Plant Cell Reports (2024) 43:202
https://doi.org/10.1007/s00299-024-03291-y

ORIGINAL ARTICLE

Acidothermus cellulolyticus E1 endoglucanase expressed *in planta* undergoes extensive hydroxyproline-*O*-glycosylation and exhibits enhanced impact on biomass digestibility

Hong Fang[1,2] · Berry Dickey[3] · Daniela PerezLaguna[3] · Jacqueline Vargas Ulloa[3] · Paula PerezSanchez[3] · Jianfeng Xu[1,2]

Received: 18 May 2024 / Accepted: 20 July 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

Key Message **E1 holoenzyme was extensively Hyp-O-glycosylated at the proline rich linker region in plants, which substantially increased the molecular size and improved the enzymatic digestibility of the biomass of transgenic plants.**

Abstract Thermophilic E1 endo-1,4-β-glucanase derived from *Acidothermus cellulolyticus* has been frequently expressed *in planta* to reconstruct the plant cell wall to overcome biomass recalcitrance. However, the expressed holoenzyme exhibited a larger molecular size (~ 100 kDa) than the theoretical one (57 kDa), possibly due to posttranslational modifications in the recombinant enzyme within plant cells. This study investigates the glycosylation of the E1 holoenzyme expressed in tobacco plants and determines its impact on enzyme activity and biomass digestibility. The E1 holoenzyme, E1 catalytic domain (E1cd) and E1 linker (E1Lk) were each expressed in tobacco plants and suspension cells. The accumulation of holoenzyme was 2.0- to 2.3- times higher than that of E1cd. The proline-rich E1Lk region was extensively hydroxyproline-*O*-glycosylated with arabinogalactan polysaccharides. Compared with E1cd, the holoenzyme displayed a broader optimal temperature range (70 to 85 ºC). When grown in greenhouse, the expression of E1 holoenzyme induced notable phenotypic changes in plants, including delayed flowering and leaf variegation post-flowering. However, the final yield of plant biomass was not significantly affected. Finally, plant biomass engineering with E1 holoenzyme showed 1.7- to 1.8-fold higher saccharification efficiency than the E1cd lines and 2.4- to 2.7-fold higher than the wild-type lines, which was ascribed to the synergetic action of the E1Lk and cellulose binding module in reducing cell wall recalcitrance.

Keywords Endoglucanase · Thermophilic enzyme · Enzyme activity · Lignocellulosic biomass · Genetic engineering · Hydroxyproline-*O*-glycosylation

Introduction

Efficient conversion of lignocellulosic plant biomass to biofuels and biobased chemicals is plagued by high production costs associated with biomass pretreatment and enzymatic hydrolysis (Somerville, et al. 2010). This challenge is primarily attributed to the recalcitrance of the plant cell wall matrix that is predominantly composed of cellulose and hemicellulose embedded in highly cross-linked lignin polymers. These polymers act as a protective barrier, shielding the polysaccharides from both chemical and enzymatic degradation (Cosgrove 2005; Himmel, et al. 2007; Yuan, et al. 2008; Ding, et al. 2012). *In planta* expression of lignocellulose degrading enzymes, such as endoglucanase and xylanase, presents a promising strategy for reducing production costs (Brunecky, et al. 2011; Lambertz, et al. 2014; Mir, et al. 2014, 2017). This approach harnesses biomass feedstocks to simultaneously act as enzyme suppliers and substrates for biofuel production (Xiao, et al. 2018). To ensure minimal impact on the growth of plant hosts, thermophilic or hyper-thermophilic enzymes are commonly employed for

Communicated by Sukhpreet Sandhu.

✉ Jianfeng Xu
jxu@astate.edu

[1] Arkansas Biosciences Institute, Arkansas State University, Jonesboro, AR 72401, USA

[2] College of Agriculture, Arkansas State University, Jonesboro, AR 72401, USA

[3] Department of Biological Sciences, Arkansas State University, Jonesboro, AR 72401, USA

Published online: 29 July 2024

in planta expression. These enzymes exhibit minimal enzymatic activity at typical plant growth temperatures but can be strategically activated through a post-harvest temperature shift (Donohoe, et al. 2017; Mir, et al. 2017).

One such noteworthy enzyme is an endoglucanase derived from the thermophilic bacterium *Acidothermus cellulolyticus*. This bacterium, renowned for its ability to thrive in extreme environments, produces a repertoire of cellulolytic enzymes (Rubin 2008), particularly the potent and well characterized E1 endo-1,4-β-glucanase, which belongs to glycoside hydrolase family 5 (GH5) (Ziegelhoffer, et al. 2001). This endo-cellulase is well known for its remarkable thermal stability with an optimum temperature of around 80 °C and for its high catalytic efficiency over broad ranges of pH (pH5 to 9) (Baker, et al. 1994), which makes it a promising candidate for various industrial applications. Like most cellulose-degrading enzymes, E1 has a two-domain structure that consists of a catalytic domain (E1cd) and a cellulose-binding module (CBM), which are connected by a linker region (E1Lk) (Ziegelhoffer, et al. 2001). Both the holoenzyme E1 and E1cd have been expressed in different plant species, such as rice, tobacco, duckweed, and maize (Dai, et al. 2000; Ziegelhoffer, et al. 2001, 2009; Oraby, et al. 2007; Sun, et al. 2007; Makenova, et al. 2015; Fang, et al. 2020). The apoplast/cell wall space is considered as the preferred site for accumulating a large quantity of functional enzymes due to the low existence of protease and correct post-translational modification of the heterologous enzymes (Jung, et al. 2012; Klose, et al. 2015). Compared with wild-type (WT) plants, transgenic plants usually show normal growth and development. Following thermochemical pretreatment and enzyme saccharification, the transgenic plant biomass proved significantly more digestible than that of the WT, requiring lower pretreatment severity for comparable conversion levels. Notably, the reduced recalcitrance of the plant biomass was not due to post-pretreatment residual E1 activity. Instead, the deposition of substantial quantities of heterologous enzymes in the extracellular space during cell wall formation was identified as a mechanism altering the inherent recalcitrance of the cell wall. This process created dispersed voids and delamination within the cell wall matrix, contributing to enhanced digestibility (Brunecky, et al. 2011; Donohoe, et al. 2017; Fang, et al. 2020).

The E1cd domain has consistently been the preferred choice for *in planta* expression, given its ability to function independently of the CBM and exhibit enzymatic activity comparable to the holoenzyme E1 (Dai, et al. 2000; Ziegelhoffer, et al. 2009). However, there are instances where the expression of the holoenzyme targeted to the apoplast/cell wall is favored, because the complete enzyme structure may play a more effective role in altering the cell wall structure than the E1cd domain. Interestingly, when the holoenzyme was expressed *in planta*, the apparent molecular size of the recombinant enzyme, as observed on a Western blot (~ 100 kDa), was much larger than that of the enzyme produced in bacteria or chloroplasts (~ 60 kDa) (Dai, et al. 2000; Ziegelhoffer, et al. 2001). In our own study, a similar result was found (see data in the Results section). This is likely attributed to the glycosylation of the linker region of the holoenzyme (E1lk), as neither the E1cd nor CBM domains were reported to undergo extensive glycosylation (Ziegelhoffer, et al. 2009; Fang, et al. 2020). It is generally acknowledged that linkers within cellulose-degrading enzymes undergo glycosylation at serine (Ser) or threonine (Thr) residues, a process crucial for enhancing protein flexibility, preventing linker collapse, and increasing resistance to proteolysis (Payne, et al. 2013). However, the Ser/Thr-*O*-GalNAc glycosylation commonly observed in mammalian cells is notably absent in plant glycoproteins (Strasser, et al. 2021). Instead, *O*-monogalactosylation typically occurs at Ser and Thr residues in plants (Lamport, et al. 1973; Saito, et al. 2014; Marzol, et al. 2018), which would not lead to a significant increase in the observed molecular size of the recombinant E1 enzyme.

The E1Lk was unveiled as a proline-rich peptide containing repeats of the "Ser-Pro" or "Thr-Pro" motif (Fig. 1). Upon expression *in planta*, this peptide has the potential to undergo hydroxylation modifications, transforming into a hydroxyproline (Hyp)-rich peptide (HRGP), as those present in plant cell wall glycoproteins (Kieliszewski and Shpak 2001; Showalter and Basu 2016). Subsequently, a plant-specific *O*-glycosylation process, known as Hyp-*O*-glycosylation, likely introduces a substantial quantity of Hyp-*O*-glycans to the E1Lk region. This modification will dramatically increase the molecular size of the recombinant

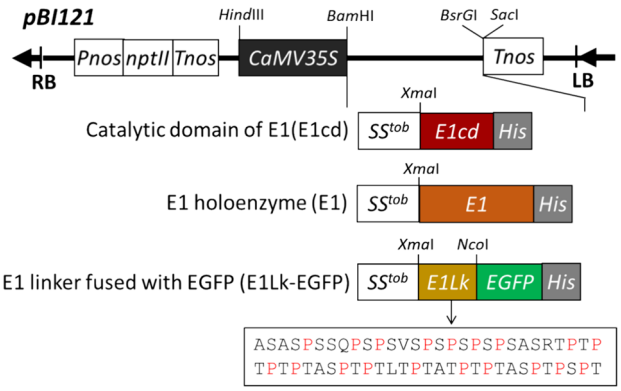

Fig. 1 Schematic representation of the *pBI121* binary vector harboring the E1, E1cd and E1Lk-EGFP gene constructs. E1Lk encodes the E1 linker consisting of 58 amino acids, including 19 clustered non-contiguous proline residues (highlighted in red). *CaMV35S*: 35S cauliflower mosaic virus promoter; SS^{tob}: tobacco extensin signal sequence; *His: 6×His tag; Pnos:* nopaline synthase gene promoter; *Tnos*: nopaline synthase gene terminator; nptII: neomycin phosphotransferase II gene; LB and RB, left border and right border

enzyme. In this study, the potential Hyp-*O*-glycosylation of E1 holoenzyme expressed *in planta* as well as its effect on the enzyme activity were investigated. Additionally, the phenotype and saccharification efficiency of transgenic plants resulted from *in planta* expression of E1 holoenzyme or E1cd were assessed and compared.

Materials and methods

Construction of expression vectors

Expression vector *pBI121-SS^{tob}-E1cd-His* and *pET28b(+)-E1cd-His*, each encoding E1cd with a 6×His tag, were constructed earlier (Fang, et al. 2020). For plant expression vector construction, the gene for E1 holoenzyme (GenBank Accession#: P54583) was amplified from the genomic DNA of *A. cellulolyticus* (ATCC 43068) by PCR using the primers: 5'-ttatccatggcgggcggctattggca-3' and 5'-tatttgtacattaatggtggtgatgatgatgacttgctgcgcaggcgactg-3', and subcloned into *pBI121-SS^{tob}-E1cd-His* plasmid at the *Nco*I and *BsrG*I sites to generate *pBI121-SS^{tob}-E1-His*. The gene encoding E1Lk was PCR-amplified using the primers: 5'-ttatcccgggccgcgtctgcatcgcctagcag-3'and 5'-ttatccatggctgacggcgtcgggcttgccg-3', and subcloned into *pBI121-SS^{tob}-EGFP-His* plasmid at the *Xma*I and *Nco*I sites to generate *pBI121-SS^{tob}-E1Lk-EGFP-His* (Fig. 1). For bacterial expression vector construction, the *E1* gene was PCR-amplified using the primer pairs: 5'-ttatccatggcgggcggctattggca-3' and 5'-ttatctcgagacttgctgcgcaggcgac-3', and then subcloned into *pET28b(+)* at the *Nco*I and *Xho*I sites to generate *pET28b(+)-E1-His.* The *E1Lk* gene was PCR-amplified using the primer pairs: 5'-ttatccatggcgtctgcatcgcctagcag-3' and 5'- ttatctcgagtgacggcgtcgggcttgccg-3', and then subcloned into *pET28b(+)-EGFP-His* at the single *Nco*I restriction site to generate *pET28b(+)-E1Lk-EGFP-His*.

Genetic transformation of tobacco plants and tobacco BY-2 cells

pBI121 vector harboring *E1, E1cd* or *E1Lk-EGFP* gene was stably transformed into tobacco plant (*Nicotiana tabacum* L. cv Xanthi) by using the *Agrobacterium*-mediated leaf disc method, as described earlier (Zhang, et al. 2019a, b; Fang, et al. 2020). The resulting transgenic plantlets were cultivated in Murashige and Skoog (MS) basal salt medium with 200 mg/L kanamycin in Magenta™ vessels at room temperature. Selected transgenic plants (T_0) were cultivated in a greenhouse for 3 to 4 months before harvesting for the analysis of the yields and saccharification efficiency of biomass.

The *E1* and *E1Lk-EGFP* genes were also each stably transformed into tobacco BY-2 cells using the *Agrobacterium*-mediated method, as recently described (Karki, et al. 2022, 2023). Selected transgenic BY-2 cells were grown in Schenk & Hildebrandt (SH) medium supplemented with 0.4 mg/L 2,4-dichlorophenoxyacetic acid (2,4-D), 0.1 mg/L kinetin, 34 g/L sucrose and 200 mg/L kanamycin. For cell suspension culture, flasks containing 80 ml SH medium were rotated at 25 °C on a gyratory shaker at 100 rpm.

Expression and purification of recombinant enzymes in *E. coli*

IPTG-induced expression of the E1cd, E1 and E1Lk-EGFP in *E. coli* BL21(DE3) was performed as described by Zhang et al. 2019a, b. The expressed enzymes were purified from cultured cells by nickel affinity chromatography using the Ni–NTA Fast Start Kit (Qiagen, CA).

Western blotting analysis and quantification of recombinant enzymes

Soluble proteins were extracted from fresh leaves or cultured plant cells by grinding in liquid nitrogen using a mortar and pestle, and then resuspended in an SDS extraction buffer (Zhang, et al. 2016a, b). Recombinant E1 and E1cd were detected with a mouse anti-E1cd primary antibody (National Renewable Energy Laboratory, Golden, CO) and a goat anti-mouse IgG horseradish peroxidase-conjugated secondary antibody (Jackson ImmunoResearch labs, West Grove, PA). Recombinant E1Lk-EGFP was detected using a rabbit anti-EGFP primary antibody (ThermoFisher Scientific, Rockford, IL) and a goat anti-rabbit IgG horseradish peroxidase-conjugated secondary antibody (Jackson ImmunoResearch labs). Blot images were captured either on X-ray film or using the Li-Cor Odyssey Fc imaging system (Li-Cor Biosciences, Lincoln, NE). Target protein products were quantified by densitometry based on Western blot using Li-Cor's Image Studio™ Software, as described earlier (Zhang, et al. 2019a, b; Fang, et al. 2020; Karki, et al. 2023).

Purification of recombinant proteins

Recombinant E1, E1cd and E1Lk-EGFP were separated from tobacco cell culture media by hydrophobic interaction chromatography (HIC), and further purified with nickel affinity chromatography (NAC) and size-exclusion chromatography (SEC), as previously described (Xu, et al. 2007, 2010; Karki, et al. 2022). The detailed purification procedure is provided in Supplementary Method 1.

Hyp content and monosaccharide composition assay

Hyp content of the purified E1 and E1Lk-EGFP proteins was assayed colorimetrically (Xu, et al. 2005). The neutral sugar

composition (galactose, rhamnose, arabinose, fucose, xylose and mannose) was analyzed as alditol acetates derivatives by gas chromatography, while uronic acids were assayed by the colorimetric method, as described previously (Xu, et al. 2005, 2008; Zhang, et al. 2019a, b). The detailed procedure for the monosaccharide composition assay is outlined in Supplementary Method 2.

Precipitation with (β-D-galactosyl)$_3$-Yariv reagent

Assay of the occurrence of Hyp-*O*-glycans (arabinogalactan polysaccharides) on the proteins by precipitation with (β-D-galactosyl)$_3$-Yariv (Biosupplies Australia Pty Ltd, Australia) was described earlier (Zhang, et al. 2019a, b). Tobacco cell produced (SP)$_{32}$-EGFP (Zhang, et al. 2016a, b) served as positive control.

Assay of enzyme activity

The activity of the recombinant E1cd and E1 was measured by its ability to cleave 4-methylumbelliferyl-β-D-cellobioside (4-MUC) to produce the fluorophore, 4-methylumbelliferone (4-MU), as described previously (Fang, et al. 2020). One unit of enzyme activity was defined as the amount of the enzyme that generates 100 pmol 4-MU product per minute. The total soluble protein (TSP) in the plant extracts was measured using the Bradford method.

Phenotype analysis

Transgenic plants (T_0) and WT tobacco were transferred into soil and cultivated in greenhouse for phenotypic analysis. The vegetative growth, flowering time and biomass yields of the cultivated plants were determined as described earlier by Fang et al. 2020.

Analysis of enzymatic saccharification

Harvested plant tissues (leaves and stems) were undergone enzymatic saccharification analysis as described earlier by Fang et al. 2020. The detailed procedure can be found in Supplementary Method 3. Saccharification efficiency was calculated as the percentage of glucose released from the glucan immobilized in biomass by enzymatic saccharification.

Statistical analysis

Experimental data were presented as the means with standard deviation (SD). One factor analysis of variance (ANOVA) followed by a *Tukey* post hoc range test was used to determine the differences among samples with $p < 0.05$ being considered as significant.

Results

Recombinant E1 and E1cd were highly expressed in tobacco plants

The E1 holoenzyme and E1cd were stably expressed in tobacco plants. Western blotting detected the expression of both enzymes in regenerated transgenic plantlets (Fig. 2a). As anticipated, E1cd migrated as a single band at ~ 38 kDa, consistent with previous findings (Ziegelhoffer, et al. 2009; Chou, et al. 2011; Fang, et al. 2020). Compared with the E1cd produced in *E. coli*, the plant produced enzyme migrated at the same position, indicating it was not glycosylated.

In the expression of the E1 holoenzyme, two distinct bands were observed. The lower dense band, measuring 38 kDa and aligning with the size of E1cd, was obviously the truncated E1 domain (E1cd) with the linker (E1Lk) and CBM being cleaved. These two domains are known to be susceptible to cleavage from the E1 holoenzyme, as reported earlier (Dai, et al. 2000; Ziegelhoffer, et al. 2001). The higher band, with a molecular size of about 100 kDa, was anticipated to be an intact holoenzyme. However, its molecular size was substantially larger than the expected size of E1 (57 kDa, calculated based on peptide sequence) and its *E. coli* produced counterpart (~ 60 kDa) (Fig. 2a). The substantial increase in molecular size suggested the occurrence of some modifications, particularly glycosylation, in the tobacco expressed E1, which was further investigated.

High expression plantlets for each gene construct were selected to grow in Magenta™ boxes and then in greenhouse. Western blot analysis using an anti-E1cd antibody detected the identical bands of transgene products (Fig. 2b, 2c) as observed in the regenerated plantlets (Fig. 2a). As for the accumulation of recombinant enzymes in transgenic plants, the content of the E1 holoenzyme, encompassing both the truncated E1cd domain and the intact enzyme, was 1.8 to 2.0 times higher than that of E1cd alone expressed in tobacco (Fig. 2d, 2e). Also, for the E1 holoenzyme, 70–75% of the expressed enzyme was truncated, leading to the predominant detection of the E1cd domain.

Recombinant enzymes expressed by tobacco cell suspension culture

To evaluate the extracellular secretion and streamline the purification processes of recombinant enzymes expressed *in planta*, we created cell suspension culture from transgenic plants expressing either E1 or E1cd. This facilitated

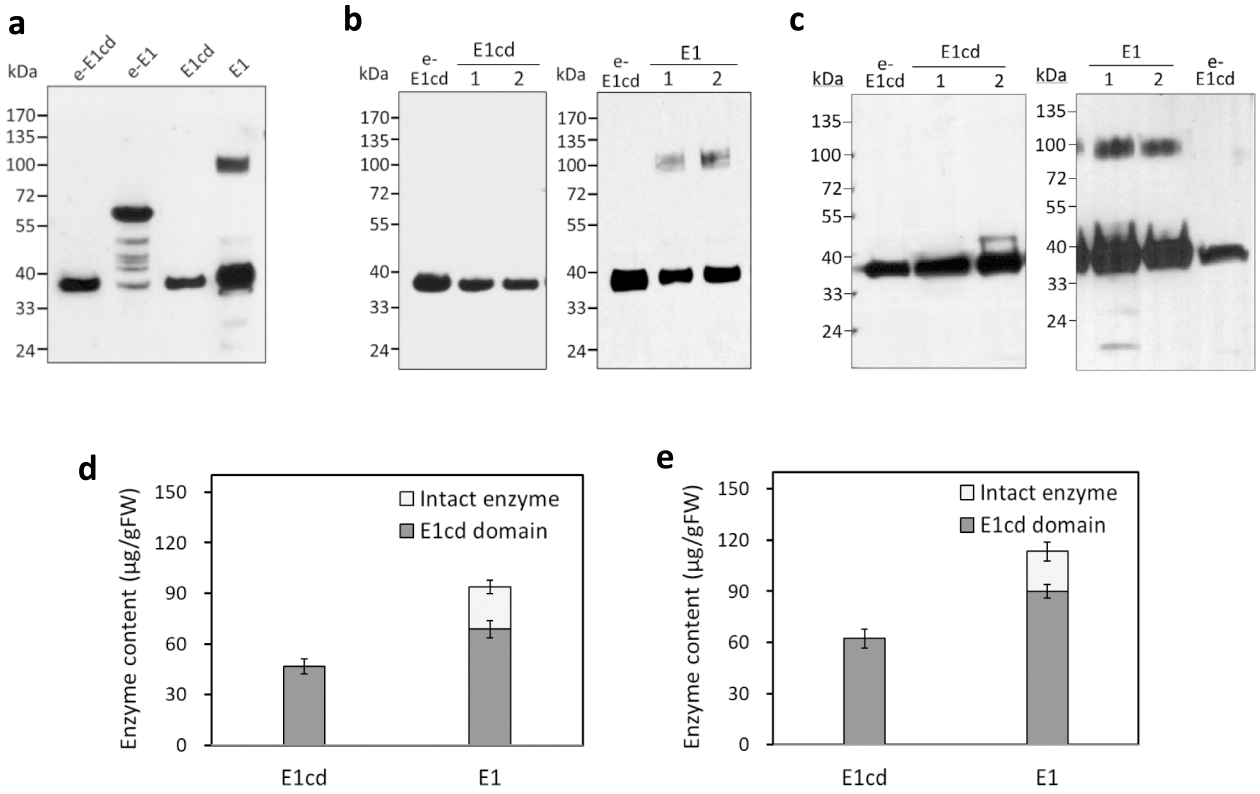

Fig. 2 Expression of recombinant enzymes in transgenic tobacco plants. **(a)** Anti-E1cd Western blotting detection of recombinant E1cd and E1 in regenerated transgenic plantlets in a petri dish. The *E. coli* expressed E1cd and E1, named as e-E1cd and e-E1, respectively, was used as a positive control (50 ng each). (**b**), (**c**) Anti-E1cd Western blotting detection of expressed enzymes in the leaves of two selected plants (#1 and 2) grown in Magenta™ boxes for 4 weeks (**b**) or grown in greenhouse for 8 weeks (**c**). Five microliters of protein extracts were loaded into each well. (**d**), (**e**) Accumulation of recombinant enzymes in the leaves of plants grown in Magenta™ boxes for 4 weeks (**d**) or grown in greenhouse for 8 weeks (**e**). For the E1 intact enzyme fraction, the E1cd equivalent amount was presented. Each data point represents the mean of measurements of five top expression transgenic lines ± SD. There was a significant difference in the enzyme accumulation between the E1 and E1cd plant lines ($p < 0.05$). FW: fresh weight

the straightforward quantification of secreted enzymes and enabled the efficient purification of enzymes directly from the culture medium. Two highly expressing transgenic tobacco lines for each gene construct were used as explants to induce calli and establish cell suspension culture. The harvested culture media was analyzed by anti-E1cd Western blotting.

Both recombinant E1cd and E1 were secreted into cell culture media (Fig. 3a). The Western blotting results obtained from the cell suspension culture closely resemble those detected in tobacco leaves. In the case of E1 expression, a cleaved 38 kDa E1cd domain and an intact enzyme of approximately 100 kDa were consistently identified. It is interesting to observe that the intact E1 enzyme constitutes a significant portion, accounting for 70 to 73% of the total secreted E1 enzyme products. This is contrary to the E1 products detected in tobacco leaves (Fig. 2). Compared with the expression of the E1cd control, the amount of secreted E1 (12 to 15 μg/ml E1cd equivalent, two bands combined) is about 2 to 2.3 times greater than that of the secreted E1cd control.

To confirm the extensive glycosylation of E1Lk and its significant contribution to the increased molecular size of E1 holoenzymes expressed in plants, we expressed the E1Lk domain fused with EGFP in tobacco BY-2 cells. As shown in Fig. 3b, E1Lk-EGFP was dramatically secreted into the culture media at a concentration ranging from 8 to 10 μg/ml (EGFP equivalent). On the anti-EGFP Western blot, it migrated as a ~ 70 kDa band, which was about 35 kDa larger than the calculated molecular weight of ~ 33 kDa. It was also larger than the bacterial-expressed non-glycosylation counterpart, which migrated at ~ 38 kDa (Fig. 3b).

Biochemical characterization of the E1 holoenzyme and E1Lk

The secreted E1 holoenzyme (intact enzyme fraction), E1cd and E1Lk-EGFP were purified from the plant cell culture

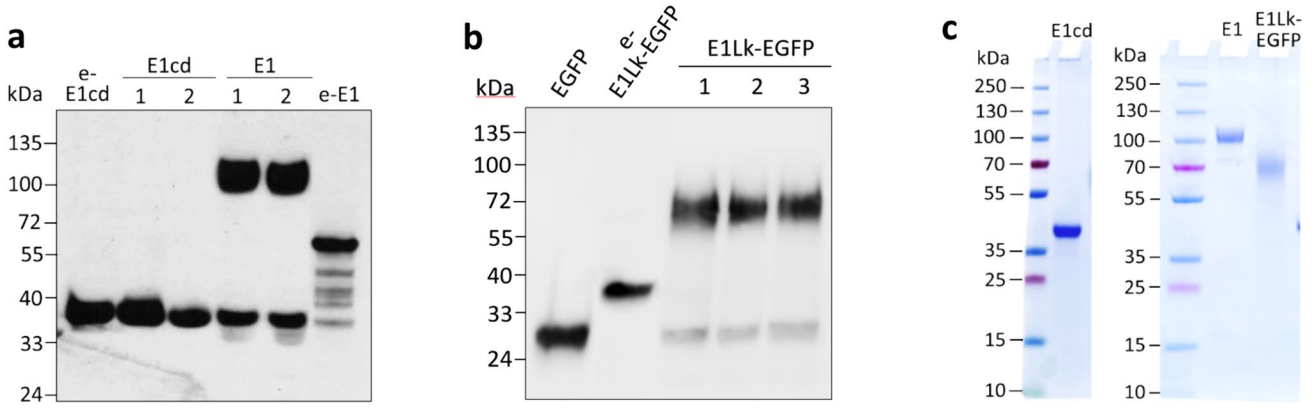

Fig. 3 Detection of secreted recombinant enzyme/protein in tobacco cell suspension culture. (**a**) Anti-E1cd Western blotting detection of recombinant E1 and E1cd secreted into cell culture medium after 12 days of culture. For each gene construct, two cell lines (#1 and 2) were created from the transgenic tobacco leaves. e-E1cd and e-E1 refers to the *E. coli* expressed E1cd and E1 (50 ng each). Ten microliters of cell culture media were loaded into each well. (**b**) Anti-EGFP Western blotting detection of E1Lk-EGFP expressed in *E. coli* and E1Lk-EGFP secreted into BY-2 cell culture medium. Three transgenic BY-2 cell lines (#1, 2 and 3) were tested. Ten microliters of cell culture media were loaded into each well. e-E1Lk-EGFP: *E. coli* expressed E1Lk-EGFP; EGFP: EGFP control (50 ng). (**c**) SDS-PAGE analysis of purified E1cd, E1 holoenzyme and E1Lk-EGFP. Two micrograms of each protein were loaded onto a 4–20% SDS-PAGE gel and subsequently stained with Coomassie Brilliant Blue R250 solution

medium for biochemical characterization. The use of three chromatographic steps (HIC, NAC, and SEC) resulted in the isolation of highly purified proteins. (Fig. 3c). A significant amount of Hyp was detected in both E1 (2.23%, w/w) and E1Lk-EGFP (2.91%, w/w) proteins, indicating that some Pro resides were hydroxylated in these proteins. In contrast, hardly any Hyp was detected in the E1cd domain. Thus, proline hydroxylation was limited to the E1Lk domain.

The monosaccharide assay revealed the presence of arabinose, galactose, rhamnose, and glucuronic acid in both the E1 and E1Lk-EGFP proteins. These sugars are characteristic components of arabinogalactan (AG) polysaccharides, known to be *O*-linked to Hyp residues (Xu, et al. 2008; Tan, et al. 2010). This indicated that the recombinant proteins underwent extensive Hyp-*O*-glycosylation in planta, with sugars constituting 41.3% and 55.2% (w/w) of E1 and E1Lk-EGFP, respectively. In addition, fucose, mannose and xylose, which are the major sugar composition of *N*-glycans, was hardly detected in the recombinant proteins.

The E1, E1cd and E1Lk-EGFP were also tested for their ability to precipitate (β-D-galactosyl)$_3$-*Yariv reagent*, which specifically binds arabinogalactan proteins (AGPs) (Kitazawa, et al. 2013). Both E1 and E1Lk-EGFP exhibited reactivity with the *Yariv* reagent, while the E1cd control did not show such interaction (Table 3). This indicates that the sugars attached to the recombinant enzymes were Type II arabinogalactans typical of classical plant AGPs (Tan, et al. 2003, 2010).

Fifty micrograms of each purified protein were tested. The data presented represent the mean of three parallel samples ± SD. $(SP)_{32}$-EGFP isolated from BY-2 cell culture medium served as a positive control. $(SP)_{32}$ refers to 32 tandem repeats of the "Ser-Pro" motif that is extensively Hyp-*O*-glycosylated with AG polysaccharides in plant cells (Zhang, et al. 2016a, b).

Enzyme activity

The plant leaf extracts were initially used for enzyme activity assay. The leaf extracts of either E1 or E1cd plants exhibited significant enzyme activity compared with that of WT plants. (Fig. 4a). The specific enzyme activity of E1 holoenzyme (intact enzyme fraction) was determined to be 14.2 U/μg, which was lower than that of the E1cd control (39.2 U/μg) (Fang, et al. 2020). This is due to the substantial increase in the molecular mass of the E1 holoenzyme compared to E1cd (~ 100 kDa vs. 38 kDa). When considering the E1cd equivalent amount of the holoenzyme for calculation, the specific enzyme activity of E1 holoenzyme was estimated as 37.4 U/μg(E1cd equivalent), closely resembling the E1cd control. This suggests that the E1Lk domain and CBM had minimal impact on the enzyme activity of the active domain (E1cd).

The temperature-activity response curve of the recombinant enzymes was also determined. Both E1cd and E1 exhibited a temperature optimum of 80 °C. Compared with E1cd, the E1 holoenzyme displayed a broad optimal temperature range of 70 to 85 °C. While E1cd experienced a sharp decline in enzyme activity beyond its optimum temperature, E1 retained high activity at 85 °C. (Fig. 4b).

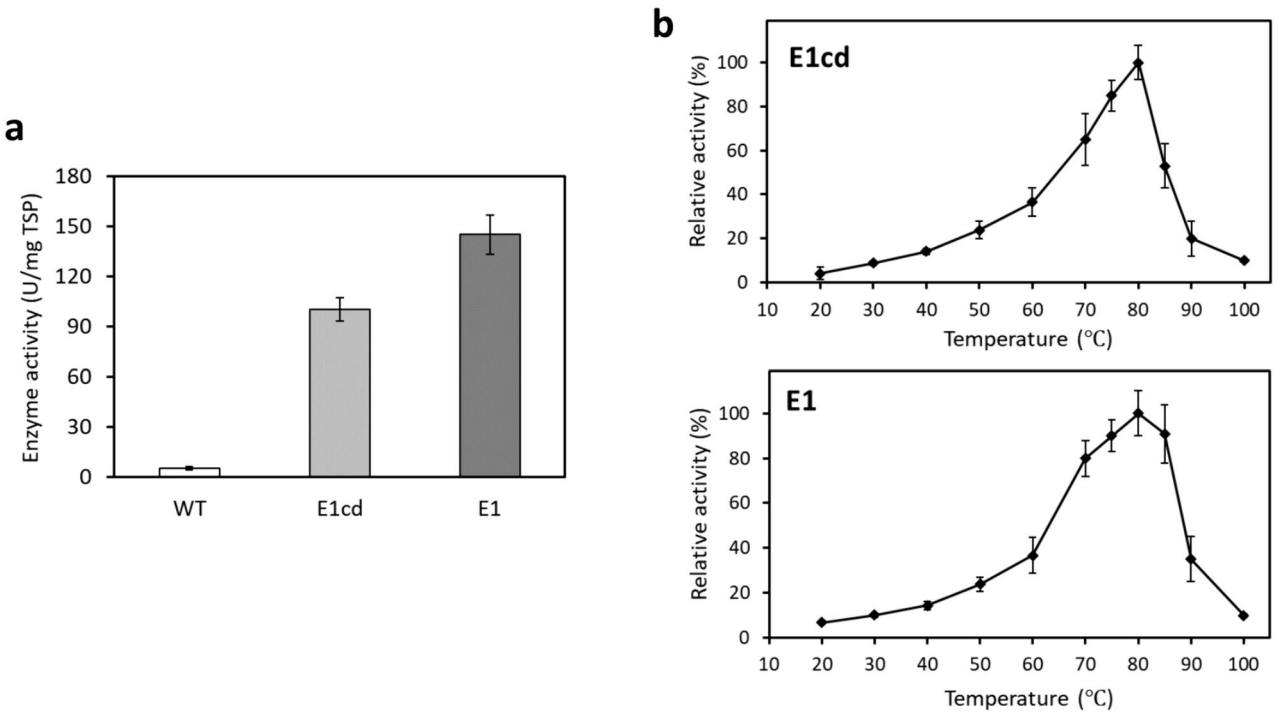

Fig. 4 Characterization of enzyme activity of the recombinant E1cd and E1 expressed in tobacco plants. (**a**) Enzyme activity contained in the leaf extracts. The transgenic tobaccos were grown in greenhouse for 8 weeks. Each data point represents the mean of measurements of three transgenic plant lines ± SD. There was significant difference in the enzyme activity contained in the E1cd and E1 plant lines ($p < 0.05$). (**b**) Temperature-activity response curve of the recombinant enzymes. The enzyme activity was determined at the optimal pH5. Error bars represent the SD of three replicate measurements

Plant growth and phenotype

T_0 transgenic tobacco plants and WT tobaccos were cultivated in greenhouse for characterization of phenotype. During the vegetative growth of plants, no significant differences were observed in terms of leaf color, leaf shape, leaf size, stem structure, and growth rate between the transgenic WT tobaccos. While both WT tobaccos and E1cd transgenic lines started flowering in the 5th week, the E1 plants exhibited a delayed flowering, occurring

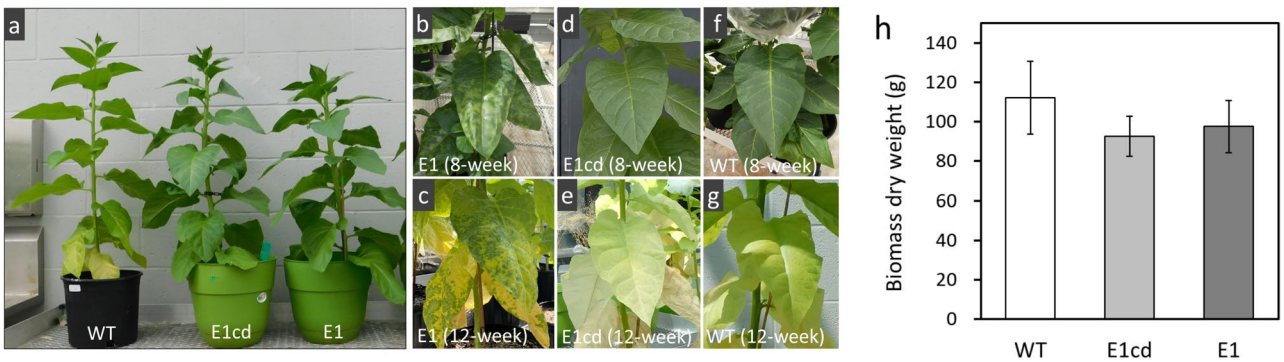

Fig. 5 Phenotypic analysis of transgenic plants. **(a)** Vegetative growth of transgenic tobacco compared with WT plants. The plants were cultivated in a greenhouse for 5 weeks. **(b), (c)** E1 tobacco leaves with faded variegation after 8-week and 12-week cultivation. **(d)** to (**g**) WT and E1cd tobacco leaves with normal leaves after 8-week and 12-week cultivation. (**h**) Biomass yields of transgenic tobacco lines and WT plants after 14-week cultivation. The above-ground biomass, including leaves, stems, pedicles, and flowers were harvested and measured. Each data point represents the mean of measurements of five individual plants ± SD. No significant differences between the transgenic plants and WT line were observed ($p > 0.05$)

two weeks later. Following flowering (8 weeks after greenhouse planting), faded variegations were noticeable on mature leaves of each E1 transgenic line, spreading between leave veins (Fig. 5b, 5c). In contrast, the leaves of WT and E1cd lines remained clean and green (Fig. 5d to 5g). The faded variegation persisted permanently on the leaves once formed, even remaining visible on dead leaves. Apart from this distinctive faded variegation on leaves, no other morphological differences were observed between the transgenic lines and WT tobacco during the vegetative growth phase. After 14 weeks of cultivation in the greenhouse, the above-ground parts of each transgenic tobacco and the wild type were harvested. In terms of total biomass, no statistical differences were found between the transgenic tobacco and WT tobacco (Fig. 5h).

Plant biomass saccharification

The leaves and stems of cultivated transgenic plants were subjected to an enzymatic saccharification assay, and the results were compared with those of WT tobacco. As shown in Fig. 6a, the E1 lines exhibited a saccharification efficiency 1.7 to 1.8 times higher than the E1cd lines and 2.4 to 2.7 times higher than the WT plants. The enhanced enzymatic digestibility in E1 plants, compared to E1cd lines, also correlated with the increased accumulation of enzymes in the E1 plants before harvesting, (Fig. 6b). Notably, the saccharification efficiency of E1cd lines (19.1%) is only slightly higher than that of the WT lines (11.8%). When comparing leaves and stems, the former generally demonstrated better enzyme digestibility than the latter.

Discussion

In planta expression of the thermophilic *A. cellulolyticus* E1 endoglucanase has been widely conducted for genetic modification of plant cell wall matrix to overcome plant biomass' recalcitrance and render cell wall materials more amenable to deconstruction. Although the catalytic domain of E1 endoglucanase (E1cd) was often chosen for *in planta* expression due to its independent functionality, experiments sometimes involved the use of the E1 holoenzyme. Interestingly, the in planta expressed E1 exhibited a significantly larger molecule size (~ 100 kDa) than the theoretic one (57 kDa). As shown in Fig. 2 and Fig. 3, the tobacco-expressed E1 is approximately 40 kDa larger than its bacteria-derived counterpart. Obviously, this was attributed to the extensive glycosylation of the recombinant enzyme expressed *in planta*, as glycosylation is usually absent in bacteria expressed proteins. Thus, this study investigated the glycosylation of E1 holoenzyme and determined how the glycosylation affected the enzyme activity and biomass digestibility.

As the E1cd and CBM domains were not previously reported to undergo extensive glycosylation when expressed *in planta* (Ziegelhoffer, et al. 2009; Fang, et al. 2020), it is plausible that the link region of the E1 holoenzyme, E1Lk, experienced significant glycosylation, leading to a substantial increase in the molecular size of the E1 enzyme. E1Lk is a proline-rich peptide featuring clustered non-contiguous proline residues, including repetitive motifs like "Ser-Pro" and "Thr-Pro" (Fig. 1). According to the Hyp-*O*-glycosylation "code" elucidated earlier (Kieliszewski and Shpak 2001), Pro residues within Pro-rich motifs are consistently hydroxylated to form Hyp in plant cells; then, non-contiguous Hyp residues, particularly when clustered as in "X-Hyp-X-Hyp…" motifs (where X = Ser, Thr, or Ara),

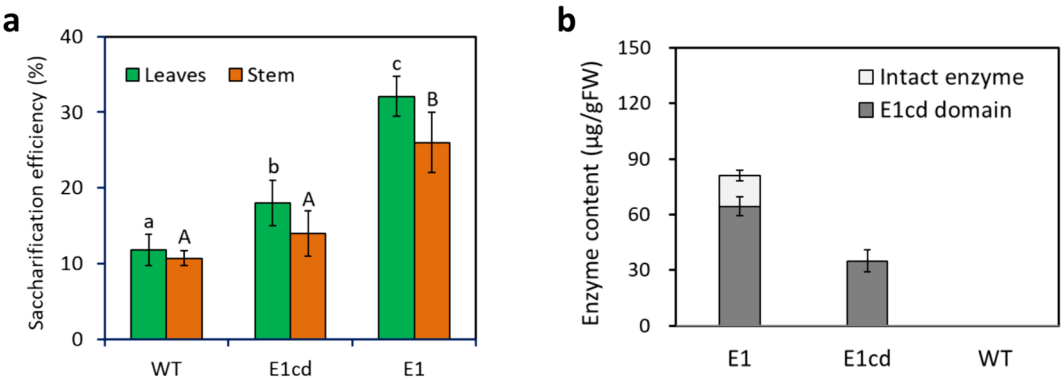

Fig. 6 Enzymatic saccharification of plant biomass and accumulation of recombinant enzymes. (**a**) Enzymatic saccharification efficiency of harvested plant biomass (growth in greenhouse for 14 weeks). (**b**) Accumulation of recombinant enzymes in the leaves of plants before harvesting. Error bars represent the SD of 5 individual plants analyzed for each gene construct. In panel (**a**), different letters ("a to c" or "A to B") indicate significant differences in saccharification efficiency ($p < 0.05$). In panel (**b**), there was a significant difference in the enzyme accumulation between the E1 and E1cd plant lines ($p < 0.05$). FW: fresh weight

serve as preferred sites for the addition of highly branched AG polysaccharides (Kieliszewski 2001; Xu, et al. 2008). Therefore, we predicted that the E1Lk, upon expressed *in planta* and targeted for secretion, should undergo extensive *O*-glycosylation on each of Pro/Hyp residues with AG polysaccharides. To validate this hypothesis, plant cell cultures expressing either E1 or E1cd were established from the transgenic plant leaves. The recombinant E1 (intact enzyme fraction) and E1cd were readily purified from the cell culture media, facilitating biochemical analysis of the enzymes. Meanwhile, the E1LK domain alone was also expressed in plant cells as a fusion with the reporter protein EGFP. This could provide additional confirmation of the posttranslational modifications in the E1LK region of the E1 holoenzyme.

In the expression of E1 in tobacco leaves, most of the expressed E1 enzyme was truncated, leaving the E1cd domain mainly detected (Fig. 2a to c). This was consistent with earlier reports indicating that the linker and CBM module are prone to cleavage from the E1 holoenzyme (Dai, et al. 2000; Ziegelhoffer, et al. 2001). Contrastingly, in the cell suspension culture, the majority of the secreted enzyme remained intact. This was due to the secretion of the enzyme into the culture media, where significantly fewer proteases were present compared to the cytoplasm or the apoplastic space within plant cells (Marconi and Alvarez 2014). Similarly, over 90% of the secreted E1Lk-EGFP in tobacco BY-2 cells retained its integrity, migrating at approximately 70 kDa on a Western blot. The substantial secretion of intact E1 and E1LK-EGFP in the cell suspension culture facilitated the purification of recombinant proteins from the culture medium for biochemical characterization, particularly the Hyp-*O*-glycosylation.

A significant amount of Hyp and sugars were detected in both E1 and E1Lk-EGFP, but not in E1cd (Tables 1, 2), indicating that the hydroxylation of Pro and subsequent Hyp-*O*-glycosylation occurred in the recombinant proteins, particularly in the E1Lk domain. The detected monosaccharide compositions rich in galactose and arabinose with lesser amounts of glucuronic acids and rhamnose is the typical structure of plant AG polysaccharides (Figs. 1, 2) (Xu, et al. 2007, 2010; Tan, et al. 2012). Furthermore, the sugars attached to these proteins were identified as Type II arabinogalactans, as they could react with the *Yariv* reagent (Table 3) (Tan, et al. 2003, 2010). These results were consistent with prior research, where synthetic Hyp-*O*-glycosylation modules, such as repetitive "Ser-Pro" or "Ala-Pro" motifs, expressed in tobacco BY-2 cells exhibited similar glycan structures (Xu, et al. 2007, 2008; Zhang, et al. 2016a, b, 2019a, 2019b). In addition, as the engineered Hyp-*O*-glycosylation modules could function as a molecular carrier in facilitating the extracellular secretion of fused proteins, high yields of recombinant proteins were consistently detected in plant cell suspension culture (Xu, et al. 2007, 2010; Zhang, et al. 2019a, b; Zhang, et al. 2019a, b; Wang, et al. 2021; Karki, et al. 2022). In this study, the increased detection of E1 in the cell culture medium, 2 to 2.5 times greater than the E1cd control, could be attributed to the highly Hyp-*O*-glycosylated linker region of the holoenzyme enzyme (Fig. 3).

The potential site of Hyp-*O*-glycosylation in the E1Lk region was further investigated. Upon analyzing the peptide sequence, it was observed that the E1Lk, E1Lk-EGFP, and EGFP contained 19, 29, and 10 Pro residues, respectively. The detection of Hyp/Pro ratio of 62.72% in E1Lk-EGFP (Table 1)

Table 1 Hyp and Pro content of recombinant enzymes/proteins expressed *in planta*

Protein samples	E1	E1cd	E1Lk-EGFP	E1Lk
Hyp (%, w/w)	2.23 ± 0.18	trace	2.91 ± 0.13	n.d.[d]
Pro number	42	21	29	19
Pro (%, w/w, calculated)[a]	8.56	6.00	10.28	39.81
Pro (%, w/w, estimated)[b]	4.84	6.36	4.64	n.d
Hyp/Pro ratio (%)[c]	46.07	0	62.72	n.d

The data presented represent the mean of three parallel samples ± SD

[a]: The Pro content was calculated based on the peptide sequence without counting the glycosylation. [b]: The Pro content was estimated based on the apparent molecular mass determined on the Western blots, ~ 100 kDa for E1, ~ 38 kDa for E1cd and ~ 70 kDa for E1Lk-EGFP. [c]: The calculation was based on the estimated Pro content. [d]: not determined

Table 2 Analysis of monosaccharide components of recombinant enzymes

Glycosyl residue	E1		E1Lk-EGFP	
	wt%	mol%	wt%	mol%
Rhamnose	2.3 ± 0.2	5.8	3.5 ± 0.1	6.6
Arabinose	11.3 ± 0.4	31.3	15.1 ± 0.4	31.3
Galactose	21.5 ± 0.2	49.6	27.8 ± 0.3	48.0
Glucuronic acid	6.2 ± 0.3	13.3	8.8 ± 0.2	14.1
Fucose	0	0	0	0
Xylose	Trace	0	0	0
Mannose	Trace	0	0	0
Total	41.3	100	55.2	100

The weight percentage (wt %) represents the mean of three measurements ± SD. The molar percentages (mol %) were calculated from the mean value of each weight percentage

Table 3 Yariv reagent co-precipitation of the recombinant enzymes

Samples	E1	E1cd	E1Lk-EGFP	$(SP)_{32}$-EGFP
Absorbance (420 nm)	0.56 ± 0.03	0	0.72 ± 0.04	0.95 ± 0.03

suggested the presence of about 18 Hyp residues in this fusion protein. Since the Pro residues in the EGFP domain are not hydroxylated (Xu, et al. 2008; Zhang, et al. 2019a, b), it can be inferred that 18 out of the 19 Pro residues in the E1Lk could be hydroxylated. Similarly, in the E1 holoenzyme, a Hyp/Pro ratio of 46.07% (Table 1) indicated that around 19 out of the 42 Pro residues were hydroxylated (here, the molecular weight difference between Hyp and Pro was ignored in calculation). These Hyp residues were possibly further *O*-glycosylated with AG polysaccharides. In E1Lk-EGFP, a 55.2% sugar detection (Table 2) suggested the attachment of approximately 38.6 kDa sugar to the fusion protein. Considering the average size of Hyp-glycan as 2.5 kDa, it was estimated that 15 to 16 glycans (38.6 kDa/2.5 kDa = 15.4) were attached to the fusion protein, specifically to the E1Lk domain. For E1, the detection of 41.3% sugar (Table 2) corresponded to ~ 41.3 kDa sugars attached to the holoenzyme, roughly equivalent to 16 to 17 Hyp-glycan (41.3 kDa/2.5 kDa = 16.5). This was consistent with the calculated result for the E1Lk-EGFP fusion protein. Based on these calculations, a Hyp-*O*-glycosylation model of the E1Lk was proposed, as illustrated in Fig. 7.

Despite the extensive Hyp-*O*-glycosylation, the potential *N*-glycosylation of the E1 holoenzyme should not be disregarded, even though there are no reports in the literature regarding the *N*-glycosylation of E1 or E1cd when expressed in plants. In fact, the bioinformatics tool NetNGlyc-1.0 predicts two potential *N*-glycosylation sites in the E1 holoenzyme, one located on the E1cd domain and the other on the CBM. However, no apparent *N*-glycosylation was observed on the E1cd domain, as the plant-expressed E1cd migrates at a similar size (~ 38 kDa) to its counterpart produced in *E. coli* (Fig. 2). The status of *N*-glycosylation on the CBM remains unknown unless the CBM alone is expressed in plants. Finally, it is essential to note that while extensive Hyp-*O*-glycosylation of the E1Lk was demonstrated, the possibility of *O*-glycosylation occurring on the Ser or Thr residue cannot be ruled out. Earlier reports have shown Ser-*O*-glycosylation of a repetitive "Ser-Pro" module (Shpak, et al. 1999; Zhang, et al. 2019a, b). However, it typically involves the addition of a single galactose to the Ser residue, thus unlikely contributing to the substantial increase in the molecular size.

Regarding enzyme activity, both plant-derived E1 and E1cd exhibited an optimal temperature of 80 °C. As anticipated, the E1Lk domain and CBM demonstrated minimal impact on the specific activity of the endoglucanase. However, compared with E1cd, the holoenzyme displayed a broader optimal temperature range of 70 °C to 85 °C (Fig. 4), which could be attributed to the Hyp-*O*-glycosylated E1LK domain. These findings align with our earlier investigations, wherein E1cd was engineered in tobacco plants using a synthetic Hyp-*O*-glycosylation module, $(SP4)_{18}$ comprising 18 repeats of a "Ser-Pro-Pro-Pro-Pro" motif. The $(SP4)_{18}$ module, much like E1Lk, exhibited limited effects on E1cd enzyme activity but extended the optimal temperature range of the enzyme (Fang, et al. 2020). The increased thermostability is presumably due to the presence of many Hyp-glycans that could increase the solubility of the E1cd and stabilize the enzyme.

There was no significant difference observed in the vegetative growth and final biomass accumulation between the transgenic plant lines (E1 and E1cd) and the WT control (Fig. 6h), which is consistent with several earlier studies, wherein the expression of E1cd in *Arabidopsis*, potato, maize, rice and tobacco did not exhibit significantly deleterious effects on overall plant growth and development (Dai, et al. 2000; Chou, et al. 2011; Jung, et al. 2012; Yennamalli, et al. 2013; Donohoe, et al. 2017; Fang, et al. 2020). This can be attributed to the thermophilic nature of the E1 endoglucanase, which demonstrates limited activity under typical plant growth temperatures (Fig. 4). Nevertheless, it was noted that plants expressing the E1 holoenzyme in this study exhibited delayed flowering and displayed variegated leaves (Fig. 6b, 6c). As the *E1cd* genotype was shown not to cause the observed phenotypes in this study and previous reports (Dai, et al. 2000; Brunecky, et al. 2011; Fang,

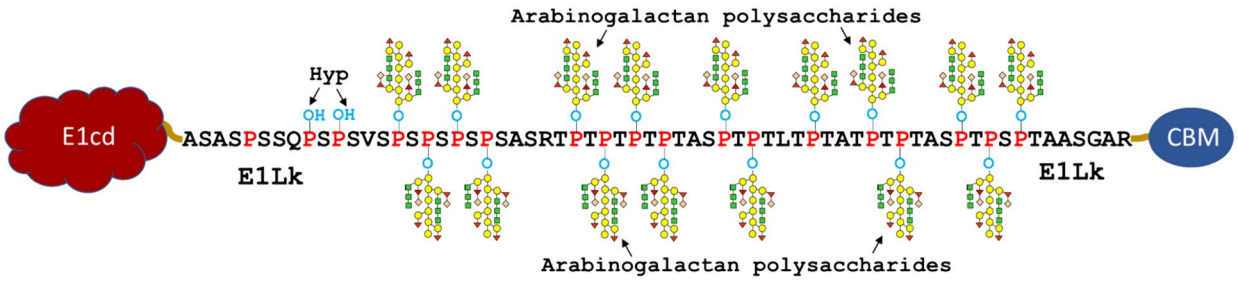

Fig. 7 Proposed model of Pro hydroxylation and Hyp-*O*-glycosylation of the Pro-rich E1Lk peptide. Of total of 19 Pro residues, 18 residues are presumably hydroxylated to be Hyp. These repetitive but non-contiguous Hyp residues, particular those occurring as "Ser-Hyp" and" Thr-Hyp" motifs, were presumably *O*-glycosylated by 16 AG polysaccharides (Hyp-glycans). The Hyp-glycan structure was elucidated earlier (Xu, et al. 2007)

et al. 2020), it is plausible that the two additional domains, CBM and E1Lk, might be accountable for the observed traits. However, our previous studies suggested that E1Lk was unlikely to exert influence on plant growth and development, because *in planta* expression of similar Hyp-*O*-glycosylation modules, such as $(SP4)_{18}$ and $(SP)_{32}$, did not induce noticeable phenotypic changes in tobacco plants (Zhang, et al. 2019a, b; Fang, et al. 2020). In contrast, earlier studies have reported delayed flower development in tobacco plants expressing the CBM of *Piromyces equi* cellulase (Obembe, et al. 2007) as well as in *Arabidopsis thaliana* expressing the CBM of strawberry expansin 2 (Nardi, et al. 2015). Despite the absence of reported correlations between leaf variegation and *in planta* CBM expression in existing literature, a substantial body of research consistently emphasized that the incorporation of CBM *in planta* leads to modifications in the composition and/or structure of the plant cell wall (Obembe, et al. 2007; Herve, et al. 2010; Nardi, et al. 2015; Bernardes, et al. 2019). These changes have been shown to influence plant development (Lampugnani, et al. 2018; De Lorenzo, et al. 2019). Hence, the observed phenotype could be reasonably attributed to the presence of the CBM in the E1 holoenzyme.

While the observed phenotypic effects on transgenic tobacco were primarily attributed to CBM, both E1Lk and CBM were recognized as contributors to the significantly enhanced saccharification efficiency compared to E1cd and WT plants (Fig. 6). The decrease in biomass recalcitrance through the *in planta* engineering of E1cd was reported across various plant species (Dai, et al. 2000; Chou, et al. 2011; Jung, et al. 2012; Yennamalli, et al. 2013; Donohoe, et al. 2017). It has been recognized that this was mainly due to the deposition of the expressed enzyme within the cell wall matrix during plant growth, which alters the cell wall structure in situ, rather than the enzyme activity retained in the harvested plant biomass (Brunecky, et al. 2011). In the case of E1 holoenzyme expression *in planta*, the extensively Hyp-*O*-glycosylated E1Lk domain not only substantially increased the molecular size of the enzyme, but also aided in the deposition and stabilization of the expressed enzyme within the cell wall matrix. Meanwhile, the CBM demonstrated its ability to modify the cell wall structure, leading to the modulation of wall loosening, thereby enhancing cellulase access to the substrate (Obembe, et al. 2007). Therefore, both E1Lk and CBM acted synergistically *in planta*, resulting in a substantial reduction in cell wall biomass recalcitrance and rendering it more susceptible to enzymatic saccharification compared to the E1cd control engineering. This was similar to our prior research, where *in planta* engineering of E1cd as a fusion with a synthetic Hyp-*O*-glycosylation module, $(SP4)_{18}$, resulted in a 2.2-fold increase in saccharification efficiency compared to E1cd engineering alone (Fang, et al. 2020).

Conclusion

This study revisited the engineering of E1 endoglucanase *in planta* to address biomass recalcitrance during enzymatic saccharification. E1 holoenzyme expressed in tobacco plants exhibited a much larger molecular size (~ 100 kDa) compared to its bacterial counterpart (~ 60 kDa) and the theoretical size (57 kDa). This difference was attributed to extensive Hyp-*O*-glycosylation of the linker region of the enzyme (E1Lk) with arabinogalactan polysaccharides, driven by the Pro-rich nature of the E1Lk peptide that comprises clustered non-contiguous Pro residues, such as clustered in "Ser-Pro" and "Thr-Pro" motif. Despite the Hyp-*O*-glycosylation, the activity of the enzyme was not negatively affected, and the holoenzyme exhibited a broader optimal temperature range than E1cd alone. Unlike E1cd, engineering E1 holoenzyme in plants led to phenotypic changes, such as delayed flowering and leaf variegation post-flowering, although these changes did not significantly impact the final yields of plant biomass. Due to the synergetic action of the E1Lk and CBM domains in reducing cell wall recalcitrance, plant biomass engineering with E1 holoenzyme demonstrated markedly higher saccharification efficiency compared to E1cd and wild-type plants. This study unveiled the post-translational modifications occurring in E1 holoenzyme when expressed *in planta* and shed light on their impact on both enzyme activity and the enzymatic digestibility of biomass.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00299-024-03291-y.

Author contributions JX and HF conceived and designed research. HF, BD, DP, JU, PP and JX conducted experiments. HF and JX analyzed data. JX wrote the manuscript. All authors read and approved the manuscript.

Funding This work was supported by the Agriculture and Food Research Initiative Grant no. 2014–04034 from the USDA National Institute of Food and Agriculture and the Arkansas Biosciences Institute, the major research component of the Arkansas Tobacco Settlement Proceeds Act of 2000.

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

Baker JO, Adney WS, Nieves RA, Thomas SR, Wilson DB, Himmel ME (1994) A new thermostable endoglucanase,

Acidothermus-Cellulolyticus E1 - synergism with *Trichoderma-Reesei* Cbh-I and comparison to *Thermomonospora-Fusca* E(5). Appl Biochem Biotechnol 45–6:245–256

Bernardes A, Pellegrini VOA, Curtolo F, Camilo CM, Mello BL, Johns MA, Scott JL, Guimaraes FEC, Polikarpov I (2019) Carbohydrate binding modules enhance cellulose enzymatic hydrolysis by increasing access of cellulases to the substrate. Carbohydr Polym 211:57–68

Brunecky R, Selig MJ, Vinzant TB, Himmel ME, Lee D, Blaylock MJ, Decker SR (2011) In planta expression of a cellulolyticus Cel5A endocellulase reduces cell wall recalcitrance in tobacco and maize. Biotechnol Biofuels. https://doi.org/10.1186/1754-6834-4-1

Chou HL, Dai Z, Hsieh CW, Ku MS (2011) High level expression of Acidothermus cellulolyticus beta-1, 4-endoglucanase in transgenic rice enhances the hydrolysis of its straw by cultured cow gastric fluid. Biotechnol Biofuels 4:58

Cosgrove DJ (2005) Growth of the plant cell wall. Nat Rev Mol Cell Bio 6:850–861

Dai ZY, Hooker BS, Anderson DB, Thomas SR (2000) Expression of *Acidothermus cellulolyticus* endoglucanase E1 in transgenic tobacco: biochemical characteristics and physiological effects. Transgenic Res 9:43–54

De Lorenzo G, Ferrari S, Giovannoni M, Mattei B, Cervone F (2019) Cell wall traits that influence plant development, immunity, and bioconversion. Plant J 97:134–147

Ding SY, Liu YS, Zeng YN, Himmel ME, Baker JO, Bayer EA (2012) How does plant cell wall nanoscale architecture correlate with enzymatic digestibility? Science 338:1055–1060

Donohoe BS, Wei H, Mittal A, Shollenberger T, Lunin VV, Himmel ME, Brunecky R (2017) Towards an understanding of enhanced biomass digestibility by in planta expression of a family 5 glycoside hydrolase. Sci Rep 7:4389

Fang H, Wright T, Jinn JR, Guo W, Zhang N, Wang X, Wang YJ, Xu J (2020) Engineering hydroxyproline-O-glycosylated biopolymers to reconstruct the plant cell wall for improved biomass processability. Biotechnol Bioeng 117:945–958

Herve C, Rogowski A, Blake AW, Marcus SE, Gilbert HJ, Knox JP (2010) Carbohydrate-binding modules promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects. Proc Natl Acad Sci USA 107:15293–15298

Himmel ME, Ding SY, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD (2007) Biomass recalcitrance: engineering plants and enzymes for biofuels production. Science 315:804–807

Jung SK, Parisutham V, Jeong SH, Lee SK (2012) Heterologous expression of plant cell wall degrading enzymes for effective production of cellulosic biofuels. J Biomed Biotechnol 2012:405842

Karki U, Wright T, Xu J (2022) High yield secretion of human erythropoietin from tobacco cells for ex vivo differentiation of hematopoietic stem cells towards red blood cells. J Biotechnol 355:10–20

Karki U, Perez Sanchez P, Chakraborty S, Dickey B, Vargas Ulloa J, Zhang N, Xu J (2023) Intracellular trafficking and glycosylation of hydroxyproline-O-glycosylation module in tobacco BY-2 cells is dependent on medium composition and transcriptome analysis. Sci Rep 13:13506

Kieliszewski MJ (2001) The latest hype on Hyp-O-glycosylation codes. Phytochem 57:319–323

Kieliszewski MJ, Shpak E (2001) Synthetic genes for the elucidation of glycosylation codes for arabinogalactan-proteins and other hydroxyproline-rich glycoproteins. Cell Mol Life Sci 58:1386–1398

Kitazawa K, Tryfona T, Yoshimi Y, Hayashi Y, Kawauchi S, Antonov L, Tanaka H, Takahashi T, Kaneko S, Dupree P, Tsumuraya Y, Kotake T (2013) beta-galactosyl Yariv reagent binds to the beta-1,3-galactan of arabinogalactan proteins. Plant Physiol 161:1117–1126

Klose H, Gunl M, Usadel B, Fischer R, Commandeur U (2015) Cell wall modification in tobacco by differential targeting of recombinant endoglucanase from *Trichoderma reesei*. BMC Plant Biol 15:54

Lambertz C, Garvey M, Klinger J, Heesel D, Klose H, Fischer R, Commandeur U (2014) Challenges and advances in the heterologous expression of cellulolytic enzymes: a review. Biotechnol Biofuels 7:135

Lamport DT, Katona L, Roerig S (1973) Galactosylserine in extensin. Biochem J 133:125–132

Lampugnani ER, Khan GA, Somssich M, Persson S (2018) Building a plant cell wall at a glance. J Cell Sci. https://doi.org/10.1242/jcs.207373

Makenova AT, Scholthof HB, Ramankulov EM, Manabayeva SA (2015) Transient expression of Acidothermus cellulolyticus endoglucanase E1 by a tomato bushy stunt virus-based plant expression vector. J Biotechnol 208:S29–S30

Marconi PL, Alvarez MA (2014) The expression of the 14D9 catalytic antibody in suspended cells of *Nicotiana tabacum* cultures increased by the addition of protein stabilizers and by transference from erlenmeyer flasks to a 2-L bioreactor. Biotechnol Prog 30:1185–1189

Marzol E, Borassi C, Bringas M, Sede A, Rodriguez Garcia DR, Capece L, Estevez JM (2018) Filling the gaps to solve the extensin puzzle. Mol Plant 11:645–658

Mir BA, Mewalal R, Mizrachi E, Myburg AA, Cowan DA (2014) Recombinant hyperthermophilic enzyme expression in plants: a novel approach for lignocellulose digestion. Trends Biotechnol 32:281–289

Mir BA, Myburg AA, Mizrachi E, Cowan DA (2017) In planta expression of hyperthermophilic enzymes as a strategy for accelerated lignocellulosic digestion. Sci Rep 7:11462

Obembe OO, Jacobsen E, Timmers J, Gilbert H, Blake AW, Knox JP, Visser RG, Vincken JP (2007) Promiscuous, non-catalytic, tandem carbohydrate-binding modules modulate the cell-wall structure and development of transgenic tobacco (Nicotiana tabacum) plants. J Plant Res 120:605–617

Oraby H, Venkatesh B, Dale B, Ahmad R, Ransom C, Oehmke J, Sticklen M (2007) Enhanced conversion of plant biomass into glucose using transgenic rice-produced endoglucanase for cellulosic ethanol. Transgenic Res 16:739–749

Payne CM, Resch MG, Chen L, Crowley MF, Himmel ME, Taylor LE 2nd, Sandgren M, Stahlberg J, Stals I, Tan Z, Beckham GT (2013) Glycosylated linkers in multimodular lignocellulose-degrading enzymes dynamically bind to cellulose. Proc Natl Acad Sci U S A 110:14646–14651

Rubin EM (2008) Genomics of cellulosic biofuels. Nature 454:841–845

Saito F, Suyama A, Oka T, Yoko OT, Matsuoka K, Jigami Y, Shimma YI (2014) Identification of novel peptidyl serine alpha-galactosyltransferase gene family in plants. J Bio Chem 289:20405–20420

Showalter AM, Basu D (2016) Extensin and arabinogalactan-protein biosynthesis: glycosyltransferases, research challenges, and biosensors. Front Plant Sci 7:814

Shpak E, Leykam JF, Kieliszewski MJ (1999) Synthetic genes for glycoprotein design and the elucidation of hydroxyproline-O-glycosylation codes. Proc Natl Acad Sci U S A 96:14736–14741

Somerville C, Youngs H, Taylor C, Davis SC, Long SP (2010) Feedstocks for lignocellulosic biofuels. Science 329:790–792

Strasser R, Seifert G, Doblin MS, Johnson KL, Ruprecht C, Pfrengle F, Bacic A, Estevez JM (2021) Cracking the "Sugar Code": a snapshot of N- and O-Glycosylation pathways and functions in plants cells. Front Plant Sci 12:640919

Sun Y, Cheng JJ, Himmel ME, Skory CD, Adney WS, Thomas SR, Tisserat B, Nishimura Y, Yamamoto YT (2007) Expression and characterization of *Acidothermus cellulolyticus* E1 endoglucanase

in transgenic duckweed *Lemna minor* 8627. Bioresource Technol 98:2866–2872

Tan L, Leykam JF, Kieliszewski MJ (2003) Glycosylation motifs that direct arabinogalactan addition to arabinogalactan-proteins. Plant Physiol 132:1362–1369

Tan L, Varnai P, Lamport DT, Yuan C, Xu J, Qiu F, Kieliszewski MJ (2010) Plant O-hydroxyproline arabinogalactans are composed of repeating trigalactosyl subunits with short bifurcated side chains. J Bio Chem 285:24575–24583

Tan L, Showalter AM, Egelund J, Hernandez-Sanchez A, Doblin MS, Bacic A (2012) Arabinogalactan-proteins and the research challenges for these enigmatic plant cell surface proteoglycans. Front Plant Sci 3:140

Wang X, Karki U, Abeygunaratne H, UnnoldCofre C, Xu J (2021) Plant cell-secreted stem cell factor stimulates expansion and differentiation of hematopoietic stem cells. Process Biochem 100:39–48

Xiao Y, He X, Ojeda-Lassalle Y, Poovaiah C, Coleman HD (2018) Expression of a hyperthermophilic endoglucanase in hybrid poplar modifies the plant cell wall and enhances digestibility. Biotechnol Biofuels 11:225

Xu J, Okada S, Tan L, Goodrum KJ, Kopchick JJ, Kieliszewski MJ (2010) Human growth hormone expressed in tobacco cells as an arabinogalactan-protein fusion glycoprotein has a prolonged serum life. Transgenic Res 19:849–867

Xu J, Shpak E, Gu T, Moo-Young M, Kieliszewski M (2005) Production of recombinant plant gum with tobacco cell culture in bioreactor and gum characterization. Biotechnol Bioeng 90:578–588

Xu J, Tan L, Goodrum KJ, Kieliszewski MJ (2007) High-yields and extended serum half-life of human interferon alpha2b expressed in tobacco cells as arabinogalactan-protein fusions. Biotechnol Bioeng 97:997–1008

Xu J, Tan L, Lamport DTA, Showalter AM, Kieliszewski MJ (2008) The O-Hyp glycosylation code in tobacco and Arabidopsis and a proposed role of Hyp-glycans in secretion. Phytochem 69:1631–1640

Yennamalli RM, Rader AJ, Kenny AJ, Wolt JD, Sen TZ (2013) Endoglucanases: insights into thermostability for biofuel applications. Biotechnol Biofuels 6:136

Yuan JS, Tiller KH, Al-Ahmad H, Stewart NR, Stewart CN (2008) Plants to power: bioenergy to fuel the future. Trends Plant Sci 13:421–429

Zhang N, Dolan M, Wu D, Phillips GC, Xu J (2016a) Dramatic secretion of recombinant protein expressed in tobacco cells with a designer glycopeptide tag is highly impacted by medium composition. Plant Cell Rep 35:2513–2522

Zhang N, Gonzalez M, Savary B, Xu J (2016b) High-yield secretion of recombinant proteins expressed in tobacco cell culture with a designer glycopeptide tag: Process development. Biotechnol J 11:497–506

Zhang N, Wright T, Caraway P, Xu J (2019a) Enhanced secretion of human alpha1-antitrypsin expressed with a novel glycosylation module in tobacco BY-2 cell culture. Bioengineered 10:87–97

Zhang N, Wright T, Wang X, Karki U, Savary B, Xu J (2019b) Engineering "designer" glycomodules for boosting recombinant protein secretion in tobacco hairy root culture and studying hydroxyproline-O-glycosylation process in plants. Plant Biotechnol J 17:1130–1141

Ziegelhoffer T, Raasch JA, Austin-Phillips S (2001) Dramatic effects of truncation and sub-cellular targeting on the accumulation of recombinant microbial cellulase in tobacco. Mol Breed 8:147–158

Ziegelhoffer T, Raasch JA, Austin-Phillips S (2009) Expression of *Acidothermus cellulolyticus* E1 endo-beta-1,4-glucanase catalytic domain in transplastomic tobacco. Plant Biotechnol J 7:527–536

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.