Plant Mol Biol (2011) 76:345–355
DOI 10.1007/s11103-011-9743-7

An efficient downstream box fusion allows high-level accumulation of active bacterial beta-glucosidase in tobacco chloroplasts

Benjamin N. Gray · Huijun Yang · Beth A. Ahner · Maureen R. Hanson

Received: 5 July 2010 / Accepted: 17 January 2011 / Published online: 30 January 2011

Abstract Production of enzymes for lignocellulose hydrolysis *in planta* has been proposed as a lower-cost alternative to microbial production, with plastid transformation as a preferred method due to high foreign protein yields. An important regulator of chloroplast protein production is the downstream box (DB) region, located immediately downstream of the start codon. Protein accumulation can vary over several orders of magnitude by altering the DB region. Experiments in bacteria have suggested that these differences in protein accumulation may result from changes in translation efficiency, though the precise mechanism of DB function is not known. In this study, three DB regions were fused to the *bglC* ORF encoding a β-glucosidase from the thermophilic bacterium *Thermobifida fusca* and inserted into the tobacco (*Nicotiana tabacum*) plastid genome. More than a two order of magnitude of difference in BglC protein accumulation was observed, dependent on the identity of the DB fusion. Differential transcript accumulation explained some the observed differences in protein accumulation, but in addition, less 3′ degradation of *bglC* transcripts was observed in transgenic plants that accumulated the most BglC enzyme. Chloroplast-produced BglC was active against both pure cellobiose and against tobacco lignocellulose. These experiments demonstrate the potential utility of transplastomic plants as a vehicle for heterologous β-glucosidase production for the cellulosic ethanol industry.

Keywords Transplastomic · Beta-glucosidase · Tobacco · *Thermobifida fusca* · Cellulosic ethanol · Downstream box

Electronic supplementary material The online version of this article (doi:10.1007/s11103-011-9743-7) contains supplementary material, which is available to authorized users.

B. N. Gray · H. Yang · B. A. Ahner (✉)
Department of Biological and Environmental Engineering, Cornell University, 202 Riley Robb, Ithaca, NY 14853, USA
e-mail: baa7@cornell.edu

Present Address:
B. N. Gray
Agrivida, Inc., 200 Boston Ave., Ste. 3100, Medford, MA 02155, USA

M. R. Hanson
Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853, USA

Introduction

Cellulosic ethanol has been promoted as a promising gasoline substitute with the potential to significantly reduce fossil fuel dependence and the environmental problems associated with fossil fuel usage. The preferred method of cellulosic ethanol production proceeds via enzymatic hydrolysis of a lignocellulosic substrate followed by fermentation of the resulting hydrolysate to produce ethanol. Typical microbial cellulase systems contain a complex mixture of endo- and exo-glucanases, β-glucosidase, accessory enzymes, and non-hydrolytic proteins for efficient cellulose hydrolysis (reviewed in Zhang and Lynd 2004). The most common industrial source of cellulases is the cell culture supernatant of the fungus *Trichoderma reesei*, which has a high cellulase activity, but it is often deficient in β-glucosidase activity (Juhász et al. 2005). Supplementation of *T. reesei* cellulase preparations with β-glucosidase increases glucose and/or ethanol concentrations after cellulose hydrolysis (e.g. Schell et al. 1990; Lamed et al. 1991; Spindler et al. 1989). A low-cost source of β-glucosidase for addition to *T. reesei* cellulase cocktails would be of interest for industrial enzymatic cellulose hydrolysis.

Transgenic plants have been proposed as low-cost sources of foreign proteins, with projected order of magnitude cost savings relative to microbial protein production systems (Twyman et al. 2003). Previous reports of nuclear expression of β-glucosidases in tobacco include one from *Aspergillus niger* that resulted in a maximum accumulation of 2.3% total soluble protein (% TSP) when targeted to the vacuole (Wei et al. 2004), and one from *Thermotoga maritima* that resulted in accumulation of up to 5.8% TSP (Jung et al. 2010). Chloroplast transformation has resulted in reports of extraordinarily high levels of foreign protein, up to 70% TSP (Oey et al. 2009), and a number of proteins have been expressed at greater than 10% TSP from the chloroplast genome (reviewed in Maliga 2003), including two cellulases (Gray et al. 2009; Ziegelhoffer et al. 2009). Though the enzyme concentration *in planta* was not determined, active cellulolytic enzymes were obtained from transplastomic tobacco plants carrying transgenes with coding regions from *Clostridium thermocellum* and *T. reesei* (Verma et al. 2010).

High-level foreign protein expression in chloroplasts depends on the proper selection of foreign gene regulatory elements such as 5′ and 3′ untranslated regions (UTRs), promoters, and terminators. One regulatory region that has been shown to be important is the downstream box (DB) region (Ye et al. 2001; Kuroda and Maliga 2001a, b; Lenzi et al. 2008; Gray et al. 2009; Farran et al. 2010), composed of the 10–15 codons immediately downstream of the start codon. Both silent and non-silent changes in this region have been shown to affect foreign protein production. Though the exact mechanism of these effects is not known, the DB may influence RNA stability. Inefficient DB regions in *E. coli* result in abortive translation in which the ribosomal subunits dissociate and the nascent polypeptide drops off from the transcript (Gonzalez de Valdivia and Isaksson 2005). Decreased association of ribosomes with mRNA results in degradation of the mRNA, implying a feedback between translation efficiency and mRNA stability (Nilsson et al. 1987; Rapaport and Mackie 1994). Studies of transformed plant chloroplasts have revealed that the DB can influence both translation efficiency (Kuroda and Maliga 2001b) and RNA levels (Kuroda and Maliga 2001a; Gray et al. 2009). Similarities between DB function in *E. coli* and chloroplasts are consistent with the bacterial origin of the transcription and translation machinery found in plastids.

Also as in prokaryotes, protein expression can be achieved through multi-gene operons in chloroplasts. Native plastid ORFs are translated from both monocistronic and unprocessed polycistronic mRNA species (Barkan 1988); likewise it does not appear that mRNA processing is strictly required for translation of foreign proteins. For example, high levels of the CryAa2 protein accumulated in transplastomic tobacco leaves when the most actively translated transcript was the full-length polycistronic message (Quesada-Vargas et al. 2005). In contrast, polycistronic mRNA processing was reported to be required for efficient production of yellow fluorescent protein (YFP; Zhou et al. 2007). We previously reported that accumulation of monocistronic *cel6A* mRNA correlated with Cel6A protein accumulation in transplastomic tobacco, though a causative relationship was not established (Gray et al. 2009).

Our previous study examined variation in chloroplast expression of Cel6A, a *Thermobifida fusca* endoglucanase, resulting from the fusion of three different DB regions to the *cel6A* ORF (Gray et al. 2009). The current report describes the fusion of the same DB regions, originating from the genes that encode the TetC, NPTII, and GFP proteins, to the *bglC* ORF encoding a *T. fusca* β-glucosidase (BglC). The effects of DB fusion to the *bglC* ORF are explored at both the RNA and protein levels.

Results

Identification of fully transformed DB-BglC tobacco transformants

Transplastomic tobacco lines containing the *bglC* ORF were generated via biolistic bombardment of tobacco seedlings, resulting in the transformed *trnI*/*trnA* locus diagrammed in Fig. 1a. The vectors used for these transformations are identical to those described previously (Gray et al. 2009), except that the *cel6A* ORF used previously was replaced with the *bglC* ORF. The complete fusion protein was prepared by inserting an *Nhe*I restriction site (encoding Ala and Ser) and the first 13 codons of the genes that encode the TetC, NPTII, and GFP proteins between the start codon and the rest of the *bglC* ORF. These three DB regions were chosen because the full-length *tetC* and *neo* genes, encoding the TetC and NPTII proteins, respectively, have been expressed at high levels in chloroplasts previously (Tregoning et al. 2003; Kuroda and Maliga 2001a), while the GFP DB region was used previously to enhance accumulation of EPSPS protein in tobacco chloroplasts (Ye et al. 2001). The DB-BglC transgenes were inserted between the chloroplast *trnI* and *trnA* genes. Following three rounds of tissue culture regeneration on spectinomycin, DNA was isolated from resistant and WT shoots. DNA was *Xho*I digested and hybridized with a radiolabeled *trnI*-specific probe (Fig. 1c). This DNA blot showed the expected 3.0 and 0.64 kb *Xho*I fragments from WT and transformed plants, respectively, but also showed a faint 3.0 kb fragment in all of the DB-BglC transformed plants analyzed. Tobacco DNA was

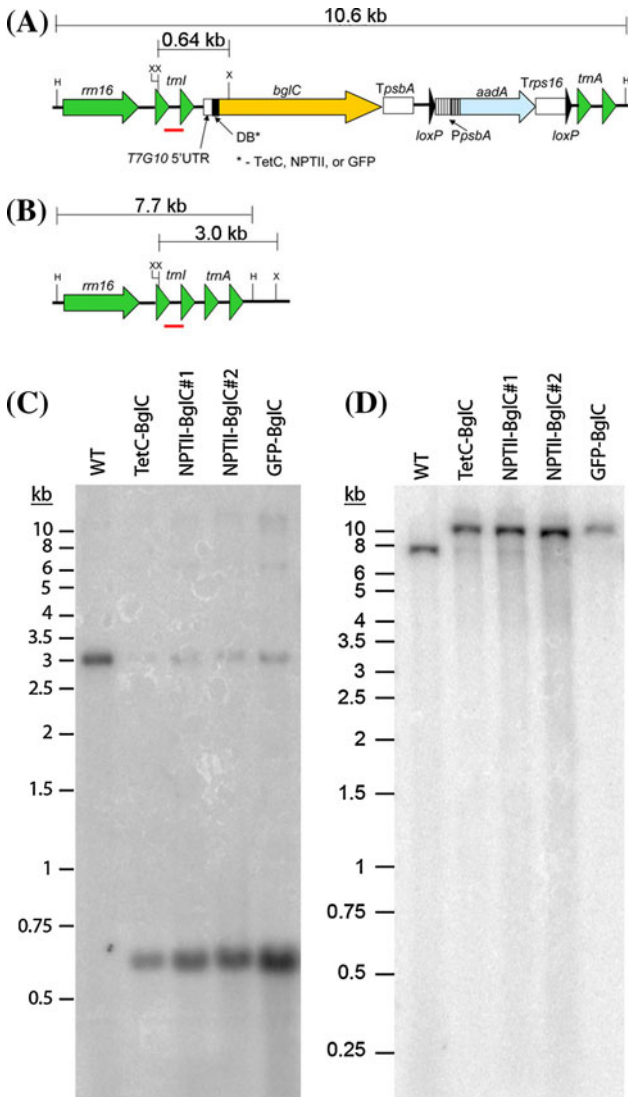

Fig. 1 Transformed and wild-type plastid genomes and DNA blots from transplastomic tobacco. **a** Schematic diagram of the DB-BglC chloroplast transformation vectors. **b** Schematic diagram of the wild-type *trnI*/*trnA* region. The *Xho*I (X) and *Hind*III (H) sites relevant to DNA blotting experiments are shown, along with the predicted fragment sizes. The location of the *trnI* probe used for DNA blotting is shown as a *red line* below each schematic diagram. **c** *Xho*I digested DNA, with 3.0 kb and 640 bp *trnI*-containing WT and DB-BglC fragments, respectively. **d** *Hind*III digested DNA, with 7.7 kb and 10.6 kb *trnI*-containing WT and DB-BglC fragments, respectively

also digested with *Hind*III and hybridized with the same *trnI*-specific probe (Fig. 1d). This blot showed the expected 7.7 and 10.6 kb *Hind*III fragments in WT and DB-BglC tobacco, respectively, but again a faint WT band is observed in the transformed plants. Faint WT bands in both digestions are attributed to extraplastidic copies of the *trnI* gene present in the nuclear genome. No segregation of seedlings or sectors lacking antibiotic resistance were observed when progeny of the transplastomic lines were plated on spectinomycin medium (data not shown). NPTII-BglC#1 and NPTII-BglC#2 plants are derived from independent transformation events. Analysis at the DNA, protein, and phenotypic levels revealed no differences between these two plant lines (data not shown), so NPTII-BglC#1 was used for all subsequent analyses.

Chloroplast-produced DB-BglC protein accumulation

When the T0 plants each had approximately 30 leaves, soluble leaf protein was extracted from young, mature, and old (i.e., approximate leaf numbers 28, 15, and 2) NPTII-BglC, TetC-BglC, and GFP-BglC tobacco leaves for immunoblotting. Figure 2 shows that NPTII-BglC and TetC-BglC accumulated to 8.0–12 and 1.6–2.6% of total soluble protein (% TSP), respectively; whereas levels were below detection in GFP-BglC plants. BglC protein concentration was stable as leaves aged, with similar BglC concentrations in young, mature, and old leaves for both DB-BglC constructs that contained measurable protein. Subsequent immunoblots done with middle aged leaves from T1 plants resulted in average values (±SD) of $9.6 \pm 3.7\%$, $3.5 \pm 0.5\%$ and $0.05 \pm 0.01\%$ for NPTII-BglC, TetC-BglC and GFP-BglC, respectively, all of which were statistically different from each other based on a student's *t*-test ($P < 0.05$).

Differential *bglC* transcript processing and abundance

Total leaf RNA was isolated from young leaves of T1-generation DB-BglC tobacco and separated by electrophoresis for RNA blotting (Fig. 3). All *bglC*-containing transcripts were more abundant in the NPTII-BglC plants than in either TetC-BglC or GFP-BglC plants whereas no major quantitative or qualitative differences were seen between the *tetC-bglC* and *gfp-bglC* transcripts. Figure 3 shows a complex pattern of *DB-bglC* transcript accumulation, with polycistronic RNAs transcribed from the native plastid ribosomal promoter located upstream of the *trnI*/*trnA* insertion site and various shorter processed RNAs produced from the primary polycistronic transcripts. The full-length polycistronic transcript is expected to span approximately 9.9 knt and contain the *16*, *23*, *4.5*, and *5 s rrn*, *trnI*, *bglC*, *aadA*, and *trnA* transcripts. There are visible bands at roughly 5, 7.5, 9 and 10 knt, corresponding to polycistronic transcripts containing the *bglC* ORF. Major bands at approximately 3.0 and 1.7 knt, respectively, are likely the dicistronic *bglC-aadA* and the monocistronic *bglC* species, based on predicted sizes and hybridization with the *bglC* probe. The *trnI-bglC* dicistron is located at approximately 2.5 knt, where a faint band is observed. Tricistronic *16 s rrn-trnI-bglC*, *trnI-bglC-aadA* and *bglC-aadA-trnA* transcripts would generate bands of approximately 4.3, 3.7 and 3.8 knt,

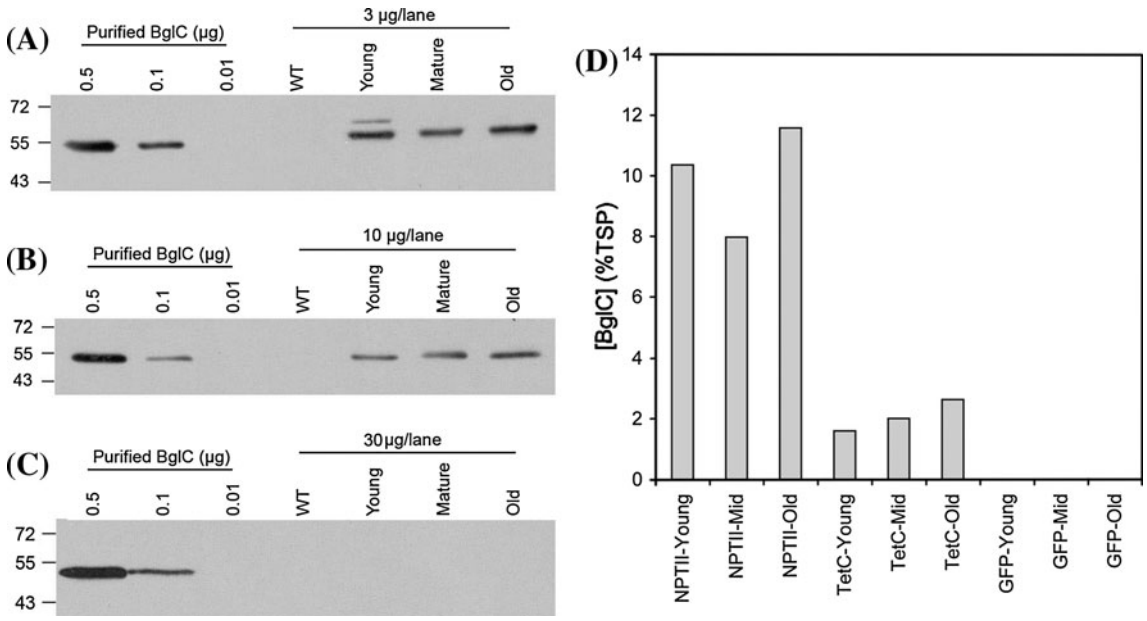

Fig. 2 Immunoblotting with protein extracted from DB-BglC leaves. **a** NPTII-BglC. **b** TetC-BglC. **c** GFP-BglC. **d** Quantification of the immunoblots shown in **a**–**c**

respectively, but none of these are observed. Three sub-ORF sized bands were detected, ranging from 1.0 to 1.4 knt, that accumulate in all three DB-BglC plant lines. They are likely intermediates in the *bglC* mRNA degradation pathway formed by endonucleolytic cleavage at specific sites within the *bglC* monocistron.

Mapping of 5′ and 3′ ends of monocistronic *bglC* mRNAs

The RNA blot of *bglC* transcripts cannot reveal the precise 5′ and 3′ ends of the transcripts in the three different transgenic lines. To determine whether the 5′ and 3′ transcript termini of the monocistronic *DB-bglC* mRNAs were similar in lines carrying different DBs, circular RT–PCR was performed to simultaneously identify the 5′ and 3′ ends produced in plants transformed with each of the three DB-BglC constructs. Figure 4a shows a schematic diagram of the *DB-bglC* mRNA ends determined with this method. The most common 5′ end for monocistrons from all three DB-BglC tobacco lines was at −124 (relative to the +1 start codon), in the intergenic region between *trnI* and *bglC* (Fig. 4b). RNA from all three *DB-bglC* constructs contained a small number of untranslatable 5′ ends within the *bglC* ORF. The distribution of observed 5′ ends observed in monocistrons produced in each DB-BglC plant line was remarkably similar. Figure 4c shows a similar analysis of the observed 3′ transcript ends, where greater differences were observed among the three DB-BglC constructs. Over 80% of observed *nptII-bglC* monocistrons had a 3′ end at least 51 nt downstream of the *bglC* stop codon. Over 75% of monocistrons in TetC-BglC tobacco contained 0–99 nt downstream of the *bglC* stop codon, and 75% of observed monocistrons from GFP-BglC tobacco were degraded into the *gfp-bglC* ORF from the 3′ end or retained fewer than 50 nt downstream of the stop codon. Significant fractions (19 and 42%, respectively) of observed *tetC-bglC* and *gfp-bglC* monocistrons could not be translated into a full-length enzyme due to 3′ ends within the *bglC* ORF. The precise 5′ and 3′ termini determined by circular RT–PCR are shown in supplementary Table 2. The circular RT–PCR employed here to determine 5′ and 3′ RNA ends was not capable of detecting monocistronic RNAs that had been degraded more than approximately 40 nt downstream of the start codon from the 5′ end or approximately 50 nt upstream of the stop codon from the 3′ end due to the annealing sites of the primers used for these experiments.

Figures 5a and b show the lowest-energy predicted structures of the 5′ and 3′ termini, respectively, of *DB-bglC* monocistrons (Zuker 2003). The commonly observed −160 (within *trnI*) and −124 (in the intergenic region between *trnI* and *bglC*) 5′ termini are marked in Fig. 5a. Similarly, the commonly observed +1598 (T*psbA*) and +1645 (*loxP*) 3′ termini are marked in Fig. 5b. These figures show that the most commonly observed 5′ and 3′ termini are found at the ends of predicted hairpin structures, consistent with previous reports of chloroplast

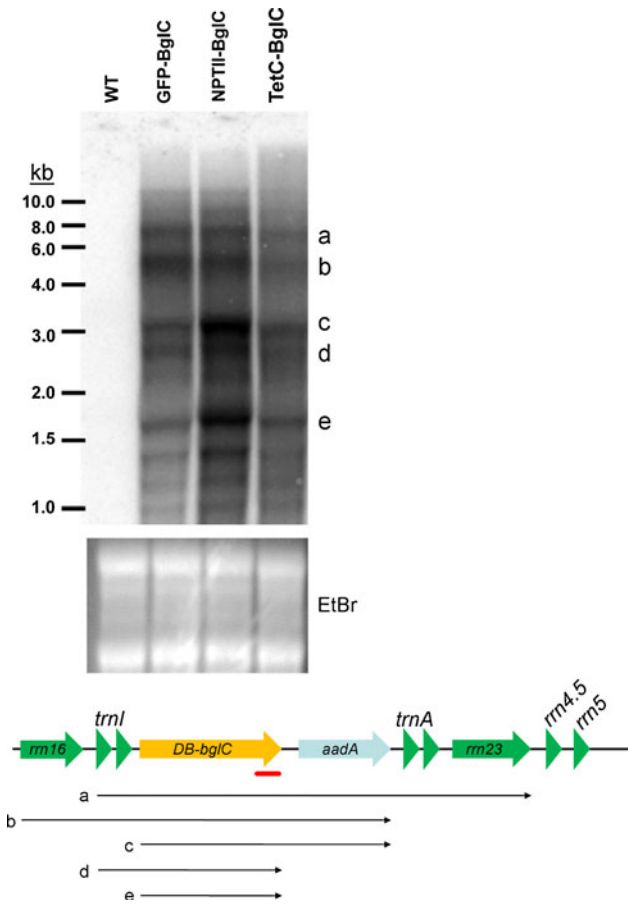

Fig. 3 RNA blotting with RNA extracted from DB-BglC leaves. RNA was hybridized with a *bglC*-specific probe comprised of a PCR product corresponding to the final ~500 nt of the *bglC* ORF (shown as a *red line* below the schematic diagram). Letters to the right of the RNA blot correspond to the transcripts shown in the diagram below the RNA blot. Ethidium bromide-stained ribosomal RNA bands are shown below the RNA blot to indicate relative loading. The schematic diagram below the RNA blot shows predicted transcripts, with letters corresponding to the major transcripts detected by RNA blotting experiments

mRNA maturation by exonuclease trimming up to a hairpin structure (Hayes et al. 1999; Monde et al. 2000).

Identification of polyadenylated *bglC* transcripts

Chloroplast mRNAs have been reported to undergo polyadenylation prior to exonucleolytic degradation (Hayes et al. 1999). Nevertheless, polyadenylated transcripts were not detected in the circular RT–PCR experiments described above, despite the detection of transcripts that appear to have been degraded by exonucleases. In order to determine whether *bglC* transcripts were polyadenylated in the chloroplast, reverse transcription was performed using forward primer BglCint-fwd and an oligo(dT)$_{17}$ primer as the reverse primer. Two primer pairs, BglCint-fwd/oligo(dT)$_{17}$ and BglCint-fwd2/oligo(dT)$_{17}$, were used for PCR amplification of polyadenylated *bglC* transcripts. Three polyadenylated *bglC* species were identified by this method (Fig. 6). Two of the polyadenylation sites were located within the *bglC* ORF. A third polyadenylation site was located near the 3′ end of T*psbA*, though this site could also be located within *loxP*; the exact polyadenylation site is difficult to determine more precisely than ± 30 nucleotides.

All three polyadenylation sites were amplified from all three DB-BglC plants, though the relative abundances appeared to differ among the three constructs using this semi-quantitative approach. NPTII-BglC plants contained the least amount of RNA polyadenylated within the *bglC* ORF, and TetC-BglC plants contained more of this polyadenylated RNA species than GFP-BglC plants. The amount of RNA polyadenylated downstream of T*psbA* appeared to be roughly equal among the three constructs.

Measurement of chloroplast-produced NPTII-BglC activity

In order to determine whether chloroplast-produced NPTII-BglC protein was active against cellobiose, soluble NPTII-BglC tobacco leaf protein was extracted and incubated with cellobiose. Glucose concentration was measured after a 10 min incubation at 50°C. NPTII-BglC tobacco leaf protein extracts contain an enzyme activity that was able to produce glucose from cellobiose (Fig. 7a). Wild-type tobacco protein extract did not hydrolyze an appreciable amount of cellobiose in this assay (data not shown), indicating that the cellobiose hydrolysis observed was a result of NPTII-BglC protein. Figure 7a shows that quantification of NPTII-BglC protein concentration against a calibration curve of known amounts of BglC added to WT tobacco protein was in good agreement with quantification of NPTII-BglC protein concentration from the immunoblots. This indicates that all, or nearly all, chloroplast-produced NPTII-BglC was enzymatically active against cellobiose.

Hydrolysis of tobacco leaf tissue by chloroplast-produced NPTII-BglC

In order to determine whether NPTII-BglC leaf protein extracts were suitable for use with commercial cellulase preparations for hydrolysis of complex lignocellulosic substrates, WT tobacco leaf tissue was pre-treated in NaOH, then hydrolyzed with Spezyme CP, a commercially produced cellulase preparation. Wild-type or NPTII-BglC tobacco leaf protein was added to the hydrolysis reaction and glucose concentration was measured at various time points during hydrolysis. Figure 7b shows that glucose concentration was significantly higher after 4 h of hydrolysis at 50°C when both Spezyme CP and NPTII-BglC

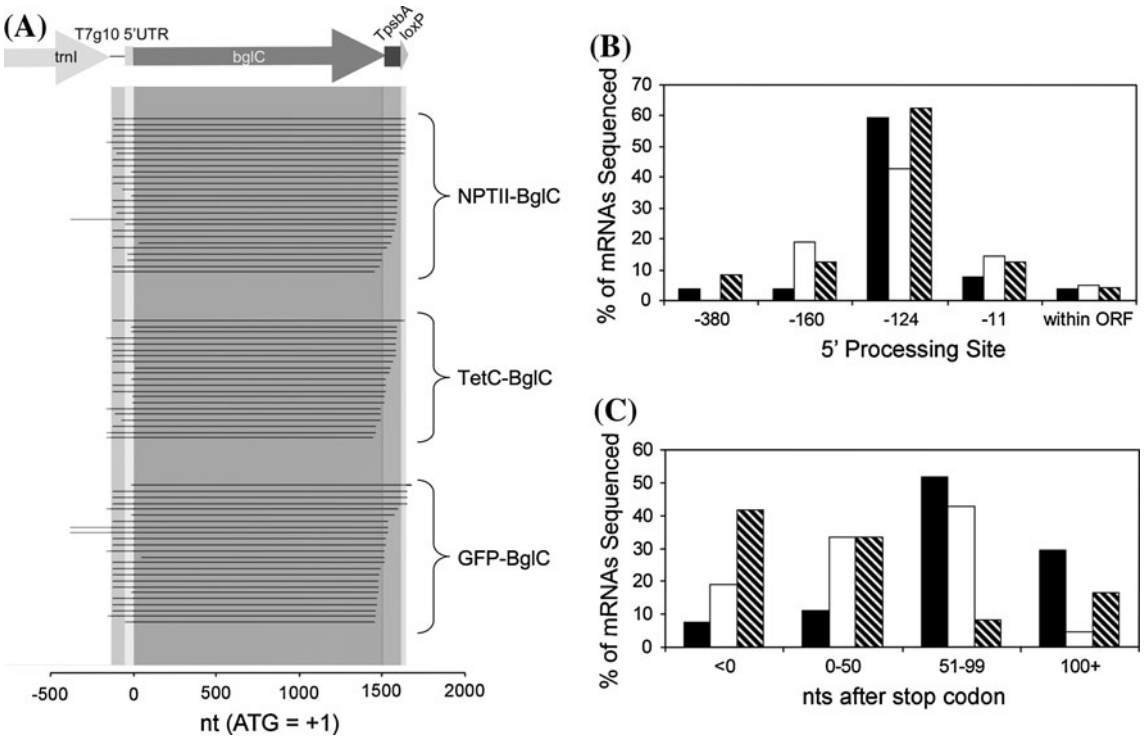

Fig. 4 Identification of 5′ and 3′ ends of monocistronic *DB-bglC* mRNAs by circular RT–PCR. **a** Schematic diagram (to scale) showing the transcripts from each construct detected by circular RT–PCR. **b** Analysis of 5′ ends detected in each DB-BglC plant, relative to the AUG start codon at +1. **c** Analysis of 3′ ends detected in each DB-BglC plant. In **b** and **c**, NPTII-BglC is indicated by *black bars*, TetC-BglC is indicated by *white bars*, and GFP-BglC is indicated by *striped bars*

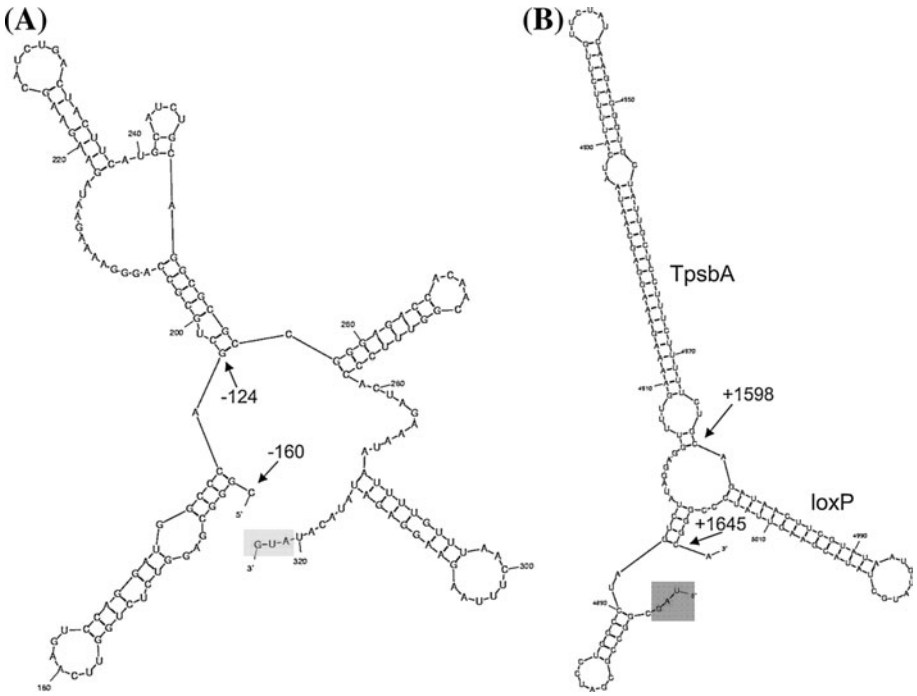

Fig. 5 Lowest-energy structure predictions around the observed 5′ and 3′ ends of monocistronic *DB-bglC* transcripts. **a** Structure prediction of the 5′ terminus, showing the commonly observed −160 and −124 5′ termini. The AUG start codon is shaded. **b** Structure prediction of the 3′ terminus, showing the UAG stop codon (*shaded*), the T*psbA* hairpin-loop, and the *loxP* hairpin-loop. Commonly observed +1598 and +1645 3′ termini are indicated by *arrows*

protein were added to the hydrolysis reaction than when Spezyme CP was omitted or when WT tobacco protein was used. Glucose concentrations increased over the 24 h testing period and were 3- to 8-fold higher after 24 h when both Spezyme CP and NPTII-BglC protein were added than in the other 3 hydrolysis mixtures tested.

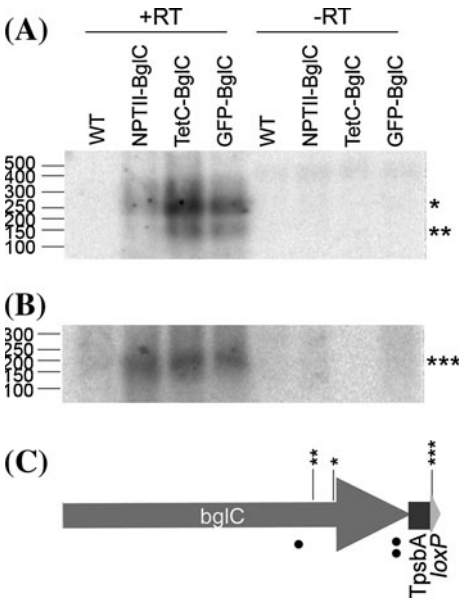

Fig. 6 RT-PCR amplification and DNA blotting detection of polyadenylated *DB-bglC* transcripts. **a** PCR product with primers BglCint-fwd/oligo(dT)$_{17}$. **b** PCR product with primers BglCint-fwd2/oligo(dT)$_{17}$. **c** Schematic diagram (to scale) showing the observed polyadenylation sites in the *bglC* ORF and near the 3′ end of T*psbA*. Both PCRs followed reverse transcription using primers BglCint-fwd and oligo(dT)$_{17}$. The annealing sites for BglCint-fwd and BglCint-fwd2 are shown below the figure as *one* and *two dots*, respectively

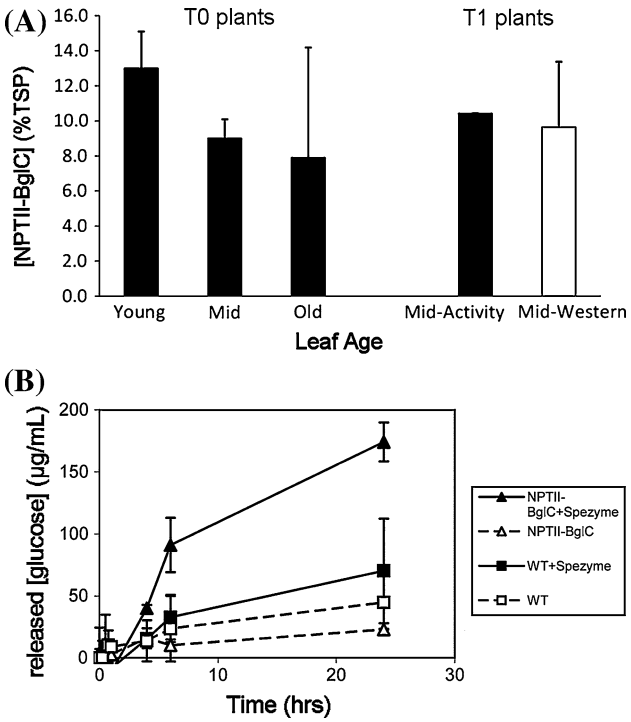

Fig. 7 Demonstration of activity of plastid-derived NPTII-BglC protein. **a** Cellobiose hydrolysis by chloroplast-produced NPTII-BglC. Quantification of NPTII-BglC concentration based on cellobiose hydrolysis was performed by incubating known amounts of purified BglC with cellobiose to generate a calibration curve. Activity-based quantification of NPTII-BglC accumulation in T1 leaf tissue is compared with immunoblot-based quantifications from multiple immunoblots ($n = 9$). **b** Hydrolysis of WT tobacco tissue by Spezyme CP (Genencor) with and without the addition of NPTII-BglC protein extract. Pre-treated WT leaf tissue was hydrolyzed by Spezyme CP along with NPTII-BglC leaf protein (*filled triangles*) or with WT leaf protein (*filled squares*). *Open triangles* and *squares* show tobacco hydrolysis by NPTII-BglC and WT protein extracts, respectively, without the addition of Spezyme CP

Discussion

The DB-BglC chloroplast expression experiments described here demonstrate that the fusion of three different DB regions to an ORF of interest can result in variation of protein accumulation over more than two orders of magnitude (Fig. 2). Similar variation in protein accumulation was seen in our previous DB-Cel6A chloroplast expression experiments (Gray et al. 2009). With both of these ORFs, the GFP DB region resulted in the lowest observed expression, on the order of 0.05% TSP. This is in contrast to the successful use of the GFP DB region to stimulate EPSPS production in chloroplasts that resulted in EPSPS accumulation of over 10% TSP (Ye et al. 2001). Farran et al. (2010) also fused a GFP DB region to a fibronectin domain and stimulated protein accumulation in tobacco chloroplasts, though the nucleotide composition of this DB region differed significantly from that used by Ye et al. (2001) and Gray et al. (2009 and this report). In addition to the coding differences among the *gfp* DB regions used by Ye et al. (2001); Farran et al. (2010), and Gray et al. (2009 and this report), the *Nhe*I site inserted immediately upstream of the *gfp* DB region in our experiments may also affect protein accumulation. Surprisingly, we observed a switch in the efficacy of the TetC and NPTII downstream boxes to stimulate the highest levels of Cel6A and BglC accumulation, respectively (Gray et al. 2009). These results demonstrate that context is important to DB function because the NPTII, TetC, and GFP DB regions tested in our Cel6A and BglC chloroplast expression experiments were identical at both the nucleotide and the amino acid levels. In other words, perhaps the particular ORF downstream of the DB may alter the folding of the mRNA and thus affect expression mediated by a DB. In the absence of a clearer mechanistic understanding of downstream box function, it appears that empirical downstream box optimization is required for each ORF of interest for expression in transplastomic plants.

Differences in accumulation of proteins expressed from chloroplast transgenes could, in principle, result from differences in transcription rates, RNA turnover rates, translation rates, and/or protein turnover rates. Differences in transcription rates among the three DB-BglC constructs tested here are not likely, as all three *DB-bglC* ORFs are likely to be transcribed from the native plastid ribosomal

promoter upstream of the *trnI*/*trnA* insertion site. The abundance of *nptII-bglC* transcripts was higher than that of either *tetC-bglC* or *gfp-bglC* transcripts, even though the T7g10 5′UTR and the *psbA* 3′UTR sequences were present on all three transgenes. The 5′ and 3′UTRs can affect the susceptibility of transcripts to degradation by ribonucleases (Monde et al. 2000), but any such degradation mediated or prevented by the UTRs alone would be expected to be comparable among all three transgenic lines.

Though higher protein levels in the NPTII-BglC plants correlated with higher transcript abundance, in the other two transgenic plant lines very different amounts of BglC protein accumulated despite similar levels of transcripts (Figs. 2 and 3). When we previously expressed DB-Cel6A protein in transplastomic tobacco, accumulation of the monocistronic *cel6A* transcript correlated with accumulation of Cel6A protein (i.e., both monocistronic *tetC-cel6A* mRNA and TetC-Cel6A protein accumulated to the highest levels of the three constructs tested), suggesting a link between accumulation of the monocistron and protein production (Gray et al. 2009). Zhou et al. (2007) observed a similar increase in translation efficiency of a *yfp* ORF following RNA processing to produce a monocistronic transcript. In the current study, the differences in the accumulation of monocistronic *DB-bglC* transcripts relative to polycistronic transcripts as observed by RNA blotting do not entirely explain the observed differential DB-BglC protein accumulation. The most significant relationship between *DB-bglC* transcripts and protein accumulation that we observed is the relative intactness of the transcripts at the 3′ end. A significant proportion of monocistronic *tetC-bglC* and *gfp-bglC* transcripts were partially degraded, as revealed by circular RT–PCR (Fig. 4).

Most of the monocistronic mRNA for all three constructs had a 5′ terminus in the intergenic region between *trnI* and the *bglC* ORF. The 3′ termini of the three different DB-BglC constructs, however, displayed striking differences. Many *tetC-bglC* and *gfp-bglC* monocistrons degraded into the *loxP* and T*psbA* hairpins located downstream of the *bglC* ORF, and some monocistrons degraded into the *bglC* ORF such that the transcript could no longer be translated into a full-length enzyme. The detection of some intact *tetC-bglC* and *gfp-bglC* monocistrons suggests that the observed differences are not a result of differential transcription or endonucleolytic processing of the primary transcript, but of differential transcript degradation. Because the three DB-BglC constructs differed only at the 5′ end of the *DB-bglC* ORF, and were regulated by identical 5′ and 3′ UTRs, these results demonstrate that the DB region plays a role in regulating mRNA turnover. Whether protein stability may also be a factor in the observed differential protein accumulation has not been investigated. A recent report on protein stability determinants in chloroplasts, however, suggests that the N-terminal part of the protein encoded by the DB region plays an important role in protein stability (Apel et al. 2010).

Chloroplast-produced NPTII-BglC protein was assayed for activity against cellobiose (Fig. 7a), and was used in conjunction with Spezyme CP, a commercial cellulase preparation, to hydrolyze tobacco leaf tissue to glucose (Fig. 7b). These assays demonstrated that most, if not all, of the NPTII-BglC protein produced in chloroplasts is active. Importantly, NPTII-BglC was able to hydrolyze not only pure cellobiose, but also the complex mixture of glucose oligomers produced by enzymatic hydrolysis of tobacco leaf tissue. Commercial cellulase preparations are typically made from the supernatant liquid of *Trichoderma reesei* cell cultures, which have been shown to contain low levels of β-glucosidase activity relative to their cellulase activities (Juhász et al. 2005). Supplementation of commercial cellulases with β-glucosidase has been shown to improve the performance of lab-scale cellulosic ethanol production (e.g., Schell et al. 1990; Spindler et al. 1989), and may be required for efficient glucose production on a commercial scale. The NPTII-BglC plants generated in this study contain over two times more β-glucosidase than the maximal yield of chloroplast-targeted β-glucosidase expressed from the nucleus (Jung et al. 2010). This study illustrates the potential for a significant improvement in plant-based β-glucosidase production by expressing β-glucosidase genes from the plastid genome. In the long term, enzyme cocktails prepared entirely from transplastomic plants may be more economical and even more effective that those produced commercially today (Verma et al. 2010). Transplastomic plants expressing NPTII-BglC could be a useful source of low-cost β-glucosidase for the cellulosic ethanol industry.

Experimental procedures

DB-BglC plasmid vector construction

The *bglC* open reading frame (ORF) was PCR-amplified from plasmid pNS6 (Spiridonov and Wilson 2001). In conjunction with reverse primer BglC-rev, forward primers TetCBglC-fwd, NPTIIBglC-fwd, and GFPBglC-fwd (primer sequences are shown in supplementary Table 1) were used to generate modified *bglC* ORFs containing downstream box (DB) fusions from the *tetC* (Tregoning et al. 2003), *neo* (Kuroda and Maliga 2001a), and *gfp* (Ye et al. 2001) genes, respectively. These PCR reactions added *Nde*I and *Nhe*I sites at the 5′ end of the *bglC* ORF and a *Not*I site at the 3′ end. The resulting PCR products were *Nhe*I/*Not*I digested and ligated into the *Nhe*I/*Not*I backbone of plasmid pGFPCel6A (Gray et al. 2009) to generate pTetCBglC, pNPTIIBglC, and pGFPBglC, respectively.

Generation of transplastomic plants

Transplastomic tobacco was generated by the biolistic and tissue culture methods described by Svab and Maliga (1993). Briefly, two-weeks old tobacco seedlings (*N. tabacum* cv. Samsun) grown in sterile MS agar medium were bombarded with 0.6 micron gold beads (Bio-Rad, Hercules, CA) coated with the appropriate plasmid DNA (i.e., pTetCBglC, pNPTIIBglC, or pGFPBglC). Two days after bombardment, leaves of bombarded seedlings were cut in half and transferred to RMOP agar medium containing 500 mg/l spectinomycin. Antibiotic-resistant shoots containing the desired gene insertions were subjected to two to three additional rounds of regeneration on spectinomycin-containing RMOP medium to generate fully transformed shoots. These plants were transferred to MS medium containing 500 mg/l spectinomycin for rooting, then to soil for greenhouse growth and seed collection.

DNA blotting

DNA was extracted from tobacco leaves as described previously (Gray et al. 2009). Extracted DNA was thoroughly digested by *Xho*I or *Hind*III and electrophoresed in a 1% (w/v) agarose gel. Following electrophoresis, DNA was transferred to a Hybond N+ membrane (Amersham Biosciences, Piscataway, NJ). A portion of the chloroplast *trnI* gene was amplified from WT tobacco DNA using primers Iprobe-fwd/Iprobe-rev. This PCR product was radiolabeled with the DECAprime II Random Primed DNA Labeling Kit (Ambion, Austin, TX) according to the manufacturer's instructions and then hybridized with the DNA-containing membrane. Following hybridization, membranes were washed and exposed to a Phosphorimager screen for visualization (Molecular Dynamics, Sunnyvale, CA).

BglC production and purification

BL21(DE3) *E. coli* cells (Invitrogen, Carlsbad, CA) harboring the pNS6 plasmid (Spiridonov and Wilson 2001) were grown in LB medium containing 50 μg/ml kanamycin. BglC production was induced by adding 0.5 mM IPTG to the cell culture. Approximately 6 h after IPTG induction, cells were harvested by centrifugation. Cells were re-suspended in Hepes (50 mM, pH 7.0) containing 1 mM PMSF, then lysed in Hepes (50 mM, pH 7.0) containing 0.5% (w/v) SDS and 1 mM dithiothreitol. Cell debris was pelleted by centrifugation following lysis and supernatant fluid was transferred to a fresh container. Supernatant protein was concentrated using a MacroSep column (MWCO 30,000; Pall, East Hills, NY) and then separated by ion affinity chromatography on a Q-Sepharose column (Sigma, St. Louis, MO). After loading onto the column, the protein was washed in three column volumes of Hepes (50 mM, pH 7.0) and eluted in a step gradient of sodium chloride (0–1 M NaCl, 0.1 M steps) in Hepes (50 mM, pH 7.0). Purity of the eluted protein was assessed by Coomassie staining of a 12% polyacrylamide gel. Purified BglC concentration was determined by measuring the spectrophotometric absorption at 280 nm.

SDS–PAGE and immunoblotting

Protein was extracted from tobacco leaves as described previously (Gray et al. 2009). Protein samples were electrophoresed in 12% (w/v) polyacrylamide gels, then transferred to a nitrocellulose membrane (Pierce, Rockford, IL). The membrane was incubated in 5% (w/v) milk in TBST (100 mM Tris, pH 7.6, 685 mM sodium chloride, 0.5% [w/v] Tween-20), then exposed to primary antibody. Polyclonal anti-BglC antibody (kindly provided by David Wilson, Cornell University, Ithaca, NY) was diluted 1:250 in 5% (w/v) milk in TBST. Secondary antibody was horseradish peroxidase-conjugated anti-rabbit polyclonal antibody (Sigma, St. Louis, MO) diluted 1:25,000 in 5% (w/v) milk in TBST. Following incubation with secondary antibody, the membrane was incubated with SuperSignal West Dura Extended Duration Substrate (Pierce) and visualized on CL-Xposure film (Pierce). Immunoblots were quantified using Scion Image software (Scion Corporation, Frederick, MD).

RNA blotting

Total RNA was extracted from tobacco leaves using Trizol reagent (Invitrogen) according to the manufacturer's instructions but with the added step of extracting with chloroform twice before isopropanol precipitation. RNA concentration was quantified by measuring the spectrophotometric absorption at 260 nm. RNA was electrophoresed in a 1% (w/v) agarose gel, then transferred to a Hybond N+ membrane (Amersham Biosciences). A *bglC*-specific radiolabeled probe was generated as described above from PCR product BglCint-fwd/BglC-rev. Following hybridization with the membrane, the membrane was washed and exposed to a PhosphorImager screen (Molecular Dynamics).

Circular RT–PCR and determination of mRNA processing sites

Total leaf RNA was extracted as described above. Five μg of leaf RNA were simultaneously circularized and DNase-treated by the addition of 10 U T4 RNA ligase (Fermentas, Glen Burnie, MD) and 2.5 U DNase (Invitrogen). The RNA circularization/DNase reaction was allowed to proceed for 1 h at 37°C. Following RNA circularization, RNA was re-extracted using the Trizol reagent (Invitrogen)

according to the manufacturer's instructions, then re-suspended in 25 µl RNase-free water. RNA concentration and quality were determined by measuring spectrophotometric absorbance at 260 and 280 nm. Gene-specific RT–PCR was performed with the Superscript III One-Step RT–PCR kit (Invitrogen), using 0.5 µg circularized RNA as a substrate and primers BglCint-fwd2 and BglCint-rev. Reverse transcription occurred at 50°C for 30 min, followed by 40 PCR cycles consisting of 15 s denaturation at 94°C, 30 s annealing at 55°C, and 90 s extension at 68°C. RT–PCR products were cloned into the pCR2.1 TOPO vector using the TOPO TA cloning kit (Invitrogen) according to the manufacturer's instructions. Plasmids were isolated from kanamycin-resistant clones for sequencing, which was performed at the Life Sciences Core Laboratory Center (Cornell University, Ithaca, NY). Sequences were manually aligned with known DB-BglC vector sequences to determine the 5′ and 3′ ends of *DB-bglC* mRNAs.

Amplification of polyadenylated mRNAs

To amplify polyadenylated *bglC* mRNAs, 2.5 µg total RNA from young leaves of WT, NPTII-BglC, TetC-BglC, and GFP-BglC plants were thoroughly treated with DNase (Invitrogen). Following DNase treatment, polyadenylated *bglC* mRNAs were reverse transcribed using Sensiscript reverse transcriptase (Qiagen, Valencia, CA), with 30 ng DNase-treated RNA and primers BglCint-fwd and oligo$(dT)_{17}$. The reaction was also carried out without the addition of reverse transcriptase as a negative control. Following reverse transcription, polyadenylated *bglC* cDNAs were amplified using the reverse transcription products as a template and either BglCint-fwd/oligo$(dT)_{17}$ or BglCint-fwd2/oligo$(dT)_{17}$ as primer pairs. Biomix Taq polymerase (Bioline, Taunton, MA) was used according to the manufacturer's instructions for PCR amplification. Following PCR amplification, 7.5 µl of each 25 µl PCR reaction was electrophoresed in a 1.5% (w/v) agarose gel for DNA blotting, as described above. A ^{32}P-labelled probe was synthesized from PCR BglCint-fwd/BglC-rev for detection of polyadenylated *bglC* cDNAs by DNA blotting.

BglC activity assays

Protein was extracted from tobacco leaves as described previously (Gray et al. 2009). Protein extracts were incubated with 50 mM cellobiose (Sigma) for 10 min at 50°C. Two different amounts of NPTII-BglC tobacco protein were assayed in triplicate, containing 750 and 300 ng total protein, respectively. A standard curve was generated by adding known amounts of purified BglC protein (0–120 ng BglC) to WT tobacco protein and incubating with 50 mM cellobiose. Following the 10 min incubation with cellobiose, all samples were transferred to a 95°C heating block for 5 min to denature BglC and stop the reaction. Glucose was measured using a glucose assay kit (Sigma) essentially according to the manufacturer's instructions. The glucose assay kit protocol was modified to accommodate a 96-well plate format, and spectrophotometric absorbance at 540 nm was measured using a plate reader.

Tobacco hydrolysis

Soluble protein was extracted from 12 g (fresh weight) WT tobacco leaf tissue as described previously (Gray et al. 2009). Following extraction of the soluble protein, the leaf tissue was pre-treated for 24 h at room temperature in 125 mM NaOH while mixing at 200 rpm. After 24 h, the pre-treated leaf tissue was thoroughly washed in double-distilled water to remove NaOH. One gram of pre-treated and washed tobacco leaf tissue was added to each of four 15-ml centrifuge tubes. Two of the tubes received 2 mg WT protein and the other two tubes received 2 mg NPTII-BglC protein. Forty microliters of Spezyme CP cellulase (Genencor, Rochester, NY) were added to one of the tubes containing WT protein and to one of the tubes containing NPTII-BglC protein. Sodium acetate buffer (50 mM, pH 5.0) was added to all four tubes to a final volume of 2 ml. All tobacco enzyme mixtures were incubated at 50°C for 24 h while mixing. Samples were taken from the tubes immediately after adding enzyme to the tubes (time zero), as well as at 15, 30, 45 min, 1, 4, 6, and 24 h after the start of hydrolysis. Samples were stored at −80°C until analysis. Glucose content of the samples was measured using a glucose assay kit (Sigma) as described above.

Acknowledgments We thank David Wilson and Diana Irwin for their kind donation of anti-BglC antibodies and the pNS6 plasmid for cloning the *bglC* gene, and Deborah Sills for her kindly donated Spezyme CP for tobacco hydrolysis. BNG was the recipient of an NSF Graduate Research Fellowship. This work was supported by a USDA Grant (USDA NRI 2007-02133) to BAA and MRH. This research was also supported in part by the Cornell University Agricultural Experiment Station federal formula funds, Project No. NYC-165425 received from Cooperative State Research, Education and Extension Service, US Department of Agriculture. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the US Department of Agriculture.

References

Apel W, Schulze WX, Bock R (2010) Identification of protein stability determinants in chloroplasts. Plant J 63:636–650

Barkan A (1988) Proteins encoded by a complex chloroplast transcription unit are each translated from both monocistronic and polycistronic mRNAs. EMBO J 7:2637–2644

Farran I, McCarthy-Suárez I, Río-Manterola F, Mansilla C, Lasarte JJ, Mingo-Castel AM (2010) The vaccine adjuvant extra domain a from fibronectin retains its proinflammatory properties when expressed in tobacco chloroplasts. Planta 231:977–990

Gonzalez de Valdivia EI, Isaksson LA (2005) Abortive translation caused by peptidyl-tRNA drop-off at NGG codons in the early coding region of mRNA. FEBS J 272:5306–5316

Gray BN, Ahner BA, Hanson MR (2009) High-level bacterial cellulase accumulation in chloroplast-transformed tobacco mediated by downstream box fusions. Biotechnol Bioeng 102: 1045–1054

Hayes R, Kudla J, Gruissem W (1999) Degrading chloroplast mRNA: the role of polyadenylation. Trends Biochem Sci 24:199–202

Juhász T, Egyházi A, Réczey K (2005) β-glucosidase production by *Trichoderma reesei*. Appl Biochem Biotechnol 121–124: 243–254

Jung S, Suyeon K, Bae H, Lim H-S, Bae H-J (2010) Expression of thermostable bacterial β-glucosidase (BglB) in transgenic tobacco plants. Bioresource Technol 101:7144–7150

Kuroda H, Maliga P (2001a) Complementarity of the 16S rRNA penultimate stem with sequences downstream of the AUG destabilizes the plastid mRNAs. Nucleic Acids Res 29:970–975

Kuroda H, Maliga P (2001b) Sequences downstream of the translation initiation codon are important determinants of translation efficiency in chloroplasts. Plant Physiol 125:430–436

Lamed R, Kenig R, Morgenstern E, Calzada JF, De Micheo F, Bayer EA (1991) Efficient cellulose solubilization by a combined cellulosome-β-glucosidase system. Appl Biochem Biotechnol 27:173–183

Lenzi P, Scotti N, Alagna F, Tornesello ML, Pompa A, Vitale A, De Stradis A, Monti L, Grillo S, Buonaguro FM, Maliga P, Cardi T (2008) Translational fusion of chloroplast-expressed human papillomavirus type 16 L1 capsid protein enhances antigen accumulation in transplastomic tobacco. Transgenic Res 17: 1091–1102

Maliga P (2003) Progress towards commercialization of plastid transformation technology. Trends Biotechnol 21:20–28

Monde RA, Schuster G, Stern DB (2000) Processing and degradation of chloroplast mRNA. Biochimie 82:573–582

Nilsson G, Belasco JG, Cohen SN, von Gabain A (1987) Effect of premature termination of translation on mRNA stability depends on the site of ribosome release. Proc Natl Acad Sci USA 84: 4890–4894

Oey M, Lohse M, Kreikemeyer B, Bock R (2009) Exhaustion of the chloroplast protein synthesis capacity by massive expression of a highly stable protein antibiotic. Plant J 57:436–445

Quesada-Vargas T, Ruiz ON, Daniell H (2005) Characterization of heterologous multigene operons in transgenic chloroplasts. Transcription, processing, and translation. Plant Physiol 138: 1746–1762

Rapaport LR, Mackie GA (1994) Influence of translational efficiency on the stability of the mRNA for ribosomal protein S20 in *Escherichia coli*. J Bacteriol 176:992–998

Schell DJ, Hinman ND, Wyman CE, Werdene PJ (1990) Whole broth cellulase production for use in simultaneous saccharification and fermentation. Appl Biochem Biotechnol 24–25:287–297

Spindler DD, Wyman CE, Grohmann K, Mohagheghi A (1989) Simultaneous saccharification and fermentation of pretreated wheat straw to ethanol with selected yeast strains and β-glucosidase supplementation. Appl Biochem Biotechnol 20–21: 529–540

Spiridonov NA, Wilson DB (2001) Cloning and biochemical characterization of BglC, a beta-glucosidase from the cellulolytic actinomycete *Thermobifida fusca*. Curr Microbiol 42:295–301

Svab Z, Maliga P (1993) High-frequency plastid transformation in tobacco by selection for a chimeric *aadA* gene. Proc Natl Acad Sci USA 90:913–917

Tregoning JS, Nixon P, Kuroda H, Svab Z, Clare S, Bowe F, Fairweather N, Ytterberg J, van Wijk KJ, Dougan G, Maliga P (2003) Expression of tetanus toxin fragment C in tobacco chloroplasts. Nucleic Acids Res 31:1174–1179

Twyman RM, Stoger E, Schillberg S, Christou P, Fischer R (2003) Molecular farming in plants: host systems and expression technology. Trends Biotechnol 21:570–578

Verma D, Kanagaraj A, Jin S, Singh ND, Kolattukudy PE, Daniell H (2010) Chloroplast-derived enzyme cocktails hydrolyse lignocellulosic biomass and release fermentable sugars. Plant Biotechnol J 8:332–350

Wei S, Marton I, Dekel M, Shalitin D, Lewinsohn E, Bravdo B-A, Shoseyov O (2004) Manipulating volatile emission in tobacco leaves by expressing *Aspergillus niger* beta-glucosidase in different subcellular compartments. Plant Biotechnol J 2: 341–350

Ye GN, Hajdukiewicz PT, Broyles D, Rodriguez D, Xu CW, Nehra N, Staub JM (2001) Plastid-expressed 5-enolpyruvylshikimate-3-phosphate synthase genes provide high level glyphosate tolerance in tobacco. Plant J 25:261–270

Zhang Y-HP, Lynd LR (2004) Toward an aggregated understanding of enzymatic hydrolysis of cellulose: Noncomplexed cellulase systems. Biotechnol Bioeng 88:797–824

Zhou F, Karcher D, Bock R (2007) Identification of a plastid intercistronic expression element (IEE) facilitating the expression of stable translatable monocistronic mRNAs from operons. Plant J 52:961–972

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

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31:3406–3415