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 Eerden¹, Aniko Várnaí², John Kristian Jameson², Lisa Paruch¹, Anders Moen³, Jan Haug Anonsen³, Piotr Chylenski⁴, Hege Særvold Steen⁵, Inger Heldal⁶, Ralph Bock⁷, Vincent G. H. Eijsink² and Jihong Liu-Clarke⁹

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

*Correspondence (Tel +47 9959 4790; email Jihong.liu-clarke@nibio.no (JLC); Tel +47 4517 4259; email aniko.varna@nmbu.no (AV); Tel +47 2285 7295; email j.h.anonsen@lbv.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 post-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*nat was correctly processed by *N. benthamiana* to a pyrogul-tamate, 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* *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 production platforms, including thermotolerant 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.

\textit{TrCel7A} forms the major component of the \textit{Trichoderma reesei} secretome. When the fungus grows on cellulose, \textit{TrCel7A} comprises ca. 60% of the total secreted proteins (Gritzali and Brown, 1979), highlighting its importance in plant cell wall degradation. \textit{TrCel7A}, also known as celllobiohydrolase I, is a part of the cellulolytic machinery of \textit{T. reesei} and depolymerizes cellulose by cleaving off cellobiose units from the reducing end of cellulose chains in a processive manner. \textit{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 \textit{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 \textit{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 (Fägerström 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 \textit{TrCel7A}, and cellulases in general. Therefore, in the present study, we expressed \textit{TrCel7A} from \textit{T. reesei} in \textit{Nicotiana benthamiana} (\textit{TrCel7A}\textsuperscript{rec}) using vacuum agroinfiltration. We mapped the PTMs and compared the properties of the recombinant enzyme to those of the native \textit{TrCel7A} (\textit{TrCel7A}\textsuperscript{nat}).

\textbf{Results}

\textbf{\textit{TrCel7A} production in \textit{N. benthamiana} and protein analysis}

In order to assess the potential of \textit{N. benthamiana} as a production platform for fungal cellulases, we cloned the gene coding for \textit{TrCel7A} and expressed it in \textit{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 \textit{N. benthamiana} leaves with antibodies recognizing the C-terminal His\textsubscript{6}-tag suggested that a construct containing the barley \textit{α}-amylase signal sequence, which has a proven efficiency in the production of \textit{TrCel7A} in \textit{N. benthamiana} (Hahn et al., 2015), exhibited the highest accumulation of \textit{TrCel7A} and was therefore chosen for further study. Full-length \textit{TrCel7A} 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 His\textsubscript{6}-tag. 0.4 mg pure \textit{TrCel7A}\textsuperscript{rec} could be isolated from 200 g transformed plant material.

To investigate \textit{TrCel7A}\textsuperscript{rec} accumulation and purity, plant extracts were analysed with SDS-PAGE, staining with Coomassie Brilliant Blue (Figure 1b). The band corresponding to full-length \textit{TrCel7A}\textsuperscript{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 \textit{TrCel7A}, lacking the CBM and the His-tag. This shows that a high level of \textit{TrCel7A}\textsuperscript{rec} is produced in \textit{N. benthamiana}, but only a small fraction is present as full-length \textit{TrCel7A}\textsuperscript{rec}.

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

\textbf{Analysis of post-translational modifications (PTMs)}

Post-translational modifications have been shown to affect enzyme stability, catalytic efficiency and substrate binding of \textit{TrCel7A}. We therefore analysed and compared the nature and relative abundance of PTMs decorating \textit{TrCel7A} produced in \textit{N. benthamiana} (\textit{TrCel7A}\textsuperscript{rec}) and native \textit{TrCel7A} purified from the secretome of \textit{T. reesei} strain QM 9414 (\textit{TrCel7A}\textsuperscript{nat}), using LC-MS\textsuperscript{5} (Table 1, Figure 2). To maximize protein sequence coverage, and thereby increase the likelihood of PTM detection, we employed a multi-enzyme approach (digesting the proteins by trypsin and chymotrypsin) that allowed the detection of 93% of \textit{TrCel7A}\textsuperscript{rec} and 100% of \textit{TrCel7A}\textsuperscript{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 TrCel7Arec and TrCel7A\textsuperscript{cat} 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 glycosylation) and quantification of their relative abundance are summarized in Appendices S3–S7.

PyroQ analysis

Analysis of the N-terminal peptides revealed that both TrCel7A\textsuperscript{cat} and TrCel7A\textsuperscript{rec} were modified by pyroQ (Table 1 and Figure S3); MS\textsuperscript{2} fragmentation of the doubly charged ion at m/z 1047.999 from both TrCel7A\textsuperscript{cat} and TrCel7A\textsuperscript{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)SethPLWTQK\textsuperscript{18}. The C-terminal part of the peptide (reflected by y1-y11 in the y-ion series) contained only unmodified amino acid residues up to the y11 position (i.e. SETHPLWTQK\textsuperscript{18}), 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. 1QS\textsuperscript{2} and 1QSA\textsuperscript{2}, 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\textsuperscript{cat} and TrCel7A\textsuperscript{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\textsuperscript{cat} (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 completely modified (RGA 100%), 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\textsuperscript{cat}, with seven types of high-mannose structures detected in total (Figure S6).

Similar to TrCel7A\textsuperscript{cat}, the Asn45, Asn270 and Asn384 sites of plant-produced TrCel7A\textsuperscript{rec} were almost completely modified (RGA > 99% vs. 99.5%–100% in TrCel7A\textsuperscript{cat}) in TrCel7A\textsuperscript{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\textsuperscript{cat} (100%, 37.6% and 100%, respectively), a single GlcNAc was the major glycoform only at the Asn384 site (61.3%) in TrCel7A\textsuperscript{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 GlcNAc\textsubscript{3}Man\textsubscript{3}Fuc\textsubscript{1}Xyl\textsubscript{1} (36.2%) and GlcNAc\textsubscript{3}Man\textsubscript{3}Fuc\textsubscript{2}Xyl\textsubscript{1} (30.1%). The N-glycans on the Asn270 site showed the highest microheterogeneity in TrCel7A\textsuperscript{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 (GlcNAc\textsubscript{2}Man\textsubscript{3}Fuc\textsubscript{1}Xyl\textsubscript{1}, GlcNAc\textsubscript{2}Man\textsubscript{2}Fuc\textsubscript{1}Xyl\textsubscript{1}, GlcNAc\textsubscript{2}Man\textsubscript{2}Fuc\textsubscript{2}Xyl\textsubscript{1}, GlcNAc\textsubscript{2}Man\textsubscript{2}Xyl\textsubscript{1}, GlcNAc\textsubscript{2}Man\textsubscript{3}Xyl\textsubscript{1}).

N-glycan analysis

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\textsuperscript{cat} catalytic domain, and (6) purified full-length TrCel7A\textsuperscript{cat}. Samples 1, 5 and 6 contained 1 μg of protein, whereas samples 2–4 contained 2 μg. Unstained Protein MW Marker (Thermo) was used as marker.
### 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 | Q5ACTLQSETHPLTWQK | Trypsin | 1047.999 M2H4⁺ | Fungus | – | – | 100 |
| N-glycosylation | Asn45 | RWHATNNSTNcY | Chymotrypsin | 1597.681 MH⁺ | Fungus | 100 | GlcNAc₁ | 100.0 |
| | Asn64 | SSTcPD2NETcAKncLDGAAY | Chymotrypsin | 2506.994 MH⁺ | Fungus | 2 | GlcNAc₂Man₉Gal₇Fuc₂Xyl₁ | 93.1 |
| | | | | | Plant | 5.1 | GlcNAc₂Man₉Gal₇Fuc₂Xyl₁ | 51.1 |
| | Asn270 | RLGNTSFY | Chymotrypsin | 794.416 MH⁺ | Fungus | 99.7 | GlcNAc₁ | 37.6 |
| | | | | | Plant | 99.1 | GlcNAc₁Man₉Gal₉Fuc₂Xyl₁ | 18.5 |
| | Asn384 | WLDYPNTSSSTPGAVR | Chymotrypsin | 2081.977 MH⁺ | Fungus | 99.9 | GlcNAc₁ | 100.0 |
| | | | | | Plant | 100 | GlcNAc₁Man₉Gal₇Fuc₂Xyl₁ | 61.3 |
To determine the O-linked glycosylations of TrCel7Anat and TrCel7Arec, a similar strategy as for the detection of peptides modified with N-glycans was pursued. For TrCel7Anat, 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]+, charge-adjusted mass 5799.339 Da [M+H]+; corresponding to the peptide 433GGNPPGNNPPTTTTRPATTGSPGPTOSHY 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 TrCel7Anat, indicating a strong correlation between linker O-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, \(474\text{SGPTVCASGTTC}485\), through the doubly charged precursor ion \(m/z 680.274 [M + 2H]^2^+\); corresponding to the doubly carbamidomethylated \(474\text{SGPTVCASGTTC}485\) 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).

O-glycosylation was found at low levels also at the C-terminal part of the catalytic domain, in the peptide \(395\text{GSCSTSGVPAQVESQPNAK}415\), containing potential O-glycosylation patches at

---

**Figure 2** Summary of PTMs detected in the plant-produced \(Tr\text{Cel7A}_{\text{rec}}\) and the native fungal \(Tr\text{Cel7A}_{\text{nat}}\). 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 \(Tr\text{Cel7A}_{\text{rec}}\) (A) and the MS\(^2\) fragmentation of the most abundant O-glycoform (B). The MS\(^2\) fragmentation for all O-glycopeptides is given in Appendix S7.
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 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]+), 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+2H]+) at the Pro458 position (Figure S18). The analyses showed that this linker peptide carries multiple plant-specific O-glycan structures and allowed the assembly of the most common O-glycan structure in the TrCel7Arec linker as shown in Figure 2.

In contrast with the abundance of O-glycosylation in TrCel7Arec, only a single O-glycosylation site in the linker region (427TRRPATTTGGSPQGPTSH)465 was detected despite 95% coverage of the protein sequence (Figure S2A). In addition to O-glycans, M5² fragmentation of the precursor ion belonging to the non-glycosylated peptide (m/z 505.247; observed mass of 2017.969 Da [M+4H]⁴) revealed a hydroxyproline (pGly) 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). M5² fragmentation of the corresponding reporter ions revealed intermediate glycan structures and allowed the assembly of the most common O-glycan structure in the TrCel7Arec linker as shown in Figure 2. As an example, fragmentation of the quadruple charged precursor ion at m/z 732.818 ([M+4H]+²); observed mass of 2925.703 Da, [M+3H]⁴) 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), with a combined RGA (including mono- and di-glycosylated versions) of <2%.

In Figure 4, Activity of the TrCel7A variants on methylumbelliferyl cellobioside (MUC) and Avicel. (a) Activity of TrCel7Arec (green lines) and TrCel7Anat (black lines) on MUC. MUC (0.4 mM) was incubated with 20–31 μg/ml plant-produced TrCel7Arec and fungal TrCel7Anat 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 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 100 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 TrCel7Arec or fungal TrCel7Anat. Reactions were incubated in 50 mM Na-acetate buffer 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.

© 2019 The Authors. Plant Biotechnology Journal published by Society for Experimental Biology and The Association of Applied Biologists and John Wiley & Sons Ltd., 18, 631–643
attached to the C-terminal CBM, are unlikely to affect the linker region and the CBM as well as the His-tag, which is visualized 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.

![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.](image)

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, which could affect substrate-binding properties. The binding assay revealed impaired binding efficiency of the plant-expressed TrCel7A compared with the native fungal TrCel7A (Figure 5). The results show that the enzyme cocktail with the plant-expressed TrCel7A (100% TrCel7A) gave approximately 25% lower yield after 48 h of incubation than the enzyme cocktail containing the fungal TrCel7A (100% TrCel7Anat, reference cocktail) (Figure 5). Compared to the reactions without TrCel7A (No TrCel7A), both TrCel7Arec and TrCel7Anat increased the glucan yields. For the plant version, however, increasing the amount of TrCel7Arec (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 TrCel7Arec compared with the native fungal TrCel7Anat for all three substrates. TrCel7Anat bound to Avicel to a high extent, leaving only 5% of the total loaded protein free in solution, whereas some 40% of TrCel7Arec 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 TrCel7Anat and 38% of TrCel7Arec 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 Lomonossoff, 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/phyllozyme; Daniell et al., 2019). Leaf Expression Systems is a transient expression-based system using Agrobacterium-mediated transient expression in leaves, while PhyloZyme 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 production in yeast (Dana 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<sup>nat</sup> 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<sup>rec</sup> in N. benthamiana yields 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<sup>nat</sup> and TrCel7A<sup>rec</sup>. Asn45 and Asn384, the residues located at the entrance and exit of the catalytic tunnel, respectively, carried only a single GlcNAc in TrCel7Arec and thus lower catalytic efficiency. Limited accessibility of the catalytic tunnel would result in lower catalytic efficiency of TrCel7A<sup>rec</sup> as compared with TrCel7A<sup>nat</sup>, 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<sup>nat</sup> was identified only as a single, heavily glycosylated peptide. We could not obtain smaller fragments of the linker peptide in TrCel7A<sup>rec</sup> with performed an extensive comparison of the catalytic and substrate-binding properties of the plant-produced TrCel7A<sup>rec</sup> and the native TrCel7A<sup>nat</sup> 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 pyroglycocolate. Here we showed that the N-terminus of TrCel7A<sup>rec</sup> was faithfully converted to pyroglycocolate 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). Asn384, the residues located at the entrance and exit of the catalytic tunnel, respectively, carried only single GlcNAc in TrCel7Arec. Asn45 and Asn384, the residues located at the entrance and exit of the catalytic tunnel, respectively, carried only single GlcNAc in TrCel7Arec, which could indicate that sufficient levels of glutaminyl 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.
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 TrCel7Anat 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 TrCel7Anat protein, TrCel7Arec 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 TrCel7Arec, both individually and as part of a minimal enzyme cocktail, than that of TrCel7Anat 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 His6 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 TrCel7Arec warrants further investigations.

It is important to note here that the accumulation of a truncated form of TrCel7Arec 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 TrCel7Arec 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 pyrogglutamate 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 TrCel7Arec 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. Glican 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 (His6) 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 (MANKHLSLFLVLLGGLSA-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 TrCel7Arec 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 5–6-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 μm, 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 phenyl Sepharose 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 pl 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 (Stahlberg 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 Amten 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-acetylhexosaminne-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 TrCel7Arec and the fungal TrCel7A-alt 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 μmol 4-methylumbelliferone (4-MU) formed per minute in the reaction. For details, see the Appendix S8.

Activity on Avicel, a cellulose model substrate

Reaction mixtures containing 2% (w/v) Avicel (PH-101, Sigma-Aldrich, St. Louis, MO) and 75 mM fungal or plant-expressed TrCel7A were prepared in 12.5 mM Na-acetate buffer (pH 5.0). Reactions, in a total volume of 200 μL, 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-μm 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 Amplex® Red glucose/glucose oxidase assay (https://assets.thermofisher.com/TFS-Assets/LG/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 μL, 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 μmol 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 μL 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-μm 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-C18 (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 μL. 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 Haugslien and Astrid Svvertsen for technical support.

Conflict of interest

There is no conflict of interest for the current study.

© 2019 The Authors. Plant Biotechnology Journal published by Society for Experimental Biology and The Association of Applied Biologists and John Wiley & Sons Ltd., 18, 631–643.
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; HS 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., Supaker, N.T., Shahajain, 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.

Anonson, J.H., Vik, A., Egge-Jacobsen, W. and Koomen, 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.

Amtzen, M.O., Várnai, A., Mackie, R.I., Eijisink, V.G.H. and Pope, P.B. (2017) Outer membrane vesicles from Fibrobacter succinogenes SBS 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. Trends Plant Sci. 17, 1408–1417.

Børgestad, H.G. and Petersson, L.G. (1980) The 1,4-β-glucan cellobiohydrolases of Trichoderma reesi QM 9414. A new type of cellulolytic synergism. FEBS Lett. 119, 97–100.

Goto, M. (2007) Protein O-glycosylation in fungi: diverse structures and multiple functions. Biosci. Biotechnol. Biochem. 71, 1415–1427.

Hahn, S., Giritch, A., Bartels, D., Bortesi, L. and Gleba, Y. (2015) A novel and easy strategy for protein purification using “teabags”. J. Vis. Exp. e51711.

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, 708–716.

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.

Huang, 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.
Kont, R., Kari, J., Borch, K., Westh, P. and Väijäriä, 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 heterogeneous expression of cellulolytic enzymes: a review. *Biootechnol. 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 straw overproducing glucosidase. *Green Chem.* **20**, 2047–2056.

Li, Y., Sun, H., Fan, C., Hu, H., Wu, L., Jin, X., Lu, Z. et al. (2019) Overproduction of fungal endo-B-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-γ-glutaminyl peptides. *Nature* **197**, 1299.

Mewono, L., Ngeuma-Ona, E., Gotte, M., Koroney, A.S., Follet-Gueye, M.-L., Azeddine, D., Malit, 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**, 43–51.

von Osowski, I., Teeri, T., Kalkkinen, N. and Oker-Blom, C. (1997) Expression of Trichoderma reesei QM 9414. Analysis of domain function in two cellobiohydrolases by limited proteolysis. *Eur. J. Biochem.* **170**, 575–581.

Várnai, A., Vilkar, L., Marjamäe, K. and Sika-aho, M. (2011) Adsorption of monoclonal enzyme components 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.

Wu, V.W., Dana, C.M., Iavarone, A.T., Clark, D.S. and Glass, N.L. (2017) Additional supporting information may be found online in the Supporting information section at the end of the article.

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*AT*.

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*AT*.

Appendix S7 (incl. Figs. S18-S19): LC-MS analysis of O-glycans in TrCel7A*REC*.

Appendix S8 Additional experimental procedures.