



In-depth characterization of *Trichoderma reesei* cellobiohydrolase *TrCel7A* produced in *Nicotiana benthamiana* reveals limitations of cellulase production in plants by host-specific post-translational modifications

André van Eerde¹, Anikó Várnai^{2,*}, John Kristian Jameson², Lisa Paruch¹, Anders Moen³, Jan Haug Anonsen^{3,*}, Piotr Chylenski², Hege Særvold Steen¹, Inger Heldal¹, Ralph Bock⁴ , Vincent G. H. Eijsink² and Jihong Liu-Clarke^{1,*} 

¹NIBIO, Norwegian Institute of Bioeconomy Research, Ås, Norway

²Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences (NMBU), Ås, Norway

³Department of Biosciences, Faculty of Mathematics and Natural Sciences, University of Oslo (UiO), Oslo, Norway

⁴Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, Germany

Received 14 May 2019;

revised 1 July 2019;

accepted 26 July 2019.

*Correspondence (Tel +47 9959 4790; email Jihong.liu-clarke@nibio.no (JLC); Tel +47 4517 4259; email aniko.varnai@nmbu.no (AV); Tel +47 2285 7295; email j.h.anonsen@ibv.uio.no (JHA)

Keywords: *Nicotiana benthamiana*, *Trichoderma reesei*, plant molecular farming, cellobiohydrolase, enzyme activity, plant glycosylation, substrate binding, protein stability.

Summary

Sustainable production of biofuels from lignocellulose feedstocks depends on cheap enzymes for degradation of such biomass. Plants offer a safe and cost-effective production platform for biopharmaceuticals, vaccines and industrial enzymes boosting biomass conversion to biofuels. Production of intact and functional protein is a prerequisite for large-scale protein production, and extensive host-specific post-translational modifications (PTMs) often affect the catalytic properties and stability of recombinant enzymes. Here we investigated the impact of plant PTMs on enzyme performance and stability of the major cellobiohydrolase *TrCel7A* from *Trichoderma reesei*, an industrially relevant enzyme. *TrCel7A* was produced in *Nicotiana benthamiana* using a vacuum-based transient expression technology, and this recombinant enzyme (*TrCel7A*^{rec}) was compared with the native fungal enzyme (*TrCel7A*^{nat}) in terms of PTMs and catalytic activity on commercial and industrial substrates. We show that the N-terminal glutamate of *TrCel7A*^{rec} was correctly processed by *N. benthamiana* to a pyroglutamate, critical for protein structure, while the linker region of *TrCel7A*^{rec} was vulnerable to proteolytic digestion during protein production due to the absence of *O*-mannosylation in the plant host as compared with the native protein. In general, the purified full-length *TrCel7A*^{rec} had 25% lower catalytic activity than *TrCel7A*^{nat} and impaired substrate-binding properties, which can be attributed to larger *N*-glycans and lack of *O*-glycans in *TrCel7A*^{rec}. All in all, our study reveals that the glycosylation machinery of *N. benthamiana* needs tailoring to optimize the production of efficient cellulases.

Introduction

Non-edible lignocellulose biomass, such as wood, straw and other forest and agricultural residues as well as residues from pulp mills, is currently being explored for the production of second-generation biofuels. The major challenge in converting lignocellulose biomass feedstocks to biofuels is to overcome the recalcitrance of the plant cell wall polysaccharides. Enzymatic conversion of lignocellulosic biomass to fermentable sugars requires large quantities of complex enzyme cocktails with high-performance enzymes. State-of-the-art commercial enzyme cocktails are most often produced in fungi (e.g. *Trichoderma reesei*, *Aspergillus niger* and *Myceliophthora thermophila*) using fermenter-based technologies. *T. reesei* is widely used due to its high secretion capacity and excellent performance in industrial fermentations. Despite progress with optimizing the efficiency of enzyme

production, enzymes still make up a major cost factor of the biomass-to-ethanol conversion, and the price of enzymes is one of the bottlenecks of the economically feasible implementation of large-scale commercial lignocellulose biorefineries (Himmel *et al.*, 2007; Klein-Marcuschamer *et al.*, 2012). Several techno-economic studies suggest that the production costs of industrial enzymes for biorefining are rather high, ranging from ~\$0.34 to \$1.47 per gallon of cellulosic ethanol produced (using cocktails of cellulolytic enzymes), with enzyme companies stating ~\$0.50/gallon of ethanol (Klein-Marcuschamer *et al.*, 2012; Park *et al.*, 2016).

Different avenues have been explored to reduce enzyme and consequently biofuel production costs, such as optimization of the enzymatic saccharification process, on-site enzyme production, genetic engineering of the enzyme-producing strains to obtain more efficient enzymes, and using alternative enzyme

Please cite this article as: van Eerde, A., Várnai, A., Jameson, J. K., Paruch, L., Moen, A., Anonsen, J. H., Chylenski, P., Steen, H. S., Heldal, I., Bock, R., Eijsink, V. G. H. and Liu-Clarke, J. (2019) In-depth characterization of *Trichoderma reesei* cellobiohydrolase *TrCel7A* produced in *Nicotiana benthamiana* reveals limitations of cellulase production in plants by host-specific post-translational modifications. *Plant Biotechnol. J.*, <https://doi.org/10.1111/pbi.13227>

production platforms, including thermostable fungi, yeast, bacteria, insect cells and plants (Garvey *et al.*, 2013; Klein-Marcuschamer *et al.*, 2012; Lambertz *et al.*, 2014; von Ossowski *et al.*, 1997). Plants as a recombinant protein production platform can be used for the production of vaccines, antibodies and enzymes (Clarke *et al.*, 2017; Garvey *et al.*, 2013; Jin and Daniell, 2015; Lambertz *et al.*, 2014; Park *et al.*, 2016; Petersen and Bock, 2011; Verma *et al.*, 2010, 2013). Plants have several advantages compared with the traditional platforms for recombinant protein production in that they are less expensive, scalable and extremely versatile (Clarke and Zhang, 2013; Jin and Daniell, 2015; Lambertz *et al.*, 2014). Most importantly, as complex higher eukaryotes, plants are known to have a flexible protein production machinery that can properly assemble complex proteins. Plants are also able to conduct numerous types of post-translational modifications (Daniell *et al.*, 2015).

To date, a number of recombinant cellulases have been produced successfully in plants, both by transgenic (i.e. by nuclear or chloroplast engineering) (Dai *et al.*, 1999; Harrison *et al.*, 2011; Li *et al.*, 2018, 2019; Taylor *et al.*, 2008) and transient expression (Garvey *et al.*, 2014; Hahn *et al.*, 2015). In general, nuclear and transient expressions are favoured in the case of fungal cellulases in order to benefit from post-translational modifications (PTMs) that occur only in higher eukaryotes and may be essential for enzyme function and stability. PTMs, indeed, are a major factor in the cellulolytic performance of fungal cellulases (Amore *et al.*, 2017; Beckham *et al.*, 2012; Dana *et al.*, 2014), highlighting the importance of mapping the modifications introduced by heterologous expression hosts, such as plants. There are marked differences in the PTM machinery between fungi and plants, in particular concerning protein glycosylation (Deshpande *et al.*, 2008; Goto, 2007; Strasser, 2016). While *N*-glycosylation in both filamentous fungi (Deshpande *et al.*, 2008) and plants (Bosch *et al.*, 2013; Strasser, 2016) has been thoroughly characterized, the *O*-glycosylation machinery is less well studied (Mewono *et al.*, 2015). It is remarkable that glycosylation patterns of recombinant plant-produced proteins have been investigated in detail only in a handful of studies (Castilho *et al.*, 2014; Dicker *et al.*, 2016; Schneider *et al.*, 2014). In order to utilize plant-based enzyme production platform effectively for large-scale and sustainable future production of industrial enzymes, understanding the impact of alternative PTMs on the catalytic performance and stability of the recombinant enzymes is essential.

TrCel7A forms the major component of the *Trichoderma reesei* secretome. When the fungus grows on cellulose, *TrCel7A* comprises ca. 60% of the total secreted proteins (Gritzali and Brown, 1979), highlighting its importance in plant cell wall degradation. *TrCel7A*, also known as cellobiohydrolase I, is a part of the cellulolytic machinery of *T. reesei* and depolymerizes cellulose by cleaving off cellobiose units from the reducing end of cellulose chains in a processive manner. *TrCel7A* contains a GH7 catalytic domain (responsible for the enzymatic activity) and a CBM1 carbohydrate-binding module (increasing enzymatic efficiency by enabling substrate binding); these two domains are adjoined by an *O*-glycosylated linker peptide. The linker and the CBM1 module are critical for *TrCel7A*'s high activity on crystalline substrates (Kont *et al.*, 2016; Payne *et al.*, 2013; Tomme *et al.*, 1988). In addition, PTMs have been found to play an important role in the activity and stability of *TrCel7A*. On the one hand, the catalytic GH7 domain carries naturally four *N*-linked glycans, which enhance proteolytic stability and thermostability (Amore

et al., 2017). On the other hand, the linker and CBM are *O*-glycosylated, which not only protects the enzyme from proteolytic cleavage but also contributes to adsorption onto cellulose (Amore *et al.*, 2017; Payne *et al.*, 2013). Another important PTM concerns the conversion of the N-terminal glutamine to a pyroglutamate (Fägerstam and Pettersson, 1980), which has an important role in stabilizing the structure of fungal cellulases (Dana *et al.*, 2014). Thus, not only the cost-effectiveness but also the compatibility of PTMs needs to be considered when selecting an expression platform alternative to filamentous fungi for the production of *TrCel7A*, and cellulases in general. Therefore, in the present study, we expressed *TrCel7A* from *T. reesei* in *Nicotiana benthamiana* (*TrCel7A^{rec}*) using vacuum agroinfiltration. We mapped the PTMs and compared the properties of the recombinant enzyme to those of the native *TrCel7A* (*TrCel7A^{nat}*).

Results

TrCel7A production in *N. benthamiana* and protein analysis

In order to assess the potential of *N. benthamiana* as a production platform for fungal cellulases, we cloned the gene coding for *TrCel7A* and expressed it in *N. benthamiana* by vacuum-based agroinfiltration as described by Clarke *et al.* (2017). Different gene constructs were designed and cloned into the pEAQ-HT-DEST1 expression vector (Sainsbury *et al.*, 2009) to investigate the effect of subcellular targeting (Figures 1 and S1). Immunoblotting of protein extracts from infiltrated *N. benthamiana* leaves with antibodies recognizing the C-terminal His₈-tag suggested that a construct containing the barley α -amylase signal sequence, which has a proven efficiency in the production of *TrCel7A* in *N. benthamiana* (Hahn *et al.*, 2015), exhibited the highest accumulation of *TrCel7A* and was therefore chosen for further study. Full-length *TrCel7A^{rec}* protein could be detected at approximately 68 kDa (Figure 1a), a size considerably bigger than the predicted molecular weight of 53.5 kDa for the non-glycosylated protein, including the His₈-tag. 0.4 mg pure *TrCel7A^{rec}* could be isolated from 200 g transformed plant material.

To investigate *TrCel7A^{rec}* accumulation and purity, plant extracts were analysed with SDS-PAGE, staining with Coomassie Brilliant Blue (Figure 1b). The band corresponding to full-length *TrCel7A^{rec}* was barely detectable in samples from infiltrated leaves. However, the gel showed clear accumulation of a broad band around 60 kDa to high levels in the samples from infiltrated leaves, which was not visible in an extract from non-infiltrated leaves (Figure 1b). The protein could be isolated and was identified by mass spectrometry as a truncated form of *TrCel7A^{rec}*, lacking the CBM and the His-tag. This shows that a high level of *TrCel7A^{rec}* is produced in *N. benthamiana*, but only a small fraction is present as full-length *TrCel7A^{rec}*.

Isoelectric focusing of purified samples of *TrCel7A^{rec}* and *TrCel7A^{nat}* revealed differences in the pI of the protein variants. The pI of the plant-expressed *TrCel7A* (*TrCel7A^{rec}*; 4.0) was lower than that of the native, fungal *TrCel7A* (*TrCel7A^{nat}*; 4.4), suggesting a difference in the PTM patterns of the two proteins.

Analysis of post-translational modifications (PTMs)

Post-translational modifications have been shown to affect enzyme stability, catalytic efficiency and substrate binding of *TrCel7A*. We therefore analysed and compared the nature and relative abundance of PTMs decorating *TrCel7A* produced in *N. benthamiana* (*TrCel7A^{rec}*) and native *TrCel7A* purified from

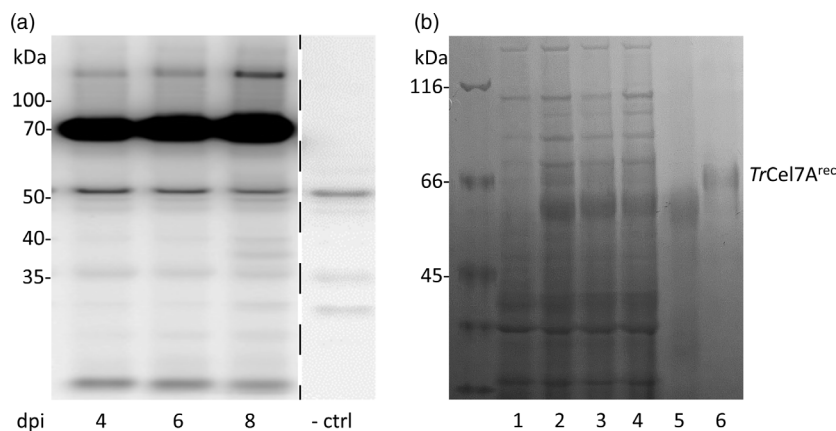


Figure 1 (a) Immunoblot analysis of protein extracts from *N. benthamiana* leaves infiltrated with the *TrCel7A* construct using a barley α -amylase signal peptide and harvested 4, 6 or 8 days post-infiltration (dpi), and a negative control. Thermo Spectra Multicolor Broad Range Protein Ladder (Thermo) was used as marker. (b) SDS-PAGE analysis of extracts from (1) negative control, infiltrated leaves harvested (2) 4 dpi, (3) 6 dpi and (4) 8 dpi, compared with (5) purified *TrCel7A*^{rec} catalytic domain, and (6) purified full-length *TrCel7A*^{rec}. Samples 1, 5 and 6 contained 1 μ g of protein, whereas samples 2–4 contained 2 μ g. Unstained Protein MW Marker (Thermo) was used as marker.

the secretome of *T. reesei* strain QM 9414 (*TrCel7A*^{nat}), using LC-MS² (Table 1, Figure 2). To maximize protein sequence coverage, and thereby increase the likelihood of PTM detection, we employed a multienzyme approach (digesting the proteins by trypsin and chymotrypsin) that allowed the detection of 93% of *TrCel7A*^{rec} and 100% of *TrCel7A*^{nat} (Figure S2). In addition, proteinase K was employed to reduce the peptide length near modified sites to ensure glycopeptide precursor ion masses conformed to the MS detection window. Both *TrCel7A*^{rec} and *TrCel7A*^{nat} displayed a high degree of PTMs, summarized in Table 1 and Figure 2, and discussed in detail below. Identification of each type of PTM (pyroQ, *N*- and *O*-linked glycoforms) and quantification of their relative abundance are summarized in Appendices S3–S7.

PyroQ analysis

Analysis of the N-terminal peptides revealed that both *TrCel7A*^{nat} and *TrCel7A*^{rec} were modified by pyroQ (Table 1 and Figure S3); MS² fragmentation of the doubly charged ion at *m/z* 1047.999 from both *TrCel7A*^{nat} and *TrCel7A*^{rec} (Figure S3) was consistent with a carbamidomethylated (at the Cys; marked as lowercase 'c' in the sequence) and pyroQ-modified N-terminal peptide, that is Q(pyro)SAcTLQSETHPPLTWQK¹⁸. The C-terminal part of the peptide (reflected by γ 1- γ 11 in the γ -ion series) contained only unmodified amino acid residues up to the γ 11 position (i.e. ⁸SETHPPLTWQK¹⁸), indicating that the pyroQ residue is located towards the N-terminus of the peptide. Consistent with this, the N-terminal part of the peptide (i.e. ¹QS² and ¹QSA³, as reflected by b2-b3 in the b-ion series, respectively) was found to carry the modification, locating the pyroQ modification to the absolute N-terminus of both *TrCel7A*^{nat} and *TrCel7A*^{rec}.

N-linked glycosylation

Next, we looked at the glycosylation status of the four theoretical *N*-linked glycosylation sites located in the catalytic domain of *TrCel7A*, namely Asn45, Asn64, Asn270 and Asn384. As expected, most sites displayed a varying level of microheterogeneity (i.e. minor variations in the composition of the glycans at individual glycosylation sites (Table 1, Figures S4–S7). For *TrCel7A*^{nat} (Table 1, Figures 2, S4 and S7), the Asn45 and

Asn384 sites were detected as completely modified with a single GlcNAc residue (relative glycoform abundance, RGA, 100% and 99.9%, respectively). In contrast, the Asn64 site was found mostly unmodified; only 2% of this residue were detected as modified with high-mannose glycans (Figure S5). The major glycoform (93%) of this lowly modified site contained 8 mannose (Man) units and the minor glycoform (7%) contained 7 Man units. The Asn270 site was nearly completely modified (RGA 99.5%), with the two major glycoforms being a single GlcNAc modification (37%) or a high-mannose structure with 8 Man units (27%). This *N*-glycan site showed the highest degree of microheterogeneity in *TrCel7A*^{nat}, with seven types of high-mannose structures detected in total (Figure S6).

Similar to *TrCel7A*^{nat}, the Asn45, Asn270 and Asn384 sites of plant-produced *TrCel7A*^{rec} were almost completely modified (RGA > 99% vs. 99.5%–100% in *TrCel7A*^{nat}) in *TrCel7A*^{rec}, although there was a large variation in the types of *N*-glycans attached (Table 1, Figures 2, S8, S10, S11). While the major glycoform was a single GlcNAc on the Asn45, Asn270 and Asn384 sites in *TrCel7A*^{nat} (100%, 37.6% and 100%, respectively), a single GlcNAc was the major glycoform only at the Asn384 site (61.3%) in *TrCel7A*^{rec}, and the remaining 38.7% were complex *N*-glycans (Figure S11). The Asn45 and the Asn270 sites carried predominantly plant-specific complex and paucimannosidic *N*-glycans (50.3% and 30.1% for Asn45 and 39.0% and 47% for Asn270, respectively; Figures S8 and S10) and harboured a single GlcNAc residue only in 17.4% and 0.5% of the glycosylated peptides, respectively. The major glycoforms at the Asn45 site were GlcNAc₂Man₃Fuc₁Xyl₁ (36.2%) and GlcNAc₂Man₃Fuc₁Xyl₁ (30.1%). The *N*-glycans on the Asn270 site showed the highest microheterogeneity in *TrCel7A*^{rec}. A broad variety of plant-specific complex *N*-glycans were detected attached to the Asn270 site (Figure S10), with six main glycoforms each representing between 8% and 19% of all glycoforms detected (GlcNAc₂Man₃Xyl₁, GlcNAc₂Man₃Fuc₁Xyl₁, GlcNAc₂Man₃, GlcNAc₃Man₃Fuc₁Xyl₁, GlcNAc₃Man₃Xyl₁, GlcNAc₃Man₃). The major glycoform detected for this site contains the *N*-glycan GlcNAc₂Man₃Xyl₁ (18.5%). *N*-glycosylation of the Asn64 site in *TrCel7A*^{rec} was similar to that in *TrCel7A*^{nat} in the sense that the Asn64 site was found mostly unmodified (RGA

Table 1 Post-translational modifications (PTMs) detected with LC-MS for the fungal and plant-produced *TrCel7A* variants

PTM type	Site	Peptide	Protease	Peptide mass	Expression host	RGA (%)	Substitution	Relative abundance (%)
Pyroglutamate	Gln1	<u>Q</u> SACTLQSETHPPLTWQK	Trypsin	1047.999 M2H ²⁺	Fungus	–	–	100
					Plant	–	–	100
N-glycosylation	Asn45	RWTHAT <u>N</u> SSTNcY	Chymotrypsin	1597.681 MH ⁺	Fungus	100	GlcNAc₁	100.0
					Plant	100	GlcNAc ₄ Man ₃ Gal ₁ Fuc ₂ Xyl ₁	2.6
							GlcNAc ₄ Man ₃ Fuc ₁ Xyl ₁	11.5
							GlcNAc₃Man₃Fuc₁Xyl₁	36.2
							GlcNAc ₂ Man ₃ Fuc ₁ Xyl ₁	30.1
							GlcNAc ₂ Man ₉	2.2
							GlcNAc ₁	17.4
	Asn64	SSTLcPD <u>N</u> ETcAKNccLDGAAY	Chymotrypsin	2506.994 MH ⁺	Fungus	2	GlcNAc ₂ Man ₇	6.9
					Plant	5.1	GlcNAc₂Man₈	93.1
							GlcNAc ₄ Man ₃ Fuc ₁ Xyl ₁	2.2
							GlcNAc ₃ Man ₃ Fuc ₁ Xyl ₁	4.7
							GlcNAc ₂ Man ₃ Fuc ₁ Xyl ₁	1.2
							GlcNAc ₄ Man ₃	1.6
							GlcNAc ₃ Man ₃	2.8
							GlcNAc ₂ Man ₅	10.8
							GlcNAc ₂ Man ₆	11.4
							GlcNAc₂Man₇	51.1
							GlcNAc ₂ Man ₉	14.0
	Asn270	RLG <u>N</u> TSFY	Chymotrypsin	794.416 MH ⁺	Fungus	99.7	GlcNAc ₂ Man ₄	0.1
							GlcNAc ₂ Man ₆	3.4
							GlcNAc ₂ Man ₇	11.2
							GlcNAc ₂ Man ₈	27.7
							GlcNAc ₂ Man ₉	16.6
							GlcNAc ₂ Man ₁₀	1.1
							GlcNAc₁	37.6
		RLG <u>N</u> TSF, RLG <u>N</u> TSFY			Plant	99.1	GlcNAc ₃ Man ₃ Gal ₁ Fuc ₁ Xyl ₁	0.3
								GlcNAc ₃ Man ₃ Gal ₁ Xyl ₁
							GlcNAc ₄ Man ₃ Fuc ₁ Xyl ₁	2.0
							GlcNAc ₃ Man ₃ Fuc ₁ Xyl ₁	9.8
							GlcNAc ₂ Man ₃ Fuc ₁ Xyl ₁	12.4
							GlcNAc ₄ Man ₃ Fuc ₁	1.4
							GlcNAc ₃ Man ₃ Fuc ₁	5.0
							GlcNAc ₂ Man ₃ Fuc ₁	5.3
							GlcNAc ₄ Man ₃ Xyl ₁	1.1
							GlcNAc ₃ Man ₃ Xyl ₁	9.4
							GlcNAc₂Man₃Xyl₁	18.5
							GlcNAc ₄ Man ₃	1.8
							GlcNAc ₃ Man ₃	8.0
							GlcNAc ₂ Man ₃	10.8
							GlcNAc ₂ Man ₄	0.2
							GlcNAc ₂ Man ₅	1.0
							GlcNAc ₂ Man ₆	1.2
							GlcNAc ₂ Man ₇	1.5
							GlcNAc ₂ Man ₈	3.8
							GlcNAc ₂ Man ₉	5.8
							GlcNAc ₁	0.5
	Asn384	WLDSTYP <u>N</u> ETSSTTPGAVR	Chymotrypsin	2081.977 MH ⁺	Fungus	99.9	GlcNAc₁	100.0
					Plant	100	GlcNAc ₄ Man ₃ Fuc ₁ Xyl ₁	6.3
							GlcNAc ₃ Man ₃ Fuc ₁ Xyl ₁	16.1
							GlcNAc ₂ Man ₃ Fuc ₁ Xyl ₁	16.3
							GlcNAc₁	61.3

Table 1. Continued

PTM type	Site	Peptide	Protease	Peptide mass	Expression host	RGA (%)	Substitution	Relative abundance (%)						
O-glycosylation	Gly395-Lys415	GScSTSSGVPAQVESQSPNAK	Trypsin	2077.946 MH ⁺	Fungus	15	Hex₁	73.3						
							Hex ₂	11.7						
							Hex ₃	5.1						
							Hex ₄	6.5						
							Hex ₅	3.0						
							Hex ₆	0.3						
							Hex ₇	0.03						
	Val416-Lys422	VTFS NIK	Trypsin	808.456 MH ⁺	Fungus	2	Hex₁	98.2						
	Gly433-Tyr465	GGNPPGGNPPG TTTT RRPA TTT G SS PG TQ SHY	Proteinase K	3206.521 MH ⁺	Fungus	100	Hex ₁₄	6.1						
							Hex ₁₅	20.4						
							Hex₁₆	40.0						
							Hex ₁₇	14.8						
							Hex ₁₈	6.5						
							Hex ₁₉	4.0						
							Hex ₂₀	2.8						
							Hex ₂₁	2.8						
							Hex ₂₂	2.2						
							Hex ₂₃	1.0						
							Hex ₂₄	0.3						
							Thr447-Tyr465	TRRPATTTGSS P^{OH} GPTQSHY	Chymotrypsin	2017.9686 MH ⁺	Plant	1.9	Hex ₄ dHex ₁ HexA ₁ Pen ₃	7.1
													Hex ₄ dHex ₁ HexA ₁ Pen ₂	7.4
													Hex ₃ dHex ₁ HexA ₁ Pen ₂	9.0
													Hex ₃ dHex ₁ HexA ₁ Pen ₁	12.7
													Hex ₃ dHex ₁ HexA ₁	6.6
Hex ₂ dHex ₁ HexA ₁ Pen ₃													6.5	
Hex₂dHex₁HexA₁Pen₂	19.1													
Hex ₂ dHex ₁ HexA ₁ Pen ₁	16.7													
Ser474-Cys485	S GPTVcASGTTc	Chymotrypsin	1197.487 MH ⁺	Fungus	2	Hex₁	100							
				Plant		No substitution								

Amino acid residues potentially carrying a modification are in bold and underlined; oxidized Pro (hydroxyproline; a common modification in plant proteins) is marked as P^{OH}; carbamidomethyl-Cys (a modification due to sample processing) is indicated as c. Alterations from the native peptide masses caused by PTMs and the possible PTM forms are indicated. The relative glycoform abundance (RGA) indicates the amount of glycosylated peptides as a percentage of the total number of peptides detected. The types of substitution (i.e. glycoforms) are listed, and the most abundant substitution types are highlighted in bold. The glycan residues are abbreviated as follows: *N*-acetylhexosamine, HexNAc; *N*-acetylglucosamine, GlcNAc; hexose, Hex; mannose, Man; galactose, Gal; deoxyhexose, dHex; fucose, Fuc; hexuronic acid, HexA; pentose, Pen; xylose, Xyl. The relative abundance of each glycoform is given as the percentage of the total number of glycosylated peptides detected. Figure 2 provides an overview of the localization of the glycosylations.

5.1% vs. 2% in *TrCel7A*^{nat}), and the major glycoforms (90.2% in total) were high-mannose glycans, with 5–7 and 9 Man units (Figure S9).

O-linked glycosylation

To determine the O-linked glycosylations of *TrCel7A*^{nat} and *TrCel7A*^{rec}, a similar strategy as for the detection of peptides modified with *N*-glycans was pursued. For *TrCel7A*^{nat}, four peptides carrying O-glycosylation were detected: one containing the linker region and part of the CBM1 domain, one at the CBM1 and two at the C-terminus of the catalytic domain (Figure 2, Table 1). In line with previous work (Amore *et al.*, 2017), one of these peptides, containing the three patches of putative O-glycosylation sites (Thr445–Thr448, Thr453–Thr455 and Ser457–Ser458) in the linker and two O-glycosylation sites (Thr462 and

Ser464) in the CBM1 showed a broad diversity in the extent of glycosylation (Figure S12); the overall number of substituting hexose (Hex) units ranged from 14 to 24, with the most abundant glycopeptide signal corresponding to 16 substituting Hex units. This glycopeptide was identified from the quadruple charged precursor ion (*m/z* 1450.590 [M+4H]⁴⁺, charge-adjusted mass 5799.339 Da [M+H]⁺; corresponding to the peptide ⁴³³GGNPPGGNPPGTTTTRRPATTTGSSPGPTQSHY⁴⁶⁵ with the theoretical mass 3206.5210 Da [M+H]⁺; Figure S11C) with a mass addition corresponding to 16 Hex residues (162.0528 Da). The glycan microheterogeneity of this peptide was identified through manual investigation of MS precursor masses with loss or gain of charge-adjusted accurate Hex masses (for more details, see Figure S12). No unmodified linker peptides were detected for *TrCel7A*^{nat}, indicating a strong correlation between linker O-

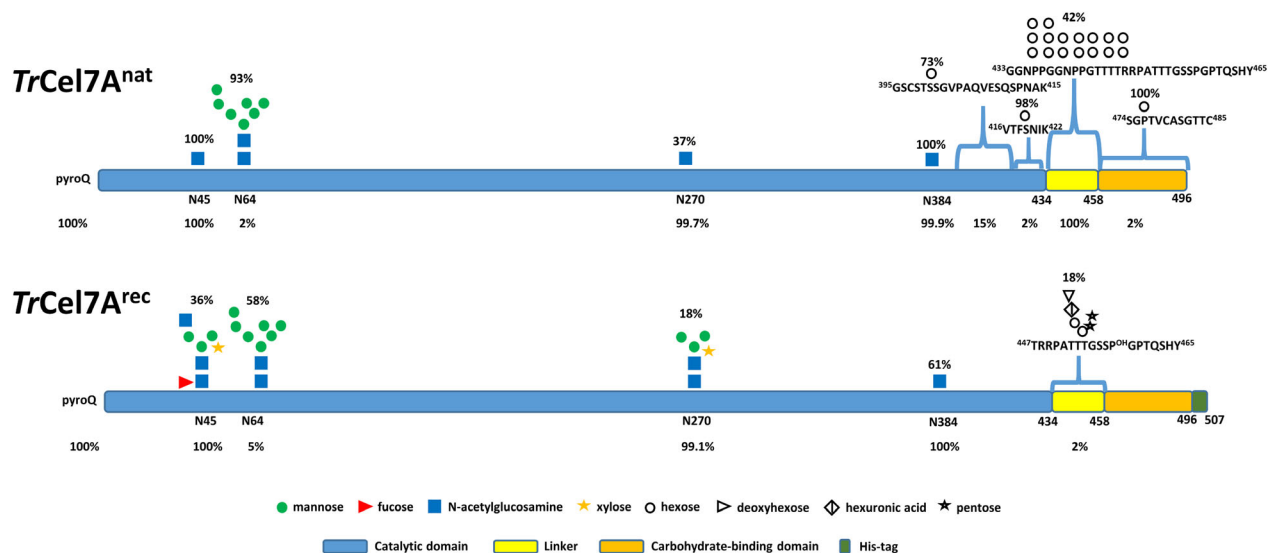


Figure 2 Summary of PTMs detected in the plant-produced *TrCel7A*^{rec} and the native fungal *TrCel7A*^{nat}. The pictures show the N-terminal pyroQ modification and the most abundant glycoforms of *N*- and *O*-glycans. Numbers above the glycoforms denote the relative glycoform abundances (in percent); numbers below each PTM site denote the overall relative abundance of glycosylation for the given site (in percent). *O*-glycosylations, which could not be pinpointed to a specific amino acid, are indicated by blue curly brackets; see text for further details.

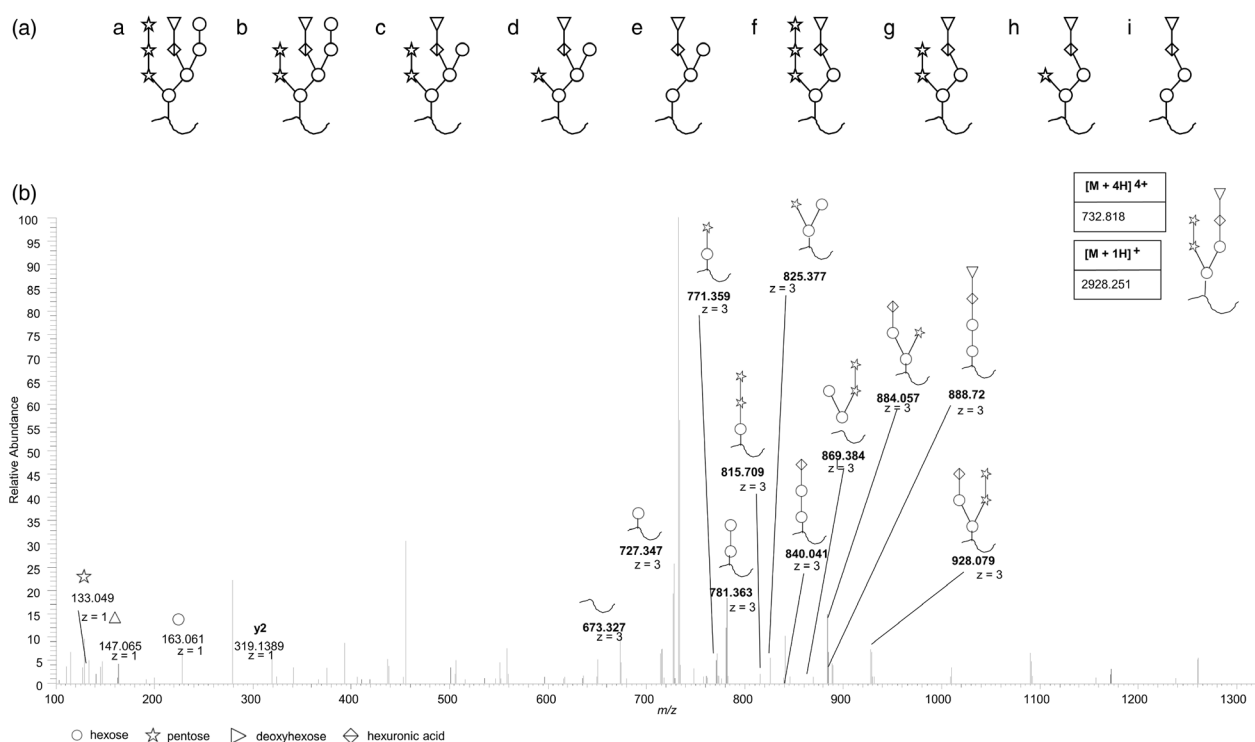


Figure 3 *O*-glycopeptide structures found in *TrCel7A*^{rec} (A) and the MS² fragmentation of the most abundant *O*-glycoform (B). The MS² fragmentation for all *O*-glycopeptides is given in Appendix S7.

glycosylation and maturity of *TrCel7A* in the native host *T. reesei*. We were unable to detect shorter fragments of this peptide, for example parts of the CBM1 lacking the linker region, most probably because *O*-glycosylation protected this peptide from further cleavage. PTM at the third *O*-glycosylation site of the CBM1 was, on the other hand, identified in a separate peptide,

⁴⁷⁴SGPTVCASGTTc⁴⁸⁵, through the doubly charged precursor ion (m/z 680.274 [M+2H]²⁺; corresponding to the doubly carbamidomethylated ⁴⁷⁴SGPTVCASGTTc⁴⁸⁵ peptide with the theoretical mass 1359.540 Da [M+H]⁺), with a single Hex modification (Figure S13). Unexpectedly, glycosylation was seen in only 2% of the peptides detected (Figure S14).

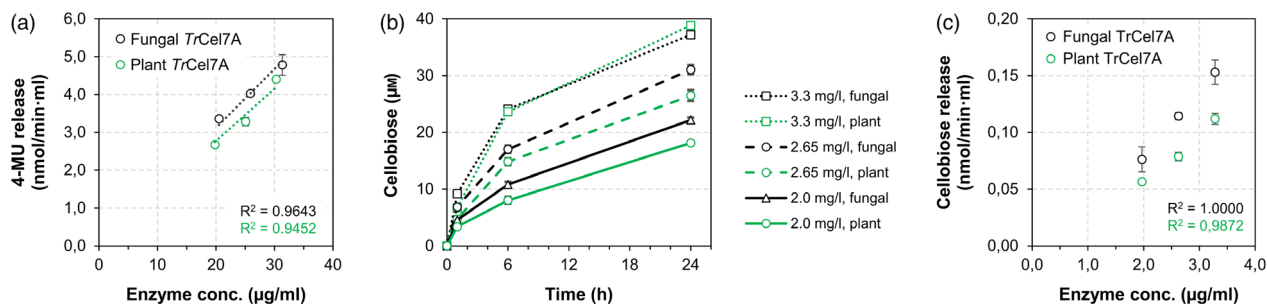


Figure 4 Activity of the *TrCel7A* variants on methylumbelliferyl cellobioside (MUC) and Avicel. (a) Activity of *TrCel7A*^{rec} (green lines) and *TrCel7A*^{nat} (black lines) on MUC. MUC (0.4 mM) was incubated with 20–31 $\mu\text{g}/\text{mL}$ plant-produced *TrCel7A*^{rec} and fungal *TrCel7A*^{nat} for 10 min; the y-axis shows the amount of 4-methylumbelliferone (4-MU) released in the reaction mixture within 10 min. The linear correlation between the release of 4-MU and enzyme concentration in the reaction indicates that the enzymes are working at the initial, linear range of the reaction. Reactions were performed in 50 mM Na-acetate buffer pH 5.0 at 50 °C, in a total volume of 100 μL . (b) Cellobiose release from Avicel by plant-produced *TrCel7A*^{rec} (green lines) and fungal *TrCel7A*^{nat} (black lines) over time. Avicel (2%, w/v) was incubated with 2.0 (circles with solid line), 2.65 (triangles with dashed line) and 3.3 $\mu\text{g}/\text{mL}$ (squares with dotted line) *TrCel7A*. (c) Activity of *TrCel7A*^{rec} (green lines) and *TrCel7A*^{nat} (black lines). Avicel (2%, w/v) was incubated with 2.0–3.3 $\mu\text{g}/\text{mL}$ *TrCel7A* for 60 min. The linear correlation between cellobiose release and enzyme concentration indicates that the enzymes are working at the initial, linear range of the reaction. Reactions were performed in 12.5 mM Na-acetate buffer pH 5.0 at 50 °C, in a total volume of 200 μL . Reactions were run in triplicates. In panel B, each time point represents an individual sample: Standard deviations, from three individual experiments, are shown.

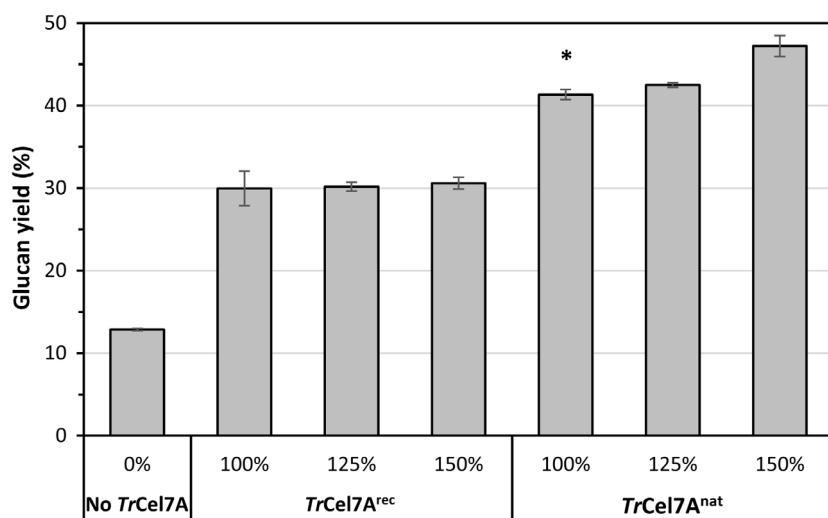


Figure 5 Hydrolysis of BALI-pretreated spruce with a minimal enzyme cocktail. BALI-pretreated spruce (5%, w/w) was incubated with a minimal enzyme cocktail containing either plant-produced *TrCel7A*^{rec} or fungal *TrCel7A*^{nat}. Reactions were incubated in 50 mM Na-acetate pH 5.0 at 50 °C for 48 h, with a total enzyme loading of 7.064 mg enzyme/g glucan. The asterisk marks the reference enzyme cocktail with enzyme ratios optimized by Chylenski *et al.* (2017). ‘No *TrCel7A*’ indicates that *TrCel7A* was omitted from the enzyme cocktail.

O-glycosylation was found at low levels also at the C-terminal part of the catalytic domain, in the peptide ³⁹⁵GSCSTSSGVPAQVESQSPNAK⁴¹⁵, containing potential O-glycosylation patches at Ser396–Ser401 and Ser409–Ser411, and in the peptide ⁴¹⁶VTFSNIK⁴²², containing Thr417–Ser419. Of note, the latter glycosylation sites were not annotated by the NetOGlyc prediction algorithm. Although the ³⁹⁵GSCSTSSGVPAQVESQSPNAK⁴¹⁵ peptide was identified as carrying glycosylations ranging from one to seven Hex units, the most abundant glycopeptide carried only a single Hex (98.2%) as revealed by the triply charged precursor ion at *m/z* 747.34 ([M+3H]³⁺; observed mass of 2240.006 Da, [M+H]⁺), corresponding to the carbamidomethylated peptide with a single Hex modification (Figure S15). Glycosylation with one or two Hex units of ⁴¹⁶VTFSNIK⁴²² (theoretical mass 808.456 Da [M+H]⁺) was identified from doubly charged precursor ions at *m/z* 485.758 and *m/z* 566.784 (observed masses of 970.509 and 1132.551 Da [M+H]⁺, respectively; Figure S16). Interestingly, two versions of the singly substituted ⁴¹⁶VTFSNIK⁴²² peptide were detected in the chromatogram (Figure S17), indicating that both hydroxyl bearing

amino acids (i.e. Thr417 and Ser419) may be glycosylated, although at different abundance. The relative abundance of O-glycosylated ⁴¹⁶VTFSNIK⁴²² peptides, however, was very low, with a combined RGA (including mono- and di-glycosylated versions) of <2%.

In contrast with the abundance of O-glycosylation in *TrCel7A*^{nat}, for *TrCel7A*^{rec} only a single O-glycosylation site in the linker region (⁴⁴⁷TRRPATTTGSSP^{OH}GPTQSHY⁴⁶⁵, Figure 2) was detected despite 95% coverage of the protein sequence (Figure S2A). In addition to O-glycans, MS² fragmentation of the precursor ion belonging to the non-glycosylated peptide (*m/z* 505.247; observed mass of 2017.969 Da [M+H]⁺) revealed a hydroxyproline (P^{OH}) at the Pro458 position (Figure S18). The analyses showed that this linker peptide carries multiple plant-specific O-linked glycoforms (Table 1; Figures 3 and S19). MS² fragmentation of the corresponding reporter ions revealed intermediate glycan structures and allowed the assembly of the most common O-glycan structure in the *TrCel7A*^{rec} linker as shown in Figure 2. As an example, fragmentation of the quadruple charged precursor ion at *m/z* 732.818 ([M+4H]⁴⁺; observed mass of

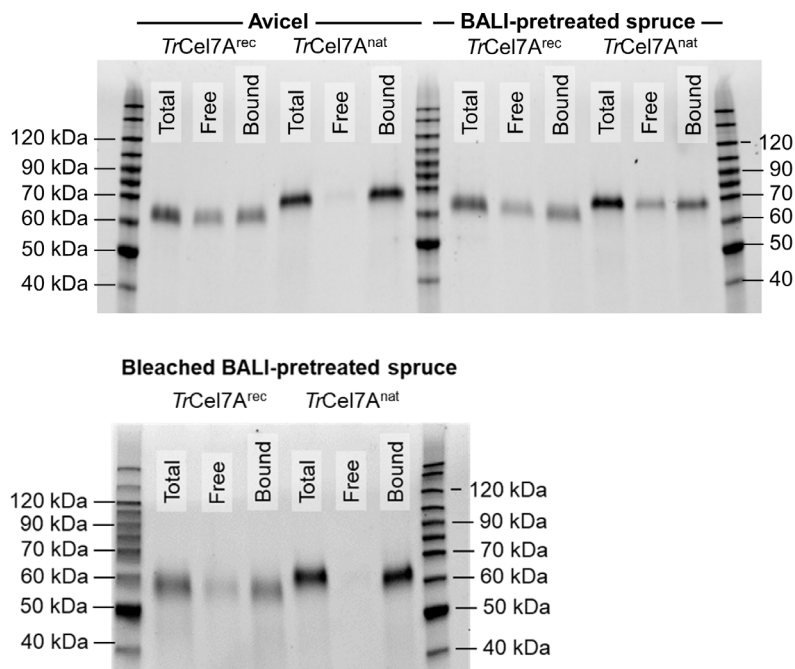


Figure 6 Adsorption of *TrCel7A* variants on Avicel, BALI-pretreated spruce and bleached BALI-pretreated spruce. The gel visualizes the total amounts of protein used, the amounts found free in solution and the amounts bound to the substrate for both full-length *TrCel7A* variants. A reaction mixture without substrate was used as a reference ('Total'). Enzymes were incubated in 50 mM Na-acetate pH 5.0 at 4 °C for 30 min with either 5% (w/w) Avicel, 5% (w/w) BALI-pretreated spruce or 5% (w/w) bleached BALI-pretreated spruce.

2925.703 Da, $[M+H]^+$) revealed signals consistent with the consecutive loss of deoxyhexose (dHex), hexuronic acid (HexA) and Hex (at m/z 928.079, 869.384 and 815.709, $[M+3H]^{3+}$, respectively) and the sequential loss of two pentose (Pen) units (at m/z 928.079, 884.057 and 840.041; $[M+3H]^{3+}$) from the glycopeptide. It is noteworthy that all detected glycopeptides displayed, in addition to *O*-glycosylation, a hydroxyproline modification at Pro458. Although plant *O*-glycosylation is reported to commonly occur at hydroxyproline residues (Mewono *et al.*, 2015), we were unable to unambiguously determine the position of the *O*-glycan and specifically link it to this residue.

Activity of the plant-expressed and native *TrCel7A*

To check whether expression in *N. benthamiana* compromised the catalytic efficiency of *TrCel7A* due to different post-translational modifications (PTMs) as well as the appended C-terminal His-tag, we compared the activity of the plant-expressed *TrCel7A* (*TrCel7A*^{rec}) to that of the native *TrCel7A* (*TrCel7A*^{nat}) protein on the soluble model substrate 4-methylumbelliferyl- β -D-cellobioside (MUC). These assays revealed that the plant-expressed *TrCel7A* had 12% lower specific activity than the native *TrCel7A* (Figure 4A). The specific activities calculated for a 10-minute reaction were 137 ± 5 and 157 ± 9 U/g enzyme for *TrCel7A*^{rec} and *TrCel7A*^{nat}, respectively. Of note, the linker and CBM regions have been shown to have no effect on the catalytic efficiency of the GH7 module (Tomme *et al.*, 1988). Hence, PTMs located at the linker region and the CBM as well as the His-tag, which is attached to the C-terminal CBM, are unlikely to affect the catalytic activity of the GH7 catalytic domain towards the soluble substrate MUC.

The activities of plant-expressed *TrCel7A*^{rec} and the native *TrCel7A*^{nat} were also compared using the cellulosic model substrate, Avicel. On average, *TrCel7A*^{nat} released 20%–25% more cellobiose from Avicel than *TrCel7A*^{rec} after 1 h at all the three enzyme loadings used. This difference remained over time at the lower (2.0 and 2.65 μ g/mL) enzyme concentrations, while it disappeared at the highest (3.3 μ g/mL) enzyme concentration

(Figure 4B,C). The linear correlation between the release of cellobiose and enzyme concentration in the reaction indicates that the enzymes are working at the initial, linear range of the reaction. The specific activities for *TrCel7A*^{rec} and *TrCel7A*^{nat} could, therefore, be calculated for the first time point (i.e. 60 min) and were 32 ± 2 and 44 ± 7 U/mg enzyme, respectively.

Subsequently, we studied the performance of the *TrCel7A*^{rec} in a minimal enzyme cocktail of *T. reesei* cellulases that has been developed for spruce pretreated according to the BALI (Borregaard Advanced Lignin) process (Chylenski *et al.*, 2017) with varying loadings of *TrCel7A* (Figure 5). The results show that the enzyme cocktail with the plant-expressed *TrCel7A*^{rec} (100% *TrCel7A*^{rec}) gave approximately 25% lower yield after 48 h of incubation than the enzyme cocktail containing the fungal *TrCel7A*^{nat} (100% *TrCel7A*^{nat}; reference cocktail) (Figure 5). Compared to the reactions without *TrCel7A* (No *TrCel7A*), both *TrCel7A*^{rec} and *TrCel7A*^{nat} increased the glucan yields. For the plant version, however, increasing the amount of *TrCel7A*^{rec} (to 125% and 150%) did not lead to a further significant increase in the total glucan yield.

Enzyme adsorption to solid substrates

As substrate binding is essential for efficient enzyme catalysis, we compared the extent of binding of the two *TrCel7A* variants to both substrates, Avicel and BALI-pretreated spruce, as well as on bleached BALI-pretreated spruce (Figures 6). (Note that the plant-produced variant carries alternative PTMs and a C-terminal His-tag, which could affect substrate-binding properties.) The binding studies revealed impaired binding efficiency of the plant-expressed *TrCel7A*^{rec} compared with the native fungal *TrCel7A*^{nat} for all three substrates. *TrCel7A*^{nat} bound to Avicel to a high extent, leaving only 5% of the total loaded protein free in solution, whereas some 40% of *TrCel7A*^{rec} remained free in solution. On the other hand, the two enzyme variants bound to the more complex substrate, BALI-pretreated spruce, to a similar extent, leaving 36% of *TrCel7A*^{nat} and 38% of *TrCel7A*^{rec} in solution. Lignin removal from the BALI substrate by bleaching led

to increased adsorption of both enzyme variants, with only 3% and 30% of the enzymes, respectively, remaining in solution (Figure 6).

Discussion

Significant progress has been made in the development of plant expression systems for recombinant enzymes, including downstream processing, in order to overcome bottlenecks associated with protein yields. To date, the main approaches used for expression of recombinant proteins in plants are stable expression of transgenes in the nuclear genome of transgenic plants (or plant cell lines) or in the chloroplast genome of transplastomic plants, and transient expression of transgenes in plants (Bock, 2015; Daniell *et al.*, 2015; Peyret and Lomonosoff, 2015). Plants have been shown to be able to produce cellulases, such as *TrCel7A* (Dai *et al.*, 1999; Hahn *et al.*, 2015; Harrison *et al.*, 2011; Hussain *et al.*, 2015). The large demand for enzymes for the biorefinery of forest biomass has driven worldwide efforts to produce cell wall-degrading enzymes more cost-effectively. Plant-based enzyme production offers a highly promising approach, due to low production costs, high attainable expression levels and approved good manufacturing practices (GMPs). Moreover, two plant expression systems have been launched for commercial production of recombinant enzymes, vaccines and biopharmaceuticals (<https://www.leafexpressionsystems.com/> and <https://www.pennovation.upenn.edu/the-community/innovators/phyllzyme>; Daniell *et al.*, 2019). Leaf Expression Systems is a transient expression-based system using *Agrobacterium*-mediated transient expression in leaves, while PhylloZyme is an *Agrobacterium*-free chloroplast genome engineering technology for commercial production of recombinant enzymes (Daniell *et al.*, 2019). These systems can contribute to the cost-effective, large-scale production of enzymes for various industrial applications, including the processing of lignocellulosic biomass. When comparing the production cost of plants versus fungal hosts, plant-based enzyme production systems have several advantages: (i) no fermentation facility is required, eliminating one of the largest cost factors, (ii) easy up-scaling by simply increasing the cultivation area and (iii) the possibility to use non-food and non-feed plants (e.g. tobacco, for which agricultural practices are fully established and alternative uses are currently sought). Recently, it was calculated that the costs of dried tobacco leaves ranged between \$1.48 and \$1.85/lb in the past decade, and operating and machinery costs are \$3.21. Thus, the cost of enzyme production in tobacco leaves is much lower than in any microbial fermentation facility requiring costly construction, operation and maintenance (Daniel *et al.* 2019). Leaf Expression Systems (<https://www.leafexpressionsystems.com/>) has marketed some plant-made products commercially, such as recombinant human triosephosphate isomerase (TPI) at a price of 100 µg for £200 and 1 mg for £1500, thus providing a concrete figure for the cost comparison with fermentation-based enzyme production systems. However, the yield of recombinant protein varies between plant production systems and depends on many factors, including the choice of the host plant, expression method, the protein of interest and its properties, the plant cultivation system and the steps involved in downstream purification. For example, in a leaf-based production platform (Daniell *et al.*, 2019), the yield of leaf biomass was reported to be much higher (approximately 10-fold) in soil-grown plants than in hydroponic plants. While some data are available on recombinant protein stability, the catalytic

properties and the effects of PTMs on enzyme activity have not been studied in detail and more investigations are needed. Therefore, in this study, we analysed the PTMs in detail and performed an extensive comparison of the catalytic and substrate-binding properties of the plant-produced *TrCel7A^{rec}* and the native *TrCel7A^{nat}* on both artificial (MUC and Avicel) and industrial (BALI-pretreated and bleached BALI-pretreated spruce) substrates.

Pyroglutamate formation at the N-terminus of GH7 cellulases is essential for proper folding and, consequently, for enzyme stability and activity (Dana *et al.*, 2014; Divne *et al.*, 1994; Wu *et al.*, 2017). Although the presence of glutaminyl cyclase activity has been found in papaya (Messer, 1963) and glutaminyl cyclases from potato and *Arabidopsis thaliana* have been characterized (Schilling *et al.*, 2007), it has not been reported if *N. benthamiana* has a glutaminyl cyclase-encoding gene or is able to convert the N-terminal glutamate to pyroglutamate. Here we showed that the N-terminus of *TrCel7A^{rec}* was faithfully converted to pyroglutamate when the enzyme was produced in *N. benthamiana*, indicating that sufficient levels of glutamate cyclase are present in *N. benthamiana* leaves for the correct N-terminal processing of *TrCel7A* upon transient expression. Glutaminyl cyclase is present in certain fungi such as *T. reesei*, but is absent from yeast (Dana *et al.*, 2014; Wu *et al.*, 2017), which makes the plant production platform especially appealing compared with yeast-based expression systems.

In *T. reesei*, the type of *N*-linked glycans of *TrCel7A* differs depending on the strain and culture conditions (Adney *et al.*, 2009; Hui *et al.*, 2001; Jeoh *et al.*, 2008; Stals *et al.*, 2004a,b). The most systematic analysis of the role of glycosylation in *TrCel7A^{nat}* to date has been performed by Amore *et al.* (2017). The lack of *N*-glycans at the catalytic module of *TrCel7A* can affect folding (Qi *et al.*, 2014) and both thermal and proteolytic stability of the cellulase (Amore *et al.*, 2017; Qi *et al.*, 2014). Reports on the effects of *N*-glycosylation on activity are mixed; while most studies report that *N*-glycans (and their removal) do not affect catalytic activity (Amore *et al.*, 2017; Dana *et al.*, 2014; Qi *et al.*, 2014), some have found that removing larger *N*-glycans appended to recombinant *TrCel7A* in hyper-glycosylating expression hosts leads to increased enzyme activity (Adney *et al.*, 2009; Ranaei Siadat *et al.*, 2016). In our present study, the *N*-glycans detected were in agreement with previous reports listing three major *N*-glycosylation sites (Amore *et al.*, 2017; Harrison *et al.*, 1998; Wang *et al.*, 2019). Expression of *TrCel7A^{rec}* in *N. benthamiana* yielded an enzyme that carried PTMs at all four *N*-glycosylation sites. While the most abundant *N*-glycans were similar in the fungal and plant-expressed variants (see Figure 2 and Table 1), we observed interesting differences in the *N*-glycosylation pattern of *TrCel7A^{nat}* and *TrCel7A^{rec}*. Asn45 and Asn384, the residues located at the entrance and exit of the catalytic tunnel, respectively, carried only a single GlcNAc in *TrCel7A^{nat}*, which is in agreement with previous reports (Harrison *et al.*, 1998; Hui *et al.*, 2001). The corresponding residues were found to carry more complex glycans in *TrCel7A^{rec}*, which could potentially hinder access to the entrance and exit of the catalytic tunnel and thus lower catalytic efficiency. Limited accessibility of the catalytic tunnel would result in lower catalytic efficiency of *TrCel7A^{rec}* as compared with *TrCel7A^{nat}*, as observed on the soluble model substrate MUC.

There were marked differences in *O*-linked glycosylation between the two *TrCel7A* variants, especially concerning the linker region. The linker peptide in *TrCel7A^{nat}* was identified only

as a single, heavily glycosylated peptide. We could not obtain smaller fragments of the linker peptide in *TrCel7A^{nat}* with proteolytic digestion, corroborating the importance of *O*-glycosylation for proteolytic stability and indicating that glycans were distributed along the linker as short-chain glycans at multiple positions (Amore *et al.*, 2017). Importantly, we found that the third *O*-glycosylation site, Ser474, of the CBM1 was unmodified in *TrCel7A^{nat}* in 98% of the peptides detected, suggesting that *O*-glycosylation by a single Hex at this position in the CBM1 is not important and thus corroborating the findings by Amore *et al.* (2017). In contrast to the extensively *O*-glycosylated *TrCel7A^{nat}* protein, *TrCel7A^{rec}* was modified only to a limited extent (i.e. 2% substitution at a single detected peptide). The presence and extent of *O*-linked glycosylation on the linker have previously been shown to affect binding to cellulose (Jeoh *et al.*, 2008; Payne *et al.*, 2013). In fact, the ca. 25% lower efficiency of *TrCel7A^{rec}*, both individually and as part of a minimal enzyme cocktail, than that of *TrCel7A^{nat}* can be attributed not only to lower catalytic activity (as seen on the soluble model substrate MUC) but also to impaired binding properties, which presumably relate to the lower level of *O*-glycosylation. In addition to the lack of *O*-glycosylation, the incorporation of the His₈ affinity tag at the C-terminal end of the CBM1 could affect cellulose binding negatively (Dana *et al.*, 2014). The impact of the aberrant *O*-linked glycosylation in *TrCel7A^{rec}* warrants further investigations. It is important to note here that the accumulation of a truncated form of *TrCel7A^{rec}* during production suggests that low level of *O*-glycosylation in the linker region correlates with an (expected) decrease in proteolytic stability.

In general, the purified *TrCel7A^{rec}* performed similarly to the native enzyme. This finding demonstrates that exploiting plants as expression platform for the production of cellulases is worthwhile, with stabilization of the N-terminus by pyroglutamate formation representing a particularly attractive feature for those cellulases that are dependent on this PTM. Expression of cellulases by plants, however, needs to be optimized for improved protein stability and, consequently, production efficiency, before it can be adopted in large-scale applications. The types of *O*-glycans detected in *TrCel7A^{rec}* represent a step forward in our understanding of the *O*-glycosylating machinery of *N. benthamiana*, which is essential for optimizing the stability of cellulases (and other recombinant proteins) upon expression in plants. Glycan types likely have significant effects on enzyme properties, and modification of glycosylation patterns could potentially yield cellulases with improved properties (Beckham *et al.*, 2012; Payne *et al.*, 2015). Transient expression in plants could be a method to study this in more detail, as glyco-engineering in plants is possible (Castilho and Steinkellner, 2012).

Experimental procedures

Construction of a plant expression vector for *TrCel7A*

The *Trichoderma reesei cel7A* coding sequence (*TrCel7A*, CBH1, Uniprot P62694) appended with a polyhistidine (His₈) tag at the C-terminus was used to design expression vectors for plants. Four different expression constructs were generated: a) the basic construct as described above, b) a *cel7A* gene version with the additional four residues KDEL functioning as an ER retention signal and added to the C-terminus, c) a gene version without signal peptide and d) a version in which the native signal peptide was replaced by a plant signal peptide (MANKHLSLSLFLVLLGLSA-SLASG of barley α -amylase). The sequences were codon

optimized for *N. benthamiana* expression and chemically synthesized (GeneArt, Thermo Fisher Scientific, Regensburg, Germany). The coding regions were integrated into the plant expression vector pEAQ-HT-DEST1 using Gateway cloning technology, as described previously (Dobrica *et al.*, 2017).

Transient expression in *N. benthamiana* and *TrCel7A^{rec}* analysis

The plant expression vectors were introduced into ElectroMAX™ *Agrobacterium tumefaciens* LBA4404 cells (Invitrogen, Carlsbad, CA) by electroporation as described before (Clarke *et al.*, 2008). An in-house assembled vacuum infiltration system was used to facilitate agroinfiltration on leaves of 5- to 6-week-old *N. benthamiana*, as described (Dobrica *et al.*, 2017). Total protein extraction and immunoblotting analysis were carried out essentially as described by van Eerde *et al.* (2019) and are detailed in Appendix S8.

Purification of *TrCel7A* from *N. benthamiana*

For purification of *TrCel7A*, *N. benthamiana* leaves harvested seven days after infiltration were used. Frozen leaves were ground to powder using a liquid nitrogen-cooled mortar and pestle. Ground plant material (200 g) was suspended in 1 L 0.1 M Na-lactate buffer pH 3.5 containing 0.02 M β -mercaptoethanol, incubated for 10 min at room temperature and filtered through four layers of Miracloth (Merck, Darmstadt, Germany). The filtrate was centrifuged for 20 min at 4 °C at 25 000 g. 5 M NaCl was added to the supernatant to a 0.5 M final concentration, and solid imidazole was added to a final concentration of 0.01 M imidazole. The pH of the solution was adjusted to 7.0 using a 0.5 M KH₂PO₄ solution, stirred for an hour and centrifuged for 20 min at 4 °C at 25 000 g. Next, *TrCel7A* was extracted from the total protein solution with His-tag purification. To this end, 1 mL Ni-NTA agarose beads (Qiagen, Hilden, Germany) in a sealed bag prepared according to Castaldo *et al.* (2016) from a polyester woven mesh (0.43 μ m, Spectrum, Rancho Dominguez, CA) were incubated in the protein solution under stirring overnight at 7 °C. The next day, the Ni-NTA beads were removed and washed with wash buffer containing 0.05 M potassium phosphate, 0.5 M NaCl and 0.01 M imidazole at pH 7. The absorbed proteins were finally eluted using a buffer containing 0.05 M potassium phosphate, 0.5 M NaCl and 0.25 M imidazole at pH 7. The eluate was dialysed against 0.1 M sodium acetate pH 5 containing 0.25 M NaCl, after which ammonium sulphate was added to a final concentration of 1 M. The solution was applied to a 5-mL HiTrap phenylsepharose column (GE Healthcare Bio-sciences, Uppsala, Sweden), washed with a buffer containing 0.05 M sodium lactate and 1 M ammonium sulphate at pH 3.5, and eluted with 0.05 M Na-lactate buffer pH 3.5. The fractions containing *TrCel7A* were pooled and then concentrated with a 6-mL centrifugal ultrafiltration tube (10 kDa cut-off, Pall, Ann Arbor, MI). The same ultrafiltration tube was used to exchange the buffer to 0.1 M sodium acetate and 0.25 M NaCl, pH 5.

The pI of the plant-expressed *TrCel7A* was determined with isoelectric focusing using the Criterion system of Bio-Rad Laboratories. For details, see the Appendix S8.

Analysis of post-translational modifications (PTMs) with LC-MS

Post-translational modifications (PTMs) of *TrCel7A* by *N. benthamiana* were compared to naturally occurring PTMs of *TrCel7A* expressed by *Trichoderma reesei* QM 9414 (VTT Culture

Collection, D-74075, Finland) and purified as described previously (Ståhlberg *et al.*, 1996). The purified fungal and plant-expressed *TrCel7A* variants were digested proteolytically, using trypsin (Promega), Proteinase K (Sigma-Aldrich) or chymotrypsin (Roche), and the peptides were analysed for post-translational modifications, following the methods described by Arntzen *et al.* (2017) and Anonsen *et al.* (2012). The peptides were separated and analysed using a reverse phase (C18) nano-LC-MS system (Dionex Ultimate 3000 UHPLC; Thermo Scientific, Bremen, Germany) connected to a Q-Exactive mass spectrometer (Thermo Scientific) and operated in data-dependent mode to switch automatically between orbitrap-MS and higher-energy collisional dissociation (HCD) orbitrap-MS² acquisition. The data were recorded with Xcalibur versions 2.0.7 and 2.2, and the peptide masses were identified with the Mascot search engine (Perkins *et al.*, 1999). A search for the appearance of specific low mass reporter ions (oxonium ions) at *m/z* 204.086 (*N*-acetylhexosamine), 366.139 (*N*-acetylhexosamine-hexose), 163.061 (hexose) and 133.049 (pentose) in Xcalibur software (Thermo Scientific) was carried out. For more details, see the Appendix S8.

Activity measurements of the plant-expressed and native *TrCel7A* on MUC, a soluble model substrate

Cellobiohydrolase activity of the plant-expressed *TrCel7A*^{tec} and the fungal *TrCel7A*^{nat} was measured against 4-methylumbelliferyl- β -D-cellobioside (MUC, Sigma-Aldrich, Darmstadt, Germany), a water-soluble model compound. Specific activities were calculated as U/mg protein, 1 U corresponding to 1 μ mol 4-methylumbelliferone (4-MU) formed per minute in the reaction. For details, see the Appendix S8.

Activity on Avicel, a cellulosic model substrate

Reaction mixtures containing 2% (w/v) Avicel (PH-101, Sigma-Aldrich, St. Louis, MO) and 75 nM fungal or plant-expressed *TrCel7A* were prepared in 12.5 mM Na-acetate buffer (pH 5.0). Reactions, in a total volume of 200 μ L, were incubated at 50 °C for 24 h; reactions were performed in triplicates. Samples were harvested after 1, 6 and 24 h, and the reaction was stopped by filtering the whole reaction mixture through a 96-well filter plate equipped with 0.45- μ m filter membrane (Merck Millipore Ltd., Tullagreen Carrigtwohill, Ireland). The released cellobiose was converted to glucose by adding β -glucosidase (*AnCel3A*) from *Aspergillus niger* (Megazyme, Bray, Ireland) to the sample supernatant at 0.01 g/L β -glucosidase concentration and incubating the mixture at 50 °C for 15 min. The amount of glucose was then determined using the AmplexTM Red glucose/glucose oxidase assay (<https://assets.thermofisher.com/TFS-Assets/LSG/manuals/mp22188.pdf>) by Thermo Fisher Scientific. D-glucose in the samples was reacted with 0.2 U/mL glucose oxidase from *Aspergillus niger* (G7141, Sigma-Aldrich) for 1 h at room temperature, and the released H₂O₂ was used to convert Amplex Red (Cayman Chemical, Ann Arbor, MI) to resorufin using horseradish peroxidase (P8250, Sigma-Aldrich). Absorbance was measured at 540 nm; H₂O₂ (Sigma-Aldrich) was used as standard.

Specific activity of the *TrCel7A* variants was determined by incubating the enzyme with 4% (w/v) Avicel in 12.5 mM Na-acetate buffer (pH 5.0) at 50 °C for 60 min at enzyme concentrations in the range of 75–125 nM. Reactions were performed in a total volume of 200 μ L, in triplicates. Cellobiose was quantified with the Amplex Red assay as described above. Linear correlation

between the enzyme concentration and cellobiose release indicated that the enzymes were working at the initial, linear range of the reaction. Specific activities were calculated as U/mg protein, 1 U corresponding to 1 μ mol cellobiose formed per minute in the reaction.

Activity on BALI-pretreated spruce, an industrially relevant substrate

The efficiency of the plant-expressed *TrCel7A* was compared to that of the native, fungal *TrCel7A* in a minimal enzyme cocktail that had been developed for BALI-pretreated spruce by Chylenski *et al.* (2017). The minimal enzyme cocktail (marked as '100%') was loaded based on the glucan content of the substrate and was composed of 1.97 mg *TrCel7A*, 1.18 mg *TrCel6A* and 3.40 mg *TrCel7B* from *T. reesei* and 0.50 mg β -glucosidase (*AnCel3A*) from *A. niger* per gram glucan (Chylenski *et al.*, 2017); the enzyme cocktail contained either the plant-expressed or the fungal *TrCel7A*. In this experiment, we either omitted *TrCel7A* from the enzyme cocktail ('0%', i.e. 0 mg *TrCel7A*/g glucan) or tested three levels of *TrCel7A*: '100%' (1.97 mg *TrCel7A*/g glucan, as dosed in the minimal enzyme cocktail), '125%' (2.47 mg *TrCel7A*/g glucan, i.e. adding 25% more *TrCel7A*) and '150%' (2.9 mg *TrCel7A*/g glucan, i.e. adding 50% more *TrCel7A*) to the cocktail while maintaining the loading of other cellulases per g substrate constant. In 500 μ L reaction volumes, BALI-pretreated spruce (5%, w/v) was incubated in 50 mM Na-acetate (pH 5.0) at 50 °C for 48 h in an Eppendorf thermomixer (Eppendorf AG, Hamburg, Germany) with shaking at 1000 rpm, with a total enzyme loading of 8 mg enzyme/g glucan. Reactions were performed in triplicates. The reactions were terminated by boiling the samples for 15 min; subsequently, 1 mL ultrapure water was added to each tube to decrease the error due to high solids loading. The samples were centrifuged at 11 000 *g* for 15 min and then filtered through a 96-well filter plate equipped with 0.45- μ m filter membrane (Millipore, MA). Glucose release was analysed by HPLC using the Dionex Ultimate 3000 system (Dionex, Sunnyvale, CA) coupled with a Shodex RI-101 Refractive Index (RI) detector (Showa Denko KK, Japan). Separation of hydrolysis products was performed on a Rezex RFQ-fast acid H⁺ (8%) 100 \times 7.8 mm column (Phenomenex, Torrance, CA), operated at 85 °C at 1 mL/min flow rate, with 5 mM H₂SO₄ as mobile phase.

Adsorption experiments

Avicel, BALI-pretreated spruce and bleached BALI-pretreated spruce (5% (w/w) dry weight) were suspended in 50 mM Na-acetate buffer (pH 5.0) containing 0.1 g/L *TrCel7A* in a total volume of 200 μ L. *TrCel7A* solution (0.1 g/L) without substrate was used as reference. Supernatants and solid suspensions were run on a 10% Mini-PROTEAN TGX Stain-Free precast gel (Bio-Rad Laboratories). The amount of protein in the gel was quantified based on fluorescence following the method described previously (Várnai *et al.*, 2011). Finally, the extent of adsorption was calculated based on *TrCel7A* band intensities.

Acknowledgements

This work was supported by the Research Council of Norway, through grants 243974, 256766 and 257622, as well as by NIBIO core funding and the Max Planck Society. The authors thank Sissel Haugslie and Astrid Sivertsen for technical support.

Conflict of interest

There is no conflict of interest for the current study.

Author contributions

AvE, AV, PC, VGHE, RB and JLC conceived and designed the study. LP carried out vector design, sequencing, supervised technical staff and participated in agroinfiltration; HSS and IH participated in the experimental plan and conducted all the agroinfiltration work. AvE performed protein purification and analysis; JKJ analysed enzyme properties; JKJ, AM and JHA analysed PTMs. AvE, AV, AM, JHA and JLC drafted the manuscript; all authors reviewed and edited the manuscript.

References

- Adney, W.S., Jeoh, T., Beckham, G.T., Chou, Y.-C., Baker, J.O., Michener, W., Brunecky, R. et al. (2009) Probing the role of *N*-linked glycans in the stability and activity of fungal cellobiohydrolases by mutational analysis. *Cellulose*, **16**, 699–709.
- Amore, A., Knott, B.C., Supekar, N.T., Shajahan, A., Azadi, P., Zhao, P., Wells, L. et al. (2017) Distinct roles of *N*- and *O*-glycans in cellulase activity and stability. *Proc. Natl Acad. Sci. USA*, **114**, 13667–13672.
- Anonsen, J.H., Vik, A., Egge-Jacobsen, W. and Koomey, M. (2012) An extended spectrum of target proteins and modification sites in the general *O*-linked protein glycosylation system in *Neisseria gonorrhoeae*. *J. Proteome Res.* **11**, 5781–5793.
- Arntzen, M.O., Várnai, A., Mackie, R.I., Eijsink, V.G.H. and Pope, P.B. (2017) Outer membrane vesicles from *Fibrobacter succinogenes* S85 contain an array of carbohydrate-active enzymes with versatile polysaccharide-degrading capacity. *Environ. Microbiol.* **19**, 2701–2714.
- Beckham, G.T., Dai, Z., Matthews, J.F., Momany, M., Payne, C.M., Adney, W.S., Baker, S.E. et al. (2012) Harnessing glycosylation to improve cellulase activity. *Curr. Opin. Biotechnol.* **23**, 338–345.
- Bock, R. (2015) Engineering plastid genomes: methods, tools, and applications in basic research and biotechnology. *Annu. Rev. Plant Biol.* **66**, 211–241.
- Bosch, D., Castilho, A., Loos, A., Schots, A. and Steinkellner, H. (2013) *N*-glycosylation of plant-produced recombinant proteins. *Curr. Pharm. Des.* **19**, 5503–5512.
- Castaldo, M., Barilind, L., Mauritzson, F., Wan, P.T. and Snijder, H.J. (2016) A fast and easy strategy for protein purification using “teabags”. *Sci. Rep.* **6**, 28887.
- Castilho, A. and Steinkellner, H. (2012) Glyco-engineering in plants to produce human-like *N*-glycan structures. *Biotechnol. J.* **7**, 1088–1098.
- Castilho, A., Windwarder, M., Gattinger, P., Mach, L., Strasser, R., Altmann, F. and Steinkellner, H. (2014) Proteolytic and *N*-glycan processing of human α 1-antitrypsin expressed in *Nicotiana benthamiana*. *Plant Physiol.* **166**, 1839–1851.
- Chylenski, P., Forsberg, Z., Ståhlberg, J., Várnai, A., Lersch, M., Bengtsson, O., Sæbo, S. et al. (2017) Development of minimal enzyme cocktails for hydrolysis of sulfite-pulped lignocellulosic biomass. *J. Biotechnol.* **246**, 16–23.
- Clarke, J.L. and Zhang, P. (2013) Plant biotechnology for food security and bioeconomy. *Plant Mol. Biol.* **83**, 1–3.
- Clarke, J.L., Spetz, C., Haugslie, S., Xing, S., Dees, M.W., Moe, R. and Blystad, D.R. (2008) *Agrobacterium tumefaciens*-mediated transformation of poinsettia, *Euphorbia pulcherrima*, with virus-derived hairpin RNA constructs confers resistance to *Poinsettia mosaic virus*. *Plant Cell Rep.* **27**, 1027–1038.
- Clarke, J.L., Paruch, L., Dobrica, M.O., Caras, I., Tucureanu, C., Onu, A., Ciulean, S. et al. (2017) Lettuce-produced hepatitis C virus E1E2 heterodimer triggers immune responses in mice and antibody production after oral vaccination. *Plant Biotechnol. J.* **15**, 1611–1621.
- Dai, Z., Hooker, B.S., Quesenberry, R.D. and Gao, J. (1999) Expression of *Trichoderma reesei* exo-cellobiohydrolase I in transgenic tobacco leaves and calli. *Appl. Biochem. Biotechnol.* **77–79**, 689–699.
- Dana, C.M., Dotson-Fagerstrom, A., Roche, C.M., Kal, S.M., Chokhawala, H.A., Blanch, H.W. and Clark, D.S. (2014) The importance of pyroglutamate in cellulase Cel7A. *Biotechnol. Bioeng.* **111**, 842–847.
- Daniell, H., Streatfield, S.J. and Rybicki, E.P. (2015) Advances in molecular farming: key technologies, scaled up production and lead targets. *Plant Biotechnol. J.* **13**, 1011–1012.
- Daniell, H., Ribeiro, T., Lin, S., Saha, P., McMichael, C., Chowdhary, R. and Agarwal, A. (2019) Validation of leaf and microbial pectinases: commercial launching of a new platform technology. *Plant Biotechnol. J.* **17**, 1154–1166.
- Deshpande, N., Wilkins, M.R., Packer, N. and Nevalainen, H. (2008) Protein glycosylation pathways in filamentous fungi. *Glycobiology*, **18**, 626–637.
- Dicker, M., Tschofen, M., Maresch, D., König, J., Juarez, P., Orzaez, D., Altmann, F. et al. (2016) Transient glyco-engineering to produce recombinant IgA1 with defined *N*- and *O*-glycans in plants. *Front. Plant Sci.* **7**, 18.
- Divne, C., Ståhlberg, J., Reinikainen, T., Ruohonen, L., Pettersson, G., Knowles, J.K., Teeri, T.T. et al. (1994) The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. *Science*, **265**, 524–528.
- Dobrica, M.O., Lazar, C., Paruch, L., Skomedal, H., Steen, H., Haugslie, S., Tucureanu, C. et al. (2017) A novel chimeric Hepatitis B virus S/preS1 antigen produced in mammalian and plant cells elicits stronger humoral and cellular immune response than the standard vaccine-constituent, S protein. *Antiviral Res.* **144**, 256–265.
- van Eerde, A., Gottschamel, J., Bock, R., Hansen, K.E.A., Munang'andu, H.M., Daniell, H. and Liu Clarke, J. (2019) Production of tetravalent dengue virus envelope protein domain III based antigens in lettuce chloroplasts and immunologic analysis for future oral vaccine development. *Plant Biotechnol. J.* **17**, 1408–1417.
- Fägerstam, L.G. and Pettersson, L.G. (1980) The 1,4- β -glucan cellobiohydrolases of *Trichoderma reesei* QM 9414. A new type of cellulolytic synergism. *FEBS Lett.* **119**, 97–100.
- Garvey, M., Klose, H., Fischer, R., Lambert, C. and Commandeur, U. (2013) Cellulases for biomass degradation: comparing recombinant cellulase expression platforms. *Trends Biotechnol.* **31**, 581–593.
- Garvey, M., Klinger, J., Klose, H., Fischer, R. and Commandeur, U. (2014) Expression of recombinant cellulase Cel5A from *Trichoderma reesei* in tobacco plants. *J. Vis. Exp.* **88**, e51711.
- Goto, M. (2007) Protein *O*-glycosylation in fungi: diverse structures and multiple functions. *Biosci. Biotechnol. Biochem.* **71**, 1415–1427.
- Gritzali, M. and Brown, R.D. (1979) The cellulase system of *Trichoderma*. In *Hydrolysis of Cellulose: Mechanisms of Enzymatic and Acid Catalysis* (Brown, R.D. and Jurasek, L., eds), pp. 237–260. Washington: American Chemical Society.
- Hahn, S., Giritch, A., Bartels, D., Bortesi, L. and Gleba, Y. (2015) A novel and fully scalable *Agrobacterium* spray-based process for manufacturing cellulases and other cost-sensitive proteins in plants. *Plant Biotechnol. J.* **13**, 708–716.
- Harrison, M.J., Nouwens, A.S., Jardine, D.R., Zachara, N.E., Gooley, A.A., Nevalainen, H. and Packer, N.H. (1998) Modified glycosylation of cellobiohydrolase I from a high cellulase-producing mutant strain of *Trichoderma reesei*. *Eur. J. Biochem.* **256**, 119–127.
- Harrison, M.D., Geijskes, J., Coleman, H.D., Shand, K., Kinkema, M., Palupe, A., Hassall, R. et al. (2011) Accumulation of recombinant cellobiohydrolase and endoglucanase in the leaves of mature transgenic sugar cane. *Plant Biotechnol. J.* **9**, 884–896.
- Himmel, M.E., Ding, S.Y., Johnson, D.K., Adney, W.S., Nimlos, M.R., Brady, J.W. and Foust, T.D. (2007) Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science*, **315**, 804–807.
- Hui, J.P.M., Lanthier, P., White, T.C., McHugh, S.G., Yaguchi, M., Roy, R. and Thibault, P. (2001) Characterization of cellobiohydrolase I (Cel7A) glycoforms from extracts of *Trichoderma reesei* using capillary isoelectric focusing and electrospray mass spectrometry. *J. Chromatogr. B*, **752**, 349–368.
- Hussain, H.I., Nugent, G.D., Stevenson, K., Stalker, D.M. and Stevenson, T.W. (2015) Co-expression of microbial cellulases in transgenic wheat as a potential source for cellulosic ethanol production. *Plant Biotechnol. J.* **32**, 291–297.
- Jeoh, T., Michener, W., Himmel, M.E., Decker, S.R. and Adney, W.S. (2008) Implications of cellobiohydrolase glycosylation for use in biomass conversion. *Biotechnol. Biofuels*, **1**, 10.
- Jin, S. and Daniell, H. (2015) The engineered chloroplast genome just got smarter. *Trends Plant Sci.* **20**, 622–640.

- Klein-Marcuschamer, D., Oleskiewicz-Popiel, P., Simmons, B.A. and Blanch, H.W. (2012) The challenge of enzyme cost in the production of lignocellulosic biofuels. *Biotechnol. Bioeng.* **109**, 1083–1087.
- Kont, R., Kari, J., Borch, K., Westh, P. and Välijamäe, P. (2016) Inter-domain synergism is required for efficient feeding of cellulose chain into active site of cellobiohydrolase Cel7A. *J. Biol. Chem.* **291**, 26013–26023.
- Lambertz, C., Garvey, M., Klinger, J., Heesel, D., Klose, H., Fischer, R. and Commandeur, U. (2014) Challenges and advances in the heterologous expression of cellulolytic enzymes: a review. *Biotechnol. Biofuels*, **7**, 135.
- Li, Y., Liu, P., Huang, J., Zhang, R., Hu, Z., Feng, S., Wang, Y. *et al.* (2018) Mild chemical pretreatments are sufficient for bioethanol production in transgenic rice straws overproducing glucosidase. *Green Chem.* **20**, 2047–2056.
- Li, Y., Sun, H., Fan, C., Hu, H., Wu, L., Jin, X., Lv, Z. *et al.* (2019) Overproduction of fungal endo- β -1,4-glucanase leads to characteristic lignocellulose modification for considerably enhanced biomass enzymatic saccharification and bioethanol production in transgenic rice straw. *Cellulose*. <https://doi.org/10.1007/s10570-019-02500-2>
- Messer, M. (1963) Enzymatic cyclization of L-glutamine and L-glutamyl peptides. *Nature*, **197**, 1299.
- Mewono, L., Nguema-Ona, E., Gotte, M., Koroney, A.S., Follet-Gueye, M.-L., Azeddine, D., Maité, V.-G. *et al.* (2015) O-glycosylation in plant and mammal cells: the use of chemical inhibitors to understand the biosynthesis and function of O-glycosylated proteins. *Plant Sci. Today*, 43–51.
- von Ossowski, I., Teeri, T., Kalkkinen, N. and Oker-Blom, C. (1997) Expression of a fungal cellobiohydrolase in insect cells. *Biochem. Biophys. Res. Commun.* **233**, 25–29.
- Park, S.H., Ong, R.G. and Sticklen, M. (2016) Strategies for the production of cell wall-deconstructing enzymes in lignocellulosic biomass and their utilization for biofuel production. *Plant Biotechnol. J.* **14**, 1329–1344.
- Payne, C.M., Resch, M.G., Chen, L.Q., Crowley, M.F., Himmel, M.E., Taylor, L.E. II, Sandgren, M. *et al.* (2013) Glycosylated linkers in multimodular lignocellulose-degrading enzymes dynamically bind to cellulose. *Proc. Natl Acad. Sci. USA*, **110**, 14646–14651.
- Payne, C.M., Knott, B.C., Mayes, H.B., Hansson, H., Himmel, M.E., Sandgren, M., Ståhlberg, J. *et al.* (2015) Fungal cellulases. *Chem. Rev.* **115**, 1308–1448.
- Perkins, D.N., Pappin, D.J., Creasy, D.M. and Cottrell, J.S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, **20**, 3551–3567.
- Petersen, K. and Bock, R. (2011) High-level expression of a suite of thermostable cell wall-degrading enzymes from the chloroplast genome. *Plant Mol. Biol.* **76**, 311–321.
- Peyret, H. and Lomonosoff, G.P. (2015) When plant virology met *Agrobacterium*: the rise of the deconstructed clones. *Plant Biotechnol. J.* **13**, 1121–1135.
- Qi, F., Zhang, W., Zhang, F., Chen, G. and Liu, W. (2014) Deciphering the effect of the different N-glycosylation sites on the secretion, activity, and stability of cellobiohydrolase I from *Trichoderma reesei*. *Appl. Environ. Microbiol.* **80**, 3962–3971.
- Ranaei Siadat, S.O., Mollasalehi, H. and Heydarzadeh, N. (2016) Substrate affinity and catalytic efficiency are improved by decreasing glycosylation sites in *Trichoderma reesei* cellobiohydrolase I expressed in *Pichia pastoris*. *Biotechnol. Lett.* **38**, 483–488.
- Sainsbury, F., Thuenemann, E.C. and Lomonosoff, G.P. (2009) pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnol. J.* **7**, 682–693.
- Schilling, S., Stenzel, I., von Bohlen, A., Wermann, M., Schulz, K., Demuth, H.U. and Wasternack, C. (2007) Isolation and characterization of the glutamyl cyclases from *Solanum tuberosum* and *Arabidopsis thaliana*: implications for physiological functions. *Biol. Chem.* **388**, 145–153.
- Schneider, J.D., Marillonnet, S., Castilho, A., Gruber, C., Werner, S., Mach, L., Klimyuk, V. *et al.* (2014) Oligomerization status influences subcellular deposition and glycosylation of recombinant butyrylcholinesterase in *Nicotiana benthamiana*. *Plant Biotechnol. J.* **12**, 832–839.
- Ståhlberg, J., Divne, C., Koivula, A., Piens, K., Claeysens, M., Teeri, T.T. and Jones, T.A. (1996) Activity studies and crystal structures of catalytically deficient mutants of cellobiohydrolase I from *Trichoderma reesei*. *J. Mol. Biol.* **264**, 337–349.
- Stals, I., Sandra, K., Devreese, B., Van Beeumen, J. and Claeysens, M. (2004a) Factors influencing glycosylation of *Trichoderma reesei* cellulases. II: N-glycosylation of Cel7A core protein isolated from different strains. *Glycobiology*, **14**, 725–737.
- Stals, I., Sandra, K., Geysens, S., Contreras, R., Van Beeumen, J. and Claeysens, M. (2004b) Factors influencing glycosylation of *Trichoderma reesei* cellulases. I: Postsecretorial changes of the O- and N-glycosylation pattern of Cel7A. *Glycobiology*, **14**, 713–724.
- Strasser, R. (2016) Plant protein glycosylation. *Glycobiology*, **26**, 926–939.
- Taylor, L.E. II, Dai, Z., Decker, S.R., Brunecky, R., Adney, W.S., Ding, S.Y. and Himmel, M.E. (2008) Heterologous expression of glycosyl hydrolases in *planta*: a new departure for biofuels. *Trends Biotechnol.* **26**, 413–424.
- Tomme, P., Van Tilbeurgh, H., Pettersson, G., Van Damme, J., Vandekerckhove, J., Knowles, J., Teeri, T. *et al.* (1988) Studies of the cellulolytic system of *Trichoderma reesei* QM 9414. Analysis of domain function in two cellobiohydrolases by limited proteolysis. *Eur. J. Biochem.* **170**, 575–581.
- Várnai, A., Viikari, L., Marjamaa, K. and Siika-aho, M. (2011) Adsorption of monocomponent enzymes in enzyme mixture analyzed quantitatively during hydrolysis of lignocellulose substrates. *Bioresour. Technol.* **102**, 1220–1227.
- Verma, D., Kanagaraj, A., Jin, S., Singh, N.D., Kolattukudy, P.E. and Daniell, H. (2010) Chloroplast-derived enzyme cocktails hydrolyse lignocellulosic biomass and release fermentable sugars. *Plant Biotechnol. J.* **8**, 332–350.
- Verma, D., Jin, S., Kanagaraj, A., Singh, N.D., Daniel, J., Kolattukudy, P.E., Miller, M. *et al.* (2013) Expression of fungal cutinase and swollenin in tobacco chloroplasts reveals novel enzyme functions and/or substrates. *PLoS ONE*, **8**, e57187.
- Wang, M., Ma, Y., Li, L., Wang, B., Wei, X., Zhang, M., Wang, J. *et al.* (2019) The diversity of glycosylation of cellobiohydrolase I from *Trichoderma reesei* determined with mass spectrometry. *Biochem. Biophys. Res. Commun.* **508**, 818–824.
- Wu, V.W., Dana, C.M., Iavarone, A.T., Clark, D.S. and Glass, N.L. (2017) Identification of glutamyl cyclase genes involved in pyroglutamate modification of fungal lignocellulolytic enzymes. *MBio*, **8**, e02231-16.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix S1 (incl. Fig. S1): Expression of *TrCel7A* constructs in infiltrated *N. benthamiana* leaves.

Appendix S2 (incl. Fig. S2): Protein sequence and peptide coverage in the LC-MS analysis.

Appendix S3 (incl. Fig. S3): PyroQ modification in the *TrCel7A* variants.

Appendix S4 (incl. Figs. S4–S7): LC-MS analysis of *N*-glycans in *TrCel7A*^{nat}.

Appendix S5 (incl. Figs. S8–S11): LC-MS analysis of *N*-glycans in plant-expressed *TrCel7A*^{rec}.

Appendix S6 (incl. Figs. S12–S17): LC-MS analysis of *O*-glycans in *TrCel7A*^{nat}.

Appendix S7 (incl. Figs. S18–S19): LC-MS analysis of *O*-glycans in *TrCel7A*^{rec}.

Appendix S8 Additional experimental procedures.