FEBS Letters 589 (2015) 3133–3140

FEBS Letters

journal homepage: www.FEBSLetters.org

Three cellulosomal xylanase genes in *Clostridium thermocellum* are regulated by both vegetative SigA (σ^A) and alternative SigI6 (σ^{I6}) factors

Andy Sand [a,1], Evert K. Holwerda [b,1], Natalie M. Ruppertsberger [b], Marybeth Maloney [b], Daniel G. Olson [b], Yakir Nataf [a], Ilya Borovok [c], Abraham L. Sonenshein [d], Edward A. Bayer [e], Raphael Lamed [c], Lee R. Lynd [b], Yuval Shoham [a,*]

[a] *Department of Biotechnology and Food Engineering, Technion–Israel Institute of Technology, Haifa 32000, Israel*
[b] *Thayer School of Engineering at Dartmouth College, Hanover, NH 03755, United States*
[c] *Department of Molecular Microbiology and Biotechnology, Tel-Aviv University, Ramat Aviv 69978, Israel*
[d] *Department of Molecular Biology and Microbiology, Tufts University School of Medicine, Boston, MA 02111, United States*
[e] *Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot 76100, Israel*

ARTICLE INFO

Article history:
Received 12 July 2015
Revised 10 August 2015
Accepted 14 August 2015
Available online 29 August 2015

Edited by Michael Ibba

Keywords:
Regulation
Cellulosome
Biomass sensing
Xylanase
Alternative sigma factor
Anti-sigma factor
Clostridium thermocellum

ABSTRACT

Clostridium thermocellum **efficiently degrades crystalline cellulose by a high molecular weight protein complex, the cellulosome. The bacterium regulates its cellulosomal genes using a unique extracellular biomass-sensing mechanism that involves alternative sigma factors and extracellular carbohydrate-binding modules attached to intracellular anti-sigma domains. In this study, we identified three cellulosomal xylanase genes that are regulated by the σ^{I6}/RsgI6 system by utilizing *sigI6* and *rsgI6* knockout mutants together with primer extension analysis. Our results indicate that cellulosomal genes are expressed from both alternative σ^{I6} and σ^{A} vegetative promoters.**

1. Introduction

Clostridium thermocellum (*Ruminiclostridium thermocellum*) is a Gram-positive, thermophilic, anaerobic bacterium capable of utilizing cellulosic substrates [1,15,18,20,28]. The bacterium produces an extracellular supra-molecular enzyme complex, the cellulosome, that can efficiently degrade crystalline cellulosic substrates and associated plant cell wall polysaccharides [4,6–12,24,26,51,52]. The cellulosome complex consists of catalytic subunits and a non-catalytic subunit, "scaffoldin", that plays multiple roles. The scaffoldin subunit mediates both the binding to cellulose *via* an internal cellulose-binding module, CBM3 [27,56], and the attachment of the catalytic units *via* a set of closely related modules – the "cohesins". The enzymatic subunits contain in addition to their catalytic domains a docking domain – the "dockerin". *C. thermocellum* possesses over 70 different genes that encode dockerin-containing proteins and it appears that there is no specificity in the binding among the various cohesins and the various dockerins [25,34,36,45,57]. The scaffoldin also includes a special type of dockerin domain (type II dockerin) for the attachment of the cellulosome to a complementary type of non-cellulosomal cohesin (type II), which is positioned on the cell surface *via* cell-surface anchoring proteins [16,35,49]. Regulation studies by us [21–23] and others [17,46–48,53,58] demonstrated that in *C. thermocellum* the expression level and the composition of the cellulosomal proteins vary with the availability of the carbon source (cellobiose) and the presence of extracellular plant cell wall derived polysaccharides [14,31,37,41–43,47]. Somewhat

Abbreviations: CBM, carbohydrate binding module; GH, glycoside hydrolase; ECF, extracytoplasmic function; TMH, trans-membrane helix; hpt, hypoxanthine ribosyltransferase; FUDR, fluoro-deoxyuracil; 8-AZH, 8-azahypoxanthine; MOPS, morpholinepropanesulfonic acid; RT, reverse transcription

Author contributions: A.S. and Y.S. designed the experiments; A.S., E.K.H. and Y.N. performed the experiments; D.G.O., M.B.M. and N.M.R. assisted in experiments; A.S. and I.B. analyzed the data; A.S. and Y.S. wrote the manuscript; E.A.B., A.L.S., R.L., I.B. and L.R.L. made manuscript revisions.

* Corresponding author.
E-mail address: yshoham@tx.technion.ac.il (Y. Shoham).
[1] Contributed equally to this work.

http://dx.doi.org/10.1016/j.febslet.2015.08.026
0014-5793/

surprisingly, *C. thermocellum* is capable of growing efficiently only on β-glucans (β-1,4 and β-1,3 glucans), utilizing mainly cellodextrins (derived from cellulose) although it encodes and regulates numerous hemicellulolytic genes, whose end products (mainly five-carbon sugars) cannot enter the cell [39]. Thus, the bacterium should possess regulatory systems that will allow it to sense and react to the presence of high molecular weight polysaccharides in the extracellular environment without importing their low molecular weight degradation products. Indeed, we have recently identified in *C. thermocellum* a novel regulatory mechanism for sensing extracellular biomass [5,32,38]. This mechanism includes a set of six bi-cistronic operons, each of which encodes an alternative σ^I-like factor and a multi-modular trans membrane protein with the following modules: (a) a ~ 65-residue N-terminus that bears strong similarity to the N-terminal segment *of B. subtilis*'s anti-σ^I factor, RsgI, (Anti-sigma factor N-terminus, RsgI_N, pfam12791); (b) a single trans-membrane helix (TMH); and (c) a C-terminal segment positioned outside the cell membrane that contains a polysaccharides recognition function including carbohydrate-binding modules, (CBM3, CBM42) sugar-binding elements (PA14), and a glycoside hydrolase (GH) module of family 10 (GH10). Moreover, we identified an additional operon that encodes an extracytoplasmic function (ECF) family alternative sigma factor (σ^{24C}) and its cognate complex anti-sigma factor, Rsi24C, that has a C-terminal module of the GH family 5 (GH5) (Fig. 1A). In the presence of a specific polysaccharide that can interact with the corresponding extracellular RsgI-borne sensing domain (CBM or GH-like element), the cognate σ-factor is released to promote transcription of the target genes (Fig. 1B) [38]. The recently published genomes of *Clostridium clariflavum* [30] and *Acetivibrio cellulolyticus* [19] revealed similar putative biomass-sensing systems containing carbohydrate binding module (CBM) modules of families 3, 35, 42 and also PA14.

In this study, we attempted to identify genes that are regulated by the alternative σ-factor, σ^{I6}. Two knockout mutants lacking either the *sigI*6 (Clo1313_2778) or the *rsgI*6 (Clo1313_2777) genes were prepared and the expression levels of selected cellulosomal genes in the different backgrounds were determined. The expression results, in combination with the identification of the transcriptional start sites, led to identification of three xylanase genes that are regulated by σ^{I6}. In addition, primer extension analysis revealed that the expression of the tested cellulosomal genes is directed not only from an alternative σ-factor promoter but also from a vegetative σ^A-type promoter, supporting previous observations [13,21,22,40]. Batch fermentation studies also demonstrated the effect of SigI6 and RsgI6 deletion on the solubilization process of switchgrass.

2. Materials and methods

2.1. Bacterial strains, vectors and chemicals

C. thermocellum DSM 1313 was purchased from DSMZ GmbH, Braunschweig, Germany. All chemical reagents were of analytic grade and were purchased from Sigma Aldrich Chemical Co. or Fluka unless otherwise noted. The mutant strains were constructed from the genetically tractable *C. thermocellum* DSM 1313 strain according to the method described by Argyros et al. [2]. Deletion plasmids were constructed using standard cloning techniques [50]. Knocking out genes in *C. thermocellum* by markerless deletion has been described in detail elsewhere [44]. In short, after transformation cells are selected for successful uptake of the knock-out plasmid pEKH-*rsgI6* (accession number KT362976) or pEKH-*sigI6* (accession number KT362977) by selection on Thiamphenicol (Tm) (5 μg/mL). Thiamphenicol resistant colonies are picked, and subsequently plated with Tm and 5-fluoro-2′-deoxyuradine (FuDR) (10 μg/mL) to select for double cross-over events to disrupt the target gene with the antibiotic resistance cassette, as well as loss of the plasmid backbone. Colonies from the Tm and FuDR selection are picked and then grown in presence and absence of 8-azahypoxanthine (8AZH) (500 μg/mL) selecting for a third recombination event, to eliminate the antibiotic resistance cassette (*gapD-cat-hpt*) and result in a markerless disruption of the target gene. At each step of the process, colonies are checked by PCR to confirm that the selection worked as intended.

2.2. C. thermocellum growth conditions

Cells were grown in batch culture at 60 °C in GS medium (0.65 g/L K_2HPO_4, 0.5 g/L KH_2PO_4, 1.3 g/L $(NH_4)_2SO_4$, 10 g/L morpholinepropanesulfonic acid (MOPS), 5 g/L yeast extract, 1 g/L cysteine, 2 mg/L resazurin and 0.5 g/L $MgCl_2$) with the desired carbon source in Duran anaerobic bottles (Schott, Mainz, Germany) for up to 0.5 l working volume. The bottles were closed with a rubber septum and oxygen was removed by bubbling the headspace with nitrogen. Continuous cultures were performed in a BIOFLO 3000 bio-reactor (New Brunswick, USA) in a working volume of 1.0 L at 60 °C under carbon limitation (cellobiose, 2 g/L) at a dilution rate of 0.1 1/h. The cultures were considered to be at steady state after feeding five liters of fresh medium and the absorbance at 600 nm remained constant over time. It should be noted that *C. thermocellum* cannot utilize carbon from the yeast extract. Cultures that were grown in the bioreactors were kept at 60 °C, the medium contained 4 g/L MOPS, the pH was kept at 7.2 by automatic addition of 5 M NaOH, stirred at 150 rpm, and the headspace of the bioreactor was flushed continuously with nitrogen to maintain anaerobic conditions. Fermentations with 8.1 g/L switchgrass were done in Sartorius bioreactors (Sartorius Stedim, Bohemia, NY) with a 2-L working volume on MTC medium using 2 g/L of ammonium chloride as nitrogen source and 4 N KOH to maintain neutral pH [29]. Bioreactors were purged with a 20% CO_2/80% N_2 mixture for at least 3 h before inoculation. The switchgrass (*Panicum virgatum*, "cave in rock"-variety) was provided by Marvin Hall, Kay DiMarco and Dr. Tom Richard from Penn State University. It was harvested at Rocks Springs Research Farm (Spring Mills, PA) in June (mid-season), left to dry on the field for 6 days, collected and milled with a 6 mm screen. Prior to autoclaving (as suspended solids) for fermentation, the material was milled until it passed through subsequently a 200 μm and a 80 μm screen (repeated 2 times for each screen, no fractionation/separation) on a Retsch ultra centrifugal mill (Haan, Germany). For dry weight determination of residuals (cells and residual switchgrass), samples of 10 mL were centrifuged at 2800×*g* for 10 min, washed once with 10 mL of MilliQ water and collected in pre-weighed aluminum weigh-dishes, dried at 55 °C for at least 7 days until no weight decrease was observed. Determination of residual glucose, xylose and arabinose content of the residuals was done by quantitative saccharification on 5 mL samples [29].

2.3. RNA purification

Cell pellets of 10 mL culture with an absorbance of 1 at 600 nm (about 10^9 cells) were suspended in 1 mL of TRI reagent (Sigma Aldrich), frozen in liquid nitrogen and stored at −80 °C. The samples were sonicated (Sonicator model W-375; Heat System-Ultrasonics Inc., Plainview, NY), and the RNA was extracted according to the TRI reagent protocol. Following the extraction, the RNA was treated with DNaseI (Qiagen) to remove any contaminating genomic DNA, following a cleanup protocol using the RNeasy-miniKit (Qiagen).

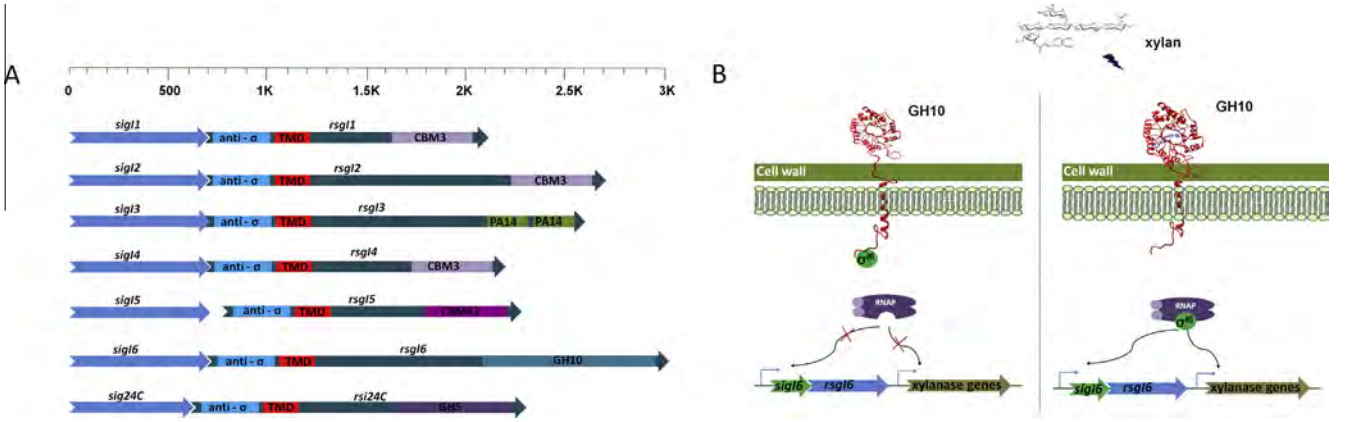

Fig. 1. Schematic representation of the biomass sensing mechanism in *C. thermocellum*. (A) The *sigI* – *rsgI* and *sig24C* – *rsi24C* operons in *C. thermocellum*. In each operon, the σ-factor gene is positioned upstream to the *rsgI* gene that encodes for a multimodular protein with an anti – σ domain at the N – terminus, a membrane spanning domain and a carbohydrate sensing module at the C – terminus. Three of these modules have been recently crystallized and their structure has been determined [55]. (B) Mechanism for the activation of σ^{I6} by extracellular xylan. The RsgI6 protein contains an intracellular anti – σ domain and an extracellular carbohydrate sensing domain. By default, the anti – σ domain strongly binds the σ factor and inactivates it. When extracellular xylan binds the sensing domain, a conformational change occurs and the σ factor is released. The free σ-factor interacts with RNA polymerase and promotes transcription of both cellulosomal genes and its own operon.

2.4. Primer extension

Oligonucleotides (5′ 6-FAM labeled) were designed in the 3′ to 5′ orientation to hybridize the gene at 80 to 150 bp from the start codon. Total RNA (50 μg) was incubated with a 5′ 6-FAM labeled oligonucleotide (10 μM) in 5× hybridization buffer (0.5 M KCl, 0.5 M Tris–HCl pH 8.3) for 1 min at 90 °C, 2 min at 60 °C and 15 min on ice. Reverse transcription was carried out using avian myelobalstosis virus reverse transcriptase (Promega), in 5× RT buffer (Promega) and 5 mM dNTPs for 2 h at 42 °C. The size of the extension product was determined using GeneScan 3130 (Applied Biosystems) and the raw trace was analyzed using PeakScanner v1.0 (Applied Biosystems).

2.5. Quantitative real-time RT-PCR

Reverse transcription (RT) of RNA was performed with the qScript cDNA synthesis kit (Quanta Scientific) following the manufacturer's instructions, using 5 μg of total RNA and random hexamers. To verify that the RNA was not contaminated with residual DNA, control reactions were performed in the absence of reverse transcriptase. Real-time RT-PCR primers were designed with the GeneRunner V4.09 (Hastings software Inc.) and Primers express 2.0 software (Applied Biosystems, Foster City, CA) to generate amplicons ranging in size from 90 to 150 bp. The Real time PCR reaction was carried out with Applied Biosystems 7300 Real Time PCR system. Each reaction (20 μL) contained cDNA, reverse and forward primers (10 μM each) and PerfeCTa SYBR green Fastmix (Quanta Scientific). The reaction for each gene was carried out in quadruplicates and the data analysis was performed with the 7300 system software by using the *rpoD* gene for normalization. For each set of primers, cDNA dilution curves were generated to calculate the real time RT PCR efficiency. Melting curves were analyzed to ensure primer annealing specificity and lack of primer secondary structures.

2.6. Sequence analyses

Sequence entries, primary analyses and ORF searches were performed using the National Center for Biotechnology Information (NCBI) server ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) and GeneRunner (Hastings software Inc.) software. BLAST analysis of protein and nucleotide sequences was performed using the NCBI server (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The different functional domains (CBM, GH, solute binding protein (SBP) and leader peptide) were identified using the CAZy (Carbohydrate-Active EnZymes) website (http://www.cazy.org/), the Pfam protein families database (http://pfam.sanger.ac.uk), integrated resource of Protein Domains (InterPro) (http://www.ebi.ac.uk/interpro/) and the database of protein families and domains PROSITE (http://www.expasy.ch/prosite/). Consensus promoter alignments were constructed by WebLogo (v.282) (http://weblogo.berkeley.edu/logo.cgi).

3. Results and discussion

3.1. Bioinformatic analysis of putative σ^{I6} promoter sequences

C. thermocellum possesses seven sigma/anti-sigma systems encoded by six *sigI-rsgI* and one *sig24C-rsi24C* operons [32,38], allowing it to regulate the expression of selected cellulosomal genes in response to the composition of extracellular polysaccharides. One of the systems is the *sigI6-rsgI6* operon, which encodes a 253 amino acid alternative σ-factor, σ^{I6}, and a 760 amino acid trans-membrane protein, RsgI6. RsgI6 is composed of a 56 amino acid N-terminal cytoplasmic domain RsgI_N (pfam12791), a hydrophobic transmembrane domain (residues 58 to 78), an RsgI conserved functional domain, and an extracellular sensing domain that resembles a xylanase GH10 catalytic domain, (residues 380 to 760) (Fig. 1A). This GH10 domain was characterized previously and was shown to both bind and hydrolyze insoluble and soluble xylan substrates [5]. The xylan specificity of the sensing domain of RsgI6 suggests that its cognate σ^{I6} regulates xylan related genes, and in fact we have previously shown that the expression of the *sigI6* gene increases tenfold when *C. thermocellum* was grown in the presence of xylan [38]. To identify which genes are regulated by σ^{I6}, we reasoned that since many alternative σ-factors auto-regulate their own expression, their promoter sequences should resemble those of their target genes. The *sigI6* promoter has been previously identified [38] and based on its sequence we searched for similar sequences in the genome of *C. thermocellum* allowing no more than two mismatches in each of the −10 and −35 sequences with a spacer of 17–19 bp. This searches yielded eight genes, seven of which encode for hemicellulose utilization-related enzymes (Table 1).

Table 1
Bioinformatic analysis of putative SigI6-regulated genes.

Locus tag	Gene	Function	Domains[b]	Putative promoter sequence[a]		Distance from start codon (bp)
				−35	−10	
Clo1313_2778	*sigI6*	Sigma factor		AAT**GCGACA**TAA N(17) TAT**ACGAAT**CGA		36
Clo1313_0521	*xyn11B*	Xylanase	GH11, CBM6, CE4, Doc1	TAA**GCGACT**TAA N(17) TTT**GCGAAT**AGA		180
Clo1313_2635	*xyn10Z*	Xylanase	GH10, CE1, CBM6, Doc1	CGA**CCGACA**CAA N(17) TTC**ACGAA**CAA		112
Clo1313_0177	*xyn10D*	Xylanase	GH10, CBM22, Doc1	AAT**GCGAC**CAAA N(17) TTC**ACGAA**ATAC		49
Clo1313_0851	*xgh74A*	Xyloglucanase	GH74, Doc1	ACG**GGTACA**TCA N(18) GGT**CCGAAT**TTA		163
Clo1313_1425	*cel5E*	Esterase	GH5, CE2, Doc1	CAA**GC**CA**A**ATAG N(17) TGT**A**AG**AAT**AAA		89
Clo1313_1564		Laminarinase	GH81, Doc1	AAA**GCGGCT**AAT N(17) ATA**TAGAA**CTG		30
Clo1313_0399	*man26A*	Mannase	GH26, CBM35, Doc1	AAT**GTTACA**ACA N(18) ATA**ATGAA**CTG		54
Clo1313_2558		Aminotransferase		TTG**CCGAC**CGTT N(17) CTG**CGA**A**A**CAC		181

[a] Nucleotides similar to the *sigI6* promoter are shown in bold. Underlined sequences represent the −10 and −35 binding sites. The spacer region, designated "N(x)", represents the number of bp between the two underlined sequences.
[b] GH, glycoside hydrolase; CBM, carbohydrate binding module; CE, carbohydrate esterase; Doc1, dockerin type 1.

3.2. The mRNA levels of the σ^{I6} – regulated genes

To directly test the involvement of σ^{I6} in regulating xylan-utilization related genes, we prepared two *C. thermocellum* knock-out strains lacking either the *sigI6* or *rsgI6* genes by the recently developed markerless knockout procedure [2]. The two mutant strains as well as the WT were grown in continuous cultures under carbon (cellobiose) limitation and the expression level of the selected genes was measured by real – time RT PCR (Table 2). The expression level of *sigI6* was 7-fold higher in the Δ*rsgI6* strain, consistent with the observation that alternative σ factors auto-regulate their own expression [3]. Of the eight genes with promoter similarity to *sigI6*, only three cellulosomal xylanases, *xyn11B, xyn10Z* and *xyn10D,* appeared to be up-regulated in the *ΔrsgI6* strain by 4-, 7- and 5-fold, respectively (Table 2). Indeed, the predicted promoters of these three genes share the highest sequence similarity to the *sigI6* promoter, with 17 bp spacer between the −10 and −35 regions (Table 1). Neither the two xylanase genes *xyn10Y* and *xyn10C* nor the cellulosome-related genes, *cel48S, cipA, cel9F, cel5B, cel9T* and *cel9U,* were affected more than 2-fold by the deletion (Table 2). The expression levels of the evaluated genes were consistent with previous studies that showed similar expression ratio between cellulosome-related genes in both batch and continuous WT cultures [33,46–48,54]. In the Δ*sigI6* strain the expression of the three xylanase genes was either similar, (*xyn10Z*), or much lower, (*xyn11B* and *xyn10D*), compared to the expression in the parental strain. The identification of three σ^{I6}-regulated genes allowed us to better define the consensus promoter sequence for σ^{I6} as SCGACH$(N)_{17}$RCGAAW which somewhat resembles the *sigI* promoter sequence in *Bacillus* sp. (Fig. 3A and B).

3.3. Identification of transcriptional start sites

To experimentally verify the promoters of the selected genes we utilized a modified primer extension procedure in which the product was obtained with fluorophore labeled primers (instead of radioactive primers) and the product size was determined using a genetic analyzer (Applied Biosystems). Primer extension analysis with total RNA from *ΔrsgI6* strain revealed two transcriptional start sites (TSSs) for the *xyn11B* and *xyn10Z* genes. Both in *xyn11B* and *xyn10Z* the first promoter shows a homology to the vegetative σ^{A} promoter, while the second promoter corresponds to σ^{I6} promoter (Fig. 2A and B). No extension products were obtained for the *xyn10D* gene, due to its low expression level under the experimental conditions. In the process of identifying additional alternative sigma factors regulated genes, we noticed that in many cases, primer extension with total RNA from WT strain grown on cellulose and the inducing polysaccharide (e.g. xylan and pectin) revealed TTSs corresponded to vegetative, σ^{A}-like promoters. The determined vegetative (σ^{A}) promoter sequences of the cellulosome-related genes, provided the consensus sequence NWDDHW$(N)_{17\text{-}19}$TWWWWT (Table 3 and Fig. 3C). Bioinformatic

Table 2
Expression analysis of the putative SigI6 – regulated genes in WT, *ΔrsgI6* and *ΔsigI6* strains grown in a continuous culture under carbon limitation at $D = 0.1\ h^{-1}$.

Locus tag	Gene	Function	Domains	WT[a]	Δ*rsgI6*[a]	FC[b]	Δ*sigI6*[a]	FC[c]
Clo1313_2778	*sigI6*	Sigma factor		1.0 ± 0.1	7.4 ± 0.8	7.4 ± 1.1	ND	ND
Clo1313_0521	*xyn11B*	Xylanase	GH11, CBM6, CE4, Doc1	6.1 ± 1.7	23 ± 2.0	3.8 ± 1.1	0.2 ± 0.1	0.03 ± 0.02
Clo1313_2635	*xyn10Z*	Xylanase	GH10, CE1, CBM6, Doc1	2.4 ± 0.2	17 ± 1.0	7.0 ± 0.7	2.0 ± 0.1	1.0 ± 0.1
Clo1313_0177	*xyn10D*	Xylanase	GH10, CBM22, Doc1	0.2 ± 0.04	1.2 ± 0.2	6.0 ± 1.5	0.03 ± 0.01	0.1 ± 0.1
Clo1313_0851	*xgh74A*	Xyloglucanase	GH74, Doc1	0.9 ± 0.1	0.7 ± 0.1	0.8 ± 0.2	0.3 ± 0.1	0.3 ± 0.1
Clo1313_1425	*cel5E*	Esterase	GH5, CE2, Doc1	0.4 ± 0.1	0.4 ± 0.1	1.0 ± 0.5	0.1 ± 0.1	0.3 ± 0.3
Clo1313_1564		Laminarinase	GH81, Doc1	0.1 ± 0.1	0.03 ± 0.01	0.3 ± 0.3	0.04 ± 0.01	0.4 ± 0.2
Clo1313_1305	*xyn10Y*	Xylanase	GH10, CBM22, CE1, Doc1	1.1 ± 0.2	2.1 ± 0.1	2.0 ± 0.4	1.3 ± 0.1	1.2 ± 0.2
Clo1313_0399	*man26A*	Mannase	GH26, CBM35, Doc1	1.2 ± 0.7	1.2 ± 0.2	1.0 ± 0.2	0.7 ± 0.2	0.6 ± 0.4
Clo1313_2530	*xyn10C*	Xylanase	GH10, CBM42, Doc1	2.3 ± 0.2	2.3 ± 0.3	1.0 ± 0.2	0.9 ± 0.1	0.4 ± 0.1
Clo1313_2747	*cel48S*	Cellulase	GH48, Doc1	19 ± 1.3	26 ± 4.0	1.4 ± 0.2	30 ± 1.0	1.6 ± 0.1
Clo1313_0627	*cipA*	Scaffoldin	CBM3	26 ± 9.0	18 ± 6.5	0.7 ± 0.3	10.9 ± 1.9	0.4 ± 0.2
Clo1313_1694	*cel9F*	Cellulase	GH9, Doc1	1.7 ± 0.2	1.4 ± 0.1	0.8 ± 0.1	1.6 ± 0.1	0.9 ± 0.1
Clo1313_1701	*cel5B*	Cellulase	GH5, Doc1	1.4 ± 0.3	1.6 ± 0.4	1.1 ± 0.3	0.8 ± 0.4	0.6 ± 0.3
Clo1313_0400	*cel9T*	Cellulase	GH9, Doc1	7.1 ± 2.2	6.8 ± 1.1	1.0 ± 0.3	1.9 ± 0.4	0.3 ± 0.1
Clo1313_3023	*cel9U*	Cellulase	GH9, CBM3 Doc1	1.3 ± 0.2	2.3 ± 0.1	1.8 ± 0.2	1.4 ± 0.1	1.1 ± 0.1
Clo1313_2558		Aminotransferase		0.1	0.1	1.0		

[a] The numbers in this columns represent the ratio of the mRNA level for each gene to *rpoD* mRNA level in the indicated strain.
[b] Fold of change in the expression level in *rsgI6* knockout strain compared to WT.
[c] Fold of change in the expression level in *sigI6* knockout strain compared to WT.

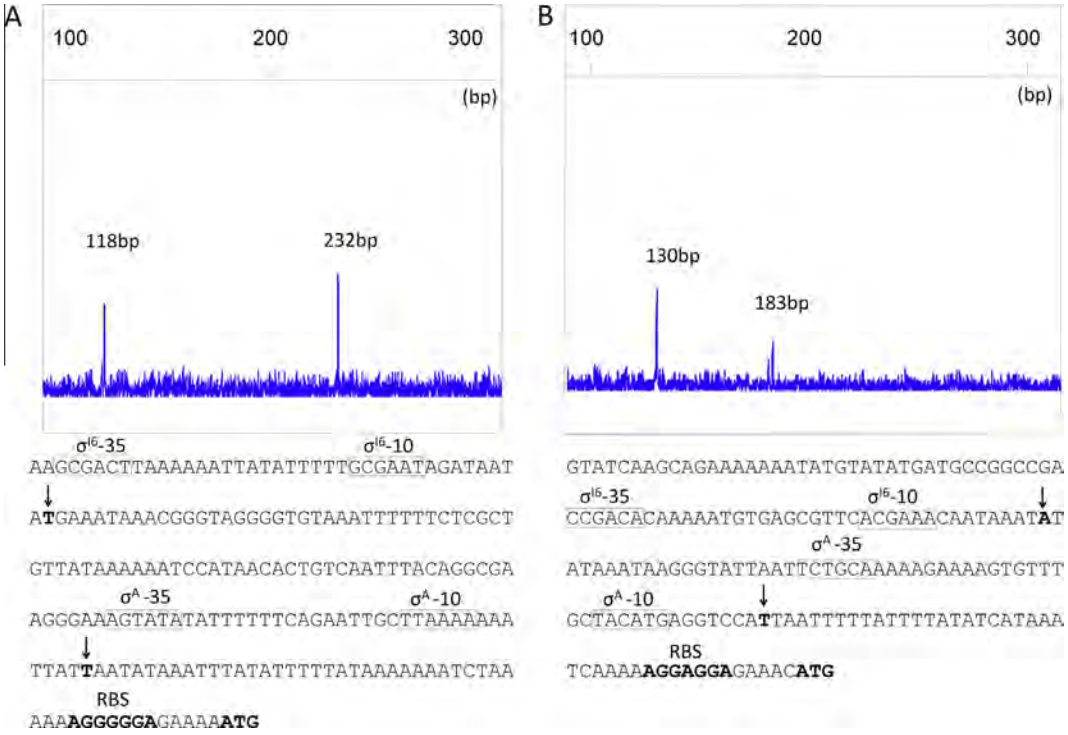

Fig. 2. Capillary electrophoresis showing extension cDNA products and promoter sequences analysis of *xyn11B* (A) and *xyn10Z* (B). Each electropherogram shows two extension products that correspond for two different transcriptional start sites. The product size was determined by GeneScan 3130 (Applied Biosystems). Based on the primer extension results, both genes contain two transcriptional start sites, corresponding to a σ^{I6} and a vegetative σ^{A} promoter.

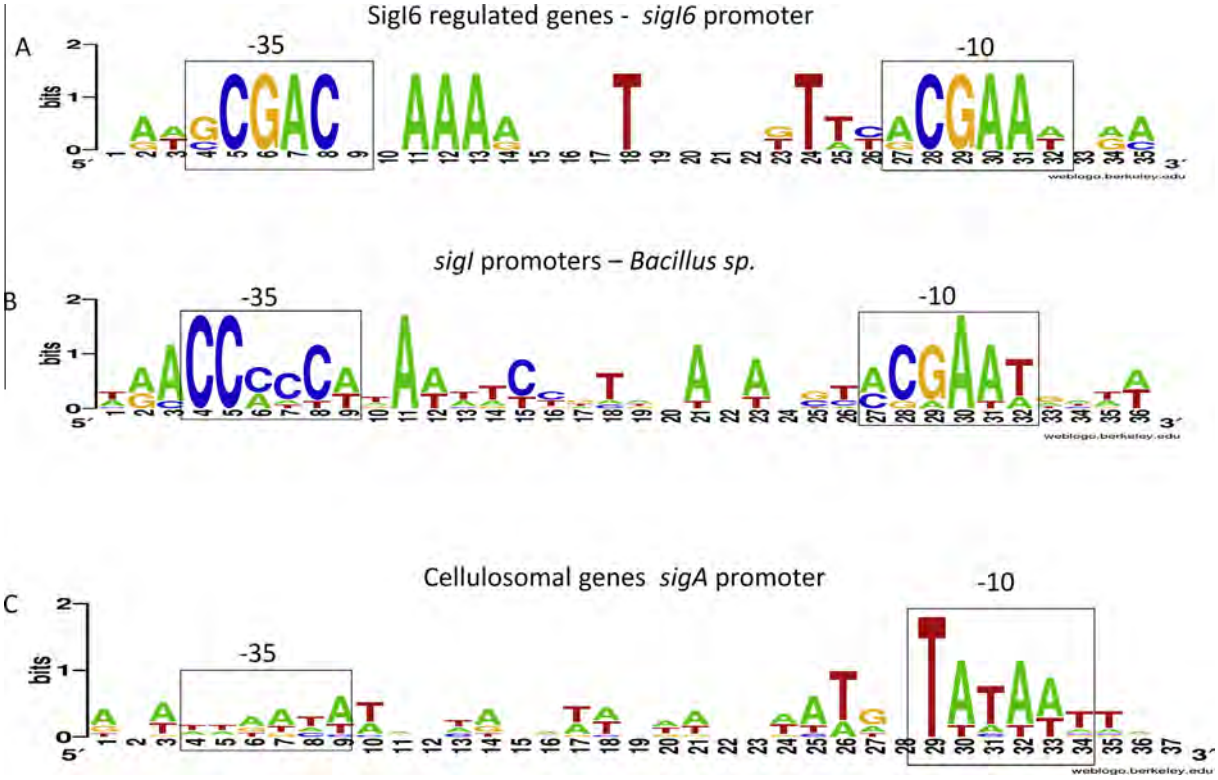

Fig. 3. Alignment of consensus σ^{I} and σ^{A} promoter sequences. (A) Consensus σ^{I6} promoter sequence in *C. thermocellum*. The alignment was constructed based on the σ^{I6}-regulated genes. (B) Consensus of σ^{I} promoter sequence in *Bacillus* sp. The alignment was constructed based on *sigI* promoters in *B. subtilis*, *B. atrophaeus*, *B. licheniformis*, *B. cereus*, *B. anthracis*, *G. stearothermophilus*, *G. caldoxylosilyticus* and *G. thermoglucosidans*. (C) Consensus σ^{A} promoters of cellulosomal genes in *C. thermocellum*. The alignment was constructed based on experimentally identified transcriptional start sites upstream of cellulosomal genes.

Table 3
σ^A promoter sequences of cellulosomal related genes.

Locus tag	Gene	Promoter sequence[a]		Distance from start codon (bp)	Refs.
		−35	−10		
Clo1313_0627	*cipA*	TTGGTTTGTATA N(17)	ATTTCAAATGCC	65	[22]
Clo1313_2747	*cel48S*	CACAAATTTATT N(17)	AAGTATTTTTTG	145	[21]
Clo1313_1396	*cel9D*	ATATTGAATTAT N(18)	TGGTATAATTAA	297	[37]
Clo1313_1960	*cel8A*	GTATAAACATGA N(17)	TGATATAATGAT	136	[13]
Clo1313_1701	*cel5B*	AGATTGGAAATC N(17)	AAGTATTACCCA	141	This study
Clo1313_2530	*xyn10C*	AATAAGACAAAC N(17)	AGTTATAATTTA	216	This study
Clo1313_0413	*cel5G*	TAAGTATTCCTG N(17)	TTGTAAATTTAC	181	This study
Clo1313_0521	*xyn11B*	GAAAGTATATAT N(17)	TGCTTAAAAAAA	62	This study
Clo1313_2635	*xyn10Z*	AATTCTGCAAAA N(17)	TGCTACATGAGG	55	This study
Clo1313_0851	*xgh74A*	AAACAGGCAGGT N(16)	ATTTTTAATCAG	31	This study
Clo1313_2805	*cel5O*	ACTATATTAACA N(17)	TGATATAATTAA	87	This study
Clo1313_1659	*cel9R*	TATAATTTATGA N(17)	TGATATAATTAT	153	This study
Clo1313_2022	*lic16B*	GGATTAAAATAC N(18)	TTATATATATTT	109	This study
Clo1313_1809	*cel9K*	AGGTTAATTTTG N(19)	TATTTTAATTGT	245	This study
Consensus		NWDDHW	TWWWWT		

The following nucleotide base codes (IUPAC) are used in this table: D = A, G or T; H = A, C or T; W = A or T; N = any nucleotide.
[a] Underlined sequences represent the −10 and −35 binding sites. The spacer region, designated "N(x)", represents the number of bp between the two underlined sequences.

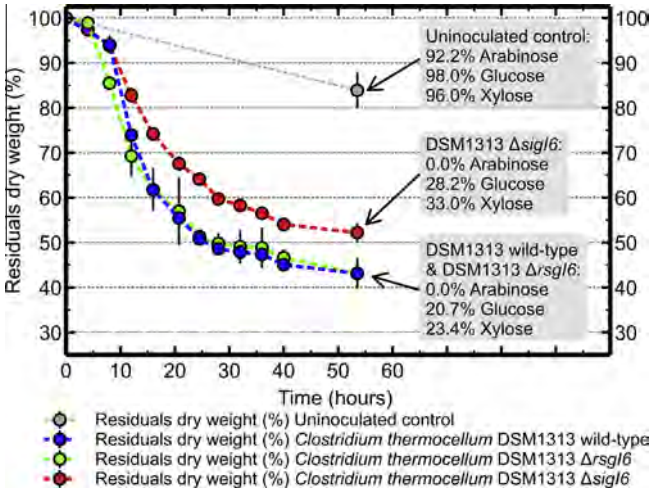

Fig. 4. Fermentation of switchgrass by *Clostridium thermocellum* DSM1313 wild-type and *sigI6* and *rsgI6* knock-out strains. Residual dry weight is shown in relative quantities to account for small differences between substrate loadings at the start of the fermentations. Values in the gray boxes represent relative carbohydrate content compared to time = 0 as determined by quantitative saccharification. An uninoculated control at similar solids loading and incubation duration resulted in 98% residual glucose equivalents, 96% residual xylose equivalents, 92% residual arabinose equivalents and 84% dry weight remaining. All reported values are averages of triplicate measurements and the error bars represent one standard deviation.

analysis of putative promoter sequences of sugar and amino acid metabolism related genes in *C. thermocellum*, resulted in the consensus sequence TTRWHN$(N)_{17\text{-}19}$WWWWNT. Both consensus sequences resemble the σ^A promoter sequence of *B. subtilis* (TTGACA$(N)_{14\text{-}16}$(TGN)TATAAT) although the sugar and amino acid metabolic gene promoter sequences are more conserved and more similar to that of *B. subtilis*. Thus, it appears that in *C. thermocellum* the cellulosomal genes are expressed from two promoters, an alternative σ^I promoter and a vegetative σ^A promoter.

3.4. Fermentations of switchgrass

To experimentally demonstrate the effect of deleting *sigI6* and *rsgI6* on batch fermentations of a non-model cellulosic substrate, wild-type and the *sigI6* & *rsgI6* knock-out strains were grown on switchgrass (chosen by the United States Department of Energy as a model herbaceous feedstock), (Fig. 4). Wild-type and *ΔrsgI6* exhibited a very similar solubilization pattern as shown by the residuals dry weight graph (residuals = cells and residual substrate). Since the growth media contained xylan, it is expected that the expression of the three SigI6 regulated xylanase genes will be similar in the WT and the *ΔrsgI6* background resulting in similar degree of solubilization. However, in the *sigI6* mutant the overall solubilization of the switchgrass was reduced (9.1% difference). It should be noted that *C. thermocellum* possesses two additional xylanases, Xyn10C and Xyn10Y that are not regulated by SigI6 and contribute to the overall switchgrass solubilization capacity. Although shown here at non-industrial conditions an effect of this magnitude for a single gene-knockout emphasizes the importance of the xylan/hemicellulose solubilizing capabilities of *C. thermocellum* and its subsequent application in industrial processes.

4. Conclusions

Our results demonstrate the advantage of using several complementary approaches for identifying alternative σ-factor regulated genes and their corresponding promoters. The regulatory mechanism of cellulosomal genes by alternative σ-factors corresponds well to the physiology of *C. thermocellum*. The bacterium expresses constitutively low levels of many cellulosomal genes utilizing presumably weak vegetative (σ^A) promoters. However, in the presence of extracellular plant cell wall derived polysaccharides, it can utilize the σ^I/RsgI biomass-sensing systems to up-regulate the required genes *via* specific promoters.

Acknowledgments

This research was supported by the US–Israel Binational Science Foundation (BSF; grant No. 2011049 to Y.S. and A.L.S.) the Israel Science Foundation (ISF; grant Nos. 715/12, 500/10 and 1349/13 to Y.S., and grant No. 24/11 to R.L. and I.B. with an additional support issued to R.L. and I.B. by The Sidney E. Frank Foundation through the ISF), the Ministry of Environmental Protection, Israel, the Israeli Center of Research Excellence (I-CORE Center No. 152/11) Program of the Planning and Budgeting Committee, and the Grand Technion Energy Program, comprising part of The Leona M. and Harry B. Helmsley Charitable Trust reporting on alternative energy series of the Technion – Israel Institute of Technology and the Weizmann Institute of Science. YS holds the Erwin and Rosl

Pollak Chair in Biotechnology at the Technion – Israel Institute of Technology. EAB is the incumbent of The Maynard I. and Elaine Wishner Chair of Bio-Organic Chemistry at the Weizmann Institute of Science. EKH, NMR, MM, DGO and LRL were supported by funding provided by the BioEnergy Science Center (BESC). The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.febslet.2015.08.026.

References

[1] Akinosho, H., Yee, K., Close, D. and Ragauskas, A. (2014) The emergence of *Clostridium thermocellum* as a high utility candidate for consolidated bioprocessing applications. Front. Chem. 2, 66.
[2] Argyros, D.A., Tripathi, S.A., Barrett, T.F., Rogers, S.R., Feinberg, L.F., Olson, D.G., Foden, J.M., Miller, B.B., Lynd, L.R., Hogsett, D.A. and Caiazza, N.C. (2011) High ethanol titers from cellulose using metabolically engineered thermophilic, anaerobic microbes. Appl. Environ. Microbiol. 77 (23), 8288–8294.
[3] Asai, K., Ootsuji, T., Obata, K., Matsumoto, T., Fujita, Y. and Sadaie, Y. (2007) Regulatory role of RsgI in sigI expression in *Bacillus subtilis*. Microbiology 153, 92–101.
[4] Bae, J., Morisaka, H., Kuroda, K. and Ueda, M. (2013) Cellulosome complexes: natural biocatalysts as arming microcompartments of enzymes. J. Mol. Microbiol. Biotechnol. 23, 370–378.
[5] Bahari, L., Gilad, Y., Borovok, I., Kahel-Raifer, H., Dassa, B., Nataf, Y., Shoham, Y., Lamed, R. and Bayer, E.A. (2011) Glycoside hydrolases as components of putative carbohydrate biosensor proteins in *Clostridium thermocellum*. J. Ind. Microbiol. Biotechnol. 38, 825–832.
[6] Bayer, E.A., Belaich, J.P., Shoham, Y. and Lamed, R. (2004) The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. Annu. Rev. Microbiol. 58, 521–554.
[7] Bayer, E.A., Kenig, R. and Lamed, R. (1983) Adherence of *Clostridium thermocellum* to cellulose. J. Bacteriol. 156, 818–827.
[8] Bayer, E.A. and Lamed, R. (1992) The cellulose paradox: pollutant *par excellence* and/or a reclaimable natural resource? Biodegradation 3, 171–188.
[9] Bayer, E.A. and Lamed, R. (1986) Ultrastructure of the cell surface cellulosome of *Clostridium thermocellum* and its interaction with cellulose. J. Bacteriol. 163, 552–559.
[10] Bayer, E.A., Lamed, R., White, B.A. and Flint, H.J. (2008) From cellulosomes to cellulosomics. Chem. Rec. 8, 364–377.
[11] Bayer, E.A., Morag, E. and Lamed, R. (1994) The cellulosome – a treasure trove for biotechnology. Trends Biotechnol. 12, 379–386.
[12] Bayer, E.A., Morag, E., Shoham, Y., Tormo, J. and Lamed, R. (1996) The cellulosome: a cell-surface organelle for the adhesion to and degradation of cellulose p. 155–182 in: Bacterial Adhesion: Molecular and Ecological Diversity (Fletcher, M., Ed.), pp. 155–182, Wiley-Liss Inc., New York.
[13] Beguin, P., Rocancourt, M., Chebrou, M.-C. and Aubert, J.-P. (1986) Mapping of mRNA encoding endoglucanase A from *Clostridium thermocellum*. Gene 202, 251–254.
[14] Bhat, S., Goodenough, P.W., Owen, E. and Bhat, K.M. (1993) Cellobiose: a true inducer of cellulosome in different strains of *Clostridium thermocellum*. FEMS Microbiol. Lett. 111, 73–78.
[15] Blumer-Schuette, S.E., Kataeva, I., Westpheling, J., Adams, M.W. and Kelly, R.M. (2008) Extremely thermophilic microorganisms for biomass conversion: status and prospects. Curr. Opin. Biotechnol. 19, 210–217.
[16] Bomble, Y.J., Beckham, G.T., Matthews, J.F., Nimlos, M.R., Himmel, M.E. and Crowley, M.F. (2011) Modeling the self-assembly of the cellulosome enzyme complex. J. Biol. Chem. 286, 5614–5623.
[17] Brown, S.D., Raman, B., McKeown, C.K., Kale, S.P., He, Z. and Mielenz, J.R. (2007) Construction and evaluation of a *Clostridium thermocellum* ATCC 27405 whole-genome oligonucleotide microarray. Appl. Biochem. Biotechnol. 137–140, 663–674.
[18] Chandel, A.K. and Singh, O.V. (2011) Weedy lignocellulosic feedstock and microbial metabolic engineering: advancing the generation of 'Biofuel'. Appl. Microbiol. Biotechnol. 89, 1289–1303.
[19] Dassa, B., Borovok, I., Lamed, R., Henrissat, B., Coutinho, P., Hemme, C.L., Huang, Y., Zhou, J. and Bayer, E.A. (2012) Genome-wide analysis of *acetivibrio cellulolyticus* provides a blueprint of an elaborate cellulosome system. BMC Genomics 13, 210.
[20] Demain, A.L., Newcomb, M. and Wu, J.H. (2005) Cellulase, clostridia, and ethanol. Microbiol. Mol. Biol. Rev. 69, 124–154.
[21] Dror, T.W., Morag, E., Rolider, A., Bayer, E.A., Lamed, R. and Shoham, Y. (2003) Regulation of the cellulosomal CelS (*cel48A*) gene of *Clostridium thermocellum* is growth rate dependent. J. Bacteriol. 185, 3042–3048.
[22] Dror, T.W., Rolider, A., Bayer, E.A., Lamed, R. and Shoham, Y. (2003) Regulation of expression of scaffoldin-related genes in *Clostridium thermocellum*. J. Bacteriol. 185, 5109–5116.
[23] Dror, T.W., Rolider, A., Bayer, E.A., Lamed, R. and Shoham, Y. (2005) Regulation of major cellulosomal endoglucanases of *Clostridium thermocellum* differs from that of a prominent cellulosomal xylanase. J. Bacteriol. 187, 2261–2266.
[24] Fontes, C.M. and Gilbert, H.J. (2010) Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. Annu. Rev. Biochem. 79, 655–681.
[25] Gal, L., Pages, S., Gaudin, C., Belaich, A., Reverbel-Leroy, C., Tardif, C. and Belaich, J.-P. (1997) Characterization of the cellulolytic complex (cellulosome) produced by *Clostridium cellulolyticum*. Appl. Environ. Microbiol. 63, 903–909.
[26] Gilbert, H.J. (2007) Cellulosomes: microbial nanomachines that display plasticity in quaternary structure. Mol. Microbiol. 63, 1568–1576.
[27] Gilbert, H.J., Knox, J.P. and Boraston, A.B. (2013) Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules. Curr. Opin. Struct. Biol. 23, 669–677.
[28] Goldemberg, J. (2007) Ethanol for a sustainable energy future. Science 315, 808–810.
[29] Holwerda, E.K., Hirst, K.D. and Lynd, L.R. (2012) A defined growth medium with very low background carbon for culturing *Clostridium thermocellum*. J. Ind. Microbiol. Biotechnol. 39, 943–947.
[30] Izquierdo, J.A., Goodwin, L., Davenport, K.W., Teshima, H., Bruce, D., Detter, C., Tapia, R., Han, S., Land, M., Hauser, L., Jeffries, C.D., Han, J., Pitluck, S., Nolan, M., Chen, A., Huntemann, M., Mavromatis, K., Mikhailova, N., Liolios, K., Woyke, T. and Lynd, L.R. (2012) Complete genome sequence of *Clostridium clariflavum* DSM 19732. Stand. Genomic Sci. 6, 104–115.
[31] Johnson, E.A., Bouchot, F. and Demain, A.L. (1985) Regulation of cellulase formation in *Clostridium thermocellum*. J. Gen. Microbiol. 131, 2303–2308.
[32] Kahel-Raifer, H., Jindou, S., Bahari, L., Nataf, Y., Shoham, Y., Bayer, E.A., Borovok, I. and Lamed, R. (2010) The unique set of putative membrane-associated anti-sigma factors in *Clostridium thermocellum* suggests a novel extracellular carbohydrate-sensing mechanism involved in gene regulation. FEMS Microbiol. Lett. 308, 84–93.
[33] Koeck, D.E., Koellmeier, T., Zverlov, V.V., Liebl, W., and Schwarz, W.H. (2015) Differences in biomass degradation between newly isolated environmental strains of *Clostridium thermocellum* and heterogeneity in the size of the cellulosomal scaffoldin. Syst. Appl. Microbiol. http://dx.doi.org/10.1016/j.syapm.2015.06.005.
[34] Kruus, K., Lua, A.C., Demain, A.L. and Wu, J.H.D. (1995) The anchorage function of CipA (CelL), a scaffolding protein of the *Clostridium thermocellum* cellulosome. Proc. Natl. Acad. Sci. U.S.A. 92, 9254–9258.
[35] Leibovitz, E. and Beguin, P. (1996) A new type of cohesin domain that specifically binds the dockerin domain of the *Clostridium thermocellum* cellulosome-integrating protein CipA. J. Bacteriol. 178, 3077–3084.
[36] Lytle, B., Myers, C., Kruus, K. and Wu, J.H.D. (1996) Interactions of the CelS binding ligand with various receptor domains of the *Clostridium thermocellum* cellulosomal scaffolding protein, CipA. J. Bacteriol. 178, 1200–1203.
[37] Mishara, S., Beguin, P. and Aubert, J.-P. (1991) Transcription of *Clostridium thermocellum* endoglucanase genes *celF* and *celD*. J. Bacteriol. 173, 80–85.
[38] Nataf, Y., Bahari, L., Kahel-Raifer, H., Borovok, I., Lamed, R., Bayer, E.A., Sonenshein, A.L. and Shoham, Y. (2010) *Clostridium thermocellum* cellulosomal genes are regulated by extracytoplasmic polysaccharides *via* alternative sigma factors. Proc. Natl. Acad. Sci. U.S.A. 107, 18646–18651.
[39] Nataf, Y., Yaron, S., Stahl, F., Lamed, R., Bayer, E.A., Scheper, T.H., Sonenshein, A. L. and Shoham, Y. (2009) Cellodextrin and laminaribiose ABC transporters in *Clostridium thermocellum*. J. Bacteriol. 191, 203–209.
[40] Newcomb, M., Millen, J., Chen, C.Y. and Wu, J.H. (2011) Co-transcription of the *celC* gene cluster in *Clostridium thermocellum*. Appl. Microbiol. Biotechnol. 90, 625–634.
[41] Nochur, S.V., Demain, A.L. and Roberts, M.F. (1992) Carbohydrate utilization by *Clostridium thermocellum*: importance of internal pH in regulating growth. Enzyme Microb. Technol. 14, 338–349.
[42] Nochur, S.V., Demain, A.L. and Roberts, M.F. (1990) Mutation of *Clostridium thermocellum* in the presence of certain carbon sources. FEMS Microbiol. Lett. 71, 199–204.
[43] Nochur, S.V., Jacobson, G.R., Roberts, M.F. and Demain, A.L. (1992) Mode of sugar phosphorylation in *Clostridium thermocellum*. Appl. Biochem. Biotechnol. 33, 33–41.
[44] Olson, D.G. and Lynd, L.R. (2012) Transformation of *Clostridium thermocellum* by electroporation. Methods Enzymol. 510, 317–330.
[45] Pages, S., Belaich, A., Belaich, J.-P., Morag, E., Lamed, R., Shoham, Y. and Bayer, E.A. (1997) Species-specificity of the cohesin-dockerin interaction between *Clostridium thermocellum* and *Clostridium cellulolyticum*: prediction of specificity determinants of the dockerin domain. Proteins 29, 517–527.
[46] Raman, B., McKeown, C.K., Rodriguez Jr., M., Brown, S.D. and Mielenz, J.R. (2011) Transcriptomic analysis of *Clostridium thermocellum* ATCC 27405 cellulose fermentation. BMC Microbiol. 11, 134–149.
[47] Raman, B., Pan, C., Hurst, G.B., Rodriguez Jr., M., McKeown, C.K., Lankford, P.K., Samatova, N.F. and Mielenz, J.R. (2009) Impact of pretreated Switchgrass and biomass carbohydrates on *Clostridium thermocellum* ATCC 27405 cellulosome composition: a quantitative proteomic analysis. PLoS One 4, e5271.
[48] Riederer, A., Takasuka, T.E., Makino, S., Stevenson, D.M., Bukhman, Y.V., Elsen, N.L. and Fox, B.G. (2011) Global gene expression patterns in *Clostridium thermocellum* as determined by microarray analysis of chemostat cultures on cellulose or cellobiose. Appl. Environ. Microbiol. 77, 1243–1253.

[49] Salamitou, S., Raynaud, O., Lemaire, M., Coughlan, M., Beguin, P. and Aubert, J.-P. (1994) Recognition specificity of the duplicated segments present in *Clostridium thermocellum* endoglucanase CelD and in the cellulosome-integrating protein CipA. J. Bacteriol. 176, 2822–2827.
[50] Sambrook, J. and Russell, D.W. (2001) Molecular Cloning: A Laboratory Manual, third ed, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y..
[51] Shoham, Y., Lamed, R. and Bayer, E.A. (1999) The cellulosome concept as an efficient microbioal strategy for the degradation of insoluble polysaccharides. Trends Microbiol. 7, 275–281.
[52] Smith, S.P. and Bayer, E.A. (2013) Insights into cellulosome assembly and dynamics: from dissection to reconstruction of the supramolecular enzyme complex. Curr. Opin. Struct. Biol. 23, 686–694.
[53] Stevenson, D.M. and Weimer, P.J. (2005) Expression of 17 genes in *Clostridium thermocellum* ATCC 27405 during fermentation of cellulose or cellobiose in continuous culture. Appl. Environ. Microbiol. 71, 4672–4678.
[54] Wei, H., Fu, Y., Magnusson, L., Baker, J.O., Maness, P.C., Xu, Q., Yang, S., Bowersox, A., Bogorad, I., Wang, W., Tucker, M.P., Himmel, M.E. and Ding, S.Y. (2014) Comparison of transcriptional profiles of *Clostridium thermocellum* grown on cellobiose and pretreated yellow poplar using RNA-Seq. Front. Microbiol. 5, 142.
[55] Yaniv, O., Fichman, G., Borovok, I., Shoham, Y., Bayer, E.A., Lamed, R., Shimon, L. J. and Frolow, F. (2014) Fine-structural variance of family 3 carbohydrate-binding modules as extracellular biomass-sensing components of *Clostridium thermocellum* anti-sigmal factors. Acta Crystallogr. D Biol. Crystallogr. 70, 522–534.
[56] Yaniv, O., Morag, E., Borovok, I., Bayer, E.A., Lamed, R., Frolow, F. and Shimon, L. J. (2013) Structure of a family 3a carbohydrate-binding module from the cellulosomal scaffoldin CipA of *Clostridium thermocellum* with flanking linkers: implications for cellulosome structure. Acta Crystallogr. Sect. F: Struct. Biol. Cryst. Commun. 69, 733–737.
[57] Yaron, S., Morag, E., Bayer, E.A., Lamed, R. and Shoham, Y. (1995) Expression, purification and subunit-binding properties of cohesins 2 and 3 of the *Clostridium thermocellum* cellulosome. FEBS Lett. 360, 121–124.
[58] Zhang, Y.H. and Lynd, L.R. (2005) Regulation of cellulase synthesis in batch and continuous cultures of *Clostridium thermocellum*. J. Bacteriol. 187, 99–106.