

# Structural and biochemical characterization of a family 7 highly thermostable endoglucanase from the fungus Rasamsonia emersonii

Schiano di Cola, Corinna; Kołaczkowski, Bartłomiej; Sørensen, Trine Holst; Christensen, Stefan Jarl; Cavaleiro, Ana Mafalda; Windahl, Michael Skovbo; Borch, Kim; Morth, Jens Preben; Westh, Peter

*Published in:*
FEBS Journal

*Link to article, DOI:*
10.1111/febs.15151

*Publication date:*
2020

*Document Version*
Peer reviewed version

Received Date : 13-Sep-2019

Revised Date : 01-Nov-2019

Accepted Date : 20-Nov-2019

Color : Fig 1-3,6-9

# Structural and biochemical characterization of a family 7 highly thermostable endoglucanase from the fungus *Rasamsonia emersonii*

Corinna Schiano-di-Cola[1], Bartłomiej Kołaczkowski[1], Trine Holst Sørensen[1,2], Stefan Jarl Christensen[1], Ana Mafalda Cavaleiro[2], Michael Skovbo Windahl[1,2,4], Kim Borch[2], Jens Preben Morth[3*] and Peter Westh[1,3*]

[1] Roskilde University, Department of Science and Environment, Universitetsvej 1, DK-4000, Roskilde, Denmark.

[2] Novozymes A/S, Biologiensvej 2, DK-2800, Lyngby, Denmark,

[3]Department of Biotechnology and Biomedicine, Technical University of Denmark, Lyngby, Denmark.

[4]Present address: Bioneer A/S, Kogle Allé 2, DK-2970 Hørsholm, Denmark

*Corresponding author: Prof. Peter Westh. E-mail: petwe@dtu.dk; Tel. +45 4525 2641 or +45 3091 5488 and Prof. Jens Preben Morth premo@dtu.dk Tel. +45 9351 1454

**Running title**: Structure and activity of a thermophilic Cel7B

**Article type**: Original article

**Database**: structural data are available in RCSB Protein Data Bank database under the accession number **6SU8**.

**Abbreviations**:
EG, endoglucanase
CBH, cellobiohydrolase
GH7, glycoside hydrolase family 7
CAZy, Carbohydrate Active Enzymes
CBM, carbohydrate-binding domain
CD, catalytic domain
DS, degree of synergy
PDB, Protein Data Bank
NAG, *N*-acetyl glucosamine
PCA, pyroglutamic acid
MS, mass spectrometry
$T_{opt}$, temperature optimum
$T_m$, thermal transition midpoint
MM, Michaelis–Menten
HPAEC-PAD, high performance anion exchange chromatography with pulsed amperometric detection
Endo-H, endoglycosidase H
PNGase F, peptide-N-glycosidase F

**Enzymes**:
*Re*Cel7B, endoglucanase (EC 3.2.1.4) from *Rasamsonia emersonii*

*Re*Cel7A, cellobiohydrolase (EC 3.2.1.176) from *Rasamsonia emersonii*

*Tr*Cel7B, endoglucanase (EC 3.2.1.4) from *Trichoderma reesei*

*Tr*Cel7A, cellobiohydrolase (EC 3.2.1.176) from *Trichoderma reesei*

**Keywords**: Thermostable cellulases; endoglucanase; enzyme kinetics; cellulose; synergy.

**Conflict of interest**: Kim Borch, Michael Skovbo Windahl, Trine Holst Sørensen and Ana Mafalda Cavaleiro work for Novozymes A/S, a major manufacturer of industrial enzymes.

**Abstract**

Thermostable cellulases from glycoside hydrolase family 7 (GH7) are the main components of enzymatic mixtures for industrial saccharification of lignocellulose. Activity improvement of these enzymes via rational design is a promising strategy to alleviate the industrial costs, but it requires detailed structural knowledge. While substantial biochemical and structural data is available for GH7 cellobiohydrolases, endoglucanases are more elusive and only few structures have been solved so far. Here we report a new crystal structure and biochemical characterization of a thermostable endoglucanase from the thermophilic ascomycete *Rasamsonia emersonii*, *Re*Cel7B. The enzyme was compared with the homologous endoglucanase from the mesophilic model ascomycete *Trichoderma reesei* (*Tr*Cel7B), which unlike *Re*Cel7B possesses an additional carbohydrate binding module (CBM). With a temperature optimum of 80°C, *Re*Cel7B displayed a number of differences in activity and ability to synergize with cellobiohydrolases compared to *Tr*Cel7B. We improved both binding and kinetics in a chimeric variant of *Re*Cel7B and a CBM, while we observe the opposite effect when the CBM was removed in *Tr*Cel7B. The crystal structure of *Re*Cel7B was determined at 2.48 Å resolution, with $R_{work}$ and $R_{free}$ factors of 0.182 and 0.206, respectively. Structural analyses revealed that *Re*Cel7B has increased rigidity in a number of peripheral loops compared to *Tr*Cel7B and fewer aromatics in the substrate binding cleft. An increased number of glycosylations were identified in *Re*Cel7B and we propose a stabilizing mechanism for one of the glycans. Global structure-function interpretations of *Re*Cel7B highlight the differences in temperature stability, turnover, binding and cellulose accessibility in GH7 endoglucanases.

**Introduction**

Cellulases from glycoside hydrolase family 7 (GH7) have widespread interest because of their applicability in upcoming biorefineries and their ecological role in the terrestrial carbon cycle [1-3]. Members of the GH7 family are mostly identified in fungal genomes [4], and can be divided into cellobiohydrolases (CBHs) and endoglucanases (EGs), based on their mode of action on cellulose [5]. CBHs (EC 3.2.1.176) are commonly described as exo-processive retaining enzymes, with high specificity for crystalline cellulose. EGs (EC 3.2.1.4) are endo-acting, and more specific

towards amorphous cellulose [6, 7]. However, this dichotomy may be an oversimplification, in as much as CBHs are able to perform endo-type initiation and some EGs show moderate processivity [8, 9]. Generally, CBHs are the most abundant enzymes in the secretome of cellulolytic fungi, while EGs are secreted to a smaller percentage [10, 11]. However, the presence of both is necessary for an efficient, synergistic degradation of cellulose [12-14]. Regarding their architecture, GH7 can be bi-modular, with a catalytic domain (CD) connected to a family 1 carbohydrate binding domain (CBM) via a flexible peptide linker, although many lack the CBM and are composed solely by the CD [15]. Even though the CBM is not necessary for catalysis, the improved enzymatic action in the presence of a CBM is often credited to increased affinity for cellulose during cellulose targeting, which increase the local enzyme concentration [16-18]. In the CD, EGs and CBHs differ significantly since the latter has eight flexible loops which form a tunnel-shaped substrate-binding site, with up to 11 carbohydrate binding subsites [19]. The EGs has fewer subsites and a more exposed catalytic cleft, as a consequence of the shorter loops covering this region [20, 21]. It has been shown earlier that the length of the loops covering the catalytic cleft play a significant role in tuning processivity, binding and endolytic activity of GH7 enzymes [22, 23].

There are more than 20 GH7 crystals structures deposited in the Protein Data Bank (PDB, www.rcsb.org), of which only four belong to the EGs [24]. Here we report the biochemical and structural investigation of *Re*Cel7B, which is the major GH7 EG from the thermophilic ascomycete *Rasamsonia emersonii* (basionym: *Talaromyces emersonii* [25]). *Re*Cel7B is composed of a single catalytic domain, and the crystal structure was determined at a resolution of 2.48 Å, with $R_{work}$ and $R_{free}$ factors of 0.182 and 0.206, respectively. To elucidate structure-function relationships, we performed detailed biochemical analyses of *Re*Cel7B and *Tr*Cel7B, the major EG from the mesophilic ascomycete *Trichoderma reesei*. The catalytic domains of *Re*Cel7B and *Tr*Cel7B have 58% amino acid sequence identity while *Tr*Cel7B includes a linker-CBM domain [20]. To account for this in the functional studies, we also expressed and characterized two variants, one of *Tr*Cel7B deprived of linker-CBM (*Tr*Cel7BΔCBM), and one of *Re*Cel7B with an added linker-CBM from *T. reesei* Cel7A (ReCel7BinsCBM). Finally, since the ability of an EG to synergize with CBHs is an important element of its function, we investigated the effect of synergistic mixtures of *Re*Cel7B in combination with the paralogous CBH (*Re*Cel7A) or the CBH from *T. reesei* (*Tr*Cel7A), and compared the results with the synergistic capability of *Tr*Cel7B.

## Results

### Phylogenetic analysis of characterized GH7

To evaluate and identify sequence differences of *Re*Cel7B with respect to other EGs, we collected 87 GH7 amino acid sequences with documented enzyme activity from the Carbohydrate Active Enzymes database (CAZy, www.cazy.org), [26]. The resulting phylogenetic tree (Fig. 1A) showed a separation between the sequences in three major clades, as supported by earlier works [27]. One clade corresponded to CBHs, mostly belonging to ascomycetes and basidiomycetes but also some metazoans and amoebozoans sequences. A second clade corresponded to the ascomycotal EGs, included *Re*Cel7B, *Tr*Cel7B and other known EG structures such as *Fusarium oxisporium Fo*Cel7B (PDB **1OVW**) [28], *Humicola insolens Hi*Cel7B (PDB **2A39**) [29] and *Trichoderma harzianum* Cel7B *Th*Cel7B ( PDB **5W0A**) [30]. *Tr*Cel7B and *Th*Cel7B share high sequence identity (~90%), while *Hi*Cel7B and *Fo*Cel7B are less similar (~60% sequence identity). The last clade corresponded to the parabasalian EGs, which diverge from their fungal equivalent, as expected by the organisms particular symbiotic lifestyle [31]. A structure-based sequence alignment with the homologous EGs is shown in Fig. 1B. With these, *Re*Cel7B shared a relatively high amino acid sequence identity (~50-60%). A conservation of the catalytic triad was observed (black stars in Fig. 1B), with E199 and E204 in *Re*Cel7B suggested as the nucleophile and the acid/base residues involved in the *retaining* mechanism, respectively [32].

### Structure solution and overall quality of *Re*Cel7B model

The structure of *Re*Cel7B (PDB entry **6SU8**) was determined at 2.48 Å resolution, with $R_{work}$ and $R_{free}$ at 0.182 and 0.206, respectively, and an average temperature factor (B-factor) of 53. The crystals have solvent content of 78% and a crystal packing suggesting that *Re*Cel7B is able to form trimers, and each trimeric unit only forms one or two crystals contacts to make up the crystal lattice. The trimer interface is mainly formed by a number of flexible loops (Fig. 2G), described in detail below. However, we believe that *Re*Cel7B is functional as a monomer and that the trimeric packing likely is an artefact inferred from the crystal conditions. The monomeric state of *Re*Cel7B in solution was confirmed by analytical size-exclusion chromatography (Fig. 3). The crystals of *Re*Cel7B belonged to the space group $C222_1$ with three molecules in the asymmetric unit.

Statistical details regarding data collection, processing and refinement of the structure are found in Table 1.

**Structure of *Re*Cel7B**

The overall fold of *Re*Cel7B display a topology typical to the GH7 family, a β-jelly roll forming a distorted β-sandwich [33]. Compared to the structural homologous found amongst the CBHs, EGs generally display shorter loops covering the catalytic area, which create an exposed cleft [34]. The nomenclature for the most relevant loops has been proposed previously [35], and we will used this formalism throughout. In *Re*Cel7B the six major loops, termed A1-A3, B1, B3 and B4 have been highlighted (Fig. 2A). To investigate the substrate-binding cleft, *Re*Cel7B was superimposed on published structures of *Tr*Cel7B and *Th*Cel7B (Fig. 2), as well as *Hi*Cel7B containing a substrate analogue in the active site [36] and *Fo*Cel7B complexed with cellobiose [37] (not shown). Two residues, W344 and W353 of *Re*Cel7B, are highly conserved in the EG structures and are probably forming the subsites -2 and +1, respectively. Close to loop B1, two tryptophan residues in the homologous CBHs play a role in ligand binding in sub-sites -4 and -7 [38]. The aromaticity is conserved in *Tr*Cel7B and *Th*Cel7B, both containing Y38 and W40 in these positions, absent in *Re*Cel7B, *Fo*Cel7B and *Hi*Cel7B. In *Re*Cel7B the aromatics are substituted into S37 and P39, respectively. Conserved residues are also R108/108, S342/318 and Y149/146 (*Re*Cel7B/*Tr*Cel7B nomenclature), and in *Tr*Cel7B these residues probably interact via hydrogen bonding to the -2 and -3 sub-sites of a cellulose chain [21]. Moreover, a histidine interacting with the ligand at subsite +1 [39] is present both in *Hi*Cel7B (H209) and *Fo*Cel7B, while it is absent in *Re*Cel7B,*Tr*Cel7B and *Th*Cel7B and they are all substituted to alanine (A211 in *Re*Cel7B). Several proline residues in *Re*Cel7B are replaced by other residues in *Tr*Cel7B, namely P14/T15, P216/S213, P239/A222, P265/T248 (*Re*Cel7B/*Tr*Cel7B nomenclature). The first N-terminal glutamine was cyclized into pyroglutamic acid (PCA), as expected for expression in a fungal system. All the cysteines formed disulfide bonds (18 cysteines in total). Despite 20 mM cellobiose in the crystallization solution, we did not observe electron density for it in the substrate-binding region.

**Loop comparisons with *Tr*Cel7B, *Th*Cel7B *Hi*Cel7B and *Fo*Cel7B**

Initial cellulose chain threading can occur at the B1 and A1 loop (Fig. 2A) [35, 40]. Both vary significantly in length and amino acid composition among the EG (Fig. 1B). As for the former,

*Hi*Cel7B and *Fo*Cel7B have the longest loops, followed by *Re*Cel7B and *Tr*Cel7B/*Th*Cel7B. The lengths of the A1 loop shows the opposite trend *Tr*Cel7B/*Th*Cel7B>*Re*Cel7B>*Hi*Cel7B>*Fo*Cel7B. The A2 loop in *Re*Cel7B is delimited by a solvent exposed alpha helix (helix α6 in Fig. 1B) and a glycosylated asparagine, N363 (Fig. 2A). Interestingly, glycosylation on N363 was not observed in the other EG structures (Fig. 1B). Closer to the active site, the A3 loop (Fig. 2A) of *Re*Cel7B contains a proline residue, (P347) which is not seen in the other known GH7 structures. The B3 loop is truncated in *Tr*Cel7B and *Th*Cel7B (Fig. 2C), as compared to *Re*Cel7B and the other EGs considered (Fig. 1B). The longer B3 loop of *Re*Cel7B is stabilized by an additional disulfide bridge (C225-C230).

Finally in the product area, the short B4 loop (Fig. 2A) connects 2 α-helices in *Fo*Cel7B and *Hi*Cel7B (α2 and α3 in Fig. 1B). In *Re*Cel7B the loop length is similar, however the presence of a proline residue (P322), absent in the other structures, impedes the formation of the α-helix α3 (Fig. 1B). On the other hand, *Tr*Cel7B and *Th*Cel7B have a different loop configuration where α2 is completely absent.

**Glycosylations in *Re*Cel7B**

The *Re*Cel7B crystals studied here retained full glycosylation (no modification of glycans prior to crystallization), and it was possible to identify five *N*-glycosylation sites in connection with asparagines from visual inspection of the electron density in positions N28, N77, N185, N277 and N363 (Fig. 4). Therefore, in the model we were able to assign one *N*-acetyl glucosamine (NAG) molecule at position N28, N77, N277 and N363, while two NAGs and one mannose were assigned in position N185. The *N*-glycosylation site N185 is conserved in *Tr*Cel7B in position N182 (Fig. 2C) and in *Th*Cel7B in the same position, while the other structures have glycosylations elsewhere (magenta boxes in Fig. 1B). The crystal lattice (Fig. 2H) showed that all the glycans are facing the solvent and are evenly distributed along the structure. The high solvent content allowed the glycan to appear in an extended conformation, in fact we were able to detect additional electron density beyond most of the glycosylation sites, however the resolution was too low to further extend the glycan structure. Instead, we analysed the glycan composition of *Re*Cel7B by intact protein mass spectrometry (MS). This revealed that *Re*Cel7B expressed in *A. oryzae* is heavily and heterogeneously glycosylated (Fig. 5). Comparison of the MS spectra between *Re*Cel7B native structure and *Re*Cel7B deglycosylated with endoglycosidase H (Endo-H), allowed us to observe a mass shift corresponding to five *N*-acetylated hexoses and 44 hexoses between the main peaks in

the MS spectra (Fig. 5). To identify and quantify the glycan composition of *Re*Cel7B, the enzyme was subjected to strong acid hydrolysis and the released monosaccharides were analysed with high performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD, Fig. 6). The detected sugars were NAG, mannose and galactose, and their amount listed in Table 2.

**Biochemical properties of *Re*Cel7B, *Re*Cel7BinsCBM and *Tr*Cel7B**

The temperature optimum of *Re*Cel7B, *Re*Cel7BinsCBM$_{CBM}$ and *Tr*Cel7B was investigated on microcrystalline cellulose Avicel (Fig. 7A). As expected by the organism of origin [25], *Re*Cel7B showed a very high temperature optimum ($T_{opt}$) at about 80 °C, while *Tr*Cel7B showed an optimum around 55 °C. *Re*Cel7BinsCBM displayed higher activity at all temperatures compared to the wild type, and the $T_{opt}$ shifted to lower values (75°C). The optimal pH on Avicel at 50 °C was also determined (Fig. 7B). *Re*Cel7B showed a rather broad pH optimum around 2.5-5, while it was more narrow for *Tr*Cel7B, at 4.5. We measured the thermal transition midpoints ($T_m$) of *Re*Cel7B, *Re*Cel7BinsCBM and *Tr*Cel7B at different pH values using Differential Scanning Fluorimetry (DSF, Fig. 7C). *Re*Cel7B and *Re*Cel7BinsCBM showed very similar and high $T_m$ at all pH values, with a maximum value of 84 °C at pH 4.5 for *Re*Cel7B. A decrease in $T_m$ is observed at pH values distant from the optimal region. *Tr*Cel7B showed similar $T_m$-pH relationships but overall lower values, with the highest $T_m$ being 62 °C at pH 4.5.

**Steady-state kinetic analysis on insoluble substrate**

We investigated *Re*Cel7B activity on microcrystalline cellulose with two different approaches [41]. The first was based on rate measurements at substrate excess, and hence followed the conventional Michaelis-Menten (MM) framework:

$$v_0 = \frac{{}^{conv}V_{\max} S_0}{{}^{conv}K_M + S_0} \quad (1)$$

As shown in Fig. 7D, we could not reach saturation within a practicable substrate load (up to ~100g/L), and we hence only report the specificity constant, η, specified by the slope in Fig. 7D for low Avicel loads. The specificity constant (Table 3) was much higher in *Tr*Cel7A compared to *Re*Cel7B (0.04 and 0.01 g $L^{-1}$ $s^{-1}$ respectively). The removal of the CBM in *Tr*Cel7BΔCBM shifted η to a value closer to *Re*Cel7B (0.015 g $L^{-1}$ $s^{-1}$), while *Re*Cel7B and *Re*Cel7BinsCBM showed similar η.

The second steady-state approach used is the so-called inverse Michaelis-Menten ($^{inv}$MM), which is well suited to describe interfacial enzymes [41]. This approach relies on rate measurements at enzyme excess and subsequent analysis of plots of rate vs. enzyme concentration. We used the so-called inverse MM equation:

$$v_0 = \frac{^{inv}V_{max} E_0}{^{inv}K_M + E_0} \quad (2)$$

Unlike the conventional enzyme-saturation case, $^{inv}V_{max}$ describes the rate when all the hydrolyzable sites, or attack sites, present on the cellulose surface become occupied with enzyme [41]. Experimental points and best fits to Eq. 2 are shown in Fig. 7E, and parameters derived from these fits are listed in Table 3. For the wild type enzymes, the inverse maximal specific rate was 2-fold higher for *Tr*Cel7B compared to *Re*Cel7B (7.5 and 3.3 nmol $g^{-1}$ $s^{-1}$ respectively). *Re*Cel7BinsCBM showed an improvement in $^{inv}V_{max}$ compared to *Re*Cel7B, while the removal of the CBM in *Tr*Cel7BΔCBM, decrease $^{inv}V_{max}$ almost three-fold compared to *Tr*Cel7B.

Finally, we made binding isotherms (Fig. 7F) by measuring the concentration of free enzyme in solution ($E_{free}$) and calculating the substrate coverage $\Gamma=(E_0-E_{free})/S_0$. This is measured as a function of $E_{free}$:

$$\Gamma = \frac{\Gamma_{max} E_{free}}{K_d + E_{free}} \quad (3)$$

Where $\Gamma_{max}$ is the saturation coverage and $K_d$ is dissociation constant. The two CBM-less enzymes, *Re*Cel7B and *Tr*Cel7BΔCBM, bound very weakly, and we were unable to determine the difference in the numerator of eq. 3 with statistical significance. Hence, binding could not be quantified for these enzymes by the current method. For the two enzymes with CBM, binding was readily detectable, and as illustrated in Fig. 7F, *Tr*Cel7B adsorbed about three-fold more on Avicel than *Re*Cel7BinsCBM.

**Synergy with cellobiohydrolases**

We made four independent synergy curves mixing the two investigated EGs with the CBH *Tr*Cel7A and *Re*Cel7A (Fig. 8). To calculate the degree of synergy, the activities of the enzyme in the mixtures (black points in Fig. 8) were compared to the activities of the CBH and EG acting alone in the same conditions (coloured points in Fig. 8). Synergy was then quantified by calculating the degree of synergy (DS):

$$DS = \frac{A_{CBH+EG}}{A_{CBH} + A_{EG}} \quad (4)$$

where $A_{CBH+EG}$ is the activity of the enzyme mixture at a specific mole fraction of CBH, $A_{CBH}$ and $A_{EG}$ are the monocomponent activities at the same molar concentration (of the monocomponent) as the mixture. Synergy was evident in all cases, and the *T. reesei* CBH-EG mixtures showed overall higher DS than *R. emersonii* (Fig. 8 A-B). The highest DS was found with 25% EG for *T. reesei*, and about 50% EG for *R. emersonii*. Synergy of *Re*Cel7B and *Tr*Cel7B was also tested in mixtures with the CBH of the opposite organism (Fig. 8 C-D). The DS, plotted as a function of the mole fraction of CBH (Fig. 9), indicated that for *Re*Cel7B mixed with *Tr*Cel7A the maximum DS was reached with high amount of EG. Surprisingly, very low quantities (1-2%) of *Tr*Cel7B were capable of significantly enhancing the activity of *Re*Cel7A and particularly *Tr*Cel7A (Fig. 9).

## Discussion

### Structural properties of *Re*Cel7B

A phylogenetic tree of all biochemically characterized GH7 sequences (Fig. 1A) shows the evolutionary distribution of the structures from this family. According to CAZy database, there are 20 unique published GH7 structures out of which 16 are CBHs. They are mostly distributed in the ascomycotal clade, although few structures are available from basidiomycetes, amoebozoans and metazoans. As for the EGs, the only four structures deposited to date are from ascomycetes, which are part of a separate clade. So far no structural information is available amongst the parabasalia EGs. *Re*Cel7B can be found in the ascomycotal EG clade, clustering with a *Talaromyces cellulolyticus* EG, from which it shares about 65% sequence identity. Considering its phylogeny and thermophilic origin, *Re*Cel7B seemed a promising candidate for expanding the structure-function knowledge of the EGs. *Re*Cel7B crystal structure was determined at 2.48 Å resolution. The overall fold was, as expected, highly conserved and, compared to *Tr*Cel7B, the most conspicuous differences were found in the so-called "exo" part of the structure, which appeared more enclosed (Fig. 2C), mainly due to the increased length of loops B3 and B4. Studies performed by Wang *et al.* [42] indicated that mutating a serine (S221) to a lysine in *Tr*Cel7B reduced the enzymatic activity more than 65%, comparable to the observed difference in activity between *Tr*Cel7B and *Re*Cel7B (Table 3). The S221K mutation in *Tr*Cel7B corresponds to K238

position in *Re*Cel7B. It is therefore tempting to propose that K238 in *Re*Cel7B plays a role for thermostability, particularly since the primary amine of K238 is engaged in a hydrogen bond network with the backbone carbonyl of C230 in the B3 loop and the backbone carbonyl of Q177 (Fig. 2E). A corresponding lysine is found in *Hi*Cel7B and in *Fo*Cel7B (2A39 and 1OWV, respectively) and these enzymes both display lower activity than *Tr*Cel7B on cellulosic substrates [43], while an alanine is present in *Th*Cel7B (A221). We propose that the added stabilization of a lysine in this region could negatively influence catalytic efficacy, since the introduction of the lysine in *Tr*Cel7B decreased the activity on Avicel and increased the processivity, as shown by Wang and co-workers [42].

Close to the B1 loop, the two aromatics of *Tr*Cel7B (W40 and Y38) are substituted to S37 and P39 in *Re*Cel7B (Fig. 2F). The same residues are also found in the thermophilic *Hi*Cel7B (S37/P39, not shown). Aromatics in this position have been suggested to interact with the substrate in an homologous CBH [38], and engineering of *Hi*Cel7B showed that it was possible to increase affinity on phosphoric acid swollen cellulose by introducing the mutations S37W/P39W [29]. Thus, *Re*Cel7B appears to have fewer or weaker binding subsites compared to *Tr*Cel7B, as in the case of *Hi*Cel7B [29], and the sequence conservation shared between them suggests that these residues could be important for the thermostability, but coming at the price of reduced substrate interactions. Close to B4 loop, another loop is also elongated in *Re*Cel7B (position S302-S304) compared to *Tr*Cel7B and *Th*Cel7B, but it is shorter than *Fo*Cel7B and *Hi*Cel7B (Fig. 1B). Sonoda and co-workers observed with molecular dynamics simulations that in *Th*Cel7B an additional subsite (+2') is well suited for accommodating branched polysaccharides typical of hemicelluloses, due to a more open cleft shape of the product site compared to *Hi*Cel7B and *Fo*Cel7B [30]. *Re*Cel7B openness in the product area is intermediate between that of *Th*Cel7B/*Tr*Cel7B and *Fo*Cel7B/*Hi*Cel7B, suggesting the possibility of a side hemicellulolytic activity.

A number of additional features were identified in *Re*Cel7B structure which could be associated with increased rigidity compared to *Tr*Cel7B. In particular, the presence of unique prolines and superficial ion pairs (D152-R335, D293-R264, D136-K139) an additional disulfide bridge and a more compact hydrophobic core (due to the presence of W203 and F395, for example) are all stability promoting elements [44, 45]. Indeed, protein engineering studies showed that thermostability can be increased by the addition of more disulphide bridges in *Tr*Cel7B [46], and by increasing hydrophobicity of cavities in an EG from *Trichoderma pseudokoningii* [47].

Interestingly, one proline residue (P239) is substituted to A222 in *Tr*Cel7B and S222 in *Th*Cel7B. Molecular simulations studies of *Th*Cel7B showed that S222 hydrogen bonds with cellotetraose and other oligosaccharides in subsite +2 [30]. The proline substitution in *Re*Cel7B might be responsible for a decrease in affinity for cellobiose in the product site.

Additionally, *Re*Cel7B displays high amount of N-glycosylation sites (Fig. 2C), five (N28, N77, N185, N277 and N363) compared to only two in *Tr*Cel7B (N56 and N182) [20, 48].We note that particularly the glycosylations promote a clear stability role in *Re*Cel7B since enzymatic deglycosylation with Endo-H or Peptide-N-glycosidase F (PNGase F) give rise a loss in $T_m$ of about 2 and 4 °C respectively. Moverover, MS and monosaccharide composition analysis indicated that *Re*Cel7B is heterogeneously glycosylated and the hexoses added onto NAG at each site were mannose and, at a smaller percentage, galactose (Table 2). It was indeed surprising that the crystallization succeeded despite the extensive amounts of glycans present.

Galactose residues have been detected before in the glycan composition of an *A. oryzae* β-galactosidase [49], and as described also for other enzymes [50], this suggests that *Re*Cel7B is mannosylated and galactose is most likely present at the end of the mannose chains to prevent further carbohydrate additions. All identified glycans were solvent exposed (Fig. 2G-H), and in most cases (particularly for N277, N28 and N77) we could not identify solvent patterns or significant intermolecular interactions with the neighbouring residues (Fig. 2A). The N363 glycosylation is located close to the A2 loop (Fig. 2A), and this could reduce loop flexibility [35] and protein unfolding [51].

Moreover, a complex stabilizing interaction was identified for the N185 glycan in loop S1 (Fig. 2D). The O6' of the NAG is in hydrogen bond distance with the carbonyl in the backbone of R159 stabilizing and interaction with loop S2. A water molecule is bridging the B3 (S221) and S2 (D154) loop via the acetyl group of NAG (Fig. 2D). Additionally, D154 interacts with S156 through another hydrogen bond further stabilizing the S2 loop. The loops B3, S1 and S2 are thus all engaging in a hydrogen bond network involving the N154 glycan. This stabilization is probably also important for *Tr*Cel7B, since a glycosylation is present in the same position (N182, Fig. 2C) and in its close homolog *Th*Cel7B [30]. Although it is widely acknowledged that a combination of several factors are responsible for stabilizing thermophilic structures [52], we note that the features listed above point as elements contributing to the high $T_m$ of *Re*Cel7B [53, 54]. In addition to this, the presence of N-glycans are probably important for the very high solubility of *Re*Cel7B, another desirable trait from an industrial perspective [55].

**Activity of *Re*Cel7B, *Tr*Cel7B and CBM deletion/insertion variants**

Biochemical characterization of *Re*Cel7B on Avicel at 50 °C revealed a broader acidic pH optimum as compared to *Tr*Cel7B (Fig. 7B) and that the high thermostability was maintained over a broad range of pH values (Fig. 7C), particularly towards the acidic range. This collectively suggests that *Re*Cel7B is a thermoacidophilic enzyme. The only other GH7 EG structure isolated from a thermophilic fungus is *Hi*Cel7B [56]. The temperature optimum for this enzyme is 62 °C and the pH optimum is 7.5 [57], much lower than the temperature optimum of 80 °C of *Re*Cel7B (Fig. 7A). The other EGs have lower temperature optima, with *Tr*Cel7B and *Th*Cel7B of about 55 °C (Fig. 7A and [58]) and *Fo*Cel7B being the least thermostable [43]. Enzymatic saccharification is usually performed at 50-55 °C for 48-72 h at pH ~5 [59, 60]. Thus, high enzyme thermostability is needed for an efficient process. It is desirable to move to higher temperatures to further accelerate the mass transfer rates, improve mixing and reduce microbial contaminations [61]. This can be achieved by the implementation of thermostable cellulases in industrial mixtures. Therefore, *Re*Cel7B is highly attractive in an industrial perspective due to its high thermostability and broad pH optimum, which allows the enzyme to remain stable under different environmental fluctuations. Moreover, among the GH7 EG structures available, *Re*Cel7B has the highest temperature optimum, thus making it a valuable tool for protein engineering studies aimed at improving endoglucanases for biomass conversion.

Kinetic investigations at 50 °C clearly showed activity of *Re*Cel7B against Avicel. This was in contrast to an earlier work where activity against this substrate was not detected [32], perhaps as a result of low substrate loads. Under substrate excess conditions (Fig. 7D), *Tr*Cel7B was significantly more active than *Re*Cel7B, with a specificity constant about four times higher (Table 3). The addition of a CBM in *Re*Cel7BinsCBM was beneficial for the activity, and the specificity constant improved by a factor of 1.5, while the removal of the CBM in *Tr*Cel7B decreased the specificity constant to the same level as *Re*Cel7BinsCBM.

We also investigated kinetics in the opposite limit of enzyme excess (Fig. 7E). In this case, the saturation rate ($^{inv}V_{max}$) is sensitive to the ability of an enzyme to locate hydrolizable sites on the cellulose surface [41]. *Tr*Cel7B showed high inverse maximal specific rate compared to *Re*Cel7B (Table 3), suggesting that it is capable of forming more catalytically competent complexes on the cellulose surface. The strong effect of the CBM on this property was evident by an increase in $^{inv}V_{max}$ of *Re*Cel7BinsCBM and a considerable decrease for *Tr*Cel7BΔCBM (compared to the

corresponding wild type). The overall binding was also promoted by the presence of the CBM since it was measurable by the current method only for *Tr*Cel7B and *Re*Cel7BinsCBM (Fig. 7F and Table 3).

Overall, the kinetic analysis showed that *Tr*Cel7B was catalytically more efficient on Avicel compared to *Re*Cel7B. The better performance of *Tr*Cel7B was partially the result of the CBM and addition of a CBM to *Re*Cel7B lessened the difference in catalytic performance. However, the chimeric enzyme *Re*Cel7BinsCBM with the CD from *Re* and linker/CBM from *Tr*Cel7A remained distinctly inferior to the *Tr*Cel7B wild type with respect to both binding and catalysis.

We conclude that the catalytic performance of the CD was poorer for the *Re* enzyme (the same conclusion may be reached by comparing the two CBM-less enzymes in Tab. 3). The biochemical results do not suggest that *Re*Cel7A have evolved stronger ligand binding in the CD to compensate for the absence of a CBM (see Tab. 3 and Fig. 7F). This observation suggests that the higher degree of coverage of the *Re*Cel7B binding cleft found in the structural part of this work does not promote substrate binding. However, since *Re*Cel7B has not evolved to have additional domains, we cannot exclude that the linker-CBM chosen in this study is not optimal, and more systematic engineering works are needed to elucidate this aspect. Other possible roles are discussed below.

**Synergy with cellobiohydrolases**

To further investigate the function of *Re*Cel7B, we considered its ability to synergize with different CBHs and compared the results with *Tr*Cel7B (Fig. 8).

Between CBHs and EG from the same organism, the *T. reesei* mixture showed higher DS than *R. emersonii*, at all investigated mole fractions (Fig. 9). This suggests that *R. emersonii* is probably not equipped with the same efficient *endo-exo* synergy system as *T. reesei*. When *Re*Cel7B and *Tr*Cel7B were tested with the CBH from the opposite organism, the amount of EG necessary to reach the maximal DS was much higher for *Re*Cel7B than *Tr*Cel7B (Fig. 9). Thus, *Re*Cel7B did not seem to be a particularly good synergistic partner, regardless of the CBH used. *Tr*Cel7B, on the other hand, was very effective, as illustrated for example in a distinctive synergy at a fraction as little as 1% (Fig. 8A).

One possible interpretation of the synergy measurements is that the activity of the EG monocomponent is directly proportional to the synergistic capability. In other words, the EG that is kinetically superior is also the most capable to synergize with a CBH. It is interesting to put this mechanistic interpretation in a more biological context. *T. reesei* and *R. emersonii* are both

saprophytic ascomycetes, capable of degrading cellulose-based biomass [62, 63]. One important difference between them is that *T. reesei* is a mesophile that is also capable of mycotrophy [64], while *R. emersonii* is a true thermophile [25]. A possible rationalization of the limited synergistic capability of the *R. emersonii* EG-CBH system is that there has been less selective pressure for this property. While *T. reesei*'s growth requirements might have exerted strong evolutionary pressure towards the most efficient cellulose-degrading system, high thermostability may have been more important for *R. emersonii*.

## Conclusions

Structural analysis revealed that *Re*Cel7B had a number of loops with increased length and rigidity, and displayed more N-glycosylations compared to *Tr*Cel7B, which overall contribute to its high thermostability. Enzyme kinetics on Avicel indicated *Re*Cel7B is a less efficient enzyme than *Tr*Cel7B, and analyses of variants with and without CBM showed that this difference could only be partially explained by the presence of a binding module in the *Tr* wild type. Tested in mixtures with cellobiohydrolases, *Re*Cel7B was not a particularly good synergistic partner, unlike the mesophilic counterpart, which distinctly improved activity even at very low percentages. These results indicate that *Re*Cel7B has not evolved specific structural features to compensate for the lack of a CBM, but instead promoted high thermostability over activity. We surmise that these new insights on the structure-function features of GH7 EGs could be useful for a better understanding of this industrially relevant enzyme family.

## Materials and Methods

### Cloning, expression and purification of the enzymes

The enzymes *Re*Cel7B, *Re*Cel7A, *Re*Cel7BinsCBM, *TrC*el7B ,*Tr*Cel7A and *Tr*Cel7BΔCBM were heterologously expressed in *Aspergillus oryzae* [65] and purified in three chromatographic steps following a previously established protocol [66]. *Re*Cel7B (GenBank database ID: **AHL20272.1**) and *Re*Cel7A (GenBank database ID: **AAL89553.1**) are respectively the wild type endoglucanase and cellobiohydrolase from *Rasamsonia emersonii*, while *TrC*el7B (GenBank database ID: **AAA34212.1**), and *Tr*Cel7A (GenBank database ID: **CAH10320.1**) are the wild type endoglucanase and cellobiohydrolase from *Trichoderma reesei*, respectively. *Re*Cel7BinsCBM is

an engineered variant of *Re*Cel7B, where linker and CBM sequences were added from *Tr*Cel7A. *Tr*Cel7BΔCBM is an engineered variant of *TrC*el7B where the linker and CBM at the C-terminal sequence was removed. The last amino acid in the sequence of the mature product of *Tr*Cel7BΔCBM is A375. The purity of the enzymes was confirmed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) using a NuPAGE 4-12% Bis-Tris gel (GE Healthcare). Enzyme concentrations were estimated by measuring the absorbance at 280 nm and using the following theoretical molar extinction coefficients [67]: 50975 $M^{-1}cm^{-1}$ for *Re*Cel7B, 74145 $M^{-1}cm^{-1}$ for *Tr*Cel7B, 74925 $M^{-1}cm^{-1}$ for *Re*Cel7A, 86760 $M^{-1}cm^{-1}$ for *Tr*Cel7A, 57185 $M^{-1}cm^{-1}$ for *Re*Cel7BinsCBM and 62310 $M^{-1}cm^{-1}$ for *Tr*Cel7BΔCBM.

**Phylogenetic tree building**

To construct the phylogenetic tree, 87 amino acid sequences from the *characterized* and *structures* pages were retrieved from GH7 CAZy database [26]. A multiple sequence alignment was performed using ClustalW [68]. The pairwise and multiple alignment gap-opening values were set to 3 and 30, respectively and all other parameters were set to default values. The amino acid sequences corresponding to signal peptide, linker and CBM were removed from the aligned sequences (cut-off before Q18 for the signal peptide and after T411 for the linker-CBM, *Re*Cel7B numbering). The sequences were subsequently re-aligned and a phylogenetic analysis was performed using MEGA7 [69]. A phylogenetic tree was built using the Maximum Likelihood method based on the Whelan And Goldman model [70]. The tree with the highest log likelihood (-19685.6796) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites. The rate variation model allowed for some sites to be evolutionarily invariable. All positions with less than 90% site coverage were eliminated. That is, fewer than 10% alignment gaps, missing data, and ambiguous bases were allowed at any position.

**X-ray crystallography**

The high solubility of *Re*Cel7B allowed us to up-concentrate the protein to >60 mg/mL using 10 k molecular weight cut-off Vivaspin 500 centrifugal concentrators (Sartorius, Concord, CA) before initiating the crystallization screening trials. The protein was solvent exchanged with water, and

20 mM cellobiose was added prior crystallization with the hanging drop method. The screenings were carried out in 96 well plates with JCSG-plus MD1-37 (Molecular Dimensions, United Kingdom), and hits were found in 2.4 M sodium malonate pH 7. Crystals were grown with the hanging drop method by mixing in 1:1:1 ratio *Re*Cel7B 60 mg/mL, sodium malonate and cellobiose 20 mM (Sigma-Aldrich, 22150). Crystals selected for data collection were mounted and frozen in liquid nitrogen. Diffraction data was collected at the Petra III synchrotron beamline P14 (Hamburg, Germany). The data were integrated and scaled using XDS, the structure was determined by Phaser molecular replacement [71] and *Hi*Cel7B as search model (PDB database ID: **2A39**). Model building was performed using Coot [72] and refinement of the structure was performed with *phenix.refine* within the phenix package [73]. For cross validation, 5% of the data was excluded for the refinement. Water molecules, sodium malonate and carbohydrate molecules were manually added by visual inspection of the electron density maps. The structure model coordinate was deposited in PDB with accession number **6SU8**.

**Temperature and pH optima**

*Re*Cel7B, *Tr*Cel7B and *Re*Cel7BinsCBM activities at different temperatures were determined on microcrystalline cellulose (Avicel, PH101, Sigma-Aldrich). Avicel was washed seven times in deionized water by decantation, then washed twice with 50 mM sodium acetate buffer pH 5 (henceforth called standard buffer). Temperature optima were determined by mixing enzyme and Avicel in standard buffer to a final concentration of 0.5 µM and 90 g/L respectively, in a microtiter plate at 16 different temperatures (15 to 90 °C with 5 °C intervals). The final reaction volume was 250 µL. The plates were first incubated without enzyme for 20 min in a thermomixer equipped with Thermotop (Eppendorf, Hamburg, Germany) at the desired temperature, then enzyme was added and incubated for 1h with shaking at 1100 rpm. Reactions were terminated by a 3 min centrifugation at 2000 ×g. The concentration of soluble reducing ends was then determined from the supernatant using the *para*-hydroxybenzoic acid hydrazide (PAHBAH) method [74], with a procedure described elsewhere [23]. The concentration of the soluble sugar reducing ends was quantified against a 6-point cellobiose calibration curve in a concentration range between 31.25-1000 µM. The pH optimum was investigated on microcrystalline cellulose by measuring enzyme activity on Avicel mixed with 11 different buffers. The buffers used were citrate/phosphate buffers (McIlvaine buffer), prepared by mixing 0.1 M citric acid (VWR chemicals) and 0.2 M disodium hydrogen phosphate $Na_2HPO_4$ (Merck, Darmstadt) in different

proportions [75]. The pH values of the buffers used were experimentally confirmed to be the following: 2.45, 3.15, 3.5, 4.2, 4.66, 5.3, 5.7, 6.25, 6.56, 7.11, and 7.55. Avicel in water was diluted to a final concentration of 204 g/L. From this, 110 µL were withdrawn and mixed with 110 µL of the buffer at the desired pH value in a microtiter plate. The reaction was then started by addition of 30 µL of enzyme to a final concentration of 100 nM and 90 g/L of Avicel. The final reaction volume was 250 µL. The plates were incubated for 1h at 50 °C in a thermomixer operating at 1100 rpm. The reaction was stopped by centrifugation and the amount of soluble reducing ends was quantified as described above.

**Thermal unfolding at different pH values**

Thermal unfolding of *Tr*Cel7B, *Re*Cel7B and *Re*Cel7BinsCBM were monitored by measurements of their intrinsic tryptophan fluorescence as a function of temperature via differential scanning fluorimetry (nanoDSF) using a Prometheus NT.48 instrument (NanoTemper Technologies, Munich, Germany). Enzymes (1 µM) were mixed with McIlvaine buffers at different pH values (11 buffers were prepared from pH 2 to pH 7 separated by 0.5 pH units), prepared by mixing 0.2 M citric acid and 0.4 M disodium hydrogen phosphate $Na_2HPO_4$ in different proportions. The final volume was 200 µL. The enzyme-buffer mixtures were transferred to a microtiter plate and equilibrated for 20 min at 25°C in an Eppendorf Thermomixer operating at 1100 rpm. Ten µL were then transferred to capillary tubes for the analysis. All measurements were done in duplicates, with a heating rate of 3.3°C/min from 20°C to 95°C. Two wavelengths (350 and 330 nm) were recorded, and their ratio was plotted against the temperature. The first derivative was then used to determine the thermal unfolding transition midpoints ($T_m$). The results were analysed with the instrument software (PR. ThermControl v2.1.5, NanoTemper).

**Enzymatic deglycosylation and intact mass determination**

*Re*Cel7B (1.5 mg/mL) was incubated at 37°C for 12 h with 50 mU of Endoglycosidase H (Endo-H, Roche), in a total volume of 1 mL in standard buffer and 200 mM NaCl and 2 mM $CaCl_2$, followed by inactivation at 100 °C for 30 min. *Re*Cel7B treated and untreated with Endo-H was analysed for their intact molecular weight using a MAXIS II electrospray mass spectrometer (Bruker Daltonik GmbH, Bremen, Germany). The samples were diluted to 0.1 mg/mL and applied to an AdvanceBio Desalting-RP column (Agilent Technologies). Samples were eluted from the column with an acetonitrile linear gradient from 5 to 95% (v/v) and introduced to the electrospray

source with a flow of 400 mL/min by an Ultimate 3000 LC system (Thermo Fisher Scientific). Data analysis was performed with DataAnalysis version 4.3 (Bruker Daltonik GmbH, Bremen, Germany).

**Thermostability of deglycosylated ReCel7B**

*Re*Cel7B at a final concentration of 0.5 mg/mL was incubated with either 62.5 mU of Endoglycosidase H (Endo-H, Roche) or 12.5 U of Peptide-N-glycosidase F (PNGase F, Roche) in 20 mM buffer HEPES pH 7.0 in a final volume of 100 µL. A control was prepared in the same conditions but without either Endo-H or PNGase F. The reaction mixtures were incubated for 18 h at 37°C in an Eppendorf thermomixer operating at 800 rpm. After incubation, thermal unfolding was immediately measured in triplicates via differential scanning fluorimetry as described before.

**Monosaccharide composition analysis of *Re*Cel7B**

Carbohydrate composition of *Re*Cel7B was determined with acid hydrolysis to release monosaccharides. *Re*Cel7B (0.74 mg) was incubated with 2.6 M trifluoroacetic acid (TFA) for 6 h at 100 °C in 500 µL. A mixture of six sugar standards (mannose, galactose, glucose, fucose, *N*-acetylglucosamine and *N*-acetylgalactosamine) with known concentrations underwent the same procedure and were used as standards for the identification. After acid hydrolysis, the samples were cooled and dried overnight in a vacuum centrifuge at room temperature. To quantify background monosaccharides in untreated *Re*Cel7B, a control was made where TFA was added to *Re*Cel7B without incubation. All samples were reconstituted in 260 µL MQ water and quantified by HPAEC-PAD using a Dionex ICS-3000 ion chromatograph (Thermo Fisher Scientific, Waltham, MA) equipped with an electrochemical detector a CarboPac PA-1 column. Samples were eluted using the following multistep gradient program at a flow rate 0.8 mL/min: 15 mM NaOH (0-4.5 min), 35 mM NaOH (4.5-7 min), 25 mM NaOAc + 75 mM NaOH (7-10 min), 175 mM NaOAc + 75 mM NaOH (10-20 min). Data was analysed using the instrument software Chromeleon 7.2. The final monosaccharide concentrations in the protein sample were corrected by a recovery factor [76], calculated from the ratio between the TFA treated and untreated sugar standards.

**Data fitting and molecular visualization**

Data was analysed and fitted using Origin 2018 (OriginLab, Northampton, MA, USA), unless otherwise stated. Graphical representation of the structures were made using PyMOL Molecular Graphics System, version 2.1.1 (Schrödinger, LLC).

**Conventional Michaelis-Menten**

Aliquots of 230 µL of washed Avicel with final loads between 0.5 and 90 g/L were transferred to 96-well microtiter plates (96F 26960 Thermo Scientific, Waltham, MA), and the enzymatic reaction was started by adding 20 µL of enzyme stock prepared in standard buffer to a final concentration of 100 nM. Each plate was sealed and mixed at 1100 rpm in a ThermoMixer equipped with a ThermoTop (Eppendorf, Hamburg, Germany). The enzyme-substrate contact time was 60 min at 50°C. The reaction was stopped by a 3-min centrifugation at 2000 ×g. Volumes of 60 µL of the supernatant were retrieved and analysed for its content of reducing sugars by using the 4-hydroxybenzoic acid hydrazide (PAHBAH) method as described before[23]. Finally, 100 µL were transferred into a microtiter plate, and the absorbance at 405 nm was measured by a plate reader (SpectraMax M2e, Molecular Devices, Sunnyvale, CA). A six-point standard curve of 32.15-1000 µM of cellobiose in standard buffer was included in each plate.

**Binding isotherms and inverse Michaelis–Menten ($^{inv}$MM)**

To quantify enzyme adsorption on microcrystalline cellulose, a constant substrate load of Avicel was used, and the enzyme load varied between 0.05 and 4 µM. Sixty µL of enzyme stock was added to microtiter plates containing 190 µL of Avicel so the final substrate concentration was 15 g/L. The plates were placed in a ThermoMixer for 1 h at 50°C and then centrifuged at 2000 ×g for 3 min to separate free from substrate-bound enzyme. Sixty µL of supernatant was then mixed with 90 µL of standard buffer in a black microtiter plate (Greiner bio-one 655079), and the enzyme concentration was measured by intrinsic protein fluorescence in a plate reader by using excitation and emission wavelengths of 280 and 345 nm, respectively. The free enzyme concentration was quantified by comparing the fluorescence signal with a calibration curve made with enzyme dissolved in standard buffer with known concentrations ranging from 0.05 to 5 µM. In the inverse MM approach, the reaction condition was the same as for the binding isotherms, except that the final Avicel concentration was 8 g/L. The reactions were incubated and analysed for the reducing sugar content with the PAHBAH method as described before. Since high enzyme concentrations affect the absorbance signal, appropriate enzyme blanks were included for each concentration.

**Synergy between EGs and CBHs**

We used a final volume of 200 µL, 10 g/L Avicel in standard buffer and 1 h contact time at 25°C. In all the assays containing synergy mixtures, one CBH was mixed with one EG and the total enzyme concentration was kept constant to 4 µM, while the molar ratio of the two components was varied systematically, ranging from 2% to 99% CBH. Reference assays were conducted with each enzyme used as monocomponent, at concentrations ranging from 0 to 4 µM. The concentrations in these experiments matched the concentration of the component in the corresponding synergy mixtures. Enzyme activity was quantified by measuring the soluble reducing-end concentrations by PAHBAH method, as described before. All experiments were performed in triplicates, unless otherwise specified.

**Analytical size-exclusion chromatography of *Re*Cel7B**

Size-exclusion chromatography of *Re*Cel7B was performed using an AKTA system (GE Healthcare, USA). The column used was a Superdex-200 10/300 GL, equilibrated with a sodium acetate 50 mM buffer 100 mM NaCl operating at a flow rate of 0.7 mL/min. The absorbance at 280 nm was monitored. The relative elution was calculated as $K_{av}=(V_e-V_0)/(V_c-V_0)$, with $K_{av}$ equilibrium distribution coefficient, $V_e$ elution volume of the sample, $V_0$ void volume (8.51 mL), $V_c$ geometric column volume (23.56 mL). A calibration curve was created by plotting the log of the molecular weight vs $K_{av}$ of the following molecular weight standards: Thyroglobulin (669 kDa), Aldolase (158 kDa), Conalbumin (75 kDa), Ovalbumin (44 kDa).

**Acknowledgement**

This work was supported by the Innovation Fund Denmark (grant number 5150-00020B), the Carlsberg Foundation and the Novo Nordisk Foundation. We also thank Senior Research Associate Gemma Lindemann at Novozymes for help in performing the analytical size-exclusion chromatography.

**Author contributions**

CSDC, PW, KB and JPM conceived and designed the study. CSDC designed and purified the enzymes, planned and carried out the experiments and analysed the data. THS, MSW and AMC carried out the cloning and expression of the enzymes. BK carried out the intact mass

spectrometry and monosaccharide composition analysis. JPM carried out the crystallization trials, diffraction data collection and structural refinement, with the support from CSDC. SJC created the phylogenetic tree. CSDC, PW and JPM wrote the manuscript. All authors reviewed and approved the manuscript.

Reference List

1. López-Mondéjar R, Zühlke D, Becher D, Riedel K & Baldrian P (2016) Cellulose and hemicellulose decomposition by forest soil bacteria proceeds by the action of structurally variable enzymatic systems, *Sci Rep.* **6**, 25279.

2. Lynd LR, Laser MS, Bransby D, Dale BE, Davison B, Hamilton R, Himmel M, Keller M, McMillan JD & Sheehan J (2008) How biotech can transform biofuels, *Nat Biotechnol.* **26**, 169.

3. Viikari L, Alapuranen M, Puranen T, Vehmaanperä J & Siika-Aho M (2007) Thermostable enzymes in lignocellulose hydrolysis in *Biofuels* pp. 121-145, Springer.

4. Hobdey SE, Knott BC, Haddad Momeni M, Taylor LE, 2nd, Borisova AS, Podkaminer KK, VanderWall TA, Himmel ME, Decker SR, Beckham GT & Stahlberg J (2016) Biochemical and Structural Characterizations of Two Dictyostelium Cellobiohydrolases from the Amoebozoa Kingdom Reveal a High Level of Conservation between Distant Phylogenetic Trees of Life, *Appl Environ Microbiol.* **82**, 3395-409.

5. Nakamura A, Watanabe H, Ishida T, Uchihashi T, Wada M, Ando T, Igarashi K & 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.

6. Zhang YH & Lynd LR (2004) Toward an aggregated understanding of enzymatic hydrolysis of cellulose: noncomplexed cellulase systems, *Biotechnol Bioeng.* **88**, 797-824.

7. Bubner P, Plank H & Nidetzky B (2013) Visualizing cellulase activity, *Biotechnol Bioeng.* **110**, 1529-49.

8. Kurašin M & Väljamäe P (2011) Processivity of cellobiohydrolases is limited by the substrate, *J Biol Chem.* **286**, 169-177.

9. Ståhlberg J, Johansson G & Pettersson G (1993) Trichoderma reesei has no true exo-cellulase: all intact and truncated cellulases produce new reducing end groups on cellulose, *Biochim Biophys Acta.* **1157**, 107-113.

10. Saloheimo M & Pakula TM (2012) The cargo and the transport system: secreted proteins and protein secretion in Trichoderma reesei (Hypocrea jecorina), *Microbiology.* **158**, 46-57.

11. Adav SS, Chao LT & Sze SK (2012) Quantitative secretomic analysis of Trichoderma reesei strains reveals enzymatic composition for lignocellulosic biomass degradation, *Molecular cellular proteomics.* **11**, M111. 012419.

12. Henrissat B, Driguez H, Viet C & Schülein M (1985) Synergism of cellulases from Trichoderma reesei in the degradation of cellulose, *Biotechnology.* **3**, 722.

13. Nidetzky B, Steiner W, Hayn M & Claeyssens M (1994) Cellulose hydrolysis by the cellulases from Trichoderma reesei: a new model for synergistic interaction, *Biochem J.* **298**, 705-710.
14. Eriksson T, Karlsson J & Tjerneld F (2002) A model explaining declining rate in hydrolysis of lignocellulose substrates with cellobiohydrolase I (Cel7A) and endoglucanase I (Cel7B) of Trichoderma reesei, *Appl Biochem Biotechnol.* **101**, 41-60.
15. Várnai A, Siika-aho M & Viikari L (2013) Carbohydrate-binding modules (CBMs) revisited: reduced amount of water counterbalances the need for CBMs, *Biotechnology for biofuels.* **6**, 30.
16. Johansson G, Ståhlberg J, Lindeberg G, Engström Å & Pettersson G (1989) Isolated fungal cellulose terminal domains and a synthetic minimum analogue bind to cellulose, *FEBS Lett.* **243**, 389-393.
17. Lehtiö J, Sugiyama J, Gustavsson M, Fransson L, Linder M & Teeri TT (2003) The binding specificity and affinity determinants of family 1 and family 3 cellulose binding modules, *Proceedings of the National Academy of Sciences.* **100**, 484-489.
18. Hilden L & Johansson G (2004) Recent developments on cellulases and carbohydrate-binding modules with cellulose affinity, *Biotechnol Lett.* **26**, 1683-1693.
19. Divne C, Stahlberg J, Reinikainen T, Ruohonen L, Pettersson G, Knowles JK, Teeri TT & Jones TA (1994) The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from Trichoderma reesei, *Science.* **265**, 524-8.
20. Kleywegt GJ, Zou JY, Divne C, Davies GJ, Sinning I, Stahlberg J, Reinikainen T, Srisodsuk M, Teeri TT & Jones TA (1997) The crystal structure of the catalytic core domain of endoglucanase I from Trichoderma reesei at 3.6 A resolution, and a comparison with related enzymes, *J Mol Biol.* **272**, 383-97.
21. Taylor CB, Payne CM, Himmel ME, Crowley MF, McCabe C & Beckham GT (2013) Binding site dynamics and aromatic–carbohydrate interactions in processive and non-processive family 7 glycoside hydrolases, *The Journal of Physical Chemistry* **117**, 4924-4933.
22. von Ossowski I, Ståhlberg J, Koivula A, Piens K, Becker D, Boer H, Harle R, Harris M, Divne C, Mahdi S, Zhao Y, Driguez H, Claeyssens M, Sinnott ML & Teeri TT (2003) Engineering the Exo-loop of Trichoderma reesei Cellobiohydrolase, Cel7A. A comparison with Phanerochaete chrysosporium Cel7D, *J Mol Biol.* **333**, 817-829.
23. Schiano-di-Cola C, Røjel N, Jensen K, Kari J, Sørensen TH, Borch K & 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.
24. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN & Bourne PE (2000) The protein data bank, *Nucleic Acids Res.* **28**, 235-242.
25. Houbraken J, Spierenburg H & Frisvad JC (2012) Rasamsonia, a new genus comprising thermotolerant and thermophilic Talaromyces and Geosmithia species, *Antonie Van Leeuwenhoek.* **101**, 403-21.

26. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM & Henrissat B (2013) The carbohydrate-active enzymes database (CAZy) in 2013, *Nucleic Acids Res.* **42**, D490-D495.
27. Watanabe H, Nakashima K, Saito H & Slaytor M (2002) New endo-β-1, 4-glucanases from the parabasalian symbionts, Pseudotrichonympha grassii and Holomastigotoides mirabile of Coptotermes termites, *Cellular Molecular Life Sciences* **59**, 1983-1992.
28. Sulzenbacher G, Driguez H, Henrissat B, Schülein M & Davies GJ (1996) Structure of the Fusarium oxysporum endoglucanase I with a nonhydrolyzable substrate analogue: substrate distortion gives rise to the preferred axial orientation for the leaving group, *Biochemistry.* **35**, 15280-15287.
29. Davies GJ, Ducros V, Lewis RJ, Borchert TV & Schülein M (1997) Oligosaccharide specificity of a family 7 endoglucanase: insertion of potential sugar-binding subsites, *J Biotechnol.* **57**, 91-100.
30. Sonoda MT, Godoy AS, Pellegrini VOA, Kadowaki MAS, Nascimento AS & Polikarpov I (2019) Structure and dynamics of Trichoderma harzianum Cel7B suggest molecular architecture adaptations required for a wide spectrum of activities on plant cell wall polysaccharides, *Biochimica et Biophysica Acta (BBA) - General Subjects.* **1863**, 1015-1026.
31. Nakashima K, Watanabe H & Azuma J-I (2002) Cellulase genes from the parabasalian symbiont Pseudotrichonympha grassii in the hindgut of the wood-feeding termite Coptotermes formosanus, *Cellular Molecular Life Sciences* **59**, 1554-1560.
32. Wang K, Luo H, Shi P, Huang H, Bai Y & Yao B (2014) A highly-active endo-1, 3-1, 4-β-glucanase from thermophilic Talaromyces emersonii CBS394. 64 with application potential in the brewing and feed industries, *Process Biochem.* **49**, 1448-1456.
33. Pearl FM, Bennett C, Bray JE, Harrison AP, Martin N, Shepherd A, Sillitoe I, Thornton J & Orengo CA (2003) The CATH database: an extended protein family resource for structural and functional genomics, *Nucleic Acids Res.* **31**, 452-455.
34. Payne CM, Knott BC, Mayes HB, Hansson H, Himmel ME, Sandgren M, Stahlberg J & Beckham GT (2015) Fungal cellulases, *Chem Rev.* **115**, 1308-448.
35. Momeni MH, Payne CM, Hansson H, Mikkelsen NE, Svedberg J, Engstrom A, Sandgren M, Beckham GT & Stahlberg J (2013) Structural, biochemical, and computational characterization of the glycoside hydrolase family 7 cellobiohydrolase of the tree-killing fungus Heterobasidion irregulare, *J Biol Chem.* **288**, 5861-72.
36. Ducros VM, Tarling CA, Zechel DL, Brzozowski AM, Frandsen TP, von Ossowski I, Schülein M, Withers SG & Davies GJ (2003) Anatomy of glycosynthesis: structure and kinetics of the Humicola insolens Cel7B E197A and E197S glycosynthase mutants, *Chem Biol.* **10**, 619-628.
37. Sulzenbacher G, Schülein M & Davies GJ (1997) Structure of the endoglucanase I from Fusarium oxysporum: Native, cellobiose, and 3, 4-epoxybutyl β-D-cellobioside-inhibited forms, at 2.3 Å resolution, *Biochemistry.* **36**, 5902-5911.

38. Taylor CB, Payne CM, Himmel ME, Crowley MF, McCabe C & Beckham GT (2013) Binding site dynamics and aromatic-carbohydrate interactions in processive and non-processive family 7 glycoside hydrolases, *J Phys Chem B.* **117**, 4924-33.
39. Blanchard S, Armand S, Couthino P, Patkar S, Vind J, Samain E, Driguez H & Cottaz S (2007) Unexpected regioselectivity of Humicola insolens Cel7B glycosynthase mutants, *Carbohydr Res.* **342**, 710-716.
40. Lin Y, Beckham GT, Himmel ME, Crowley MF & Chu J-W (2013) Endoglucanase Peripheral Loops Facilitate Complexation of Glucan Chains on Cellulose via Adaptive Coupling to the Emergent Substrate Structures, *The Journal of Physical Chemistry B.* **117**, 10750-10758.
41. Kari J, Andersen M, Borch K & Westh P (2017) An Inverse Michaelis–Menten Approach for Interfacial Enzyme Kinetics, *ACS Catalysis.* **7**, 4904-4914.
42. Wang Y, Zhang S, Song X & Yao L (2016) Cellulose chain binding free energy drives the processive move of cellulases on the cellulose surface, *Biotechnol Bioeng.* **113**, 1873-1880.
43. Vlasenko E, Schülein M, Cherry J & Xu F (2010) Substrate specificity of family 5, 6, 7, 9, 12, and 45 endoglucanases, *Bioresour Technol.* **101**, 2405-2411.
44. Li D-C & Papageorgiou AC (2019) Cellulases from thermophilic fungi: recent insights and biotechnological potential in *Fungi in Extreme Environments: Ecological Role and Biotechnological Significance* pp. 395-417, Springer.
45. Taylor TJ & Vaisman II (2010) Discrimination of thermophilic and mesophilic proteins, *BMC Struct Biol.* **10**, S5.
46. Zhang S, Wang Y, Song X, Hong J, Zhang Y & Yao L (2014) Improving Trichoderma reesei Cel7B thermostability by targeting the weak spots, *Journal of chemical information modeling.* **54**, 2826-2833.
47. Mitrovic A, Flicker K, Steinkellner G, Gruber K, Reisinger C, Schirrmacher G, Camattari A & Glieder A (2014) Thermostability improvement of endoglucanase Cel7B from Hypocrea pseudokoningii, *J Mol Catal B: Enzym.* **103**, 16-23.
48. Hui JP, White TC & Thibault P (2002) Identification of glycan structure and glycosylation sites in cellobiohydrolase II and endoglucanases I and II from Trichoderma reesei, *Glycobiology.* **12**, 837-849.
49. Nakao Y, Kozutsumi Y, Funakoshi I, Kawasaki T, Yamashina I, Mutsaers JH, Van Halbeek H & Vliegenthart JF (1987) Structures of oligosaccharides on β-galactosidase from Aspergillus oryzae, *The journal of biochemistry.* **102**, 171-179.
50. Deshpande N, Wilkins MR, Packer N & Nevalainen H (2008) Protein glycosylation pathways in filamentous fungi, *Glycobiology.* **18**, 626-637.
51. Borisova AS, Eneyskaya EV, Jana S, Badino SF, Kari J, Amore A, Karlsson M, Hansson H, Sandgren M, Himmel ME, Westh P, Payne CM, Kulminskaya AA & Stahlberg J (2018) Correlation of structure,

function and protein dynamics in GH7 cellobiohydrolases from Trichoderma atroviride, T. reesei and T. harzianum, *Biotechnol Biofuels.* **11**, 5.

52. Klibanov AM (1983) Stabilization of enzymes against thermal inactivation in *Adv Appl Microbiol* pp. 1-28, Elsevier.

53. Greene ER, Himmel ME, Beckham GT & Tan Z (2015) Glycosylation of cellulases: engineering better enzymes for biofuels in *Adv Carbohydr Chem Biochem* pp. 63-112, Elsevier.

54. Amore A, Knott BC, Supekar NT, Shajahan A, Azadi P, Zhao P, Wells L, Linger JG, Hobdey SE, Vander Wall TA, Shollenberger T, Yarbrough JM, Tan Z, Crowley MF, Himmel ME, Decker SR, Beckham GT & Taylor LE, 2nd (2017) Distinct roles of N- and O-glycans in cellulase activity and stability, *Proc Natl Acad Sci U S A.* **114**, 13667-13672.

55. Beckham GT, Dai Z, Matthews JF, Momany M, Payne CM, Adney WS, Baker SE & Himmel ME (2012) Harnessing glycosylation to improve cellulase activity, *Curr Opin Biotechnol.* **23**, 338-345.

56. Maheshwari R, Bharadwaj G & Bhat MKJMMBR (2000) Thermophilic fungi: their physiology and enzymes. **64**, 461-488.

57. Schulein M (1993) Humicola insolens alkaline cellulases, *Found Biotech Ind Ferment Res.* **8**, 109-116.

58. Pellegrini VO, Serpa VI, Godoy AS, Camilo CM, Bernardes A, Rezende CA, Junior NP, Cairo JPLF, Squina FM & Polikarpov I (2015) Recombinant Trichoderma harzianum endoglucanase I (Cel7B) is a highly acidic and promiscuous carbohydrate-active enzyme, *Applied microbiology biotechnology.* **99**, 9591-9604.

59. Kumar P, Barrett DM, Delwiche MJ & Stroeve P (2009) Methods for pretreatment of lignocellulosic biomass for efficient hydrolysis and biofuel production, *Industrial engineering chemistry research.* **48**, 3713-3729.

60. Drissen R, Maas R, Tramper J & Beeftink H (2009) Modelling ethanol production from cellulose: separate hydrolysis and fermentation versus simultaneous saccharification and fermentation, *Biocatalysis Biotransformation.* **27**, 27-35.

61. Haki G & Rakshit S (2003) Developments in industrially important thermostable enzymes: a review, *Bioresour Technol.* **89**, 17-34.

62. Coughlan MP, Folan MA, Mchale A, Considine PJ & Moloney AP (1984) The Talaromyces emersonii enzyme system, *Appl Biochem Biotechnol.* **9**, 331-332.

63. Peterson R & Nevalainen H (2012) Trichoderma reesei RUT-C30--thirty years of strain improvement, *Microbiology.* **158**, 58-68.

64. Druzhinina I & Kubicek C (2016) Familiar stranger: ecological genomics of the model saprotroph and industrial enzyme producer Trichoderma reesei breaks the stereotypes in *Adv Appl Microbiol* pp. 69-147, Elsevier.

65. Borch K, Jensen K, Krogh K, McBrayer B, Westh P, Kari J, Olsen J, Sørensen TH, Windahl MS & Xu H (2014) Cellobiohydrolase variants and polynucleotides encoding same in *World Intellectual Property Organization Patent WO2014138672*

66. Sorensen TH, Cruys-Bagger N, Windahl MS, Badino SF, Borch K & Westh P (2015) Temperature Effects on Kinetic Parameters and Substrate Affinity of Cel7A Cellobiohydrolases, *J Biol Chem.* **290**, 22193-202.

67. Pace CN, Vajdos F, Fee L, Grimsley G & Gray T (1995) How to measure and predict the molar absorption coefficient of a protein, *Protein Sci.* **4**, 2411-2423.

68. Thompson JD, Higgins DG & Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res.* **22**, 4673-4680.

69. Kumar S, Stecher G & Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets, *Molecular biology evolution.* **33**, 1870-1874.

70. Whelan S & Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach, *Molecular biology evolution.* **18**, 691-699.

71. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC & Read RJ (2007) Phaser crystallographic software, *J Appl Crystallogr.* **40**, 658-674.

72. Emsley P, Lohkamp B, Scott WG & Cowtan K (2010) Features and development of Coot, *Acta Crystallogr D Biol Crystallogr.* **66**, 486-501.

73. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ & Grosse-Kunstleve RW (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallogr Sect D Biol Crystallogr.* **66**, 213-221.

74. Lever M (1973) Colorimetric and fluorometric carbohydrate determination with p-hydroxybenzoic acid hydrazide, *Biochem Med.* **7**, 274-281.

75. McIlvaine T (1921) A buffer solution for colorimetric comparison, *J Biol Chem.* **49**, 183-186.

76. Hardy MR & Townsend RR (1994) High-pH anion-exchange chromatography of glycoprotein-derived carbohydrates in *Methods Enzymol* pp. 208-225, Elsevier.

77. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M & Söding J (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, *Mol Syst Biol.* **7**, 539.

78. Robert X & Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server, *Nucleic Acids Res.* **42**, W320-W324.

## Tables

**Table 1.** Statistics related to the X-ray data collection and refinement for *Re*Cel7B structure

| Data collection | |
|---|---|
| Protein Data Bank accession number | 6SU8 |
| Beam line | P14 Petra III |
| Wavelength (Å) | 0.9802 |
| Resolution range (Å)[a] | 198 - 2.48 (2.57 - 2.48) |
| Space group | C 2 2 21 |
| Unit cell (Å) | 113.4 263.2 197.8 |
| Molecules per asymmetric unit | 3 |
| Total reflections [a] | 762235 (75632) |
| Unique reflections [a] | 104816 (10160) |
| Multiplicity [a] | 7.3 (7.4) |
| Completeness (%)[a] | 99.7 (97.8) |
| I/σI [a] | 8.9 (1.9) |
| Wilson B-factor | 53.02 |
| $R_{merge}$[a] | 0.125 (0.839) |
| $R_{meas}$[a] | 0.135 (0.900) |
| $R_{pim}$[a] | 0.05 (0.33) |
| CC1/2[a] | 0.93 (0.58) |
| **Refinement** | |
| Resolution range (Å)[a] | 29.78 - 2.48 (2.57 - 2.48) |
| Reflections used for $R_{free}$[a] | 5113 (487) |
| $R_{work}$[a] | 0.182 (0.318) |
| $R_{free}$[a] | 0.206 (0.330) |
| CC(work)[a] | 0.96 (0.77) |
| CC(free)[a] | 0.94 (0.76) |
| Number of non-hydrogen atoms | 9359 |
| macromolecules | 8775 |
| ligands | 299 |
| solvent | 285 |
| Protein residues | 1188 |
| Root mean square deviation (bonds) (Å) | 0.010 |
| Root mean square deviation (angles) (°) | 1.18 |
| Ramachandran favored (%) | 97.63 |

| | |
|---|---|
| Ramachandran allowed (%) | 1.95 |
| Ramachandran outliers (%) | 0.42 |
| Rotamer outliers (%) | 2.13 |
| Clashscore | 5.27 |
| Average B-factor (Å$^2$) | 63.2 |
| Number of TLS groups | 13 |

[a]Statistics for the highest-resolution shell are shown in parentheses.

**Table 2.** Monosaccharide composition analysis for *Re*Cel7B. Monosaccharides were determined by HPAEC-PAD on *Re*Cel7B treated with Endo-H. Data indicates the average and standard deviation from three hydrolysis experiments.

| **Monosaccharide** | **Amount (μg $mg^{-1}$ protein)** |
|---|---|
| Mannose | 85.0 ± 1.4 |
| Galactose | 27 ± 9 |
| Glucosamine | 42 ± 1 |
| Galactosamine | ND* |
| Glucose | ND* |

*ND:not detected

**Table 3**. Kinetic and adsorption parameters for *Tr*Cel7B, *Re*Cel7B, *Tr*Cel7BΔCBM and *Re*Cel7BinsCBM on Avicel at 50 °C, obtained from the regression analyses shown in Fig. 7. Standard errors are included.

| Enzyme | Specificity constant | Inverse Michaelis-Menten | | Binding isotherms | |
|---|---|---|---|---|---|
| | $\eta$ | $^{inv}V_{Max}/S_0$ | $^{inv}K_M$ | $\Gamma_{max}$ | $K_d$ |
| | $gL^{-1}s^{-1}\times10^{-3}$ | $\mu molg^{-1}s^{-1} \times 10^{-3}$ | $\mu M$ | $\mu mol/g$ | $\mu M$ |
| *Tr*Cel7B | 41 ± 1.2 | 7.5 ± 0.14 | 0.16 ± 0.013 | 0.18 ± 0.009 | 1.4 ± 0.16 |
| *Tr*Cel7BΔCBM | 15 ± 0.5 | 2.6 ± 0.10 | 0.16 ± 0.026 | ND* | ND* |
| *Re*Cel7B | 10 ± 0.9 | 3.3 ± 0.24 | 0.58 ± 0.127 | ND* | ND* |
| *Re*Cel7BinsCBM | 15 ± 2.8 | 3.5 ± 0.19 | 0.17 ± 0.037 | 0.10 ± 0.028 | 4.0 ± 1.99 |

*ND: not detectable.

**Fig. 1**. **Phylogenetic tree of GH7 and EGs sequence alignment.** (A) Phylogenetic tree of GH7 biochemically characterized sequences from CAZy database (www.cazy.org)[26]. Each sequence is indicated with the GenBank accession number (www.ncbi.nlm.nih.gov/genbank) followed by the organism name. Only for *Phanerochaete chrysosporium* Q7LIJ0 and *Humicola insolens* P56680 the Uniprot accession number (www.uniprot.org) is indicated instead. The taxonomic classification of the source organism is indicated for each sequence in different colors, according to the legend in the corner. Bold letters indicate whether the organism is a thermophile. Sequences marked with a star have the crystal structure solved, with *Re*Cel7B highlighted in red. Bootstrap values are shown at the nodes as percentage of 100 replicates. The scale bar is 0.2 amino acid substitutions per site. **(B)** Structure-based amino acid sequence alignment of the catalytic domain of five GH7 EGs with experimentally-determined structure: *Humicola insolens* Cel7B (*Hi*Cel7B PDB: **2A39**), *Fusarium oxisporium* Cel7B (*Fo*Cel7B PDB: **1OVW**), *Trichoderma harzianum* Cel7B (*Th*Cel7B PDB: **5W0A**), *Trichoderma reesei* Cel7B (*Tr*Cel7B PDB: **1EG1**) and *Rasamsonia emersonii* Cel7B (*Re*Cel7B PDB: **6SU8**). Secondary structure elements of *Hi*Cel7B are shown on top of the alignment with the following figures and symbols: alpha helices (helices, α and η for a $3_{10}$-helix), beta strands (arrows, β), turns (T, strict β-turns as TT and strict α-turns as TTT). Strictly identical residues are marked in white characters on a red background, while chemically similar residues are shown as red characters. Regions of conserved, highly similar residues are framed in blue boxes. Green frames indicate loop regions of interest, with loop nomenclature below and residue position in *Re*Cel7B. Black stars indicate the residues involved in the catalysis (E199, D201 and E204 in *Re*Cel7B). Asparagine residues where *N*-glycosylation was identified in the different structures are framed in magenta, and the residue numbering for *Re*Cel7B is indicated below. Alignment created with Clustal Omega (www.ebi.ac.uk) [77] and rendered with ESPript 3.0 web server with default parameters (www.espript.ibcp.fr)[78].

**Fig. 2. *Re*Cel7B structure and comparison with *Tr*Cel7B and *Th*Cel7B. (A)** Cartoon structural representation of *Re*Cel7B (orange), with loops of interest highlighted (blue). Catalytic residues are marked as sticks, and glycosylations are shown in light gray. Loop nomenclature (A1-A3 B1-B3) and the one-letter code of both catalytic residues and glycosylated asparagines is indicated**. (B)** *Re*Cel7B structure rotated by -115 ° on the horizontal plane, with emphasis on the glycosylation at N185 and neighbouring loops S1 and S2 (red and cyan, respectively). **(C)** *Re*Cel7B structure superimposed with *Tr*Cel7B (green, PDB **1EG1**). Loop of interest in are highlighted in blue and magenta for *Re*Cel7B and *Tr*Cel7B, respectively. The two glycosylations in *Tr*Cel7B are indicated and the glycans shown as dark green sticks (N56 and N182, conserved both in *Tr*Cel7B and *Th*Cel7B**. (D)** The glycosylation at N185 (light gray sticks) and its stabilizing interactions. The first NAG and water molecules (red spheres) are actively engaging in stabilizing three adjacent loops S1 (red) S2 (cyan) and B3 (blue) via hydrogen bonding (represented as dotted lines). A water molecule stabilizes loop S1, S221 in loop B3 and D154 in loop S2. The 6′O of the NAG interacts with R159 of loop S2. The close residues PCA and Q2 at the N-terminal are indicated**. (E)** *Re*Cel7B (orange) superimposed to *Tr*Cel7B (green) and *Th*Cel7B (gray, PDB **5W0A**). The lysine K238 in *Re*Cel7B is in hydrogen bond distance with a cysteine (C230) and a glutamine (Q177). In *Tr*Cel7B and *Th*Cel7B, the glutamine (Q174 in both cases) is conserved, while lysine is replaced by a serine (S221) in *Tr*Cel7B and an alanine (A221) in *Th*Cel7B. The cysteine interaction is lost in *Tr*Cel7B and *Th*Cel7B because of the absence of the B3 loop. **(F)** Close to the B1 loop, the two aromatics Y38 and W40, both present in *Tr*Cel7B and *Th*Cel7B, are substituted to S37 and P39, respectively, in *Re*Cel7B. **(G)** The unit cell of *Re*Cel7B consists of a trimer decorated with glycans (gray). The trimer interface is mainly formed by the flexible loops. **(H)** The crystal lattice of *Re*Cel7B consists of repeating units of the trimer. It has high solvent content, which allows the extended glycans to remain solvent exposed. The structures were visualized and processed in PyMOL (The PyMOL Molecular Graphics System, Version 2.3.2, Schrödinger, LLC).

**Fig. 3. Analytical size-exclusion chromatogram of *Re*Cel7B (orange continuous line) and molecular weight standards (gray dotted line).** *Re*Cel7B eluted in one symmetric peak with an experimental molecular weight of ~60 kDa (arrow), corresponding to a monomeric state in solution. Peak identity: 1, thyroglobulin (669 kDa), 2, aldolase (158 kDa), 3, conalbumin (75 kDa), 4, ovalbumin (44 kDa). The apparent molecular weight of *Re*Cel7B was calculated using a calibration curve of the protein standards (insert).

**Fig. 4. Omit map of *Re*Cel7B glycosylations and active site residues.** Difference electron density ($F_0-F_C$) simulated annealed omit map contoured at 2σ level (white mesh) and contoured at 4σ level (green mesh) of *Re*Cel7B glycosylation sites shown as white sticks and the glycosylated asparagines as orange sticks for (A) N28, (B) N77, (C) N185, (D) N277, (E) N363 and (F) the active site residues E199, D201, E204 and close aromatic residues W344 and W353. Only when visible density above 4σ level was available the glycan was extended, in the case of N185 and only chain A has the additional mannose, while it was left out in chain B and C, and for N363 chain B and C has an additional N-acetylglucosamine. Electron density is only shown for chain A but equivalent densities are visible in chain B and C.

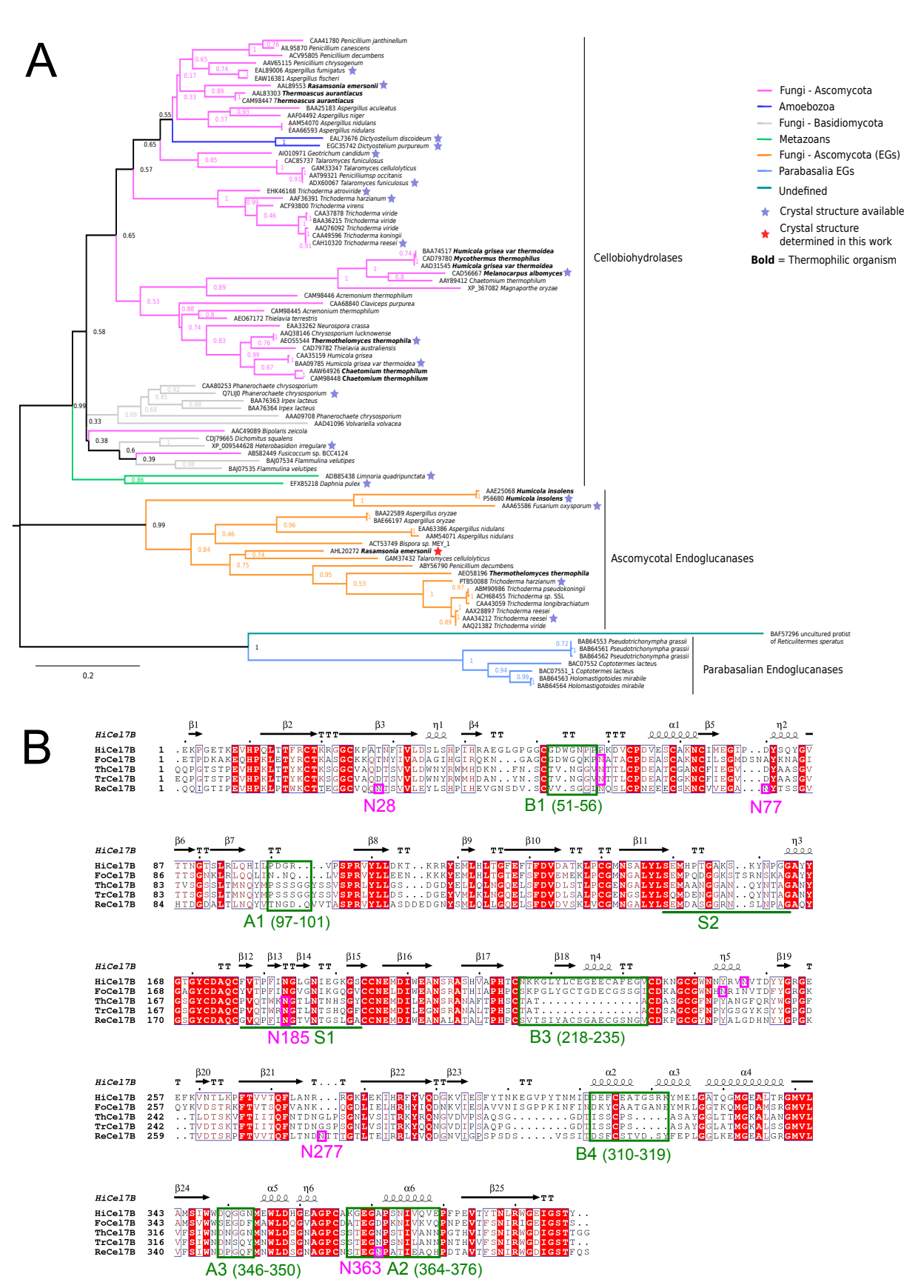
**Fig. 5.** Deconvoluted mass spectrum of deglycosylated *Re*Cel7B (upper panel) and native *Re*Cel7B (lower panel). Deglycosylation was performed enzymatically using Endo-H. Mass values for the highest intensity peaks are indicated with an arrow. The observed molecular weight of deglycosylated *Re*Cel7B corresponds to the theoretical one with the addition of 5 N-acetylated hexoses and a pyroglutamic acid modification, in accordance with the specificity of Endo-H. Difference between the main peaks of intact and deglycosylated *Re*Cel7B corresponds to a mass shift of five N-acetylated hexoses and 44 hexoses. Mass differences between the peaks in intact *Re*Cel7B correspond to different amount of hexoses additions, indicating glycosylation heterogeneity in *Re*Cel7B.

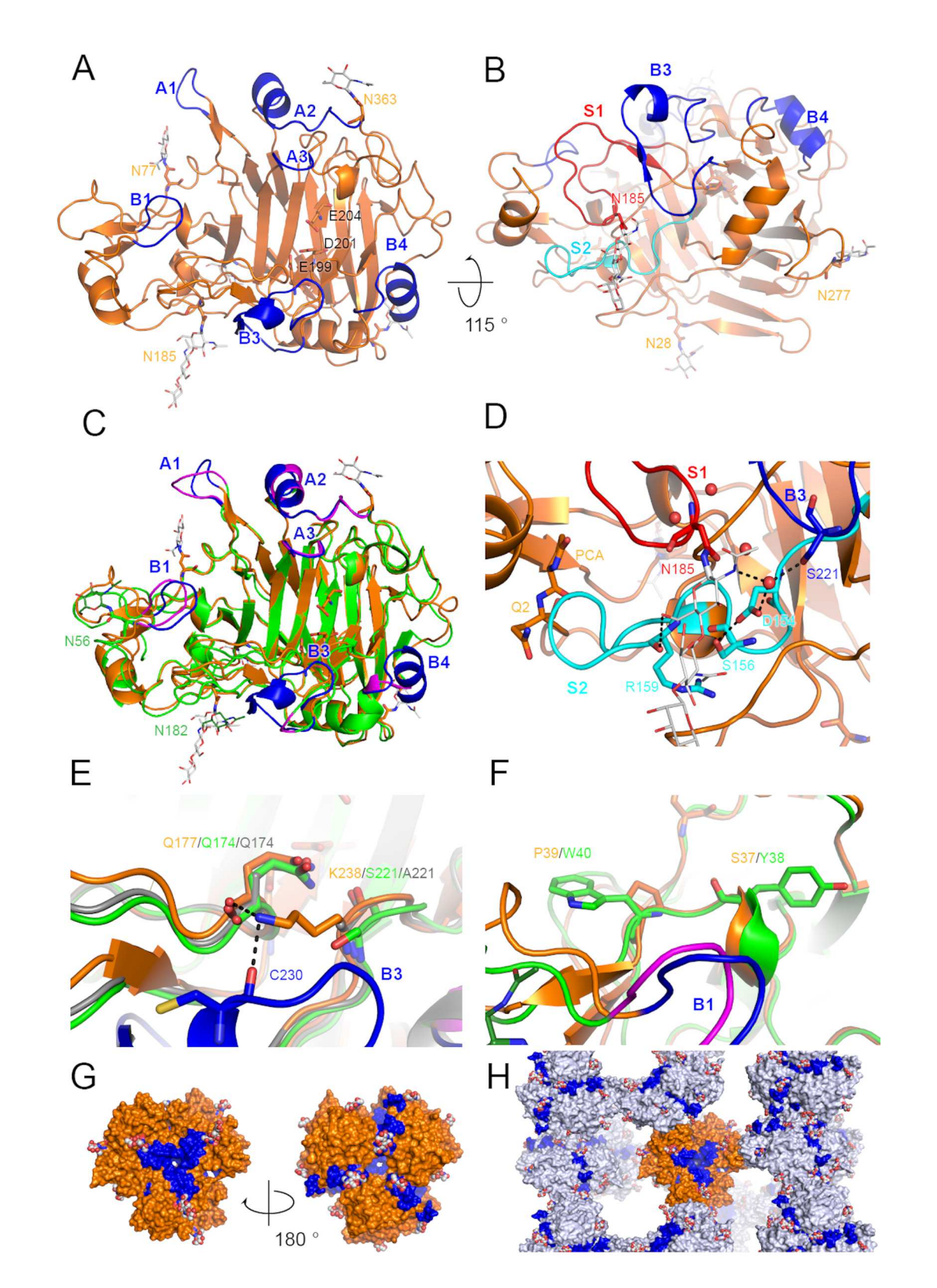
**Fig. 6.** HPAEC-PAD chromatogram profiles of *Re*Cel7B subjected to acid hydrolysis to release monosaccharides (blue continuous line) and a standard mixture of monosaccharides (black dotted line) treated in the same way. Peak identity: A, Fucose; B, *N*-acetyl galactosamine; C, *N*-acetyl glucosamine; D, Galactose; E, Glucose; F, Mannose. Peaks appeared after 12 min are considered impurities coming from the strong acid treatment of *Re*Cel7B.

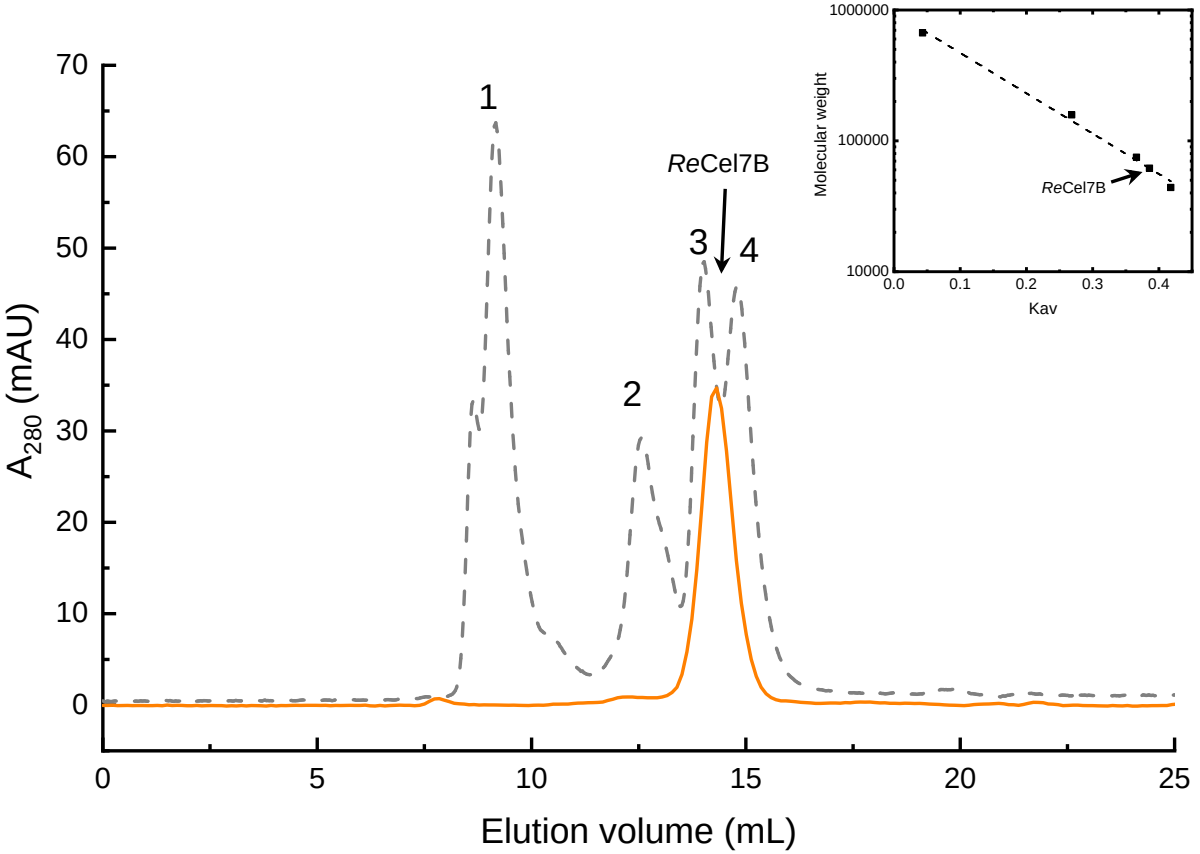
**Fig. 7. Biochemical characterization, enzyme kinetics and binding isotherms. (A)** Effect of temperature on enzyme activity for *Re*Cel7B (orange up-pointing triangles), *Tr*Cel7B (green squares) and *Re*Cel7BinsCBM (black down-pointing triangles), Conditions: Avicel 90 g/L, 0.5 µM enzymes. **(B)** Effect of pH on *Re*Cel7B and *Tr*Cel7B activity. Conditions: Avicel 90 g/L, 0.1 µM enzymes. **(C)** Thermostability, expressed as melting temperature values ($T_m$) as a function of pH for *Re*Cel7B, *Tr*Cel7B and *Re*Cel7BinsCBM. **(D)** Initial rates at low enzyme concentration and increasing substrate loads for *Re*Cel7B, *Re*Cel7BinsCBM, *Tr*Cel7B and *Tr*Cel7BΔCBM (blue circles). Conditions: 100 nM enzyme, substrate loads 0.5-90 g/L. A linear fit of the initial values (insert) was performed to extract the specificity constant. **(E)** Inverse Michaelis-Menten approach for *Re*Cel7B, *Re*Cel7BinsCBM, *Tr*Cel7B and *Tr*Cel7BΔCBM. Conditions: substrate load of 8 g/L, enzyme concentrations 0.05-4 µM. **(F)** Binding isotherms for *Re*Cel7BinsCBM and *Tr*Cel7B on Avicel 15 g/L. In all cases, symbols represent experimental data, error bars indicate standard deviation from triplicate measurements, and lines are a linear fit, best fits of Eq. 2 and Eq. 3 for panel D, E and F, respectively.

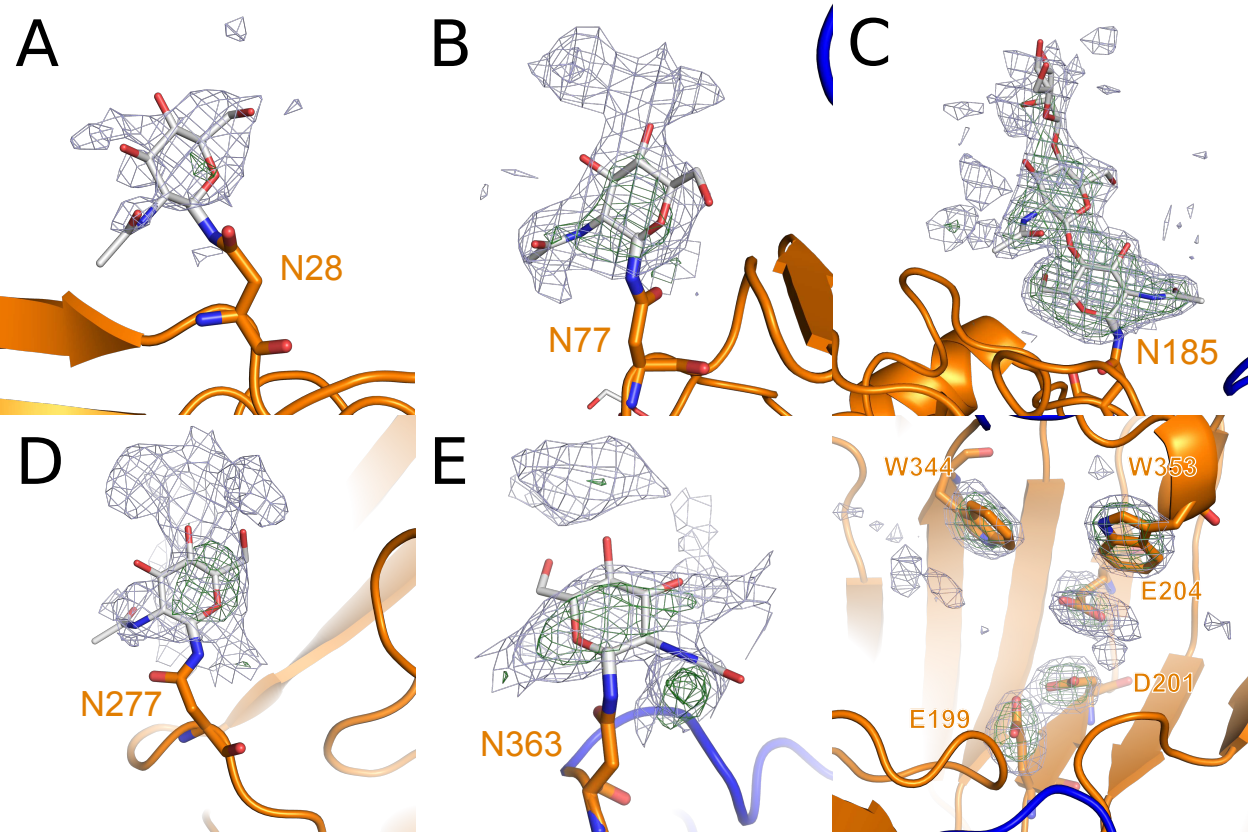
**Fig. 8. Synergy curves for different EG-CBH combinations**. Activity data for the synergy mixtures between (A) *Tr*Cel7A and *Tr*Cel7B, (B) *Tr*Cel7A and *Re*Cel7B, (C) *Re*Cel7A and *Tr*Cel7B and (D) *Re*Cel7A and *Tr*Cel7B. Black squares correspond to the total activity in the synergy mixtures at different mole fraction of CBH. The monocomponent activity is indicated with blue circles for *Tr*Cel7A, orange triangles for *Re*Cel7B, pink circles for *Re*Cel7A and green triangles for *Tr*Cel7B. Dashed lines indicate the sum of the monocomponent activities used in each mixture. Conditions: 25 °C, 10 g/L Avicel, total enzyme concentration 4 μM. For each panel, the x-axis represents the fraction of CBH used in the mixture, calculated as [CBH]/([CBH+EG]). Error bars indicate standard deviation from triplicate measurements.

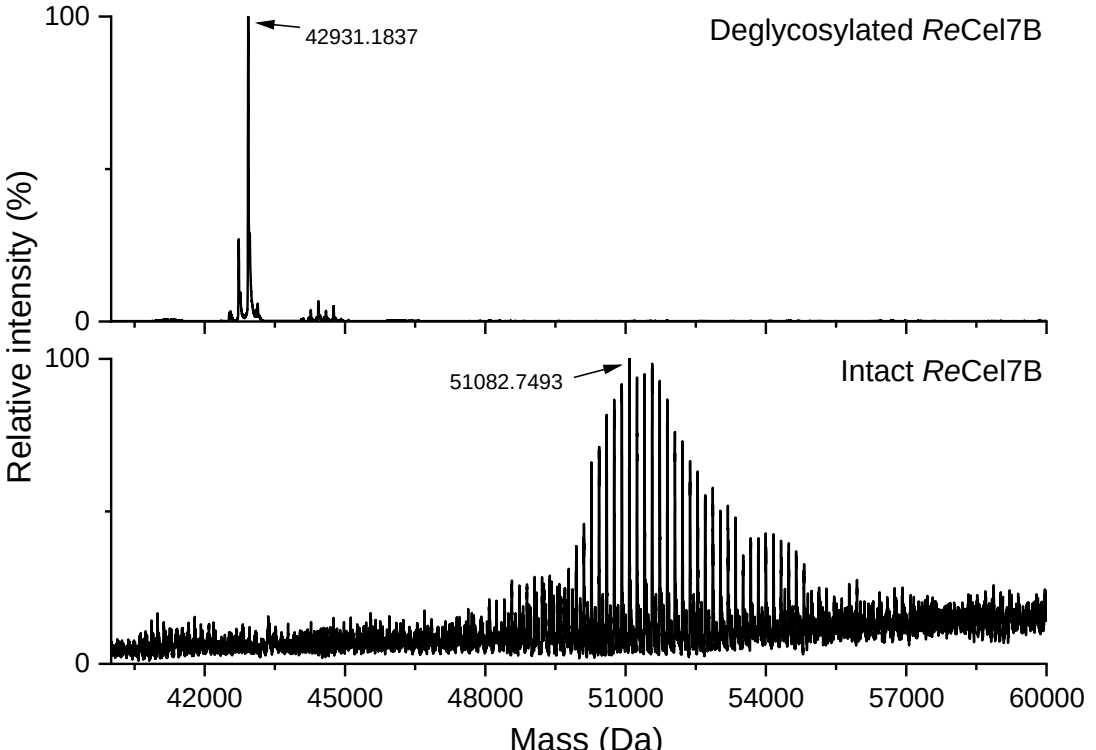
**Fig. 9. DS for different EG-CBH combinations**. DS for the mixtures *Tr*Cel7A+*Tr*Cel7B (red squares), *Re*Cel7A+*Re*Cel7B (blue circles), *Tr*Cel7A+*Re*Cel7B (green up-pointing triangles) and *Re*Cel7A+*Tr*Cel7B (purple down-pointing triangles) is plotted as a function of the mole fraction of the CBH. DS was calculated for each experimental point coming from the synergy curves shown in Fig. 8 using Eq. 4. Error bars are propagated standard deviations.

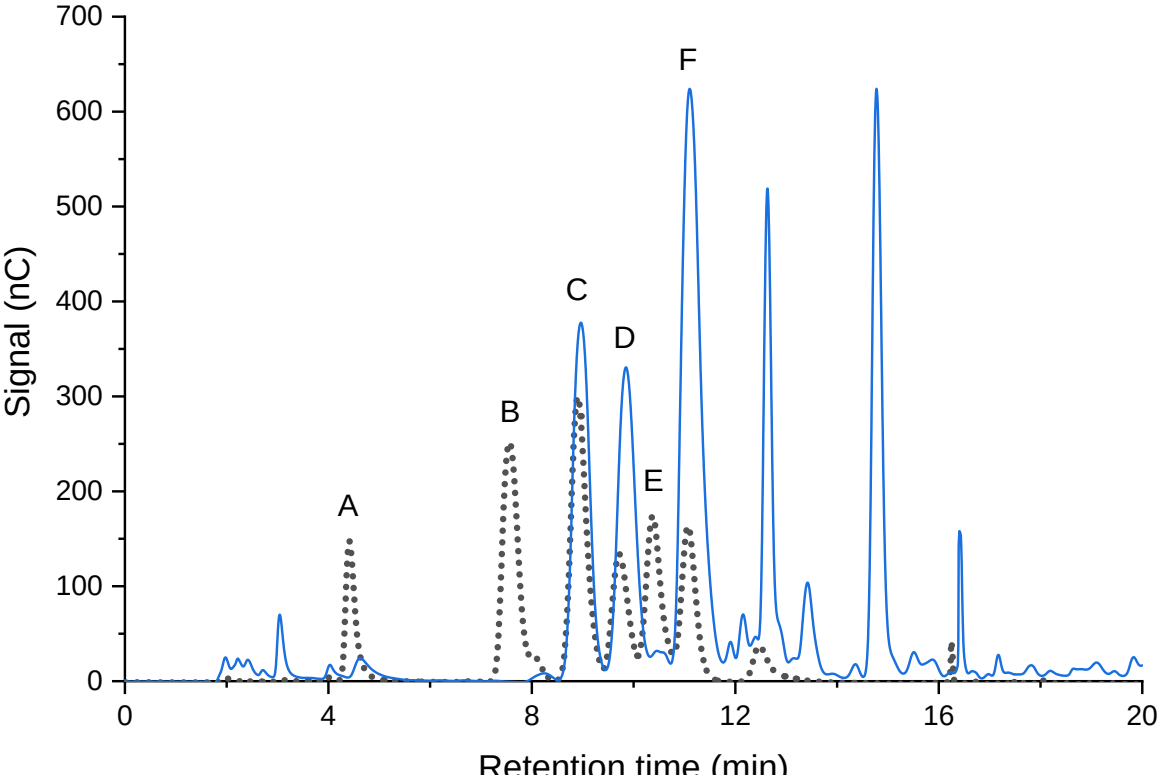

A
Fungi - Ascomycota
Amoebozoa
Fungi - Basidiomycota
Metazoans
Fungi - Ascomycota (EGs)
Parabasalia EGs
Undefined
Crystal structure available
Crystal structure determined in this work
Bold = Thermophilic organism
Cellobiohydrolases
Ascomycotal Endoglucanases
Parabasalian Endoglucanases
0.2
B
HiCel7B
FoCel7B
ThCel7B
TrCel7B
ReCel7B
N28
B1 (51-56)
N77
A1 (97-101)
S2
N185 S1
B3 (218-235)
N277
B4 (310-319)
A3 (346-350)
N363 A2 (364-376)

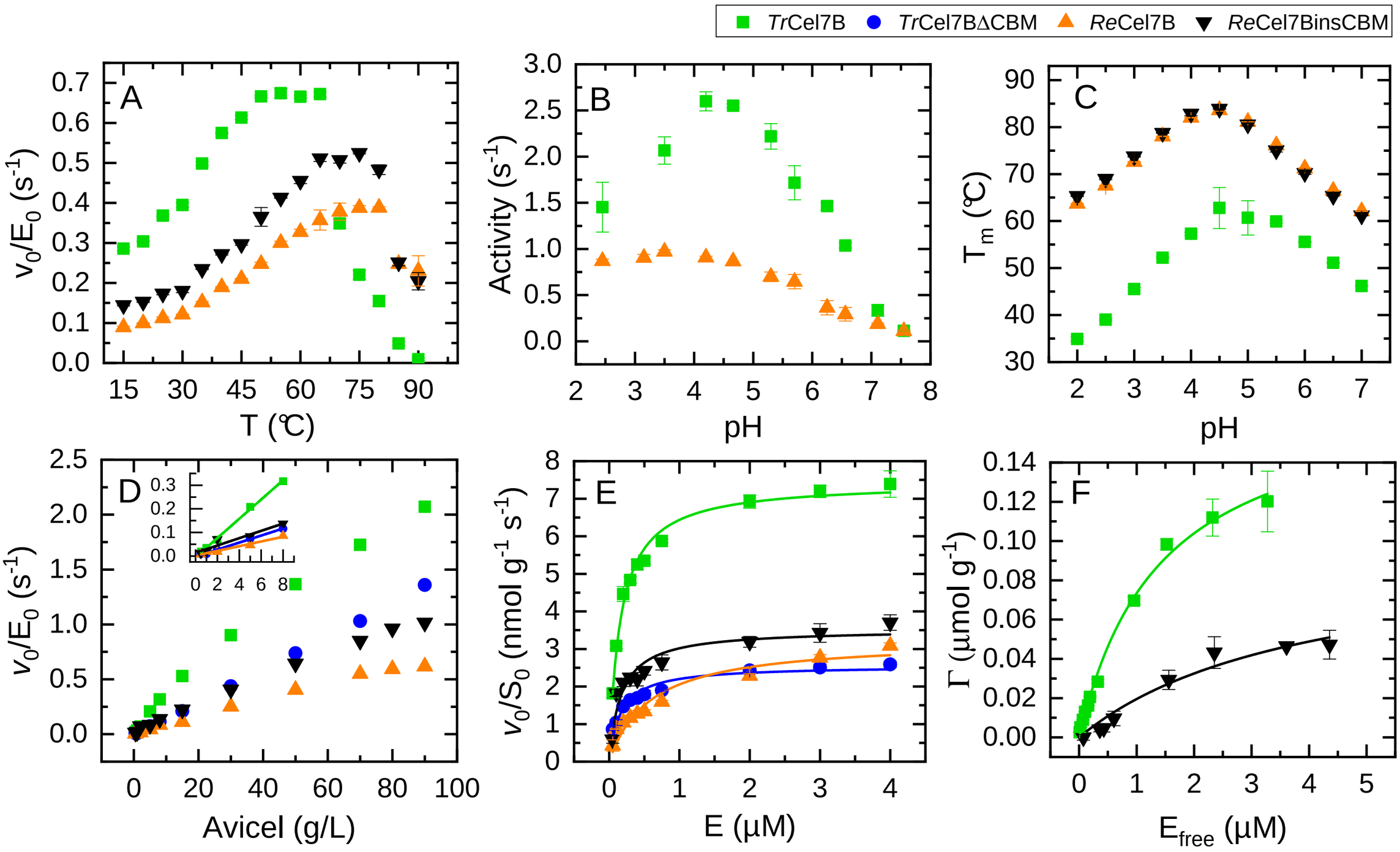

A
A1
A2
A3
B1
B3
B4
N363
N77
E204
D201
E199
N185
B
B3
S1
B4
N185
S2
N277
N28
115 °
C
A1
A2
A3
B1
N56
B3
B4
N182
D
S1
B3
PCA
N185
S221
Q2
D154
S156
S2
R159
E
Q177/Q174/Q174
K238/S221/A221
C230
B3
F
P39/W40
S37/Y38
B1
G
180 °
H

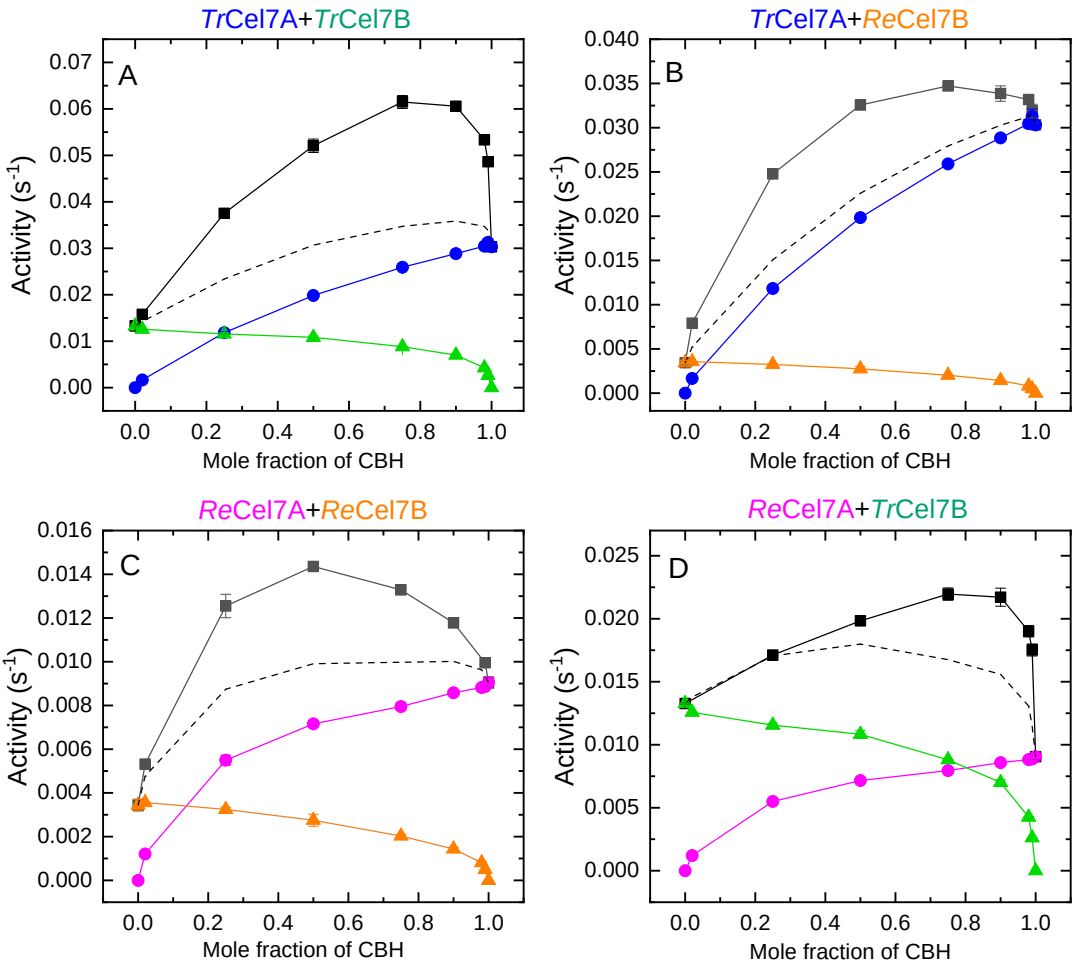

1
2
3
4
ReCel7B
A280 (mAU)
Elution volume (mL)
Molecular weight
Kav
ReCel7B

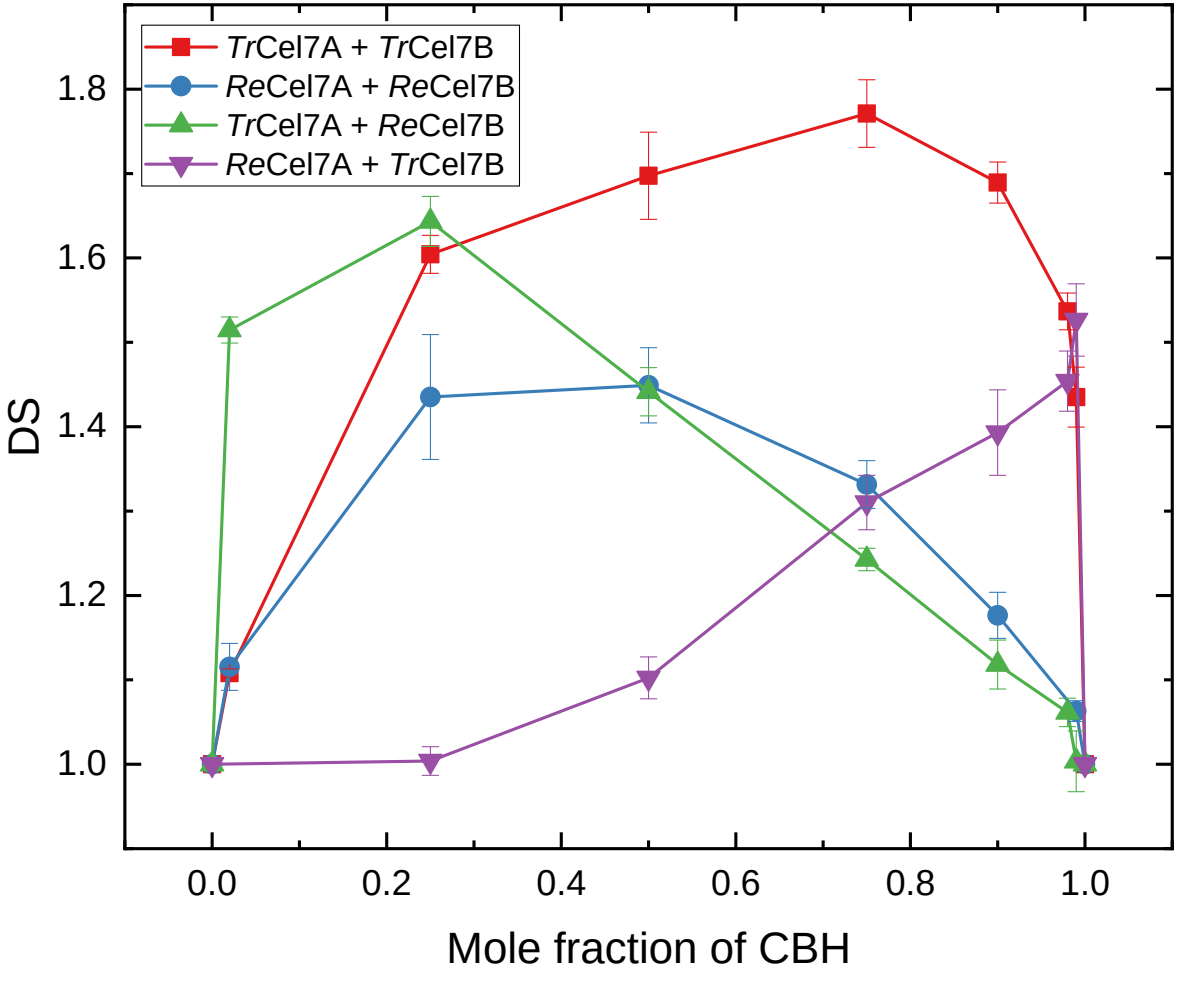

A
N28
B
N77
C
N185
D
N277
E
N363
W344
W353
E204
D201
E199


100
0
42931.1837
Deglycosylated *Re*Cel7B
Relative intensity (%)
100
0
51082.7493
Intact *Re*Cel7B
42000
45000
48000
51000
54000
57000
60000
Mass (Da)


700
600
500
400
300
200
100
0
Signal (nC)
A
B
C
D
E
F
0
4
8
12
16
20
Retention time (min)


TrCel7B
TrCel7BΔCBM
ReCel7B
ReCel7BinsCBM
A
$v_0/E_0$ ($s^{-1}$)
T (℃)
B
Activity ($s^{-1}$)
pH
C
$T_m$ (℃)
pH
D
$v_0/E_0$ ($s^{-1}$)
Avicel (g/L)
E
$v_0/S_0$ (nmol $g^{-1}$ $s^{-1}$)
E (μM)
F
Γ (μmol $g^{-1}$)
$E_{free}$ (μM)


TrCel7A+TrCel7B
A
Activity (s-1)
Mole fraction of CBH
TrCel7A+ReCel7B
B
Activity (s-1)
Mole fraction of CBH
ReCel7A+ReCel7B
C
Activity (s-1)
Mole fraction of CBH
ReCel7A+TrCel7B
D
Activity (s-1)
Mole fraction of CBH


TrCel7A + TrCel7B
ReCel7A + ReCel7B
TrCel7A + ReCel7B
ReCel7A + TrCel7B
DS
1.0
1.2
1.4
1.6
1.8
0.0
0.2
0.4
0.6
0.8
1.0
Mole fraction of CBH