, while the high affinity in the +3 subsite is likely due to a similar interaction with Trp^{414} (Figure 4). Even though product binding sites, which are those that bind to the cleaved of products after a hydrolysis has been performed, are thought to be less important for processive ability than substrate binding sites, strong binding in the product sites may provide a "pushing potential" to remove obstacles, which limits the

processive ability of GHs,[46-48] occurring on the crystalline surfaces. This pushing potential arises when a productive binding occurs and the new-formed product dissociates leaving product binding sites available for the binding of new sugar moieties facilitating the moving of the polymer chain with i.e. two units for *Tf*Cel48A and four units for *Tf*Cel9A. This has previously been suggested for *Sm*ChiA.[49] Moreover, it is in line with calculations showing that binding to product sites are stronger for processive cellulases compared to those that are nonprocessive.[50]

Previously, it has been suggested that the main product of *Tf*Cel9A-catalyzed hydrolysis of cellulose is $(Glc)_4$ instead of cellobiose in line with what is observed here (Figure 2).[20] It is worth noting that it was necessary to have 5 μM of the enzyme to degrade 20 % of $(Glc)_4$ equally fast as 25 nM was able to degrade $(Glc)_5$ or $(Glc)_6$ (Table 2). Altogether, the results show the importance of the –4 subsite in the mechanism of *Tf*Cel9A. Trp^{313}, Trp^{209}, and Trp^{256} stack with a sugar moiety in subsites –2, –3 and –4, respectively, creating a significant binding free energy contribution to overcome the decrystallization penalty of 12 kcal/mol for removing a cellotetraose unit from the crystalline lattice of cellulose.[4] Mutation of either of these residues results in reduced processive ability and activity on crystalline bacterial cellulose.[14] Again, this demonstrates the importance of having an inherent pushing/pull potential in processive GHs, especially since *Tf*Cel9A needs to move across four glucose units along the cellulose crystal. Interestingly, when the substrate is switched from crystalline bacterial cellulose to soluble substrates, the activity of the mutants increases corresponding to observations for GH18 chitinases.[38, 45, 51] As an example, *Tf*Cel9A-D261A experienced a ten-fold reduction in activity on bacterial cellulose, a two-fold reduction in processive ability, and a five-fold increase in activity on carboxymethyl cellulose compared to the wild-type.[14] The GH18 *Sm*ChiB-W97A experienced a ten-fold reduction in activity on chitin, a minimum two-fold reduction in processive ability, and a 30-fold increase in the activity on the water soluble chitosan compared

to the wild-type.[38, 46] Processive ability is essential to overcome the recalcitrance of the crystalline substrate, but comes at a large cost in enzyme speed.[38] The results show that the +2 subsite also provides substantial binding free energy to the substrate. Interestingly, it appears that there are mainly two polar residues (His^{376} and Arg^{378}) that participate in binding (Figure 4). In this regard, Arg^{378} plays an important role in the processive ability of *Tf*Cel9A.[14]

*Tf*Cel5A is an endo-nonprocessive cellulase.[15] The crystal structure of the enzyme in the presence of a $(Glc)_5$ molecule has been determined (pdb code 2ckr, Figure 3 and 4) without an accompanying paper discussing the enzyme-substrate interactions. Moreover, it is likely that the observed complex is not the Michaelis complex as all five glucose are in the $^4C_1$ chair conformation. It is common that the Michaelis complex have its –1 glucose residue in the $^{1,4}B$ boat conformation.[52, 53] In addition, the catalytic acid (Glu^{355}) have been exchanged to a Gln-residue and this does not point towards a glycosidic bond. Combined, this makes it difficult to assign individual subsites in the active site. Still, inspection of the crystal structure and the observation that there is an equal preference for productive binding of $(Glc)_6$ from –4 to +2 and –3 to +3 (Figure 2) and suggest that the active site of *Tf*Cel5A is relative short and contains five subsites. Moreover, the productive binding from –2 to +2 for $(Glc)_4$ and –3 to +2 for $(Glc)_5$ suggest that the –3, –2, and +2 subsites are relative strong binding.

Correlation between Substrate Binding Free Energy and Processive Ability. As mentioned previously, *Tf*Cel48A is exo-processive, *Tf*Cel9A is endo-processive, and *Tf*Cel5A is endo-nonprocessive. The processive ability of a GH has been described in two different ways.[54] The first is apparent processivity (P^{app}), equaling the average number of catalytic acts (N_{catal}) an enzyme performs per one initiation of a processive run (N_{init}) (Equation 3).

$$P^{app} = \frac{N_{catal}}{N_{init}} \quad (3)$$

Here, the ratio of N_{catal} vs. N_{init} is normally assessed by either determining the ratio of soluble vs. non soluble reducing ends products or soluble even-numbered oligosaccharide products vs. odd-numbered products using different methods after GH action. Please read Horn *et al*. for details.[54] P^{app} is dependent on the nature of the substrate and is limited by length of the obstacle-free path available for the GH.[47, 48] The second describes the "processivity potential," that is, the true or intrinsic processivity (P^{intr}). P^{intr} can also be described as an average number of consecutive catalytic acts performed before dissociation (Equation 4) where k_{cat} is the catalytic rate constant and k_{off} is the dissociation rate constant.[55] Here, k_{cat} and k_{off} are determined in individual experimental set-ups as described by i.e. Kurasïn and Väljamäe or Kurasïn *et al*.[48, 49]

$$P^{intr} = \frac{k_{cat}}{k_{off}} \quad (4)$$

Based on calculated binding free energies of oligo-saccharide binding to several industrially important GH family 7 processive cellulases, Payne *et al.* hypothesized that there is a correlation between P^{intr} and binding free energy ($\Delta G_r°$) (Equation 5).[10]

$$-\frac{\Delta G_r^o}{RT} = ln\left(\frac{P^{intr} k_{on}}{k_{cat}}\right) \quad (5)$$

Indeed, Hamre *et al*. showed that increased processive ability corresponds to more favorable binding free energy and that this likely is a general feature of GHs.[37] Moreover, there are often strong binding amino acids that confer the majority of this binding free energy.[26, 39] P^{app} has previously been determined for *Tf*Cel5A, *Tf*Cel9A, and *Tf*Cel48A as well as for *Tr*Cel7A from *Trichoderma reesei* by measuring the ratio of soluble vs. insoluble reducing sugars (Table 3).[13] Here, the exo-processive *Tf*Cel48A displays equal P^{app} as the industrially relevant *Tr*Cel7A (23.4 vs. 22.0). P^{app} for the endo-processive *Tf*Cel9A is smaller (7.0) than for *Tf*Cel48A, but higher than that observed for the endo-nonprocessive *Tf*Cel5A (2.2). Still, it is

worth noting that when *Tf*Cel48A and *Tr*Cel7A performs processive runs, a distance of "only" two glucose units are travelled before catalysis while this distance is four glucose units for *Tf*Cel9A. In this regard, it is interesting to observe that k_{off} ($3.4 \cdot 10^{-3}$ s^{-1}) on BMCC is 22 times lower for *Tf*Cel9A than for the exo-processive *Tf*Cel6B (working from the nonreducing end), which has a P^{app} of 12.1. This value is also similar to that observed for *Tr*Cel7A on bacterial cellulose (BC) from *Acetobacter xylinum* ($0.7 \cdot 10$ s^{-1}).[48, 56] With a k_{cat} of 2.8 s^{-1} on the bacterial cellulose, *Tr*Cel7A has a P^{intr} of 4000. Unfortunately, a k_{cat} for *Tf*Cel9A has not been determined on the BMCC substrate allowing an estimation of P^{intr}. Still, this enzyme has the highest activity of any individual *T. fusca* enzyme on crystalline substrates, particularly bacterial cellulose (BC).[14] Altogether, this suggests that *Tf*Cel9A has significant processive ability. It is interesting, then, to observe that in a set of co-evolved GHs from a single organism there exists a correlation between binding free energy and processive ability (Table 3). The two processive GHs have ~2.2 kcal/mol stronger binding affinity towards $(Glc)_6$ compared to the nonprocessive GH (Table 3). Even though *Tr*Cel7A is not co-evolved with *Tf*Cel5A, it is interesting to observe that the difference in binding free energy and processive ability also is of same magnitude between these enzymes as between *Tr*Cel48A and *Tf*Cel5A.

Correlation of Thermodynamic Signatures to Mode of Action. It is tempting to relate some of the binding energies to function although great caution must be shown when interpreting the obtained values. The largest contribution to the binding free energy for all three cellulases is desolvation upon ligand binding ($-T\Delta S_{solv}° = -12.8$, -17.3, and -15.2 kcal/mol for *Tf*Cel48A, *Tf*Cel9A, and *Tf*Cel5A, respectively). The same was observed for the co-evolved GH18s of *Serratia marcescens* (two exo-processive and one endo-nonprocessive chitinase).[37] Furthermore, it is interesting to observe that the two endo-active cellulases, with their cleft topology (Figure 3), have a more favorable solvation entropy change than the exo-active cellulase, having a tunnel topology.[8, 20, 24] This corresponds with observations for the

two GH18s human chitotriosidase (HCHT) and *Sm*ChiA. HCHT has a more open cleft topology, a higher probability of endo-initiation (P_{endo} = 0.95) on α-chitin, and displayed a more favorable solvation entropy change ($-T\Delta S_{solv}°$ = –10.0 kcal/mol) upon binding with the inhibitor allosamidin compared to *Sm*ChiA ($-T\Delta S_{solv}°$ = –4.9 kcal/mol) that has a more closed cleft topology, and a lower probability of endo-initiation (P_{endo} = 0.76).[49, 57, 58] It is likely, and somewhat intuitive, that a shallow and open substrate-binding cleft that is highly exposed to the solvent will result in more water molecules being displaced upon ligand binding.

The two processive cellulases have less favorable changes in binding enthalpy ($\Delta H_r°$ = −2.7 and –1.1 kcal/mol for *Tf*Cel48A and *Tf*Cel9A, respectively than the nonprocessive *Tf*Cel5A ($\Delta H_r°$ = –6.4 kcal/mol). The same trend was also observed for the *S. marcescens* GH18s, where the two exo-processive *Sm*ChiA and *Sm*ChiB have less favorable binding enthalpy than the endo-nonprocessive *Sm*ChiC.[37] This suggest that binding free enthalpy for processive GHs is tuned to accommodate both initial binding to the substrate as well as being able to remain close to the polysaccharide chain enabling a slide to catalyze another hydrolysis. Support for this is found in the work of Meyer and Schultz who have studied how malto-oligosaccharides "slide" across the pores of the protein maltoporin. Here, the pores have surface exposed aromatic residues that interact with the oligosaccharides during "sliding". To address these interaction, a combination of computational and structural analysis was undertaken.[59] The presence of the aromatic residues yielded a binding-profile with less high-energy barriers than if binding involved hydrogen bonds alone. Such interactions are the most prevalent in this system. The combination of these two types interaction promotes sliding of the substrate in the active site. Support is also found in the work of Varrot *et al*. Here, five high-resolution structures of the processive *Hi*Cel6A in complex with non-hydrolysable thio-oligosaccharides were solved to detail the interactions between the enzyme and substrate both at productive as well as intermediate non-productive binding modes.[43] Key findings were that that the flexibility of the

tryptophan residues in the active site allows for the accommodation of both binding modes, and that the number of direct hydrogen bonds and solvent mediated interactions vary for productive and non-productive binding mode. Combined, this allows the polymeric substrate to slide through the active site.

The two processive cellulases have also a smaller penalty on the conformational entropy change ($-T\Delta S_{conf}^{\circ}$ = 4.5 and 7.3 kcal/mol for *Tf*Cel48A and *Tf*Cel9A, respectively) compared to the nonprocessive cellulase ($-T\Delta S_{conf}^{\circ}$ = 12.8 kcal/mol for *Tf*Cel5A). It is interesting to observe that *Tf*Cel48A with its tunnel topology has the lowest penalty of conformational entropy change. This, again, is in line with what is observed for the GH18s of *S. marcescens*.[37, 41] A detailed thermodynamic study revealed that ligand binding to the processive GH18 *Sm*ChiB, which has a tunnel topology, has 11 and 7.3 kcal/mol more favorable conformational entropy change than *Sm*ChiA and *Sm*ChiC, respectively, with the relative open active site clefts. Moreover, it was also shown that ligand binding amino acids are central for forming the tunnel and keeping this rigid before binding of substrate, resulting in a low penalty in conformational entropy change. Both for *Tf*Cel48A and *Sm*ChiB, desolvation of the enzyme – ligand complex more than compensates the relatively low loss in conformational entropy change (Table 3).[41]

In summary, our results show that mapping productive binding of substrate combined with understanding the thermodynamics of substrate association provides detailed information on how a set of co-evolved GHs from a single organism approaches the degradation of a recalcitrant polysaccharide substrate. Independent on the mode of action, desolvation of the enzyme – ligand complex is a major driving force of substrate binding. Processive GHs bind the substrate stronger than nonprocessive GHs (~2.2 kcal/mol stronger binding affinity). The binding free energy from the enzymes come from both substrate as well as product binding sites to remain attached to the substrate between catalytic cycles and to provide a pushing potential for removal of obstacles, respectively. In *Tf*Cel48A and *Tf*Cel9A, the second substrate subsite

away from the catalytic acid (–2 and +2, respectively, Figure 3) is the most important to remain bound while it is second and fourth product subsite (+2 and –4, respectively, Figure 3) that is fundamental for the pushing potential. The latter occurrence is also likely the reason why $(Glc)_2$ and $(Glc)_4$ are the initial main product from the respective enzyme actions. Moreover, it appears that processive GHs have lower contributions of binding enthalpy the overall binding free energy than nonprocessive GHs in order to balance both initial binding to the substrate as well as being able to slide down the substrate to set up another catalysis of hydrolysis. Furthermore, it seems that a tunnel topology at the active site results in less favorable solvation entropy change upon substrate binding than a cleft topology. Still, a tunnel topology yields a smaller conformational entropy penalty than the cleft topology. Finally, GHs degrading cellulose and chitin face similar challenges. The results from this study and others point towards that microorganisms have evolved suits of enzymatic machineries with similar complementary modes of action and initial substrate binding to degrade these polysaccharides.

ACCESSION CODES

*Tf*Cel5A (Uniprot ID: Q47RH8), *Tf*Cel9A (Uniprot ID: Q47MW0), and *Tf*Cel48A- (Uniprot ID Q47NH7).

ACKNOWLEDGMENT

This work was supported by grant from the Norwegian Research Council 209335 (MS). Financial support was also provided by the National Science Foundation (1552355 to CMP). This material is also based upon work supported by (while CMP is serving at) the NSF. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

Reference List

[1] Cocinero, E. J., Gamblin, D. P., Davis, B. G., and Simons, J. P. (2009) The Building Blocks of Cellulose: The Intrinsic Conformational Structures of Cellobiose, Its Epimer, Lactose, and Their Singly Hydrated Complexes, *J. Am. Chem. Soc. 131*, 11117-11123.
[2] Brown, R. M. (2004) Cellulose Structure and Biosynthesis: What Is in Store for the 21st Century?, *J. Polym. Sci. A Polym. Chem. 42*, 487-495.
[3] O'Sullivan, A. C. (1997) Cellulose: The Structure Slowly Unravels, *Cellulose 4*, 173-207.
[4] Beckham, G. T., Matthews, J. F., Peters, B., Bomble, Y. J., Himmel, M. E., and Crowley, M. F. (2011) Molecular-Level Origins of Biomass Recalcitrance: Decrystallization Free Energies for Four Common Cellulose Polymorphs, *J. Phys. Chem. B 115*, 4118-4127.
[5] Himmel, M. E., Ding, S. Y., Johnson, D. K., Adney, W. S., and Nimlos, M. R. (2007) Biomass Recalcitrance: Engineering Plants and Enzymes for Biofuels Production, *Science 315*, 804-807.
[6] Matthews, J. F., Skopec, C. E., Mason, P. E., Zuccato, P., Torget, R. W., Sugiyama, J., Himmel, M. E., and Brady, J. W. (2006) Computer Simulation Studies of Microcrystalline Cellulose Iβ, *Carbohydr. Res. 341*, 138-152.
[7] Nishiyama, Y., Langan, P., and Chanzy, H. (2002) Crystal Structure and Hydrogen-bonding System in Cellulose Iβ from Synchrotron X-ray and Neutron Fiber Diffraction, *J. Am. Chem. Soc. 124*, 9074-9082.
[8] Davies, G. J., and Henrissat, B. (1995) Structures and Mechanisms of Glycosyl Hydrolases, *Structure 3*, 853-859.
[9] Breyer, W. A., and Mattthews, B. W. (2001) A structural basis for processivity, *Prot. Sci. 10*, 1699-1711.
[10] Payne, C. M., Jiang, W., Shirts, M. R., Himmel, M. E., Crowley, M. F., and Beckham, G. T. (2013) Glycoside Hydrolase Processivity is Directly Related to Oligosaccharide Binding Free Energy, *J. Am. Chem. Soc. 135*, 18831-18839.
[11] Lombard, V., Ramulu, H. G., Drula, E., Coutinho, P. M., and Henrissat, B. (2014) The Carbohydrate-Active Enzymes Database (CAZy) in 2013, *Nucleic Acids Res. 42*, D490-D495.
[12] Tomme, P., Warren, R. A. J., Miller, R. C., Kilburn, D. G., and Gilkes, N. R. (1996) Cellulose-Binding Domains: Classification and Properties, In *Enzymatic Degradation of Insoluble Carbohydrates*, pp 142-163, American Chemical Society.
[13] Irwin, D. C., Spezio, M., Walker, L. P., and Wilson, D. B. (1993) Activity Studies of Eight Purified Cellulases: Specificity, Synergism, and Binding Domain Effects, *Biotechnol. Bioeng. 42*, 1002-1013.
[14] Li, Y., Irwin, D. C., and Wilson, D. B. (2007) Processivity, Substrate Binding, and Mechanism of Cellulose Hydrolysis by *Thermobifida Fusca* Cel9A, *Appl. Environ. Microbiol. 73*, 3165-3172.
[15] Wilson, D. B. (2004) Studies of *Thermobifida fusca* plant cell wall degrading enzymes, *Chem. Rec. 4*, 72-82.
[16] Posta, K., Beki, E., Wilson, D. B., Kukolya, J., and Hornok, L. (2004) Cloning, Characterization and Phylogenetic Relationships of Cel5B, a New Endoglucanase Encoding Gene from *Thermobifida Fusca*, *J. Basic Microbiol. 44*, 383-399.
[17] Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition, *Biochem. J. 382*, 769-781.
[18] Schulein, M. (2000) Protein Engineering of Cellulases, *Biochim. Biophys. Acta 1543*, 239-252.
[19] Barr, B. K., Hsieh, Y. L., Ganem, B., and Wilson, D. B. (1996) Identification of Two Functionally Different Classes of Exocellulases, *Biochemistry 35*, 586-592.

[20] Sakon, J., Irwin, D., Wilson, D. B., and Karplus, P. A. (1997) Structure and Mechanism of Endo/Exocellulase E4 from *Thermomonospora Fusca*, *Nat. Struct. Mol. Biol. 4*, 810-818.
[21] Sakon, J., Irwin, D., Wilson, D. B., and Karplus, P. A. (1997) Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*, *Nat. Struct. Biol. 4*, 810-818.
[22] Davies, G. J., Wilson, K. S., and Henrissat, B. (1997) Nomenclature for sugar-binding subsites in glycosyl hydrolases, *Biochem. J. 321*, 557-559.
[23] Irwin, D. C., Zhang, S., and Wilson, D. B. (2000) Cloning, Expression and Characterization of a Family 48 Exocellulase, Cel48A, from *Thermobifida Fusca*, *Eur. J. Biochem. 267*, 4988-4997.
[24] Kostylev, M., Alahuhta, M., Chen, M., Brunecky, R., Himmel, M. E., Lunin, V. V., Brady, J., and Wilson, D. B. (2014) Cel48A From *Thermobifida Fusca*: Structure and Site Directed Mutagenesis of Key Residues, *Biotechnol. Bioengineer. 111*, 664-673.
[25] Eide, K. B., Lindbom, A. R., Eijsink, V. G. H., Norberg, A. L., and Sørlie, M. (2013) Analysis of Productive Binding Modes in the Human Chitotriosidase, *Febs Lett. 587*, 3508-3513.
[26] Hamre, A. G., Jana, S., Reppert, N. K., Payne, C. M., and Sørlie, M. (2015) Processivity, Substrate Positioning, and Binding: The Role of Polar Residues in a Family 18 Glycoside Hydrolase, *Biochemistry 54*, 7292-7306.
[27] Hekmat, O., Lo Leggio, L., Rosengren, A., Kamarauskaite, J., Kolenova, K., and Stålbrand, H. (2010) Rational Engineering of Mannosyl Binding in the Distal Glycone Subsites of *Cellulomonas Fimi* Endo-β-1,4-mannanase: Mannosyl Binding Promoted at Subsite −2 and Demoted at Subsite −3, *Biochemistry 49*, 4884-4896.
[28] Pechsrichuang, P., Lorentzen, S. B., Aam, B. B., Tuveng, T. R., Hamre, A. G., Eijsink, V. G. H., and Yamabhai, M. (2018) Bioconversion of Chitosan into Chito-oligosaccharides (CHOS) using Family 46 Chitosanase from *Bacillus Subtilis* (*Bs*Csn46A), *Carbohydr. Polym. 186*, 420-428.
[29] Wiseman, T., Williston, S., Brandts, J. F., and Lin, L. N. (1989) Rapid Measurement of Binding Constants and Heats of Binding using a New Titration Calorimeter, *Anal. Biochem. 179*, 131-137.
[30] Fukada, H., and Takahashi, K. (1998) Enthalpy and Heat Capacity Changes for the Proton Dissociation of Various Buffer Components in 0.1 M Potassium Chloride, *Proteins 33*, 159-166.
[31] Cederkvist, F. H., Saua, S. F., Karlsen, V., Sakuda, S., Eijsink, V. G. H., and Sørlie, M. (2007) Thermodynamic Analysis of Allosamidin Binding to a Family 18 Chitinase, *Biochemistry 46*, 12347-12354.
[32] Zakariassen, H., and Sørlie, M. (2007) Heat Capacity Changes in Heme Protein-Ligand Interactions, *Thermochim. Acta 464*, 24-28.
[33] Baker, B. M., and Murphy, K. P. (1997) Dissecting the Energetics of a Protein-protein Interaction: The Binding of Ovomucoid Third Domain to Elastase, *J. Mol. Biol. 268*, 557-569.
[34] Baldwin, R. L. (1986) Temperature Dependence of the Hydrophobic Interaction in Protein Folding, *Proc. Natl. Acad. Sci. USA 83*, 8069-8072.
[35] Murphy, K. P., Privalov, P. L., and Gill, S. J. (1990) Common Features of Protein Unfolding and Dissolution of Hydrophobic Compounds, *Science 247*, 559-561.
[36] Colussi, F., Sørensen, 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 Cellobiohydrolase (Cel7), *J. Biol. Chem. 290*, 2444-2454.
[37] Hamre, A. G., Jana, S., Holen, M. M., Mathiesen, G., P., V., Payne, C. M., and Sørlie, M. (2015) Thermodynamic Relationships with Processivity in *Serratia Marcescens* Family 18 Chitinases, *J. Phys. Chem. B 119*, 9601-9613.

[38] Horn, S. J., Sikorski, P., Cederkvist, J. B., Vaaje-Kolstad, G., Sørlie, M., Synstad, B., Vriend, G., Vårum, K. M., and Eijsink, V. G. H. (2006) Costs and benefits of processivity in enzymatic degradation of recalcitrant polysaccharides, *Proc. Natl. Acad. Sci. U.S.A. 103*, 18089-18094.
[39] Jana, S., Hamre, A. G., Wildberger, P., Holen, M. M., Eijsink, V. G. H., Beckham, G. T., Sørlie, M., and Payne, C. M. (2016) Aromatic-mediated Carbohydrate Recognition in Processive *Serratia Marcescens* Chitinases, *J. Phys. Chem. B 120*, 1236-1249.
[40] Baban, J., Fjeld, S., Sakuda, S., Eijsink, V. G. H., and Sørlie, M. (2010) The Roles of Three *Serratia Marcescens* Chitinases in Chitin Conversion are Reflected in Different Thermodynamic Signatures of Allosamidin Binding, *J. Phys. Chem. B 114*, 6144-6149.
[41] Hamre, A. G., Frøberg, E. E., Eijsink, V. G. H., and Sørlie, M. (2017) Thermodynamics of Tunnel Formation upon Substrate Binding in a Processive Glycoside Hydrolase, *Arch. Biochem. Biophys. 620*, 35-42.
[42] Zolotnitsky, G., Cogan, U., Adir, N., Solomon, V., Shoham, G., and Shoham, Y. (2004) Mapping Glycoside Hydrolase Substrate Subsites by Isothermal Titration Calorimetry, *Proc. Natl. Acad. Sci. U.S.A 101*, 11275-11280.
[43] Varrot, A., Frandsen, T. P., von Ossowski, I., Boyer, V., Cottaz, S., Driguez, H., Schulein, M., and Davies, G. J. (2003) Structural Basis for Ligand Binding and Processivity in Cellobiohydrolase Cel6A from *Humicola Insolens*, *Structure 11*, 855-864.
[44] Hult, E. L., Katouno, F., Uchiyama, T., Watanabe, T., and Sugiyama, J. (2005) Molecular Directionality in Crystalline *β*-chitin: Hydrolysis by Chitinases A and B from *Serratia Marcescens* 2170, *Biochem. J. 388*, 851-856.
[45] Zakariassen, H., Aam, B. B., Horn, S. J., Vårum, K. M., Sørlie, 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.
[46] Hamre, A. G., Lorentzen, S. B., Väljamäe, P., and Sørlie, M. (2014) Enzyme Processivity changes with the Extent of Recalcitrant Polysaccharide Degradation, *FEBS Lett. 588*, 4620-4624.
[47] Igarashi, K., Uchihashi, T., Koivula, A., Wada, M., Kimura, S., Okamoto, T., Penttilä, M., Ando, T., and Samejima, M. (2011) Traffic Jams reduce Hydrolytic Efficiency of Cellulase on Cellulose Surface, *Science 333*, 1279-1282.
[48] Kurašin, M., and Väljamäe, P. (2011) Processivity of Cellobiohydrolases is Limited by the Substrate, *J. Biol. Chem. 286*, 169-177.
[49] Kurašin, M., Kuusk, S., Kuusk, P., Sørlie, M., and Väljamäe, P. (2015) Slow Off-rates and Strong Product Binding are Required for Processivity and Efficient Degradation of Recalcitrant Chitin by Family 18 Chitinases, *J. Biol. Chem. 290*, 29074-29085.
[50] Bu, L., Nimlos, M. R., Shirts, M. R., Ståhlberg, J., Himmel, M. E., Crowley, M. F., and Beckham, G. T. (2012) Product Binding varies Dramatically between Processive and Nonprocessive Cellulase Enzymes, *J. Biol. Chem. 287*, 24807-24813.
[51] Krokeide, I. M., Synstad, B., Gåseidnes, S., Horn, S. J., Eijsink, V. G. H., and Sørlie, M. (2007) Natural substrate assay for chitinases using high-performance liquid chromatography: A comparison with existing assays, *Anal. Biochem. 363*, 128-134.
[52] Biarnes, X., Ardevol, A., Planas, A., Rovira, C., Laio, A., and Parrinello, M. (2007) The Conformational Free Energy Landscape of Beta-D-glucopyranose. Implications for Substrate Preactivation in Beta-glucoside Hydrolases, *J. Am. Chem. Soc. 129*, 10686-10693.
[53] Koshland, D. E. (1953) Stereochemistry and the Mechanism of Enzymatic Reactions, *Biol. Rev. Camb. Philos. Soc. 28*, 416-436.
[54] Horn, S. J., Sørlie, M., Vårum, K. M., Väljamäe, P., and Eijsink, V. G. H. (2012) Measuring Processivity, In *Meth. Enzymol.*, pp 69-95, Academic Press.

[55] Lucius, A. L., Maluf, N. K., Fischer, C. J., and Lohman, T. M. (2003) General Methods for Analysis of Sequential "n-step" Kinetic Mechanisms: Application to Single Turnover Kinetics of Helicase-Catalyzed DNA Unwinding, *Biophys. J. 85*, 2224-2239.
[56] Moran-Mirabal, J. M., Bolewski, J. C., and Walker, L. P. (2011) Reversibility and Binding Kinetics of *Thermobifida Fusca* Cellulases Studied through Fluorescence Recovery after Photobleaching Microscopy, *Biophys. Chem. 155*, 20-28.
[57] Eide, K. B., Lundmark, S. T., Sakuda, S., and Sørlie, M. (2013) Thermodynamic Analysis of Allosamidin Binding to the Human Chitotriosidase, *Thermochim. Acta 565*, 146-150.
[58] Kuusk, S., Sørlie, M., and Väljamäe, P. (2017) Human Chitotriosidase Is an Endo-processive Enzyme, *Plos One 12*, 20.
[59] Meyer, J. E., and Schulz, G. E. (1997) Energy Profile of Maltooligosaccharide Permeation through Maltoporin as Derived from the Structure and from a Statistical Analysis of Saccharide-protein Interactions, *Protein Sci. 6*, 1084-1091.