

RESEARCH

Open Access



Efficient whole-cell-catalyzing cellulose saccharification using engineered *Clostridium thermocellum*

Jie Zhang^{1,2,3,4}, Shiyue Liu^{1,2,3,4}, Renmin Li^{1,2,3,4}, Wei Hong⁵, Yan Xiao^{1,2,3}, Yingang Feng^{1,2,3}, Qiu Cui^{1,2,3*} and Ya-Jun Liu^{1,2,3*} 

Abstract

Background: Cost-efficient saccharification is one of the main bottlenecks for industrial lignocellulose conversion. *Clostridium thermocellum* naturally degrades lignocellulose efficiently using the cellulosome, a multiprotein supermolecular complex, and thus can be potentially used as a low-cost catalyst for lignocellulose saccharification. The industrial use of *C. thermocellum* is restrained due largely to the inhibition of the hydrolysate cellobiose to its cellulosome. Although the supplementation of beta-glucosidase may solve the problem, the production of the enzymes greatly complicates the process and may also increase the cost of saccharification.

Results: To conquer the feedback inhibition and establish an efficient whole-cell catalyst for highly efficient cellulose saccharification, we constructed a recombinant strain of *C. thermocellum* $\Delta pyrF::CaBglA$ which produced a secretory exoglucanase CelS-bearing heterologous BGL using a newly developed seamless genome editing system. Without the extra addition of enzymes, the relative saccharification level of $\Delta pyrF::CaBglA$ was stimulated by over twofolds compared to its parent strain $\Delta pyrF$ through a two-stage saccharification process with 100 g/L Avicel as the carbon source. The production of reducing sugars and the relative saccharification level were further enhanced to 490 mM and 79.4%, respectively, with increased cell density.

Conclusions: The high cellulose-degrading ability and sugar productivity suggested that the whole-cell-catalysis strategy for cellulose saccharification is promising, and the *C. thermocellum* strain $\Delta pyrF::CaBglA$ could be potentially used as an efficient whole-cell catalyst for industrial cellulose saccharification.

Keywords: β -Glucosidase, Cellulosome, CelS, Fermentable sugar, Genome editing, Lignocellulose

Background

Lignocellulosic biomass is the most abundantly available raw material on the Earth. Its sustainability and effective cost make it an attractive feedstock to substitute fossil resources [1–3]. Because of the recalcitrant structure, the main obstacle of lignocellulose bioconversion is the high cost of deconstruction [4–6]. *Clostridium thermocellum* has previously been demonstrated to have outstanding potential in lignocellulose bioconversion [7], because it

produces a cellulosome, a multiprotein supermolecular complex, for highly efficient degradation of cellulose [8, 9]. By assembling various enzymatic subunits and multiple structural scaffoldings together, the cellulosome makes full use of the synergy effects resulting from the interactions to promote the lignocellulose hydrolysis process [10, 11]. Thus, the cellulosome-producing *C. thermocellum* is naturally suitable for lignocellulose bioconversion [12–14]. Nevertheless, the wild-type *C. thermocellum* cannot be directly used as a whole-cell industrial catalyst so far. One of the uppermost problems is the feedback inhibition caused by the end-product cellobiose to the cellulosome greatly limits the continuous cellulose saccharification [15, 16].

*Correspondence: cuiqiu@qibebt.ac.cn; liuyj@qibebt.ac.cn

³ Qingdao Engineering Laboratory of Single Cell Oil, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao 266101, People's Republic of China

Full list of author information is available at the end of the article

Currently, the feedback inhibition is generally relieved by supplying beta-glucosidases (BGL) into the hydrolysis system. For example, the targeted integration of BGL into the *C. thermocellum* cellulosome could enhance the degradation of cellulosic substrates [17]. Prawitwong et al. obtained high production of glucose using a *C. thermocellum* culture supplemented with a BGL from *Thermoanaerobacter brockii* [18]. Many efforts have been made to obtain low-cost BGLs, including the screening and engineering of BGLs for higher activity [19–23], and the development of novel processes for protein recycling [24, 25]. However, a relatively high load of the enzymes is still unavoidable because the activity and stability of the added BGLs are continuously decreasing during the hydrolysis process. In addition, the production and supplementation of BGL may greatly complicate the saccharification process.

To avoid the extra addition of enzymes in the saccharification system, we consider one of the most convenient and efficient way is to construct a recombinant *C. thermocellum* simultaneously producing secretory BGL as a whole-cell catalyst for cellulose saccharification. To relieve the feedback inhibition and promote the cellulose hydrolysis process, the expressed BGL should function synergistically with the cellulosome by assembling and form a substrate-coupled catalyzing pathway (cellulose–cellobiose–glucose) with other cellulosomal enzymes. The exoglucanase CelS (also known as Cel48S) is the most abundant enzymatic subunit in the cellulosome of *C. thermocellum* [26, 27], and a previous study supports it has the major contribution to cellulosome function [28]. Since CelS is the main producer of cellobiose in the cellulosome, relieving the cellobiose inhibition to CelS should be a priority. Thus, the fusion of BGL with CelS may be a preferred way for BGL supplementation. To achieve the fused expression of CelS with BGL in *C. thermocellum*, precise genome-editing methods, such as the markerless gene-deletion approach [29], are required.

Results

Selection of beta-glucosidases

Beta-glucosidases from *Caldicellulosiruptor* sp. F32 (*CaBglA*), *C. thermocellum* DSM1313 (*CtBglA*) and *T. brockii* DSM1457 (*CglT*), and a compost microbial metagenome (Td2f2) were preliminarily chosen as the candidates according to previous biochemical analyses [23, 25, 30–32], and were characterized under the optimal conditions of the cellulosome of *C. thermocellum* (pH 5.5 and 55 °C) for better coordination (Additional file 1). Under such reaction conditions, *CaBglA* exhibited the highest specific activity (346.0 ± 3.7 U/mg) and thermal stability. *CtBglA* had moderate activity (208.2 ± 4.8 U/mg) and sensitivity to glucose. *CglT* showed the highest glucose tolerance but relatively low specific activity (79.3 ± 5.9 U/mg). Td2f2 could be stimulated with the addition of glucose, but its specific activity (9.2 ± 0.3 U/mg) and thermal stability were low (Table 1; Additional file 1). Hence, *CaBglA* and *CtBglA* were chosen to express in *C. thermocellum* DSM 1313.

Plasmid-dependent expression of BGLs in *C. thermocellum*

We initially tried to express BGLs in *C. thermocellum* using a replicating plasmid which is convenient to construct and to demonstrate the feasibility of BGL integration in cellulosome in vivo. pHK-*CtBglA*-Doc and pHK-*CaBglA*-Doc were constructed for the expression of fusion proteins *CaBglA*-Doc and *CtBglA*-Doc, respectively, under the control of the CelS promoter. The expressed proteins would contain a signal peptide for protein secretion and a dockerin (Doc) module for cellulosome assembly. Weak bands of ~65 kDa was detected by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) analysis of the cellulosomes (data not shown), but mass spectroscopy analysis failed to verify the fusion protein due to the low concentration. This result indicated very few, if any, BGL expression and assembly in the cellulosomes of the recombinant strains. BGL assay was then performed against

Table 1 Enzymatic properties of selected BGLs

Protein	Optimal temp (°C)/pH ^a	Specific activity (U/mg)	Thermal stability at 60/80 °C (%) ^b	Glucose inhibition (mM) ^c
<i>CaBglA</i>	75–80/5.0–5.6 [23]	346.0 ± 3.7	99.3/91.4	314
<i>CtBglA</i>	65/5.5	208.2 ± 4.8	116.9/4.8	205
<i>CglT</i>	60–75/6.0–7.0 [25]	79.3 ± 5.9	104.0/10.5	450
Td2f2	75/5.5 [31]	9.2 ± 0.3	71.1/40.9	None ^d

All experiments were performed in triplicate to calculate the averages and standard errors with pNPG as a substrate, and the reaction conditions were at pH 5.5 and 55 °C

^a The optimal temperature and pH of *CaBglA* and *CtBglA* were also determined in this study as shown in Additional file 1

^b The thermal stability was shown as percentage of remaining BGL activity after incubating at 60 or 80 °C for 24 h

^c The glucose inhibition of BGLs was determined by adding glucose at different concentrations (0–600 mM) to the standard reaction mixture, and calculated as the glucose concentration required to inhibit 50% of initial BGL activity (Additional file 1)

^d Instead of inhibition, Td2f2 could be stimulated by 29.2–72.0% with the addition of 100–600 mM glucose, which was consistent with previous report [31]

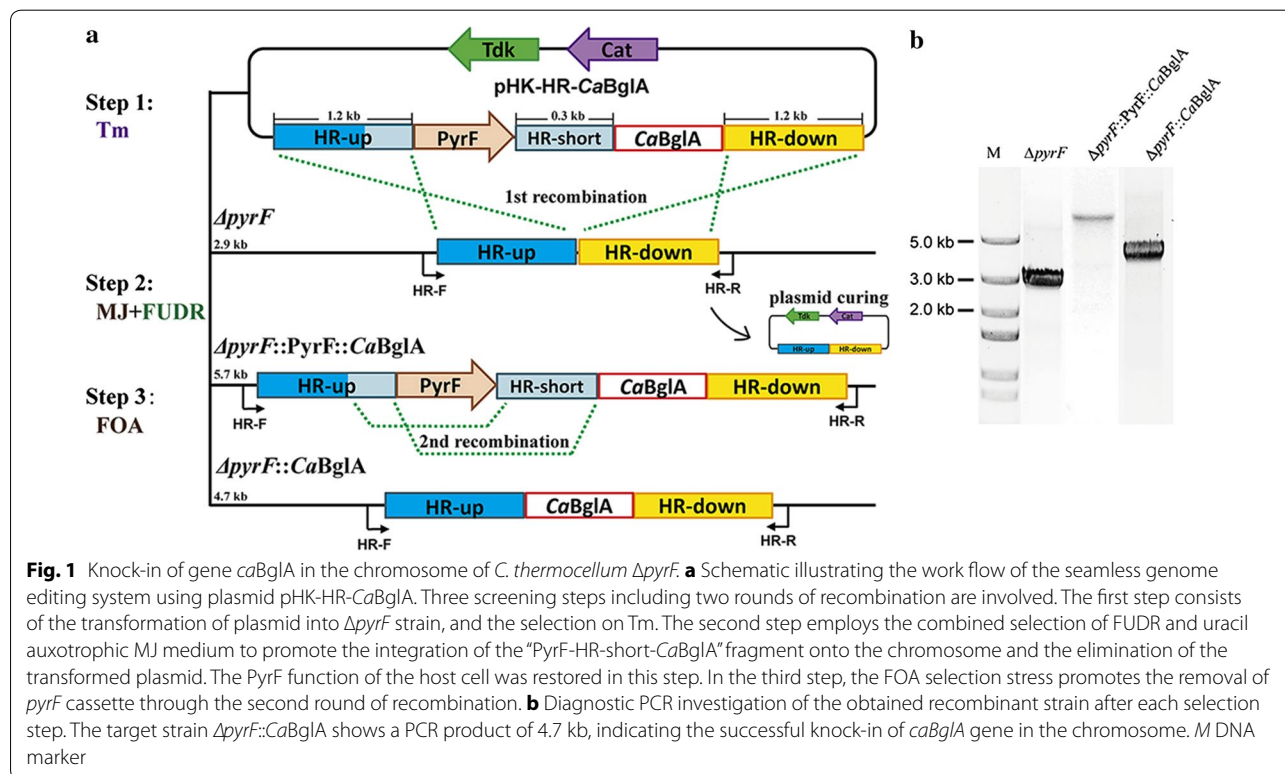
p-nitrophenyl- β -D-glucopyranoside (*p*NPG) to confirm the assembly of the expressed BGLs to the cellulosomes and their functionality. 1.15 ± 0.07 U/mg BGL activity was detected in the cellulosome of $\Delta pyrF::pHK-CaBglA$ -Doc, but no BGL activity was detected in the cellulosome of $\Delta pyrF$ and $\Delta pyrF::pHK-CtBglA$ -Doc, indicating the plasmid-based expression of the fusion protein *CaBglA*-Doc with low abundance. The failed detection of *CtBglA*-Doc indicated its failed expression or secretion. The difficulty of the plasmid-based expression in *C. thermocellum* has been discussed in previous studies [33, 34]. Nevertheless, the BGL activity of the cellulosome of $\Delta pyrF::pHK-CaBglA$ -Doc demonstrated that *C. thermocellum* can express and secrete active *CaBglA*. Thus, we decided to introduce the BGL-encoding genes into *C. thermocellum* by chromosomal integration to relieve the metabolic burden that resulted from plasmid replication and the expression of plasmid-carrying antibiotic-resistant genes. In addition, the chromosomal gene expression is also more convenient for industrial application than the plasmid-based expression.

Seamless knock-in of BGL gene in the genome of *C. thermocellum*

A seamless genome editing system was developed on the basis of the allele-coupled exchange (ACE) strategy [35], including a *pyrF*-deleted chassis strain $\Delta pyrF$ and corresponding plasmids for homologous recombination.

The ACE strategy was originally developed in mesophilic clostridia, and it had not been used in the genetic engineering of thermophilic *C. thermocellum* previously according to our knowledge [35]. Three regions of homology with different lengths (two long regions of ~ 1.2 Kb and one short region of ~ 300 bp) were involved to control the order of homologous recombination events. Two selection markers, an orotidine 5-phosphate decarboxylase encoding gene *pyrF* from *C. thermocellum* DSM1313 and a thymidine kinase encoding gene *tdk* from a thermophilic anaerobe *Thermoanaerobacter* sp. X514 [36, 37], were used to achieve the markerless manipulation (Additional file 2). With the developed genome editing system, seamless gene deletion, insertion and replacement could be achieved after three screening steps and two rounds of recombination (Fig. 1).

We opted to create a BGL-CelS fusion protein to form a substrate-coupled catalyzing channel because removing the cellobiose produced by the exoglucanase CelS, the most abundant cellulosome subunit in *C. thermocellum*, might greatly release the feedback inhibition to the whole cellulosome system. BGL was designed to locate between the catalyzing module (Cel) and dockerin module (Doc) of CelS to avoid the interference to the type I dockerin-cohesin interaction during cellulosome assembly. The produced fusion protein Cel-BGL-Doc would contain three functional modules.



The plasmid pHK-HR-*CaBglA* was constructed for the knock-in of gene *caBglA*. The termination codons of the BGL-encoding genes were eliminated during the plasmid construction to guarantee the fused expression with CelS modules. The recombinant strain $\Delta pyrF::CaBglA$ was obtained after three screening steps (Fig. 1a). In short, the $\Delta pyrF$ transformants containing pHK-HR-*CaBglA* were initially screened on solid GS-2 medium with Tm. Then, the recombinants were selected in MJ medium lacking uracil but containing FUDR. Under such selection stress, the plasmid containing *tdk* cassette must be cured to ensure the cell growth, while because the host cell was *pyrF*-deleted, the *pyrF* cassette on the plasmid was required to produce uracil. Thus, the first round recombination event occurred through two long regions of homology—HR-up and HR-down—to integrate both the selection marker *pyrF* and *CaBglA* gene onto the chromosome. The *pyrF* function of the host strain was restored in this step, and the plasmid backbone was cured. In the third step, the counter-selection function of *pyrF* was used for the removal of *pyrF* cassette through the second round of recombination between the short region of homology HR-short and the 3' region of HR-up, which harbored the same sequence with HR-short. The cells without PyrF function were selected in FOA-supplemented GS-2 medium. Colony PCRs and sequencing were performed using primer set HR-F/R to verify the recombination after each step of screening (Fig. 1b). We also tried to fuse the endogenous BGL of *C. thermocellum* DSM1313 (*CtBglA*) with CelS. However, the knock-in of *ctBglA* gene was not successful even after several attempts. The construction got stuck during the first round of recombination in the second selection step (Fig. 1a), and no positive recombinant could be detected after screening hundreds of colonies. The difficulty might be due to the preferred recombination occurred between the homologous *ctBglA* sequences, since the *ctBglA* sequence (~1.3 kb) was longer than the regions of homology (~1.2 kb).

Investigation of BGL expression by *C. thermocellum*

The cellulosomes and extracellular proteins of $\Delta pyrF::CaBglA$ were prepared and analyzed to confirm the expression of the fusion protein Cel-BGL-Doc with a theoretical size of ~135 kDa. The samples from the parent strain $\Delta pyrF$ were also analyzed as the control. SDS-PAGE analysis showed that the ~75-kDa band referring to the wild-type CelS protein was rarely detected for $\Delta pyrF::CaBglA$, but an additional band of ~135 kDa was detected (Fig. 2), indicating the expression of the fusion protein instead of the wild-type CelS. The ~135-kDa protein was further confirmed as Cel-*CaBglA*-Doc by mass spectrometry analysis (Additional file 3). Furthermore,

19.1 ± 1.2 U/mg BGL activity was detected in the cellulosome of $\Delta pyrF::CaBglA$, which was 16 times higher than that of plasmid-based *CaBglA* expression in $\Delta pyrF::pHK-CaBglA$ -Doc. These results indicated the successful expression, secretion and cellulosomal assembly of the active Cel-*CaBglA*-Doc in $\Delta pyrF::CaBglA$.

We observed lower abundance of the fused protein Cel-*CaBglA*-Doc in $\Delta pyrF::CaBglA$ compared to that of the wild-type CelS in $\Delta pyrF$, which might influence the efficiency of cellulose degradation. In order to investigate whether the decreased protein expression was because of the unsuitable codon usage of *caBglA* gene, another strain $\Delta pyrF::CaBglAm$ was constructed by replacing *CaBglA* of $\Delta pyrF::CaBglA$ with a codon-modified *CaBglAm*. The expression and cellulosomal assembly of the CelS-bearing *CaBglAm* in $\Delta pyrF::CaBglAm$ was confirmed by enzyme assay, but the expression of the fusion protein was not significantly different from that of $\Delta pyrF::CaBglA$ according to the SDS-PAGE analysis (Fig. 2).

Enhanced cellulosomal activity by CelS-bearing BGL

The cellulolytic activity of the cellulosome of $\Delta pyrF::CaBglA$ was analyzed by monitoring the concentrations of released reducing sugar after 24-h hydrolysis

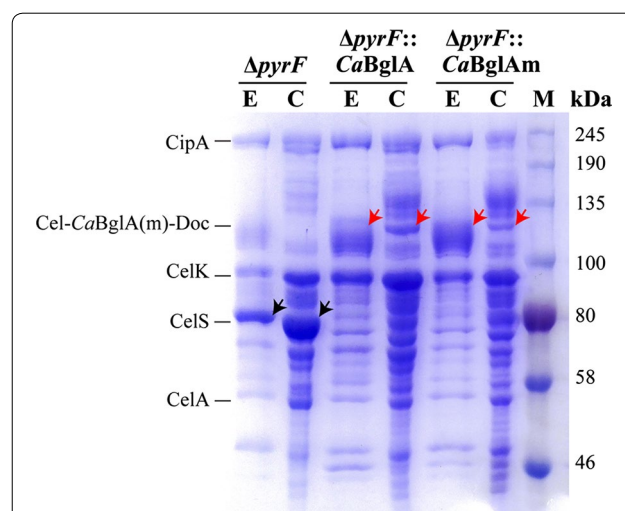


Fig. 2 SDS-PAGE analysis of cellulosomal (C) and extracellular proteins (E) of *C. thermocellum* strains. The parent strain $\Delta pyrF$ produced an intact CelS protein (black arrows). Compared to $\Delta pyrF$, an additional ~135-kDa band was observed in both cellulosomal and extracellular proteins of $\Delta pyrF::CaBglA$ and $\Delta pyrF::CaBglAm$ (red arrows), suggesting the successful expression of the fusion protein Cel-*CaBglA*(m)-Doc, and its assembly in the cellulosome. Cellulosomal and extracellular proteins of $\Delta pyrF::CaBglA$ with the size of ~135 and ~75 kDa were further identified by mass spectrometry (Additional file 3). Bands corresponding to known cellulosomal proteins are identified to the left of the Coomassie blue-stained gel. M protein standards

assay at 55 °C with Avicel as the substrate (Fig. 3). Cellobiose and glucose were also quantified by HPLC. The cellulosomal activity of $\Delta pyrF::CaBglA$ was 1.6-fold of the parent strain $\Delta pyrF$, while the glucose proportion in the released reducing sugar was increased from 34 to 78% due to the expression of $CaBglA$, suggesting the enhanced cellulolytic activity. Although the cellulosome of $\Delta pyrF$ contained no BGL activity against cellobiose, glucose was also detected in its hydrolysate against cellulose. This indicated that some endoglucanases involved in the cellulosome system of *C. thermocellum* could actively convert cello-oligosaccharides to glucose.

Cellulose saccharification using *C. thermocellum* as a whole-cell catalyst

A two-stage process, including a cell-cultivation stage and a cellulose hydrolysis stage, was employed for cellulose saccharification. To determine the cultivation time in the two-stage process, the growth patterns of $\Delta pyrF$ and $\Delta pyrF::CaBglA$ were initially analyzed using 5 g/L Avicel as a carbon source. The result showed that both strains grew in stationary phase after 28-h cultivation (Additional file 4). Thus, the cell-cultivation stage lasted for 36 h to guarantee the production of cellulosomal proteins and the complete utilization of the initial carbon source. In addition, $\Delta pyrF$ and $\Delta pyrF::CaBglA$ showed similar Avicel consumption patterns (Additional file 4). This result indicated that the expression of CelS-bearing $CaBglA$ did not influence much the cellulose degradation of *C. thermocellum* at the cell-cultivation stage because the small amounts of accumulated sugars (2.24 ± 0.05 and 1.85 ± 0.85 mM cellobiose for $\Delta pyrF$ and $\Delta pyrF::CaBglA$, respectively; no glucose was detected) would show slight, if any, inhibition effect on cellulose hydrolysis [38].

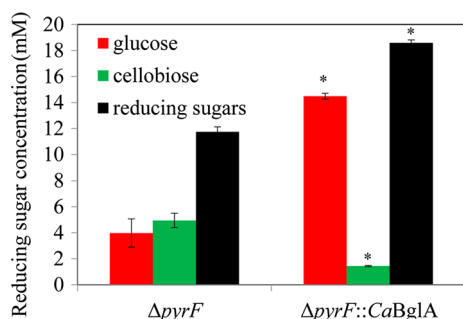
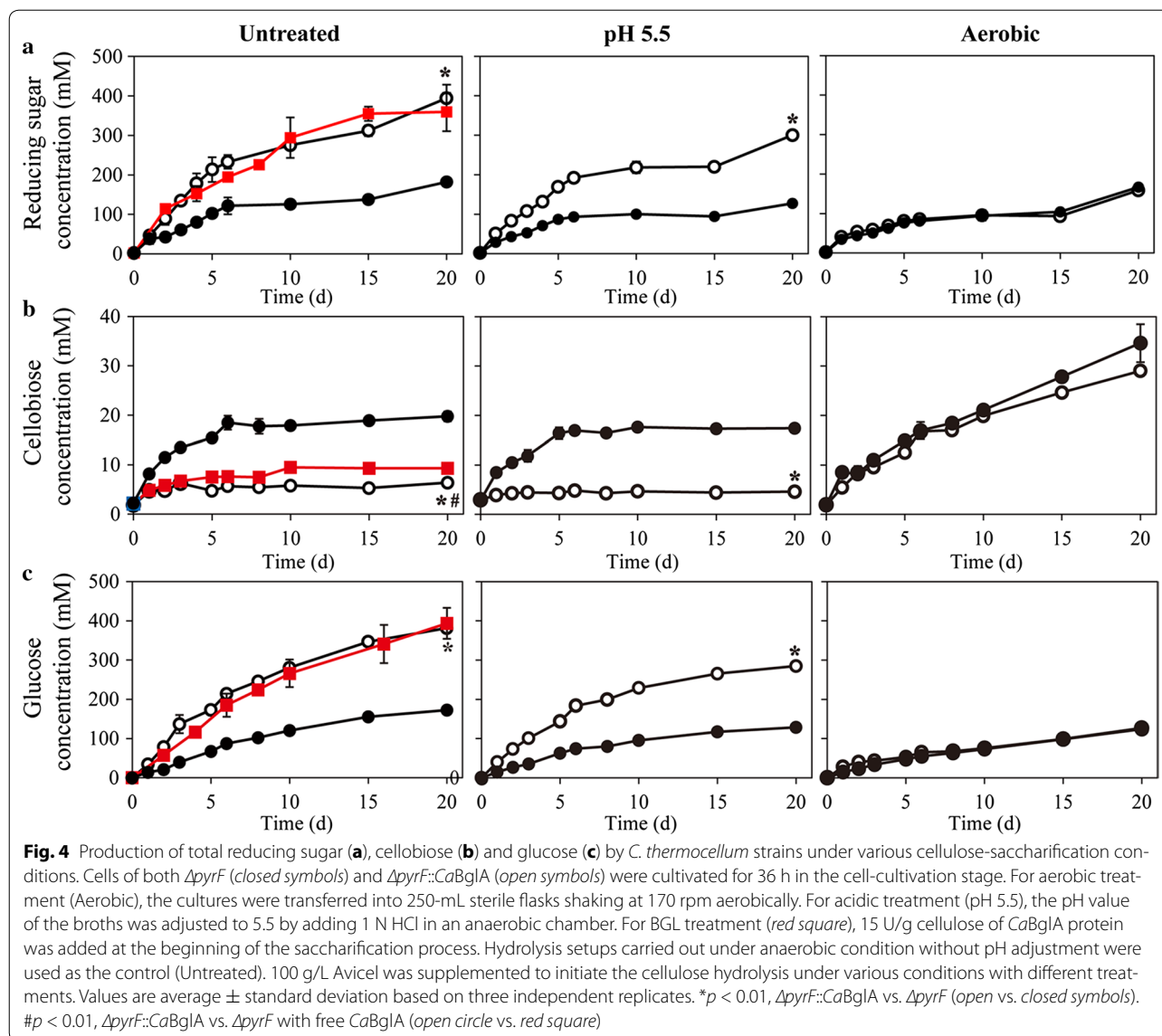


Fig. 3 Cellulolytic activity of cellulosomes from $\Delta pyrF$ and $\Delta pyrF::CaBglA$. After 24-h hydrolysis assay at 55 °C with Avicel as the substrate, the degradation activity of cellulosomes was determined by quantifying the produced reducing sugar by DNS method and cellobiose and glucose by HPLC. Values are average \pm standard deviation based on three independent replicates. * $p < 0.01$, $\Delta pyrF::CaBglA$ vs. $\Delta pyrF$

After the cell-cultivation stage, 100 g/L Avicel was supplemented to initiate the cellulose hydrolysis stage. *C. thermocellum* strains can assimilate the produced sugars [39], which may result in the decreased production of reducing sugars. As a strict anaerobe, the growth of *C. thermocellum* may cease with the presence of oxygen or low pH (pH value <6) [39]. Thus, we performed aerobic treatment (aeration) and acidic treatment (reduce the pH value to 5.5) to inhibit the cell growth as well as the assimilation process. Without any treatment, $\Delta pyrF::CaBglA$ produced 394 ± 34.7 mM reducing sugar including 381 ± 13.3 mM glucose in 20 days. Meanwhile, the parent strain $\Delta pyrF$ produced only 182 ± 8.7 mM reducing sugar including 173 ± 0.2 mM glucose. The results showed that $\Delta pyrF::CaBglA$ produced more reducing sugar than the parent strain under untreated as well as acidic conditions, but not under the aerobic condition (Fig. 4a). In addition, neither the aerobic nor acidic treatment was conducive to the production of reducing sugars by *C. thermocellum* strains, especially $\Delta pyrF::CaBglA$ (Fig. 4a). Under anaerobic (untreated or acidic) conditions, the parent strain $\Delta pyrF$ produced less glucose but accumulated more cellobiose than $\Delta pyrF::CaBglA$, indicating the expression of $CaBglA$ stimulated the conversion of cellobiose to glucose, thereby promoted the cellulose hydrolysis (Fig. 4b, c). However, $\Delta pyrF$ and $\Delta pyrF::CaBglA$ produced similar amounts of glucose and cellobiose under aerobic condition (Fig. 4b, c). This suggested that the presence of oxygen inhibited not only the activity of cellulosome but also the activity of $CaBglA$.

At the end of the cell-cultivation phase, the BGL activity in $\Delta pyrF::CaBglA$ broth was determined as 7.23 ± 0.2 U/g cellulose. Because there might be cell-bound CelS-bearing $CaBglA$ that did not release into the broth, we proposed higher BGL activity in $\Delta pyrF::CaBglA$ culture. To evaluate the contribution of expressed BGL to saccharification, a positive control was prepared by adding 15 U/g cellulose of purified $CaBglA$ protein in $\Delta pyrF$ cultures at the beginning of the saccharification phase, and 359 ± 48.5 mM reducing sugar was produced after 20-day saccharification, which was similar to the production of $\Delta pyrF::CaBglA$ (Fig. 4a). The relative levels of cellulose saccharification from $\Delta pyrF::CaBglA$ and the positive control ($\Delta pyrF$ supplemented with $CaBglA$) were 63.9 and 58.2%, respectively. These results indicated the CelS-bearing $CaBglA$ in $\Delta pyrF::CaBglA$ was active as the free $CaBglA$ protein. In addition, although the glucose production showed no significant difference at the end of saccharification process, cellobiose concentration in the culture of the positive control was slightly higher than that in $\Delta pyrF::CaBglA$ culture (Fig. 4b, c), indicating more effective removal of cellobiose by CelS-bearing



CaBglA in *C. thermocellum*. These results suggested that the fusion of *CaBglA* with *CelS* efficiently enhanced the cellulose saccharification. To investigate whether the production of reducing sugar could be further stimulated by increasing the activity of *CaBglA* in $\Delta pyrF::CaBglA$, 15 U/g cellulose of *CaBglA* were also supplemented in the culture of $\Delta pyrF::CaBglA$. However, we observed no obvious difference in the production of reducing sugars (Additional file 5).

The strain $\Delta pyrF::pHK-CaBglA-Doc$ producing plasmid-born *CaBglA-Doc* protein was also used for cellulose saccharification, since its cellulosome contained the BGL activity. However, no increase of reducing sugar was detected in this strain compared with the parent strain, which might be explained by the low expression

of *CaBglA*. In addition, the saccharification activity of the strain $\Delta pyrF::CaBglAm$ expressing a chromosome-born codon-modified *CaBglA* was compared with $\Delta pyrF::CaBglA$, and no significant change was observed, either (Additional file 6).

Improvement of the sugar production by *C. thermocellum* $\Delta pyrF::CaBglA$

Increased amount of whole-cell catalysts in the saccharification system might result in enhanced cellulose degradation and fermentable sugar production. Thus, we tried to stimulate the cell growth of *C. thermocellum* $\Delta pyrF::CaBglA$ by modifying the fermentation medium. Cellobiose and ammonium sulfate were supplemented as extra carbon and inorganic nitrogen source,

respectively [40]. Over 4.5-fold increase of the cell density ($OD_{600nm} = 5.5$) was observed when 20 g/L cellobiose and 1.3 g/L ammonium sulfate was used (Additional file 7). However, $\Delta pyrF::CaBglA$ only produced 241 ± 3.4 mM reducing sugar using such rich medium after 15-day saccharification, which was only 59% of that in the control setup using regular medium. This indicated that although *C. thermocellum* prefers cellobiose to cellulose and grows fast with increased carbon and nitrogen loading, the cellosome produced using cellobiose as the carbon source may not be suitable for cellulose degradation due to the substrate-coupled regulation mechanism [41].

To avoid the potential impact of substrate change on the composition and activity of the cellosome, concentrated $\Delta pyrF::CaBglA$ cells were inoculated in fresh medium, and the cell density was increased by 1.6- or 2.4-fold (the concentration of pellet protein increased from 0.26 to 0.46 or 0.63 mg/mL, respectively). The production of the reducing sugar was stimulated up to 490 ± 7.6 mM (Fig. 5), including 451 ± 5.7 mM (81.1 ± 1.0 g/L) glucose and 6.8 ± 0.4 mM (2.3 ± 0.1 g/L) cellobiose, and the saccharification increased to 79.4%. The pH values of the cultures were buffered and maintained at around 6 (Fig. 5). At the end of saccharification process, we observed complete degradation of the initial Avicel but no significant cell growth was detected (Additional file 8). Thus, the change of cell biomass would show slight influence on the carbon recovery. To investigate the carbon flux in the saccharification system in addition to fermentable sugars, we quantified potential intermediates and end metabolites including pyruvate, ethanol, lactate, acetate, and formate produced by $\Delta pyrF::CaBglA$. At the end of saccharification, $\Delta pyrF::CaBglA$ with 2.4-fold cell density produced 261 ± 10.3 mM of ethanol, while trace amount of other products were detected. The produced ethanol was equivalent to around 131 mM glucose. Together, 620.5 ± 9.2 mM products in glucose equivalent were produced from the initial Avicel (617 mM in glucose equivalent), and the carbon recovery was $100.6 \pm 1.6\%$.

Discussion

Clostridium thermocellum can be used as a potential whole-cell catalyst for cellulose saccharification because of its robust cellosome system [9], but proper engineering is indispensable to overcome its natural deficiencies, e.g. cellobiose inhibition, based on the industrial standard [5]. BGL supplementation is one of the most efficient strategies to relieve the cellobiose inhibition to cellosome by previous in vitro studies [17]. Prawitwong et al. developed a biological saccharification system using a *C. thermocellum* culture supplemented with BGL to produce glucose [18], and the BGL recycling was considered by fusing BGL with a cellulose binding module [24, 25],

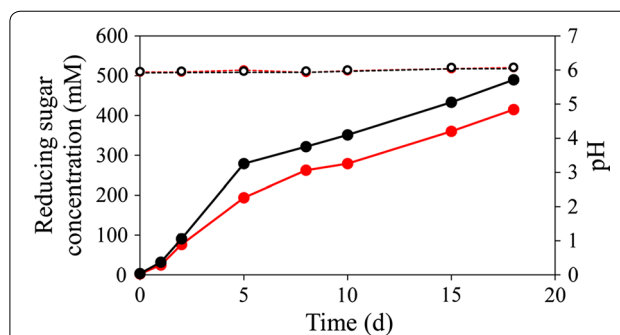


Fig. 5 Saccharification and pH curves of *C. thermocellum* strain $\Delta pyrF::CaBglA$ with increased cell density. Cells from 200- to 300-mL culture were concentrated, resuspended, and reinoculated into 100 mL fresh GS-2 medium with 100 g/L Avicel as a sole carbon source to increase the cell density by 1.6 (red) or 2.4 fold (black). The concentration of produced reducing sugars (closed circles) and the pH values (open circles) were determined based on three independent replicates, and values are shown as average \pm standard deviation

but the additional production and supplementation of BGL complicated the process and may also increase the saccharification cost. In addition, the decreasing stability and the glucose inhibition to BGL in company with the saccharification process might require relatively high or successive load of BGL proteins, which may further increase the cost. In this study, we employed the whole-cell catalyzing strategy for lignocellulose saccharification without supplementation of extra enzymes. This strategy fits to the industry requirement of low-cost, simple process, and high efficiency, but needs an efficient whole-cell catalyst producing secretory BGL.

It is known that CelS plays a key role in cellulose hydrolysis by cellosome and is the main producer of cellobiose [28, 42], thus the fusion of selected BGL with CelS would greatly release the feedback inhibition effect on cellosome. By developing a novel ACE-based seamless genome editing system in the thermophilic bacterium, we successfully inserted a BGL gene *caBglA* into the CelS-encoding sequence of *C. thermocellum* DSM 1313 precisely between its catalyzing module and dockerin module. A recombinant strain $\Delta pyrF::CaBglA$ was finally constructed producing a CelS-bearing *CaBglA* protein as an active cellosomal component. Nevertheless, we observed decreased expression of the fused protein Cel-*CaBglA*-Doc in $\Delta pyrF::CaBglA$ compared to that of the wild-type CelS in $\Delta pyrF$, which might influence the efficiency of cellulose degradation and sugar production. We thought the unsuitable codon usage of *caBglA* gene might be the problem, and optimized the codon of *caBglA* gene, but the result suggested that it was not the case. The fusion protein Cel-*CaBglA*-Doc was expressed under the control of the strong promoter

of CelS in $\Delta pyrF::CaBglA$, which might result in inherent resource shortage and therefore induce down-regulation of the expression level. Further studies are needed to reveal the reason of the low expression of the fused protein. Difference strategies to introduce BGL into cellulosome, such as fusion with other cellulosomal component, or fusion at different position of CelS, are also worth trying in future.

In addition to cellulosomal integration, we also achieved the plasmid-dependent expression of *CaBglA* in *C. thermocellum* using a multicopy-replicating plasmid pHK. Although $\Delta pyrF::pHK-CaBglA-Doc$ theoretically contained more *caBglA* gene copies than $\Delta pyrF::CaBglA$, and both employed the CelS promoter to drive the transcription, the cellulosome of $\Delta pyrF::pHK-CaBglA-Doc$ showed 16-fold lower BGL activity (1.15 ± 0.07 U/mg) than that of $\Delta pyrF::CaBglA$ (19.1 ± 1.2 U/mg). This result suggested episomal plasmid DNA might not be the proper vector for protein expression, and chromosomal integration might result in higher level of protein expression in some cases [43], and the CelS-bearing expression pattern might play key roles during the protein secretion and/or assembly process to support the functional expression of the heterologous BGL. This phenomenon should be considered when expressing other exogenous proteins in *C. thermocellum*.

Because the assimilation of celldextrins by *C. thermocellum* may result in the decreased production of reducing sugars [39], we set up aerobic or low-pH treatments to inhibit the cell metabolism at the cellulose hydrolysis stage. However, neither aeration nor pH reduction resulted in increased production of reducing sugars. In contrast, lower sugar concentrations were detected compared with the untreated control. The negative effect of aeration on the cellulose hydrolysis might be explained by the oxygen sensitivity of the cellulosome [44, 45], and *CaBglA* might also prefer anaerobic condition. Although pH 5.5 was determined as the optimal pH of both cellulosome [45, 46] and *CaBglA* [23] (Additional file 1), low-pH condition decreased the production of reducing sugars in this study. We also detected decreased pellet protein and extracellular protein concentrations when the pH value of $\Delta pyrF::CaBglA$ was decreased to 5.5 at the beginning of the cellulose hydrolysis stage (Additional file 8). These results indicated that the low-pH condition might inhibit the cell growth as well as the continuous production of cellulosomal proteins by cells, and cell lysis might also occur under acidic conditions [39]. In addition, there might be other extracellular protein components that contributed to the cellulose hydrolysis but were not in favor of acidic conditions. Thus, the control of pH value should be considered in further application of *C. thermocellum* for cellulose conversion.

Although with addition of 15 U/g cellulose of purified *CaBglA* protein, the parent strain $\Delta pyrF$ showed similar saccharification activity to that of $\Delta pyrF::CaBglA$, it accumulated more cellobiose during hydrolysis. This result indicated CelS-bearing *CaBglA* expressed by *C. thermocellum* was more effective in cellobiose conversion than supplemented free protein. In contrast to the parent strain $\Delta pyrF$, further supplementation of purified *CaBglA* in the culture of $\Delta pyrF::CaBglA$ showed no apparent effect on sugar production, suggesting that the fused expression of Cel-*CaBglA*-Doc led to the full occupation of CelS by *CaBglA*. The formed substrate-coupled catalyzing channel had relieved most of the cellobiose inhibition effect on the whole cellulosome system. Under our whole-cell-catalyzing conditions, the carbon recovery of supplemented cellulose was about 100%, i.e., 79% of supplemented carbon was converted to soluble sugars and 21% to the end product ethanol. The productivity of fermentable sugars might be further enhanced by disrupting the metabolic pathways leading to end products, especially ethanol, in the recombinant strain $\Delta pyrF::CaBglA$.

Conclusion

Lignocellulosic biomass is an attractive feedstock to substitute fossil resources, but is difficult to deconstruct. *C. thermocellum* naturally degrades lignocellulose efficiently but cannot be directly used in industry due largely to the feedback inhibition to its enzymatic system cellulosome. The supplementation of BGL could solve the problem. However, the addition of enzymes produced elsewhere complicates the process and hinders the utilization of lignocellulose in industry. Here, we constructed an efficient whole-cell catalyst producing BGL for cellulose saccharification by targeted engineering of *C. thermocellum*. Without supplementation of any other enzymes, the whole-cell catalyst showed the high cellulose-saccharification activity and sugar production. Hence, our work confirmed the feasibility of the whole-cell-catalysis strategy for cellulose saccharification, and provided a potential whole-cell catalyst for industrial cellulose saccharification.

Methods

Bacterial strains and cultivation

The bacterial strains used in this study are listed in Table 2. *Escherichia coli* strains were cultivated aerobically at 37 °C in Luria–Bertani (LB) liquid medium with shaking at 160 rpm or on solid LB plate with 1.5% agar. *C. thermocellum* strains were grown anaerobically at 55 °C in modified GS-2 [47] or MJ medium [48] with 5 g/L cellobiose, 5 to 100 g/L Avicel (PH-101, Sigma) unless otherwise stated. 30 µg/mL chloramphenicol, 50 µg/mL kanamycin and 3 µg/mL thiamphenicol (Tm) were

Table 2 Bacterial strains and plasmids used in this study

Strains/plasmids	Relevant characteristic	Sources
Strains		
<i>E. coli</i>		
DH5α	<i>f80dlacZΔM15, Δ(lacZYA-argF)U169, deoR, recA1, endA1, hsdR17(r_k⁻, m_k⁺), phoA, supE44, I⁻, thi-1, gyrA96, relA1</i>	Transgen
BL21(DE3)	<i>ompT gal dcm lon hsdS_B(r_B⁻ m_B⁻) I (DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])</i>	Transgen
<i>C. thermocellum</i>		
DSM1313	LQ8, wild type stain	DSMZ
Δ <i>pyrF</i>	Derived from DSM1313, with deleted <i>pyrF</i> gene	This work
Δ <i>pyrF</i> :: <i>CaBglA</i>	Derived from Δ <i>pyrF</i> , expressing CelS-bearing <i>CaBglA</i>	This work
Δ <i>pyrF</i> :: <i>CaBglAm</i>	Derived from Δ <i>pyrF</i> , expressing CelS-bearing <i>CaBglAm</i>	This work
Δ <i>pyrF</i> ::pHK-Ct <i>BglA</i> -Doc	Derived from Δ <i>pyrF</i> , containing plasmid pHK-Ct <i>BglA</i> -Doc	This work
Δ <i>pyrF</i> ::pHK- <i>CaBglA</i> -Doc	Derived from Δ <i>pyrF</i> , containing plasmid pHK- <i>CaBglA</i> -Doc	This work
Plasmids		
pHK	pNW33N derivative, <i>E. coli</i> - <i>C. thermocellum</i> shuttle vector, Cm ^R /Tm ^R	[51]
pHK-Δ <i>pyrF</i>	pHK derivative, containing upstream and downstream regions of <i>pyrF</i> for <i>pyrF</i> deletion	This work
pHK-HR	pHK derivative, containing <i>pyrF</i> cassette, <i>tdk</i> cassette, three regions of homology, and the target DNA sequence to knock in	This work
pHK-HR- <i>CaBglA</i>	pHK-HR derivative for markerless knock-in of <i>caBglA</i> in the chromosome of DSM 1313	This work
pHK-HR-Ct <i>BglA</i>	pHK-HR derivative for markerless knock-in of <i>ctBglA</i> in the chromosome of DSM 1313	This work
pHK-HR- <i>CaBglAm</i>	pHK-HR derivative for markerless knock-in of <i>caBglAm</i> in the chromosome of DSM 1313	This work
pHK-Ct <i>BglA</i> -Doc	pHK-HR-Ct <i>BglA</i> derivative for expression of Ct <i>BglA</i> -Doc in DSM 1313, with the promoter and signal peptide region of CelS	This work
pHK- <i>CaBglA</i> -Doc	pHK-HR- <i>CaBglA</i> derivative for expression of <i>CaBglA</i> -Doc in DSM 1313, with the promoter and signal peptide region of CelS	This work
pET28aNS	Expression vector with N-terminal hexahistidine affinity tag, with modified multiple cloning sites	[49]
pET28aNS- <i>CaBglA</i>	pET28aNS derivative for expression of <i>CaBglA</i>	This work
pET28aNS-Ct <i>BglA</i>	pET28aNS derivative for expression of Ct <i>BglA</i>	This work
pET28aNS-CgIT	pET28aNS derivative for expression of CgIT	This work
pET28aNS-Td2f2	pET28aNS derivative for expression of Td2f2	This work

supplemented to the medium when necessary. 10 μg/mL 5-fluoro-2-deoxyuridine (FUDR) or 500 μg/mL 5-fluoroorotic acid (FOA) dissolved in dimethyl sulfoxide were added for screening.

Construction of plasmids

Four recombinant plasmids, pET28aNS-*CaBglA*, pET28aNS-Ct*BglA*, pET28aNS-CgIT, and pET28aNS-Td2f2 (Table 2), were constructed for BGL expression in *E. coli* by cloning the genes *caBglA* (GenBank accession number: JX030398.1), *ctBglA* (GenBank accession number: WP_003520797.1), *cglT* (GenBank accession number: Z56279.1), and *td2f2* (GenBank accession number: HV538882.1) into the vector pET28aNS [49], respectively. The genes *caBglA*, *ctBglA*, and *cglT* were obtained by PCR with the genome DNAs of *Caldicellulosiruptor* sp. F32, *C. thermocellum* DSM1313 and *T. brockii* DSM1457 as the templates using primers *CaBglA*-F/R, *CtBglA*-F/R and *CgIT*-F/R (Additional file 9),

respectively. The gene *td2f2* was synthesized (Sangon, Beijing, China) with optimized encoding sequence to adapt the codon usage of *E. coli* [50]. Codon-modified *caBglA* gene, designated as *caBglAm* (GenBank accession number: KY418041), was synthesized based on the codon usage bias of *C. thermocellum* (Sangon, Beijing, China).

All plasmids constructed for the genetic manipulation in *C. thermocellum* were derived from pHK (GenBank accession number: KY792637) [51]. The construct the plasmid pHK-Δ*pyrF*, ~1-kb upstream and downstream homologous arms of *pyrF* were amplified from the genome DNA of *C. thermocellum* DSM1313 using primer sets PyrF5'-F/R and PyrF3'-F/R (Additional file 9), respectively, and were ligated together by overlap PCR using primers PyrF5'-F/PyrF3'-R. The obtained PCR product was cloned to pHK vector through *Pst*I and *Nhe*I restriction sites to generate pHK-Δ*pyrF* (Table 2).

The seamless editing plasmid pHK-HR contains two selection gene markers and three regions of homology

(Additional file 2). The endogenous selection marker *pyrF* (Clo1313_1266) was driven by its native promoter [37]. The selection marker *tdk* (Teth514_0091) was obtained from *Thermoanaerobacter* sp. X514 [36], and was expressed under the control of a glyceraldehyde-3-phosphate dehydrogenase (*gapDH*) promoter from *C. thermocellum* DSM1313. Three regions of homology, HR-up, HR-short, and HR-down, were amplified from the genome DNA of *C. thermocellum* DSM1313 according to the genome editing demand, in which the sequence of HR-short was the same with the 3' region of HR-up. BGL-encoding genes *ctBglA* and *caBglA* were chosen as the candidate DNA sequences to knock in, and were amplified from the genome DNAs of *C. thermocellum* DSM1313 and *Caldicellulosiruptor* sp. F32, respectively. The termination codons of the BGL-encoding genes were eliminated for fused protein expression. All fragments were cloned into the pHK vector sequentially. The Tdk expression cassette with the *gapDH* promoter was first cloned into pHK using *NheI* and *XbaI* sites. Subsequently, a fragment containing the upstream homology HR-up, the *PyrF* cassette as well as the short homology HR-short was obtained by overlap PCR and cloned into the plasmid using *XbaI* and *EagI* sites. The downstream homology HR-Down was then ligated into the plasmid using *MluI* and *BamHI* sites. We determined to integrate BGL sequences between the catalyzing module and dockerin module of CelS on the chromosome. To maintain the individual functions of the adjacent modules, repeated GGT sequence encoding glycine residuals were introduced in the upstream (3' end of HR-short) and downstream (5' end of HR-down) of the target DNA sequence. Finally, *ctBglA*, *caBglA* or *caBglAm* gene sequence was cloned into the plasmid via *MluI* and *EagI* sites, resulting in the plasmids pHK-HR-*CtBglA*, pHK-HR-*CaBglA* and pHK-HR-*CaBglAm* (Table 2), respectively, for seamless genome editing.

For the plasmid-dependent expressions of *CtBglA* and *CaBglA* in *C. thermocellum* DSM1313, the predicted promoter and signal peptide region of CelS in *C. thermocellum* DSM1313 was amplified with primer set Pcs-sigF/R (Additional file 9), and cloned into pHK-HR-*CtBglA* and pHK-HR-*CaBglA* using *NheI* and *EagI* sites to construct pHK-*CtBglA*-Doc and pHK-*CaBglA*-Doc (Table 2), respectively. The Tdk cassette, HR-Up region, *PyrF* cassette, and HR-short region of the plasmids were replaced by the promoter and signal peptide of CelS. Thus, the BGL-encoding genes would be driven by the CelS promoter, and the expressed proteins would contain the signal peptide of CelS for protein secretion. Because the HR-Down region contained the encoding sequence of the dockerin module of CelS, the expressed BGLs would bear the dockerin module for cellulosome assembly.

Heterologous expression and purification of beta-glucosidases in *E. coli*

The plasmids pET28aNS-*CaBglA*, pET28aNS-*CtBglA*, pET28aNS-CglT, and pET28aNS-Td2f2 were transformed into *E. coli* BL21(DE3) for heterologous expression of *CaBglA*, *CtBglA*, CglT, and Td2f2, respectively. Synthesis of recombinant proteins in *E. coli* BL21(DE3) cells was initiated by the addition of 1 mM IPTG, and cultivation was continued for an additional 16 h at 16 °C. Cells were harvested by centrifugation at 10,000 rpm, resuspended in 50 mM Tris-HCl buffer containing 30 mM imidazole and 300 mM NaCl, pH 8.0, and lysed by ultrasonication. The supernatants were applied onto a HisTrap™ HP Ni-affinity column (GE Healthcare). The proteins were eluted with 50 mM Tris-HCl buffer containing 500 mM imidazole and 300 mM NaCl, pH 8.0. The eluted fractions were then concentrated to 2 mL using Amicon Ultra-15 centrifugal filter units (10.0 kDa cutoff) (Merck Millipore, Billerica, MA, USA), and applied onto a Superdex 75 gel filtration column (GE Healthcare) with 50 mM K₂HPO₄-KH₂PO₄ buffer with 100 mM KCl, pH 6.0.

Preparation of cellulosomal and extracellular proteins

Clostridium thermocellum strains were cultivated in GS-2 medium with 5 g/L Avicel as the sole carbon source at 55 °C for 48 h. 200-mL cultures were centrifuged at 3000g for 30 min, the cell pellets were washed twice and resuspended in 4 mL 50 mM Tris-HCl buffer containing 5 mM DTT, pH 7.0, and lysed using a high-pressure homogenizer (Constant Systems LTD). The extracellular proteins were prepared by condensing 10 mL of the culture supernatants to 0.5 mL using Amicon Ultra-15 centrifugal filter units (10.0 kDa cutoff) (Merck Millipore, Billerica, MA, USA). The rest of the culture supernatants were used for cellulosome extraction according to a modified cellulose affinity procedure [10].

Protein analyses

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed to check the protein purity and composition as previously described [10]. The molecular weight of the protein was estimated according to the relative mobility of protein ladders (10–245 kDa, New England BioLabs). The Bradford method was used for protein quantification [52]. The mass spectroscopy analyses were achieved according to a published procedure [10]. All protein samples were stored at -80 °C for further analyses.

Enzyme assay

The BGL activity was determined against *p*-nitrophenyl-β-D-glucopyranoside (*p*NPG). Samples were incubated

in 200- μ L reaction buffer (50 mM sodium acetate, 1 mM *p*NPG, pH 5.5) at 55 °C for 5 or 10 min. The reaction was terminated by adding 1 mL of 1 M Na_2CO_3 , and the absorbance of the mixture was measured at 405 nm immediately. One unit of enzyme activity was defined as the amount of enzyme required to produce 1 μ mol of *p*-nitrophenol (*p*NP) per min under certain conditions.

The cellulase activity was tested in 1-mL final reaction volume containing 50 μ g cellulosome proteins and 15 mg Avicel or 7.5 mg cellobiose as a substrate. The reaction buffer contained 50 mM sodium acetate, 10 mM CaCl_2 and 5 mM dithiothreitol (DTT), pH 5.5. The abundance of reducing sugars was determined by the 3,5-dinitrosalicylic acid (DNS) method after incubation at 55 °C for 24 h, and the glucose and cellobiose in the hydrolysates were quantified by high-performance liquid chromatography (HPLC) as previously described [53].

Electrotransformation and screening of *C. thermocellum*

The pHK derivative plasmids were transformed into *E. coli* BL21(DE3) to remove Dcm methylation, and transformed to *C. thermocellum* DSM1313 according to published protocols [10, 51]. In brief, 200 μ L of *C. thermocellum* competent cells was added to 0.2-cm electroporation cuvettes (BioRad) with 10 μ L of DNA (~2000 ng) in sterile distilled water. A series of 40 square pulses were applied, each with an amplitude of 1.5 kV and for a duration of 50 s at 500-ms intervals. Cells were then recovered for 24 h at 51 °C in 4 mL of fresh GS-2 medium before screening on solid medium containing Tm. The Δ *pyrF* mutant was selected as described [37]. In detail, the transformants containing pHK- Δ *pyrF* were cultivated with the presence of Tm and then plated in GS-2 solid medium with 500 μ g/mL FOA. FOA-resistant colonies were screened by colony PCR using *PyrF*-F/R flanking the genomic gene *pyrF*. The colonies with 0.25-Kb PCR products were determined as *pyrF*-deleted mutants.

The mutant Δ *pyrF* was then used as the parent to construct other *C. thermocellum* recombinant strains. The Δ *pyrF* transformants containing pHK-derived plasmids were initially selected on solid GS-2 medium supplemented with Tm, and then inoculated into liquid GS-2 medium with Tm to strengthen the replication of the transformed plasmid before further screening. Transformants containing plasmid pHK-*CtBglA*-Doc or pHK-*CaBglA*-Doc were verified by detecting the *CtBglA*-Doc or *CaBglA*-Doc fragments via colony PCR and sequencing. Transformants containing pHK-HR-*CtBglA*, pHK-HR-*CaBglA* or pHK-HR-*CaBglAm* were further inoculated into MJ medium containing FUDR and cultivated until the late exponential phase. 200- μ L cultures were diluted for 1, 10, 100, and 1000 fold, plated in MJ solid medium containing FUDR, and then

cultivated at 51 °C for 5–7 days. The obtained colonies were screened by colony PCR using primer set HR-F/R (Additional file 9). Those colonies showing the PCR product of ~5.7 Kb indicated the first round of recombination was done via the long regions of homology. A band of 2.9 Kb might also be detected, indicating the mixing of the parent strain Δ *pyrF* (Additional file 10). If so, the FUDR screening in MJ medium should be repeated until single band of ~5.7 Kb was detected. The verified colonies were inoculated into GS-2 liquid medium with Tm to confirm the plasmid curing if no growth was observed, and further cultivated in GS-2 medium containing FOA to the late exponential phase, diluted, and plated in solid medium with FOA to select colonies without *PyrF* function. Colony PCRs were subsequently performed using primer set HR-F/R to confirm *pyrF* elimination. The colonies with PCR product of 4.7 Kb were finally verified as the target strain after sequencing.

Fermentation of *C. thermocellum* strains

Fermentation of *C. thermocellum* strains were in 250-mL anaerobic bottles containing 100-mL cultures using 10 g/L Avicel as the sole carbon source. Three independent fermentations were set up for each strain, and 1.5-mL cultures were sampled every 8 to 12 h with a 2.5-mL syringe. 1 mL of each sample was used to determine the amount of residual cellulose by a modified saccharification method [10], and 0.5 mL of each sample was centrifuged for pellet cells to determine the abundance of total cellular protein [10].

Cellulose saccharification

100-mL fermentations of *C. thermocellum* strains were initially performed with 5 g/L Avicel as the sole carbon source for 36 h. For aerobic treatment, the cultures were transferred into 250-mL sterile flasks shaking at 170 rpm aerobically. For acidic treatment, the pH value of the broths was adjusted to 5.5 by adding 1 N HCl in an anaerobic chamber. For BGL treatment, 15 U/g cellulose of *CaBglA* protein was added at the beginning of the saccharification process. For high-intensity treatment, cells from 200- to 300-mL culture were concentrated, resuspended, and reinoculated into 100-mL fresh GS-2 medium anaerobically. Untreated controls were prepared under consistent conditions. For all setups, 100 g/L Avicel was supplemented to initiate the cellulose-saccharification stage. The saccharification process lasted for 15–20 days, and 1-mL sample was taken from each setup with a 5-mL syringe per 1 to 5 days to determine the abundance of produced reducing sugar by the DNS method, and the concentrations of sugars (cellobiose and glucose) and other metabolites (pyruvate, ethanol, lactate, acetate and formate) by HPLC. The relative

saccharification level was determined subsequently by dividing the initial Avicel (617 mM glucose equivalents) with the amounts of the obtained reducing sugar (mM). Three independent experiments were prepared for every strain under each condition.

Additional files

Additional file 1: Figure S1. The optimal temperature (a), pH value (b), and glucose inhibition (c) of *CaBglA* and *CtBglA*. The optimal temperature was determined by incubate the reaction mixture in 50 to 80 °C water bath for 10 min. The pH value of the reaction buffer was adjusted from 4.0 to 8.0 to determine the optimal pH. The glucose inhibition was determined by adding 0 to 400 mM glucose to the standard reaction mixture, and calculated as the glucose concentration required to inhibit 50% of initial activity. Values are average \pm standard deviation based on three independent replicates.

Additional file 2: Figure S2. Map of plasmid pHK-HR used for seamless genome editing in *C. thermocellum*. The plasmid is derived from a *E. coli/C. thermocellum* shuttle vector pHK. To construct pHK-HR plasmids, the Tdk expression cassette (gapDH-F/tdk-R), the fragment containing the upstream homology HR-up (HR-up-F/R), the PyrF expression cassette (HR-pyrF-F/R), the short homology HR-short (HR-short-F/R), the downstream homology HR-Down (HR-down-F/R), and the BGL genes are ligated into the plasmid substantially.

Additional file 3: Figure S3. Identification of fusion protein Cel-BGL-Doc peptides in the parent and Δ pyrF::CaBglA strains by mass spectroscopy analysis. Cellulosomal and extracellular proteins with the size of ~135 kDa or ~75 kDa were investigated. The green highlights indicate peptides detected by mass spectroscopy. The amino acid sequences shown in black, purple, and blue belong to Cel (the catalyzing module of CelS), BGL (*CaBglA*), and Doc (the assembling module of CelS), respectively. The linker sequences are shown in bold. The insertion sites of the BGL sequence are indicated by red triangles. *CaBglA* sequences are detected in ~135-kDa but not ~75-kDa cellulosomal and extracellular proteins of Δ pyrF::CaBglA, indicating the successful expression, secretion, and cellulosomal assembly of the protein Cel-CaBglA-Doc in Δ pyrF::CaBglA.

Additional file 4: Figure S4. Growth and fermentation analysis of *C. thermocellum* strains with Avicel as a carbon source. a, cell growth represented by the abundance of total protein in cell pellets. b, Avicel consumption in mM glucose equivalents. Average values and standard deviations are calculated based on three replicates for each strain.

Additional file 5: Figure S5. Production of reducing sugars by Δ pyrF::CaBglA. The concentration of produced reducing sugar was determined by DNS method. 0 or 15 U/g cellulose of *CaBglA* were added. Three independent replicates were prepared to calculate the average values and standard deviations.

Additional file 6: Figure S6. Production of reducing sugars by Δ pyrF::CaBglA and Δ pyrF::CaBglAm. The concentration of reducing sugar was determined by DNS method. Three independent replicates were prepared to calculate the average values and standard deviations.

Additional file 7: Figure S7. Cell growth of *C. thermocellum* Δ pyrF::CaBglA in GS-2 media containing different carbon and nitrogen sources. 10 or 20 g/L cellobiose and 1.3 g/L ammonium sulfate were supplemented when required. Regular GS-2 medium was used as control. Values are average \pm standard deviation based on three independent replicates.

Additional file 8: Figure S8. Production of pellet and extracellular proteins by Δ pyrF::CaBglA with (pH 5.5) or without (Untreated) pH value adjustment at the cellulose hydrolysis stage. No dramatic change was detected during the whole saccharification process. But lower amounts of pellet and extracellular proteins were produced under pH 5.5 condition compared to the untreated control, indicating the blocked cell growth

and the reduced expression of extracellular protein, including cellulosomal proteins. Two replicates were used for mean value calculation.

Additional file 9: Table S1. List of primers used in this study.

Additional file 10: Figure S9. Colony PCR screening of the *C. thermocellum* mutants after the first round of homologous recombination. Transformants of Δ pyrF::pHK-HR-CaCglA grown on MJ solid medium with addition of FUDR were investigated using primer set HR-F/R. PCR product of ~5.7 Kb indicates the success of the first round of recombination, and a 2.9-Kb band refers to the parent strain Δ pyrF without genomic integration. Black arrows indicate the colonies showing double bands. The colonies contain both the parent strain Δ pyrF and the recombinant strain. M, DNA standards.

Authors' contributions

JZ, YF, QC, and YJL designed the research; JZ, SL, RL, WH, YX, and YJL performed the experiments. JZ, YF, and YJL analyzed the data and wrote the paper. All authors read and approved the final manuscript.

Author details

¹ Shandong Provincial Key Laboratory of Energy Genetics, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao 266101, People's Republic of China. ² CAS Key Laboratory of Biofuels, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao 266101, People's Republic of China. ³ Qingdao Engineering Laboratory of Single Cell Oil, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao 266101, People's Republic of China. ⁴ University of Chinese Academy of Sciences, Chinese Academy of Sciences, Beijing 100049, People's Republic of China. ⁵ Key Laboratory of Endemic and Ethnic Diseases (Guizhou Medical University), Ministry of Education, Guiyang 550000, People's Republic of China.

Acknowledgements

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its Additional files.

Funding

This work was supported by the National Natural Science Foundation of China (Grant Numbers 31300035, 31470210 and 31560318), the National Key Technology Research and Development Program of China (Grant Number 2015BAD15B05), and the Shandong Province Key Laboratory Union of Carbohydrate Industry Science and Technology, partially supported by the National Natural Science Foundation of China (Grant Numbers 31570029, and 31570787).

Publisher's Note

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

Received: 17 January 2017 Accepted: 19 April 2017

Published online: 12 May 2017

References

1. Demain AL. Biosolutions to the energy problem. *J Ind Microbiol Biotechnol.* 2009;36:319–32.
2. Antoni D, Zverlov VV, Schwarz WH. Biofuels from microbes. *Appl Microbiol Biotechnol.* 2007;77:23–35.
3. Gronenberg LS, Marcheschi RJ, Liao JC. Next generation biofuel engineering in prokaryotes. *Curr Opin Chem Biol.* 2013;17:462–71.

4. Hasunuma T, Okazaki F, Okai N, Hara KY, Ishii J, Kondo A. A review of enzymes and microbes for lignocellulosic biorefinery and the possibility of their application to consolidated bioprocessing technology. *Bioresour Technol.* 2013;135:513–22.
5. Olson DG, McBride JE, Shaw AJ, Lynd LR. Recent progress in consolidated bioprocessing. *Curr Opin Biotechnol.* 2012;23:396–405.
6. Gowen CM, Fong SS. Exploring biodiversity for cellulosic biofuel production. *Chem Biodivers.* 2010;7:1086–97.
7. Lynd LR, van Zyl WH, McBride JE, Laser M. Consolidated bioprocessing of cellulosic biomass: an update. *Curr Opin Biotechnol.* 2005;16:577–83.
8. Bayer EA, Belaich JP, Shoham Y, Lamed R. The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. *Annu Rev Microbiol.* 2004;58:521–54.
9. Fontes CM, Gilbert HJ. Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. *Annu Rev Biochem.* 2010;79:655–81.
10. Hong W, Zhang J, Feng Y, Mohr G, Lambowitz AM, Cui G-Z, Liu Y-J, Cui Q. The contribution of cellulosomal scaffoldings to cellulose hydrolysis by *Clostridium thermocellum* analyzed by using thermotargetrons. *Biotechnol Biofuels.* 2014;7:80.
11. Xu Q, Resch MG, Podkaminer K, Yang S, Baker JO, Donohoe BS, Wilson C, Klingeman DM, Olson DG, Decker SR, et al. Dramatic performance of *Clostridium thermocellum* explained by its wide range of cellulase modalities. *Sci Adv.* 2016;2:e1501254.
12. Schwarz WH. The cellulosome and cellulose degradation by anaerobic bacteria. *Appl Microbiol Biotechnol.* 2001;56:634–49.
13. Smith SP, Bayer EA. Insights into cellulosome assembly and dynamics: from dissection to reconstruction of the supramolecular enzyme complex. *Curr Opin Struct Biol.* 2013;23:686–94.
14. Paye JM, Guseva A, Hammer SK, Gjersing E, Davis MF, Davison BH, Olstad J, Donohoe BS, Nguyen TY, Wyman CE, et al. Biological lignocellulose solubilization: comparative evaluation of biocatalysts and enhancement via cotreatment. *Biotechnol Biofuels.* 2016;9:8.
15. Lamed R, Kenig R, Setter E, Bayer EA. Major characteristics of the cellulolytic system of *Clostridium thermocellum* coincide with those of the purified cellulosome. *Enzyme Microbiol Technol.* 1985;7:37–41.
16. Tian L, Papanek B, Olson DG, Rydzak T, Holwerda EK, Zheng T, Zhou J, Maloney M, Jiang N, Giannone RJ, et al. Simultaneous achievement of high ethanol yield and titer in *Clostridium thermocellum*. *Biotechnol Biofuels.* 2016;9:116.
17. Gefen G, Anbar M, Morag E, Lamed R, Bayer EA. Enhanced cellulose degradation by targeted integration of a cohesin-fused beta-glucosidase into the *Clostridium thermocellum* cellulosome. *Proc Natl Acad Sci USA.* 2012;109:10298–303.
18. Prawitwong P, Waeonukul R, Tachaapaikoon C, Pason P, Ratanakhanokchai K, Deng L, Sermsathanaswadi J, Septiningrum K, Mori Y, Kosugi A. Direct glucose production from lignocellulose using *Clostridium thermocellum* cultures supplemented with a thermostable beta-glucosidase. *Biotechnol Biofuels.* 2013;6:184.
19. Jabbour D, Klippel B, Antranikian G. A novel thermostable and glucose-tolerant beta-glucosidase from *Fervidobacterium islandicum*. *Appl Microbiol Biotechnol.* 2012;93:1947–56.
20. Pei J, Pang Q, Zhao L, Fan S, Shi H. *Thermoanaerobacterium thermosaccharolyticum* beta-glucosidase: a glucose-tolerant enzyme with high specific activity for cellobiose. *Biotechnol Biofuels.* 2012;5:31.
21. Teugjas H, Valjamae P. Selecting beta-glucosidases to support cellulases in cellulose saccharification. *Biotechnol Biofuels.* 2013;6:105.
22. Yi ZL, Zhang SB, Pei XQ, Wu ZL. Design of mutants for enhanced thermostability of beta-glycosidase BglY from *Thermus thermophilus*. *Bioresour Technol.* 2013;129:629–33.
23. Meng DD, Ying Y, Zhang KD, Lu M, Li FL. Depiction of carbohydrate-active enzyme diversity in *Caldicellulosiruptor* sp. F32 at the genome level reveals insights into distinct polysaccharide degradation features. *Mol bioSystems.* 2015;11:3164–73.
24. Waeonukul R, Kosugi A, Prawitwong P, Deng L, Tachaapaikoon C, Pason P, Ratanakhanokchai K, Saito M, Mori Y. Novel cellulase recycling method using a combination of *Clostridium thermocellum* cellulosomes and *Thermoanaerobacter brockii* beta-glucosidase. *Bioresour Technol.* 2012;130:424–30.
25. Waeonukul R, Kosugi A, Tachaapaikoon C, Pason P, Ratanakhanokchai K, Prawitwong P, Deng L, Saito M, Mori Y. Efficient saccharification of ammonia soaked rice straw by combination of *Clostridium thermocellum* cellulosome and *Thermoanaerobacter brockii* beta-glucosidase. *Bioresour Technol.* 2012;107:352–7.
26. Wu JD, Orme-Johnson WH, Demain AL. Two components of an extracellular protein aggregate of *Clostridium thermocellum* together degrade crystalline cellulose. *Biochemistry.* 1988;27:1703–9.
27. Raman B, Pan C, Hurst GB, Rodriguez M Jr, McKeown CK, Lankford PK, Samatova NF, Mielenz JR. Impact of pretreated Switchgrass and biomass carbohydrates on *Clostridium thermocellum* ATCC 27405 cellulosome composition: a quantitative proteomic analysis. *PLoS ONE.* 2009;4:e5271.
28. Olson DG, Tripathi SA, Giannone RJ, Lo J, Caiazza NC, Hogsett DA, Hettich RL, Guss AM, Dubrovsky G, Lynd LR. Deletion of the Cel48S cellulase from *Clostridium thermocellum*. *P Natl Acad Sci USA.* 2010;107:17727–32.
29. Olson DG, Lynd LR. Transformation of *Clostridium thermocellum* by electroporation. *Methods Enzymol.* 2012;510:317–30.
30. Breves R, Bronnenmeier K, Wild N, Lottspeich F, Staudenbauer WL, Hofemeister J. Genes encoding two different beta-glucosidases of *Thermoanaerobacter brockii* are clustered in a common operon. *Appl Environ Microbiol.* 1997;63:3902–10.
31. Uchiyama T, Miyazaki K, Yaoi K. Characterization of a novel beta-glucosidase from a compost microbial metagenome with strong transglycosylation activity. *J Biol Chem.* 2013;288:18325–34.
32. Katayeva IA, Golovchenko NP, Chuvijskaya NA, Akimenko VK. *Clostridium thermocellum* beta-glucosidases A and B: purification, properties, localization, and regulation of biosynthesis. *Enzyme Microbiol Technol.* 1992;14:407–12.
33. Olson DG, Giannone RJ, Hettich RL, Lynd LR. Role of the CipA scaffoldin protein in cellulose solubilization, as determined by targeted gene deletion and complementation in *Clostridium thermocellum*. *J Bacteriol.* 2013;195:733–9.
34. Olson DG, Maloney M, Lanahan AA, Hon S, Hauser LJ, Lynd LR. Identifying promoters for gene expression in *Clostridium thermocellum*. *Metab Eng Comm.* 2015;2:23–9.
35. Heap JT, Ehsaan M, Cooksley CM, Ng YK, Cartman ST, Winzer K, Minton NP. Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker. *Nucleic Acids Res.* 2012;40:e59.
36. Argyros DA, Tripathi SA, Barrett TF, Rogers SR, Feinberg LF, Olson DG, Foden JM, Miller BB, Lynd LR, Hogsett DA, Caiazza NC. High ethanol titers from cellulose using metabolically engineered thermophilic, anaerobic microbes. *Appl Environ Microbiol.* 2011;77:8288–94.
37. Tripathi SA, Olson DG, Argyros DA, Miller BB, Barrett TF, Murphy DM, McCool JD, Warner AK, Rajgarhia VB, Lynd LR, et al. Development of *pyrF*-based genetic system for targeted gene deletion in *Clostridium thermocellum* and creation of a pta mutant. *Appl Environ Microbiol.* 2010;76:6591–9.
38. Johnson EA, Reese ET, Demain AL. Inhibition of *Clostridium thermocellum* cellulase by end products of cellulolysis. *J Appl Biochem.* 1982;4:64–71.
39. Freier D, Mothershed CP, Wiegel J. Characterization of *Clostridium thermocellum* Jw20. *Appl Environ Microbiol.* 1988;54:204–11.
40. Garcia-Martinez D, Shinmyo A, Madia A, Demain A. Studies on cellulase production by *Clostridium thermocellum*. *Eur J Appl Microbiol Biotechnol.* 1980;9:189–97.
41. Nataf Y, Bahari L, Kahel-Raifer H, Borovok I, Lamed R, Bayer EA, Sonenshein AL, Shoham Y. *Clostridium thermocellum* cellulosomal genes are regulated by extracytoplasmic polysaccharides via alternative sigma factors. *Proc Natl Acad Sci USA.* 2010;107:18646–51.
42. Kruus K, Andreacchi A, Wang WK, Wu JH. Product inhibition of the recombinant CelS, an exoglucanase component of the *Clostridium thermocellum* cellulosome. *Appl Microbiol Biotechnol.* 1995;44:399–404.
43. Englaender JA, Jones JA, Cress BF, Kuhlman TE, Linhardt RJ, Koffas MA. Effect of genomic integration location on heterologous protein expression and metabolic engineering in *E. coli*. *ACS Synth Biol.* 2017;6:710–20.
44. Johnson EA, Demain AL. Probable involvement of sulfhydryl groups and metal as essential components of the cellulase of *Clostridium thermocellum*. *Arch Microbiol.* 1984;137:135–8.
45. Resch MG, Donohoe BS, Baker JO, Decker SR, Bayer EA, Beckham GT, Himmel ME. Fungal cellulases and complexed cellulosomal enzymes exhibit synergistic mechanisms in cellulose deconstruction. *Energy Environ Sci.* 2013;6:1858–66.
46. Ng TK, Weimer PJ, Zeikus JG. Cellulolytic and physiological properties of *Clostridium thermocellum*. *Arch Microbiol.* 1977;114:1–7.

47. Cui GZ, Hong W, Zhang J, Li WL, Feng Y, Liu YJ, Cui Q. Targeted gene engineering in *Clostridium cellulolyticum* H10 without methylation. *J Microbiol Methods*. 2012;89:201–8.
48. Johnson EA, Madia A, Demain AL. Chemically defined minimal medium for growth of the anaerobic cellulolytic thermophile *Clostridium thermocellum*. *Appl Environ Microbiol*. 1981;41:1060–2.
49. Cui Z, Li Y, Xiao Y, Feng Y, Cui Q. Resonance assignments of cohesin and dockerin domains from *Clostridium acetobutylicum* ATCC824. *Biomol NMR Assign*. 2013;7:73–6.
50. Grote A, Hiller K, Scheer M, Munch R, Nortemann B, Hempel DC, Jahn D. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res*. 2005;33:W526–31.
51. Mohr G, Hong W, Zhang J, Cui G-Z, Yang Y, Cui Q, Liu Y-J, Lambowitz AM. A targetron system for gene targeting in thermophiles and its application in *Clostridium thermocellum*. *PLoS ONE*. 2013;8:e69032.
52. Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem*. 1976;72:248–54.
53. Zhang J, Liu Y-J, Cui G-Z, Cui Q. A novel arabinose-inducible genetic operation system developed for *Clostridium cellulolyticum*. *Biotechnol Biofuels*. 2015;8:36.

Submit your next manuscript to BioMed Central
and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

