

RESEARCH ARTICLE

Decoding Biomass-Sensing Regulons of *Clostridium thermocellum* Alternative Sigma-I Factors in a Heterologous *Bacillus subtilis* Host System

Iván Muñoz-Gutiérrez¹, Lizett Ortiz de Ora¹, Inna Rozman Grinberg¹, Yuval Garty¹, Edward A. Bayer², Yuval Shoham³, Raphael Lamed¹, Ilya Borovok^{1*}

1 Department of Molecular Microbiology and Biotechnology, Tel Aviv University, Tel Aviv, 69978, Israel,

2 Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot, 76100, Israel,

3 Department of Biotechnology and Food Engineering, Technion-IIT, Haifa, 32000, Israel

☉ These authors contributed equally to this work.

✉ Current address: Department of Biochemistry and Biophysics, Stockholm University, SE-10691, Stockholm, Sweden

* ilyabo@tauex.tau.ac.il



OPEN ACCESS

Citation: Muñoz-Gutiérrez I, Ortiz de Ora L, Rozman Grinberg I, Garty Y, Bayer EA, Shoham Y, et al. (2016) Decoding Biomass-Sensing Regulons of *Clostridium thermocellum* Alternative Sigma-I Factors in a Heterologous *Bacillus subtilis* Host System. PLoS ONE 11(1): e0146316. doi:10.1371/journal.pone.0146316

Editor: Marie-Joelle Viroille, University Paris South, FRANCE

Received: October 22, 2015

Accepted: December 15, 2015

Published: January 5, 2016

Copyright: © 2016 Muñoz-Gutiérrez et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: IMG was supported with a postdoctoral fellowship by the Consejo Nacional de Ciencia y Tecnología (CONACYT proposal number 238334) [<http://www.conacyt.mx/>]. This research was supported by the Israel Science Foundation (grant No. 24/11 to RL and IB; grant Nos. 715/12, 500/10 to YS, and 1349/13 to EAB) [<http://www.isf.org.il/>] with additional support issued to RL and IB by The Sidney E. Frank Foundation through the ISF. The authors

Abstract

The Gram-positive, anaerobic, cellulolytic, thermophile *Clostridium (Ruminiclostridium) thermocellum* secretes a multi-enzyme system called the cellulosome to solubilize plant cell wall polysaccharides. During the saccharolytic process, the enzymatic composition of the cellulosome is modulated according to the type of polysaccharide(s) present in the environment. *C. thermocellum* has a set of eight alternative RNA polymerase sigma (σ) factors that are activated in response to extracellular polysaccharides and share sequence similarity to the *Bacillus subtilis* σ^1 factor. The aim of the present work was to demonstrate whether individual *C. thermocellum* σ^1 -like factors regulate specific cellulosomal genes, focusing on *C. thermocellum* σ^{16} and σ^{13} factors. To search for putative σ^{16} - and σ^{13} -dependent promoters, bioinformatic analysis of the upstream regions of the cellulosomal genes was performed. Because of the limited genetic tools available for *C. thermocellum*, the functionality of the predicted σ^{16} - and σ^{13} -dependent promoters was studied in *B. subtilis* as a heterologous host. This system enabled observation of the activation of 10 predicted σ^{16} -dependent promoters associated with the *C. thermocellum* genes: *sigI6* (itself, Clo1313_2778), *xyn11B* (Clo1313_0522), *xyn10D* (Clo1313_0177), *xyn10Z* (Clo1313_2635), *xyn10Y* (Clo1313_1305), *cel9V* (Clo1313_0349), *cseP* (Clo1313_2188), *sigI1* (Clo1313_2174), *cipA* (Clo1313_0627), and *rsgI5* (Clo1313_0985). Additionally, we observed the activation of 4 predicted σ^{13} -dependent promoters associated with the *C. thermocellum* genes: *sigI3* (itself, Clo1313_1911), *pl11* (Clo1313_1983), *ce12* (Clo1313_0693) and *cipA*. Our results suggest possible regulons of σ^{16} and σ^{13} in *C. thermocellum*, as well as the σ^{16} and σ^{13} promoter consensus sequences. The proposed -35 and -10 promoter consensus elements of σ^{16} are CNNAAA and CGAA, respectively. Additionally, a less conserved CGA sequence next to the C in the -35 element and a highly conserved AT sequence three bases

also acknowledge support from the Israeli Center of Research Excellence (I-CORE Center No. 152/11) Program of the Planning and Budgeting Committee, the Ministry of Environmental Protection 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. EAB also appreciates the support of the European Union, Area NMP.2013.1.1-2: Self-assembly of naturally occurring nanosystems: CellulosomePlus Project number: 604530 and an ERA-IB Consortium (EIB.12.022), acronym FiberFuel. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

downstream of the -10 element were also identified as important nucleotides for promoter recognition. Regarding σ^{13} , the proposed -35 and -10 promoter consensus elements are CCCYYAAA and CGWA, respectively. The present study provides new clues for understanding these recently discovered alternative σ^1 factors.

Introduction

Clostridium (Ruminiclostridium) thermocellum is a Gram-positive, anaerobic, cellulolytic thermophile that produces one of the most efficient enzymatic systems to digest cellulose [1]. The cellulolytic capacities of *C. thermocellum* have been the subject of study for many years [2], and the main motivation in these efforts has been the production of high-value products, such as ethanol, from cellulosic wastes [3]. To solubilize such carbohydrates, *C. thermocellum* secretes a multi-enzyme complex termed the cellulosome that is anchored to the cell surface [4,5]. Although during the exponential phase of growth most of the cellulosomes are cell-associated, part of them are released from the cells into the milieu [4,6,7].

The *C. thermocellum* cellulosome consists of a nonhydrolytic scaffoldin subunit CipA that integrates various catalytic subunits into the complex [8,9]. Depending on the *C. thermocellum* strain, the scaffoldin can attach 8 or 9 catalytic subunits; e.g., the CipA scaffoldin of strain DSM 1313 attaches 8 catalytic subunits, whereas that of ATCC 27405 attaches 9 catalytic subunits [10]. Additionally, the scaffoldin subunit has a family 3 carbohydrate-binding module (CBM3) that binds the cellulosome to cellulose [8,11]. *C. thermocellum* can express over 80 different cellulosomal components encoded in its genome, which include an arsenal of different saccharolytic enzymes, such as, cellulases, hemicellulases, pectin-degrading enzymes and a chitinase [12,13]. This battery of enzymes helps *C. thermocellum* to unwrap its preferred substrate, cellulose, that is covered with different types of polysaccharides in the plant cell wall [5]. During the saccharolytic process, the enzymatic content of the cellulosome is adjusted to suit the type of polysaccharide present in the biomass [14–16]. Hence, *C. thermocellum* should possess biomass-sensing mechanisms that allow the cells to detect which polysaccharide(s) is(are) present in the environment and regulate the relevant genes accordingly the enzymatic requirements. At present, however, the regulation of cellulosomal genes is poorly understood.

During the course of our efforts to gain knowledge about the biomass-sensing mechanisms in *C. thermocellum*, our research group discovered a collection of eight alternative σ factors and their cognate membrane-associated anti- σ factors that may play a role in regulating genes encoding cellulosomal enzymes and other proteins [17]. In the *C. thermocellum* genome, these alternative σ factor genes are positioned adjacent to their anti- σ factor genes in an operon-like organization [17]. This set of eight alternative σ factors (*C. thermocellum* σ^{11} to σ^{18}) are related to the *B. subtilis* σ^1 [17,18], and the expression of six of them (σ^{11} to σ^{16}) was shown to be influenced by the presence of polysaccharides (e.g., cellulose and xylan) in the growth medium [19]. Furthermore, a recent study performed by Wei and colleagues [16] showed that *C. thermocellum* *sigI3-rsgI3*, *sigI4-rsgI4* and *sigI7-rsgI7* operons are up-regulated when the bacterium was grown in dilute acid-pretreated yellow poplar. Additionally, *in vitro* experiments showed that σ^{11} directed the transcription from *sigI1* promoter and from the promoter of the gene *cel48S* [19] that encodes for the most abundant cellulosomal enzyme Cel48S [12,20].

The *C. thermocellum* anti- σ^1 factors of σ^{11} to σ^{16} (RsgI1 to RsgI6) embody three domains: (I) a C-terminal carbohydrate-binding module (CBM) localized on the outer cell surface, (II) an internal transmembrane/wall-spanning segment, and (III) an N-terminal cytoplasmic portion

(RsgI-N) which would bind the cognate σ^I factor [17,18]. The N-terminal segments (~165 residues) of the *C. thermocellum* RsgI proteins resemble *B. subtilis* RsgI, a negative regulator of its cognate σ^I factor [17,18]. Moreover, the binding capacities of the N-terminal cytoplasmic portions of RsgI1, RsgI2 and RsgI6 to their corresponding σ^I factors was demonstrated *in vitro* [19].

The C-terminal domains of the RsgIs showed binding capacities to different polysaccharides, including cellulose (RsgI1, RsgI2, RsgI4 and RsgI6), xylan (RsgI6), and pectin (RsgI3) [17,21,22]. Additionally, the crystal structures of the C-terminal CBMs of RsgI1, RsgI2 and RsgI4 were solved showing a high degree of similarity to the family 3 CBMs [22]. In the case of RsgI3, its C-terminal CBM is constituted by two tandem PA14-superfamily motifs (pfam07691, smart00758) that are found in a wide variety of other bacterial and eukaryotic proteins, which include the anthrax protective antigen (PA) [23], and the PA14 modular dyad was predicted to be a putative CBM by virtue of its binding to pectin-like polysaccharides [17]. Interestingly, the C-terminal domain of RsgI6 belongs to the glycoside hydrolase family 10 (GH10), however, its catalytic activity was shown to be very low [17,21]. Nevertheless, RsgI6-GH10 retains its binding capacity to its corresponding carbohydrates, suggesting an evolutionary adaptation to function as a polysaccharide-binding domain rather than an authentic enzymatic component [21].

The multiple *C. thermocellum* alternative σ^I factors resemble to some extent the ECF (extra-cytoplasmic function) σ factors [24–26], since they share common characteristics which include the following: (I) both kinds of σ factors autoregulate their own expression; (II) both kinds of σ factors are usually co-transcribed with another ORF encoding a transmembrane anti- σ factor that controls the activity of its cognate σ factor; (III) the anti- σ factor is composed of an extracytoplasmic sensory domain and an intracellular inhibitory domain that binds the σ factor; (IV) the activity of the σ factor is induced by inhibiting activity of the anti- σ factor [18,25]. We assume that the main difference between σ^I -like factors and ECF σ factors is related to their "architectures" (Muñoz-Gutierrez et al, unpublished). While the ECF σ factors are formed with only two of the four domains of the σ^{70} family of proteins (σ_2 and σ_4) [25,26], the σ^I -like factors have only one predictable functional domain associated with the amino-terminal sequence, σ_2 , and the sigma domain σ_4 is absent [17]. In lieu of the sigma domain σ_4 , the σ^I factors contain a novel 100-residue conserved C-terminal domain termed σ_{I-C} [17], that might serve to recognize -35 sequences of the σ^I promoters.

Until now, the knowledge we have regarding the regulation of cellulosomal genes by *C. thermocellum* σ^I -like factors is a recent report of Sand and co-workers [27] which showed that the xylanase genes *xyn10Z* (or Clo1313_2635 according to the DSM 1313 genome annotation), *xyn11B* (Clo1313_0522) and *xyn10D* (Clo1313_0177) were under the control of σ^{16} . Previously, Nataf and co-workers [19] showed that the cellulase gene *celS* (*cel48S*) was likely under the control of σ^{11} . Therefore, the present work was devoted to demonstrating whether individual *C. thermocellum* σ^I factors regulate specific cellulosomal target genes. Taking advantage of the fact that the transcription start sites of *C. thermocellum* *sigI6*, *xyn10Z* and *xyn11B* were mapped previously in our research group [19,27], we performed a bioinformatics analysis to identify σ^I -dependent promoters in the genome of *C. thermocellum* DSM 1313. This analysis allowed us to identify 40 possible σ^I -dependent promoters upstream of the *sigI*-like genes and certain cellulosomal genes of *C. thermocellum*. To corroborate the functionality of the 40 predicted promoters, we fused their DNA sequences to a promoterless *lacZ* reporter gene. To overcome the lack of genetic tools in *C. thermocellum*, we used a *B. subtilis* Δ (*sigI-rsgI*) strain as a heterologous host and studied the activation of the 40 predicted promoters by *C. thermocellum* σ^{16} and σ^{13} . This strategy allowed us to show that *C. thermocellum* σ^{16} could recognize the predicted promoters associated with *sigI6*, *sigI1*, *rsgI5*, *xyn11B*, *xyn10D*, *xyn10Z*, *xyn10Y*, *cel9V*, *cseP* and the

major scaffoldin *cipA*. Additionally, *C. thermocellum* σ^{13} could recognize the predicted promoters detected upstream of *sigI3*, *pl11* (Clo1313_1983 encodes a family 11 polysaccharide lyase (PL11) containing a CBM35 and a dockerin), *ce12* (Clo1313_0693 encodes for a protein that contains two family 12 carbohydrate esterase (CE12), a CBM35 and a dockerin) and *cipA*. The combination of these methodologies revealed a putative *C. thermocellum* σ^{16} and σ^{13} promoter consensus. Our results show that *C. thermocellum* σ^{16} and σ^{13} factors expressed in *B. subtilis* can recognize its potential promoters, supporting our hypothesis that the multiple *C. thermocellum* σ^1 -like factors might regulate cellulosomal genes.

Material and Methods

Bacterial strains, growth media and culture conditions

C. thermocellum strain DSM 1313 (LQ8) was obtained from the DSMZ (German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany). The *B. subtilis* strains constructed in this work are isogenic derivatives of the *B. subtilis* strain PY79 (laboratory stock) [28]. Additional information regarding all derivatives of *B. subtilis* PY79 that were constructed in this work is shown in S3 Table. *B. subtilis* BKE13460 was obtained from the BGSC (Bacillus Genetic Stock Center, Ohio, USA). *Escherichia coli* DH5 α (BioSuper Competent Cells, Bio-Lab Ltd, Jerusalem, Israel) was used for plasmid propagation during plasmid construction.

E. coli and *B. subtilis* were grown routinely at 37°C in liquid (at 250 rpm) or on solid LB-agar Broth (Lennox, Difco, BD Diagnostics, Maryland, USA). During β -galactosidase activity assays, *B. subtilis* was grown in Spizizen's minimal medium (SMM) employing 5 g/L fructose as carbon source and supplemented with trace elements. The SMM contained (per liter) 2 g (NH₄)₂SO₄, 14 g K₂HPO₄, 6 g KH₂PO₄, 1 g Na₃Citrate·2H₂O, and 0.2 g MgSO₄·7H₂O. The trace elements used were (per liter) 125 mg MgCl₂·6H₂O, 5.5 mg CaCl₂, 13.5 mg FeCl₂·6H₂O, 1 mg MnCl₂·4H₂O, 1.7 mg ZnCl₂, 0.43 mg CuCl₂·2H₂O, 0.6 mg CoCl₂·6H₂O, and 0.6 mg Na₂MoO₄·2H₂O. When appropriate, antibiotics were included at the following final concentrations: 100 μ g/mL ampicillin (Amp), 50 μ g/mL kanamycin (Kan), 100 μ g/mL spectinomycin (Spt), 5 μ g/mL chloramphenicol (Cam) or 3 μ g/mL erythromycin (Erm). The induction of genes under the P_{xyIA} promoter was carried out with D-xylose (10 g/L final). All chemicals were purchased from Sigma-Aldrich (Missouri, USA).

DNA manipulation techniques

The oligonucleotide primers used in the present study are shown in S1 Table. Standard procedures were employed for DNA isolation, polymerase chain reaction (PCR), restriction-enzyme digestion, dephosphorylation, transformations, and gel electrophoresis as described elsewhere [29]. Plasmids were built using a combination of standard molecular cloning techniques [29] and ligase-independent cloning using the In-Fusion HD Cloning Kit (Clontech Laboratories, Inc., California, USA). *C. thermocellum* DNA sequences were PCR-amplified using *C. thermocellum* DSM 1313 genomic DNA as template. The upstream and downstream regions of the *B. subtilis* *sigI-rsgI* operon were PCR-amplified using *B. subtilis* PY79 genomic DNA as template. The *lox71-erm-lox66* cassette was PCR-amplified using *B. subtilis* BKE13460 genomic DNA as template. Amplification of DNA for cloning was performed using TaKaRa Ex Taq (Takara Bio Inc., Shiga, Japan). Colony PCR was performed using Hy-Taq Ready Mix (Hy Laboratories Ltd, Rehovot, Israel). PCR primers were purchased from hy-labs (Hy Laboratories Ltd). Restriction enzymes, alkaline phosphatase, and ligase were purchased from Fermentas (Thermo Fisher Scientific Inc., Massachusetts, USA). PCR and agarose-gel products were isolated and purified using the hy-labs Gel/PCR Extraction Kit (Hy Laboratories Ltd). Purification of plasmids was carried out using the Presto™ Mini Plasmid Kit (Geneaid Biotech Ltd., Shijr,

Taiwan). All clones were verified by PCR and sequencing in the Instrumentation and Service Center of the Life Sciences Faculty at Tel Aviv University.

Construction of plasmids

Plasmids constructed in the present work are listed in [S2 Table](#). The pLOXErysigIrgIBs plasmid was constructed to knockout the *B. subtilis sigI-rsgI* operon including its promoter using resistance to Erm as a selective marker. The upstream (464 bp) and downstream (505 bp) regions of the *B. subtilis sigI-rsgI* operon were PCR-amplified using primer pairs P1-P2 and P3-P4, respectively. The *lox71-erm-lox66* cassette was amplified using primer pair P5-P6. Subsequently, the three PCR products were cloned simultaneously with the In-Fusion HD Cloning Kit into a linearized pUC19 vector (generated by PCR and provided with the kit) following the kit protocol, obtaining the pLOXErysigIrgIBs plasmid ([S1 Fig](#)).

To express the *C. thermocellum* σ^{16} and σ^{13} factors in *B. subtilis*, we used the pAX01 plasmid [[30](#)]. This vector was designed for integration at the *B. subtilis lacA* chromosomal locus, carries an *erm* resistance cassette as a selectable marker, and has the xylose-inducible promoter P_{xylA} . First, pAX01 was linearized with the restriction enzyme BamHI. Subsequently, the DNA sequence of *C. thermocellum sigI6* and *sigI3* were PCR-amplified using primer pair P7-P8 and P9-P10, respectively ([S1 Table](#)). Finally, the PCR products were cloned using the In-Fusion HD Cloning Kit into the linearized pAX01 vector, obtaining the pAX01-sigI6 and pAX01-sigI3 plasmids.

To study the promoters that are under the control of *C. thermocellum* σ^{16} and σ^{13} , we used the pBS1ClacZ plasmid that contains a promoterless *lacZ* reporter gene [[31](#)]. This vector was designed to integrate at the *B. subtilis amyE* locus and carries a *cat*-resistance cassette as a selectable marker. The upstream region of the *C. thermocellum* σ^1 -factor genes that contain the predicted promoter and the upstream region of some cellulosomal genes that contain predicted σ^1 -dependent promoters were PCR-amplified, using the primer pairs listed in [S1 Table](#) (primers P11 to P90). Subsequently, each PCR product was digested with restriction enzymes EcoRI and BamHI. Finally, each digested PCR was cloned into the pBS1ClacZ plasmid that was digested previously with the same restriction enzymes, thus obtaining the pBS1ClacZ derived plasmids listed in [S2 Table](#).

In order to study the important bases for promoter recognition by *C. thermocellum* σ^{16} , mutant versions of the *xyn10Z* σ^{16} -dependent promoter were created by site-directed mutagenesis. To introduce mutations in the conserved bases of the -35 element, the forward primers from P93 to P99, which contain the mutated nucleotides, were used with the reverse primer P92 ([S1 Table](#)). To introduce individual mutations in the conserved bases of the -10 element, the reverse primers from P100 to P105, which contain the mutated nucleotides, were used with the forward primer P91 ([S1 Table](#)). In order to compare the mutant version, a short version of the *xyn10Z* σ^{16} -dependent promoter with the same length of the mutant versions was PCR-amplified using the primer pair P91-P92 ([S1 Table](#)). Subsequently, each PCR product was digested with restriction enzymes EcoRI and BamHI. Finally, each digested PCR was cloned into the pBS1ClacZ plasmid that was digested previously with the same restriction enzymes, thereby obtaining the pBS1ClacZ-derived plasmids listed in [S2 Table](#).

Construction of *B. subtilis* strains

B. subtilis was transformed by using the natural competence method [[32](#)]. Chromosomal integration of plasmids by a double-crossover event was confirmed by colony PCR using the primer pairs listed in [S1 Table](#) (primers P106 to P115). The different *B. subtilis* strains obtained

were stored at -80°C in 20% (v/v) glycerol. The strains constructed in the present work are listed in [S3 Table](#).

To construct a *B. subtilis* PY79 devoid of its *sigI-rsgI* operon, *B. subtilis* PY79 was first transformed with the pLOXerysigI_{rsgI}Bs plasmid, and the cells were selected with Erm, obtaining the *B. subtilis* CO01 strain. Subsequently, *B. subtilis* CO01 was transformed with the pDR244 plasmid (obtained from the BGSC) that encodes the sequence of the Cre recombinase and has a thermosensitive origin of replication. The cells were plated on LB-agar containing Spt and were incubated at 30°C . Several individual colonies were then streaked on a plain LB-agar plate and incubated overnight at 42°C to cure pDR244. The resulting colonies were screening for plasmid curing (Spt sensitivity) and the loss of the *lox71-erm-lox66* cassette (Erm sensitivity). Finally, a single colony was streaked on plain LB-agar plate and grown at 37°C . The loss of the *lox71-erm-lox66* cassette was confirmed by PCR with the primer pair P118-P119 ([S1 Table](#)) thus obtaining the *B. subtilis* CO02 strain ([S3 Table](#)).

β -Galactosidase activity assays

To measure the β -galactosidase activity, strain samples were taken from the -80°C glycerol stock and inoculated in 5 mL of SMM with Cam. Subsequently, the cells were grown overnight at 37°C with shaking (250 rpm). The next day, the cells were inoculated in 2.5 mL of SMM to an OD_{600} between 0.1–0.2 and grown at 37°C (250 rpm). When the cells reached mid-log growth phase (approx. 0.4–0.5 OD_{600}), the culture was separated into two tubes, and one tube was supplemented with xylose (1% final concentration) whereas the other was used as a blank. Then, the cells were allowed to grow for another hour at 37°C (250 rpm). Finally, the cells were recovered by centrifuging at 16,000 g for 5 min, washed twice with Z-buffer (60 mM $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$, 40 mM NaH_2PO_4 , 10 mM KCl, 1 mM $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, and 50 mM β -mercaptoethanol, pH 7.0) and recovered in 0.5 mL of Z-buffer.

Enzymatic activity was measured with the fluorogenic substrate (4 mg/mL) 4-methylumbelliferyl β -D-galactopyranoside (4-MUG, Sigma-Aldrich) in a microplate reader (Biotek Synergy HT, Vermont, USA). The cells, recovered in Z-buffer (150 μL), were placed in a 96-well plate and 2 mL of MUG (4 mg/mL) were added to initiate the enzymatic reaction. The release of the fluorescent compound 4-methylumbelliferone (4-MU) was measured (using the excitation filter 360/40 and the emission filter 460/40) every 10 min at 30°C with medium agitation for one hour. The reaction was stopped by adding 100 μL of 1M Na_2CO_3 . To calculate the β -galactosidase activity, a standard curve with 4-MU was prepared. One unit of enzyme activity was defined as the amount of β -galactosidase that releases 1 μmol of 4-MU per minute. All the β -galactosidase activities were normalized with cell density (OD_{600}).

Bioinformatics analysis

Primary DNA sequence analyses and DNA motif searches were performed using the Clone Manager 9 Professional Edition software (Scientific & Educational Software, Durham, NC). The *B. subtilis* 168 and *C. thermocellum* DSM 1313 *sigI* genes and their promoter sequences (extracted from GenBank NZ_CP010052.1 and NC_017304.1, respectively) were used as BLAST [33] queries to mine public databases including that at the National Center for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov/>). In order to prevent a possible loss of promoter candidates during BLAST mining, we used both discontinuous megablast ("more dissimilar sequences") as well as blastn ("somewhat similar sequences") as implemented at NCBI (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Pairwise and multiple sequence alignments were performed with the CLUSTALW program [34] using either the Network Protein Sequence Analysis server (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/

[npsa_clustalw.html](#)), or the ClustalW2 at the EMBL-EBI (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>). WebLogos [35] were generated by using a public logo generator web application (<http://weblogo.berkeley.edu/>).

Results

Bioinformatics comparison of promoter sequences of alternative σ^I factor from various cellulosome-producing bacteria and *Bacillales* species

In order to identify the conserved sequence motifs that could be used for the analysis of putative *C. thermocellum* *sigI*-like gene promoters, we performed multiple sequence alignments of the experimentally detected σ^I -dependent promoter sequences. The initial analysis was performed using the *sigI* promoter sequences experimentally identified in *B. subtilis*, *Bacillus licheniformis* ATCC 14580, *Bacillus thuringiensis* serovar israelensis ATCC 35646 and *Bacillus* sp. strain NRRL B-14911 [18,36]. Additionally, the experimentally identified promoters of the σ^I -dependent genes *bcrC* and *mreBH*, which are involved in cell envelope integrity and homeostasis during heat stress in *B. subtilis* [36], were also included. The analysis was improved by including DNA sequences located immediately upstream of *sigI*-like genes in various species of the order *Bacillales*. The alignment is shown in S4 Table, and a high conservation of two short DNA sequences upstream of the *Bacillales sigI-rsgI* operons can be observed. These basic putative promoter motifs can also be observed in Fig 1A that shows a WebLogo generated with the *Bacillales* σ^I -dependent promoters shown in S4 Table. As already proposed by Tseng and Shaw [36], the suggested *Bacillales* σ^I promoter consensus sequence is ACCCCC for the -35 element and CGAA for the -10 element (Fig 1A and S4 Table). Interestingly, a conserved sequence AA downstream of the -35 element (already mentioned by Tseng and Shaw [36]), and a conserved T downstream of the -10 element can also be observed (Fig 1A and S4 Table). For future comparisons, we named the conserved sequence AA as "extended -35".

The deduced consensus sequences of the *Bacillales sigI* promoter elements were used to find sequence similarities between the predicted promoters of the different *C. thermocellum sigI* genes. The upstream intergenic regions of the eight *C. thermocellum sigI* genes were manually analyzed, focusing on potential conservation of the consensus sequences of the *Bacillales sigI* promoter elements -35 and -10. Selected promoter candidate sequences were then used for multiple sequence alignments using the ClustalW algorithm [34]. The analysis was facilitated by the fact that the transcription start sites of *C. thermocellum sigI1* and *sigI6* genes were previously identified by Nataf and co-workers [19]. Additionally, the analysis was improved by comparing the *C. thermocellum* predicted promoter sequences to those of another cellulosome-producing species, *C. straminisolvens*, whose genome (NCBI Reference Sequence: NZ_BAVR00000000.1) has a very high similarity to that of *C. thermocellum* (96.2% similar based on 16S rDNA) [37,38]. The multiple promoter sequence alignment is presented in Table 1. The putative promoters of *C. thermocellum sigI2*, *sigI3*, *sigI4*, *sigI7*, *sigI8*, and those of *C. straminisolvens sigI1*, *sigI2*, *sigI3*, *sigI4*, *sigI6*, *sigI7* and *sigI8* were predicted. As shown in Table 1, a conserved AAA sequence for the "extended -35" element and a highly conserved CGWA for the -10 element were identified. Moreover, a highly conserved C upstream of the "extended -35" was also identified (Table 1).

It is important to mention that during this analysis we failed to predict a σ^I -dependent promoter sequence for both *C. thermocellum* and *C. straminisolvens sigI5* genes, owing to the low similarity of their upstream intergenic regions. Interestingly, whereas most of the sequences of the *rsgI* genes overlap with sequences of their cognate *sigI* genes, *sigI5* and *rsgI5* genes are separated by an intergenic region of 97 nucleotides, which contains a predicted σ^I -dependent

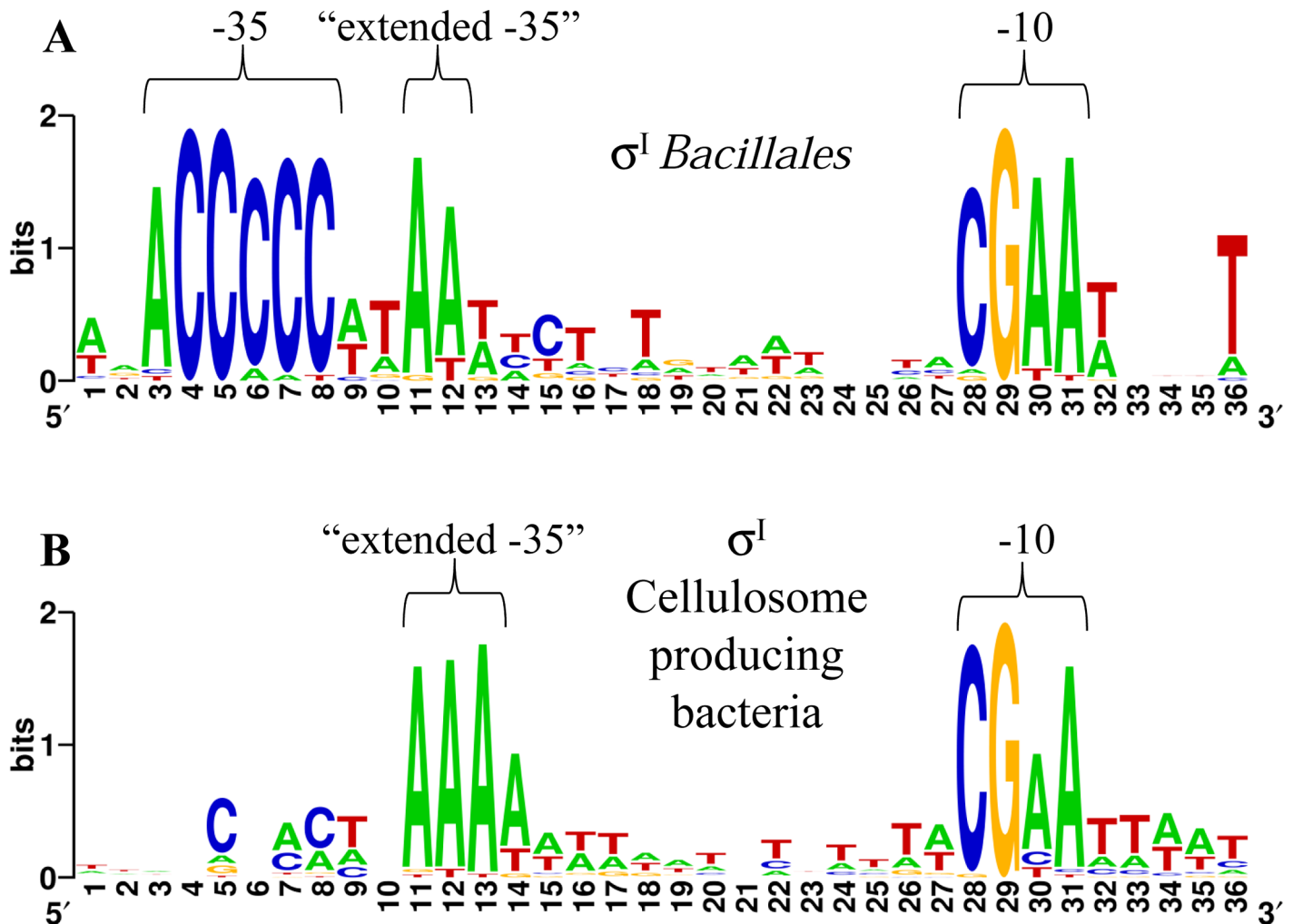


Fig 1. Identification of conserved elements of σ^I -dependent promoter sequences. (A) WebLogo generated with the *Bacillales sigI* promoters shown in [S4 Table](#). (B) WebLogo generated with the *C. thermocellum* and *C. straminisolvans sigI* promoters shown in [Table 1](#), and the *C. clariflavum*, *A. cellulolyticus* and *Pseudobacteroides cellulosolvans sigI* promoters shown in [S5 Table](#).

doi:10.1371/journal.pone.0146316.g001

promoter ([Table 2](#)). This suggests a different type of gene organization and regulation of *sigI5* and *rsgI5* in both *C. thermocellum* and *C. straminisolvans*.

To investigate how conserved are the "extended -35" and -10 elements of *C. thermocellum* and *C. straminisolvans sigI*-like gene promoters, we performed a search of *sigI-rsgI* operons using the publicly available genomic sequences of the known cellulosome-producing bacteria. The [S5 Table](#) shows the putative promoters upstream of *sigI*-like genes found during the mining. Multiple *sigI*-like genes in *Clostridium clariflavum*, *Acetivibrio cellulolyticus* and *Pseudobacteroides cellulosolvans* were found, and most of their cognate *rsgI*-like genes encode proteins containing a C-terminal CBM. As shown in [S5 Table](#), a conserved AAA sequence for the "extended -35" element and a highly conserved CGWA for the -10 element were identified in *C. clariflavum*, *A. cellulolyticus* and *P. cellulosolvans*. These results confirm the high conservation of the -35 and -10 elements in *sigI*-like promoters of taxonomically divergent cellulosome-producing bacteria. The high conservation of the "extended -35" and -10 promoter elements of cellulosome-producing bacteria is more evident in the WebLogo generated with the predicted

Table 1. Alignment of putative *C. thermocellum* and *C. straminisolvans* sigI promoters.

Gene ^a	Locus tags ^b	RsgI-C terminal domain	Promoter region 5'→3'	5' UTR
<i>Ct sigI1-rsgI1</i>	Clo1313_2174–2173, Cthe_0058–0059	CBM3	taatatac acacaa AAA aa-gcagatgtata CGaA gtaaatctact G	16
<i>Cs sigI1-rsgI1</i>	JCM21531_2023–2024	CBM3	taatatac acataa AAA aa-gcaggcttgaa CGaA gtaaatctactg	17
<i>Ct sigI2-rsgI2</i>	Clo1313_1961–1962, Cthe_0268–0267	CBM3	tggtat Cccccg AAA aaatgttcccttta CGaA ataactagtaa	147
<i>Cs sigI2-rsgI2</i>	JCM21531_2790–2789	CBM3	tgatata Ccccoct AAA atgttctctta CGaA ataacttatta	159
<i>Ct sigI3-rsgI3</i>	Clo1313_1911–1910, Cthe_0315–0316	2xPA14	tatgaa Cccctc AAA aaaatcatttgggtg CGtA caagtattgaa	13
<i>Cs sigI3-rsgI3</i>	JCM21531_2365–2366	2xPA14	tgtaaa Cccctc AAA aaa-taactttgtg CGtA caagtattaaa	15
<i>Ct sigI4-rsgI4</i>	Clo1313_1818–1817, Cthe_0403–0404	CBM3	aacgta Cagctg AAA atttctgcca CGaA attaattttttt	13
<i>Cs sigI4-rsgI4</i>	JCM21531_1812–1813	CBM3	aacgta Caacta AAA gtttgttgcacat CGaA attaatctat	13
<i>Ct sigI6-rsgI6</i>	Clo1313_2778–2777, Cthe_2120–2119	GH10	acaatg Cgacat AAA accattccggtata CGaA tcgatataa Ga	20
<i>Cs sigI6-rsgI6</i>	JCM21531_4109–4108	GH10	atgatg Cgacat AAA gctattccagtcta CGaA ttcatatagga	22
<i>Ct sigI7-rsgI7</i>	Clo1313_0104–0105, Cthe_2521–2522	UNK	attcga Ct gatg ttAtt -taaatttgggt CGaA ctttgctgatg	52
<i>Cs sigI7-rsgI7</i>	JCM21531_3721–3720	UNK	attcga C tagtgg tttg -tagatttatgt CGaA ctttgctgaca	61
<i>Ct sigI8-rsgI8</i>	Clo1313_0525–0524, Cthe_2975–2974	UNK	actttc Cgaatc AAA atgaaatccatata CGaA ttttctatagt	16
<i>Cs sigI8-rsgI8</i>	JCM21531_4043–4045	UNK	ttttac Cgaatt AAA atagaagtcata CGaA tcctctatagc	18
Consensus			C—AAA 13–14 (N) CGWA	
<i>Bacillales</i> consensus			ACCCCC—AA 15 (N) CGAA—T	

The most conserved bases (> 85%) are shown in bold capital fonts. Transcription start sites in *Ct sigI1-rsgI1* and *Ct sigI6-rsgI6*, identified by Nataf and co-workers [19], are shown in uppercase bold italics; and the promoter regions proposed by the same authors are underlined [19]. *Ct*, *Clostridium thermocellum*; *Cs*, *Clostridium straminisolvans*. W represents A or T.

^a *C. thermocellum* and *C. straminisolvans sigI5* are not included in the table. During the analysis we failed to predict a σ^1 -dependent promoter sequence for both *C. thermocellum* and *C. straminisolvans sigI5*, because of the low similarity of the upstream intergenic regions of these genes.

^b Clo1313 and Cthe are the locus tag prefixes of *C. thermocellum* strains DSM 1313 and ATCC 27405, respectively.

doi:10.1371/journal.pone.0146316.t001

promoter sequences of *C. thermocellum*, *C. straminisolvans*, *C. clariflavum*, *A. cellulolyticus* and *P. cellulosolvans* shown in Fig 1B.

Comparison between the predicted promoters of the sigI-like genes of cellulosome-producing bacteria and the promoter consensus sequence of *Bacillales sigI* genes shows a different level of similarities (Table 1 and S5 Table). For example, while the putative promoters of *C. thermocellum* and *C. straminisolvans sigI2* and *sigI3* are most similar to the promoter consensus sequence of *Bacillales sigI* genes, the putative promoters of *C. thermocellum* and *C. straminisolvans sigI7* are less similar (Table 1). This observation is quite interesting because the deduced amino acid sequence of *C. thermocellum* σ^{17} has the highest similarity to *B. subtilis* σ^1 (data not shown). Our analysis shows that the cellulosome-producing bacteria which use multiple sigI-like genes probably maintain different levels of similarity in promoter sequences to fine-tune the regulation of individual sigI-like genes, as well as cellulosomal target genes. With the predicted promoter sequences of the multiple sigI-like genes of *C. thermocellum*, *C. straminisolvans*, *C.*

Table 2. Alignment of predicted *C. thermocellum* σ^I -dependent promoters.

Gene ^a	Locus Tags ^b	Promoter region 5'→3'	5' UTR
<i>xyn10Z</i> ^c	Clo1313_2635, Cthe_1963	accgacac AAA aatgtgagcgttca CGaA acaa Ta aat At	96
<i>xyn11B</i> ^{c, d}	Clo1313_0522, Cthe_2972	agcgactt AAA aaattatatTTTTG CGaA taga Ta ata Tg	164
<i>ce8</i>	Clo1313_0500, Cthe_2949	ccccgctc AAA gttgataaacct CGaA tctta aa ataa	32
<i>ce12</i>	Clo1313_0693, Cthe_3141	taccctta AAA aaacttgcT-tctCGtAtttta at atata	51
<i>cel5E</i>	Clo1313_1425, Cthe_0797	gctgtcca AAA gaaaa-tgattttt CGaA ttaa Ta ataa	156
<i>cel8A</i>	Clo1313_1960, Cthe_0269	accctatc AAA taaccattcaattCGcAtttta Tt taag	254
<i>cel9J</i>	Clo1313_1604, Cthe_0624	gccccctt AAA aaatttta-aaatt CGaA attaa t ttttg	477
<i>cel9P</i>	Clo1313_1955, Cthe_0274	aacgtctat AA tttttt-atgataaCGataaaaa Tt aaatt	19
<i>cel9Q</i>	Clo1313_1603, Cthe_0625	accactt AAA aatgtgtatgtgcaCGgAtt t c Ta tttg	375
<i>cel9U</i> ^e	Clo1313_3023, Cthe_2360	agccccct AAA aaattttttcccttt CGaA tata Ta tagat	394
<i>cel9V</i>	Clo1313_0349, Cthe_2760	atacccat AAA atttttatgttctca CGaA tata Ta atata	124
<i>cel48S</i>	Clo1313_2747, Cthe_2089	gccccctc AAA aagtataatttttt CGaA gata Ta tatat	498
<i>cenC</i>	Clo1313_0420, Cthe_2879	cccaatcg AAA aaagaacatgtcat CGaA tcta Ta tatca	102
<i>cipA</i>	Clo1313_0627, Cthe_3077	tgccccct AAA ttccgTTtatata CGaA tata Ta ttaca	846
<i>cseP</i>	Clo1313_2188, Cthe_0044	taagccac AAA attatTT-tttctca CGaA tata Ta ttgaa	132
<i>pelB2</i>	Clo1313_0501, Cthe_2950	tcccaatg AAA tacgacccttgataCGtAttat Ta atata	67
<i>pilZ</i>	Clo1313_1490, Cthe_0733	gccccctc AAA ata-tgagaacatt CGaA atat Ta ataa	321
<i>pl11</i>	Clo1313_1983, Cthe_0246	ctaccctt AAA aaaa-ttagaattCGtAtttta Ta aaaaag	39
<i>rsgI5</i>	Clo1313_0985, Cthe_1273	atggacca AAA agtactttcaaaa CGaA atta Tt aaata	43
<i>rsgI9</i>	Clo1313_1969, Cthe_0260	ctctaaaaAtAtcgggattttttt CGaA ataac t aatag	31
<i>sdbA</i>	Clo1313_0950, Cthe_1307	caacgctcAAtacgaactctttctc CGaA ttta Tt ctatt	157
<i>xgh74A</i>	Clo1313_0851, Cthe_1398	ggtacatc AAA ggaaagtacaggtc CGaA ttta Ta tagcg	147
<i>xyn10D</i>	Clo1313_0177, Cthe_2590	tgcgacca AAA ggcgtcaaatttca CGaA atata Ta ataa	33
<i>xyn10Y</i>	Clo1313_1305, Cthe_0912	cccaacgt AAA aattcaataccttt CGaA taaa Ta acata	277
GH30-CBM6-Doc	Clo1313_0563, Cthe_3012	ccgtacat AAA aagaagttttgatt CGaA aat Ta acaca	67
GH43-2xCBM6-Doc	Clo1313_0987, Cthe_1271	cccaaccc AAA cttgccatatgTTT CGtA caaaa Ta aattg	67
HP	Clo1313_1436, Cthe_0785	atcccccttAAgaaattgacataaaaCGcAttaac t attat	106
HP-Doc	Clo1313_1494, Cthe_0729	acggaat AAA aacaactccaatta CGaA taaa Ta tacca	35
GH43-CBM42-Doc	Clo1313_2216, Cthe_0015	cccactcc AAA aaacattttaattctCGtAttat Ta taaca	46
GH39-2xCBM35-Doc	Clo1313_2793, Cthe_2137 ^f	ctcaactt AAA aaatacatcttctCGtAtatg Ta agttta	160
GH43-CBM42-Doc	Clo1313_2794, Cthe_2138	tacgtcac AAA ccaaaaaccagaa CGaA ccaa Tt aaataa	121
GH2-CBM6-Doc	Clo1313_2861, Cthe_2197 ^f	cccaacta AAA aaaataggtacttCGtAaaa Ta aaaca	163
ABC transporter	Clo1313_2866 ^f	taacccta AAA atttaaatgccgatt CGaA taaaa aa gcct	149
Consensus		AAA 13-14 (N) CGWA—TW	
<i>Bacillales</i> consensus		ACCCCC-AA 15 (N) CGAA—T	

The most conserved bases (> 85%) are shown in bold capital fonts. W represents A or T.

^a The genes without trivial names are denoted with their main protein product domain(s). GH, glycoside hydrolase; CBM, carbohydrate-binding module; Doc, dockerin. HP, hypothetical protein.

^b Clo1313 and Cthe are the locus tag prefixes of *C. thermocellum* strains DSM 1313 and ATCC 27405, respectively.

^c Transcription start sites identified by Sand and co-workers [27] are indicated in uppercase bold italics, and the underlined sequences are the promoter sequence proposed by the same authors [27].

^d In *C. thermocellum* DSM 1313 *xynB* is part of the operon *xynB-xynA*, whereas *xynB* is absent in other strains, such as ATCC 27405 and JW20.

^e All the promoter sequences shown in the table are exactly the same for both *C. thermocellum* strains, DSM 1313 and ATCC 27405. The only small difference is in the gene *cel9U* at the 5' position. Whereas strain DSM 1313 has the sequence 5' AGCCCC**TCAAAA** 3' (bold fonts are part of the consensus in the -35 element), the sequence of strain ATCC 27405 is 5' AGCTCCCT**CAAAA** 3' (underline designates an inserted T).

^f In *C. thermocellum* strain ATCC 27405 the genes Cthe_2137 and Cthe_2197 are interrupted by an IS element. Additionally, strain ATCC 27405 does not contain a gene orthologous to Clo1313_2866 (an ABC transporter ATP-binding protein).

doi:10.1371/journal.pone.0146316.t002

clariflavum, *A. cellulolyticus* and *P. cellulosolvens* we suggest that AAA of the "extended -35" and CGWA of the -10 elements represent the general motifs for σ^I -dependent promoters of cellulosome-producing bacteria.

Searching for σ^I -dependent promoter sequences of cellulosomal genes in *C. thermocellum*

Based on the assumption that σ^I factors autoregulate their own expression, and hence the genes that are under their control should have similar promoter sequences, we performed a search of putative promoter sequences of the cellulosomal genes of *C. thermocellum*. The search was performed by exploiting the conserved sequences in the "extended -35" (AAA) and -10 (CGWA) elements of the general motifs in the *sigI*-like gene promoters of cellulosome-producing bacteria. The analysis was facilitated with the recent identification of the transcriptional start sites of *C. thermocellum xyn10Z* and *xyn11B* by Sand and co-workers [27]. In Table 2 are listed the 33 putative predicted promoters that were identified during the analysis. Additionally, Table 2 shows the conserved AAA sequence for the "extended -35" element and the conserved CGWA for the -10 element. Interestingly, a highly conserved TW dinucleotide (W represents A or T), downstream of the -10 element, was also identified (Table 2).

Searching for σ^{16} -dependent promoter sequences of cellulosomal genes in *C. thermocellum*

Given the limited genetic tool set available for *C. thermocellum*, we used *B. subtilis* as a heterologous host to test the ability of *C. thermocellum* σ^{16} and σ^{13} factors to recognize the *C. thermocellum* predicted promoters. A similar strategy has been successfully used by several research groups to analyze regulatory proteins from different Firmicutes species as *Clostridium difficile*, *Enterococcus faecalis* and *Oceanobacillus iheyensis* [39–42]. Additionally, the high homology presented by RNAPs of *B. subtilis* and *C. thermocellum* (e.g., more than 67% of identical residues for subunits α , β and β' ; see S2 Fig) gave more support to this approach.

To avoid the interference of the native *B. subtilis* σ^I during the study of *C. thermocellum* σ^I factors, we constructed the *B. subtilis* CO02 strain which is devoid of its *sigI-rsgI* operon (S3 Table). The present work was first focused on the activation of putative promoter sequences by *C. thermocellum* σ^{16} factor. This initial analysis was facilitated by the fact that the *C. thermocellum* σ^{16} promoter was previously identified by Nataf and co-workers [19] and that the xylanase genes *xyn10Z*, *xyn11B* and *xyn10D* were shown to be under the control of σ^{16} by Sand and co-workers [27]. The 7 predicted promoters of *C. thermocellum sigI*-like genes (Table 1) and the 33 *C. thermocellum* σ^I -dependent predicted promoters (Table 2) were fused to a *lacZ* reporter gene (S2 Table) and integrated into the *B. subtilis amyE* locus (S3 Table). Subsequently, the *C. thermocellum* σ^{16} factor was integrated into the *B. subtilis lacA* locus, and the recognition of the predicted promoters by *C. thermocellum* σ^{16} was analyzed by measuring LacZ activity (Table 3). Our analysis showed that *C. thermocellum* σ^{16} factor recognized 10 predicted promoters that correspond to the *C. thermocellum* genes *xyn10Z*, *xyn11B*, *cipA*, *sigI6*, *xyn10Y*, *cseP* (Clo1313_2188, Cthe_0044), *sigI1*, *rsgI5*, *xyn10D* and *cel9V* (Clo1313_0349, Cthe_2760). The LacZ activities of these 10 predicted σ^{16} -dependent promoters are shown in the Table 3.

As expected, the heterologous expression of the *C. thermocellum* σ^{16} in *B. subtilis* allowed the recognition of its own promoter. This result is in agreement with the recent report of Sand and co-workers [27], which shows that *C. thermocellum* σ^{16} is autoregulated. Interestingly, five of the 10 activated promoters correspond to genes encoding the cellulosomal associated GH9, GH10 and GH11 glycoside hydrolases (*xyn10Z*, *xyn11B*, *xyn10Y*, *xyn10D* and *cel9V*). Furthermore, the highest β -galactosidase (LacZ) activities were obtained with the predicted promoters

Table 3. Quantitative evaluation of possible promoters under the control of *C. thermocellum* σ^{16} .

Gene	Locus tags ^a	Gene product ^e	Ref.	Activity (U) ^f
<i>xyn10Z</i>	Clo1313_2635, Cthe_1963	Cellulosomal xylanase:CE1-CBM6-Doc-GH10	[43]	1 126 ± 149
<i>xyn11B</i> ^b	Clo1313_0522, Cthe_2972 ^a	Cellulosomal xylanase:GH11-CBM6-Doc	[44]	919 ± 72
<i>cipA</i> ^c	Clo1313_0627, Cthe_3077	Cellulosomal scaffoldin subunit:2(Coh)-CBM3-6(Coh)-X-Doc	[45]	266 ± 13
<i>cel9V</i>	Clo1313_0349, Cthe_2760	Cellulosomal endoglucanase:GH9-2(CBM3)-Doc	[46]	223 ± 49
<i>sigI6</i>	Clo1313_2778, Cthe_2120	Alternative σ^{16} factor	[17]	139 ± 6
<i>xyn10Y</i>	Clo1313_1305, Cthe_0912	Cellulosomal xylanase:CBM22-GH10-CBM22-Doc-CE1	[47]	113 ± 10
<i>cseP</i>	Clo1313_2188, Cthe_0044	Cellulosomal component:CotH-Doc	[48]	93 ± 17
<i>sigI1</i>	Clo1313_2174, Cthe_0058	Alternative σ^{11} factor	[17]	28 ± 4
<i>rsgI5</i>	Clo1313_0985, Cthe_1273	Anti- σ^{15} factor:RsgI_N-UNK-CBM42	[17]	25 ± 5
<i>xyn10D</i>	Clo1313_0177, Cthe_2590	Cellulosomal xylanase:CBM22-GH10-Doc	[49]	20 ± 3
<i>sigI3</i> ^d	Clo1313_1911, Cthe_0315	Alternative σ^{13} factor	[17]	ND

^a Clo1313 and Cthe are the locus tag prefixes of *C. thermocellum* strains DSM 1313 and ATCC 27405, respectively.

^b In *C. thermocellum* DSM 1313 *xyn11B* is part of the operon *xyn11B-xyn11A*, whereas *xyn11B* is absent in other strains, such as ATCC 27405 and JW20.

^c In *C. thermocellum* DSM 1313 *cipA* encodes for a scaffoldin with 8 cohesins, whereas in other strains, such as ATCC 27405, *cipA* encodes for a scaffoldin with 9 cohesins [10].

^d The predicted promoter of *C. thermocellum sigI3* was used as negative control.

^e CE, carbohydrate esterase; CBM, carbohydrate binding module; Doc, dockerin; GH, glycoside hydrolase; Coh, cohesin; X, CipA X-module; CotH, spore coat protein H; UNK, unknown domain.

^f The activity was measured using MUG as substrate. One unit of enzyme activity (U) was defined as the amount of β -galactosidase that releases 1 μ mol of 4-MU per minute. Numbers in parenthesis represent the standard deviation of at least three independent experiments. ND means not detected.

doi:10.1371/journal.pone.0146316.t003

of the two xylanase genes *xyn10Z* and *xyn11B* (Table 3). Moreover, *xyn11B* is the first gene of the bicistronic operon *xyn11B-xyn11A* in *C. thermocellum* DSM 1313, whereas the *xyn11B* gene is lacking in other *C. thermocellum* strains, such as ATCC 27405 and JW20 (S6 Table). Bioinformatics analysis, performed with seven strains of *C. thermocellum* (DSM 1313, ATCC 27405, DSM 2360, YS, AD2, JW20 and BC1) and *C. straminisolvans*, showed identical predicted σ^{16} promoter upstream of a single *xyn11A* gene of strains ATCC 27405 and JW20 (S6 Table). This suggests a strong prediction for the regulation of both xylanases, Xyn11B and Xyn11A, by the *C. thermocellum* σ^{16} factor. Finally, the recognition of predicted promoters of genes encoding non-enzymatic proteins (*cipA*, *cseP*, *sigI1* and *rsgI5*) by the *C. thermocellum* σ^{16} factor suggests a more complex regulon for this kind of alternative σ factors. As already mentioned, *sigI1* and *rsgI5* are also proposed to be involved in the regulation of cellulosomal genes [17]. Hence, the recognition of the *sigI1* and *rsgI5* promoters by the *C. thermocellum* σ^{16} factor suggests the possibility of crosstalk between different *C. thermocellum* σ^1 factors and an overlap of their respective regulons.

Identification of conserved promoter elements for σ^{16} recognition

In order to identify the essential bases for σ^{16} recognition, we performed an alignment using the experimentally validated promoter sequences recognized by *C. thermocellum* σ^{16} . To improve the analysis, the 10 experimentally validated σ^{16} -dependent promoter sequences of *C. thermocellum* DSM 1313 were compared with orthologous promoter sequences of *C. straminisolvans* JCM 21531. The result is shown in Fig 2. It can be observed that the σ^{16} -dependent promoters share two highly conserved sequences. The suggested *C. thermocellum* σ^{16} -promoter consensus motifs are CNNAAA for the -35 element and CGAA for the -10 element (where N represents any base). The spacing between the suggested -35 and -10 elements is between 13

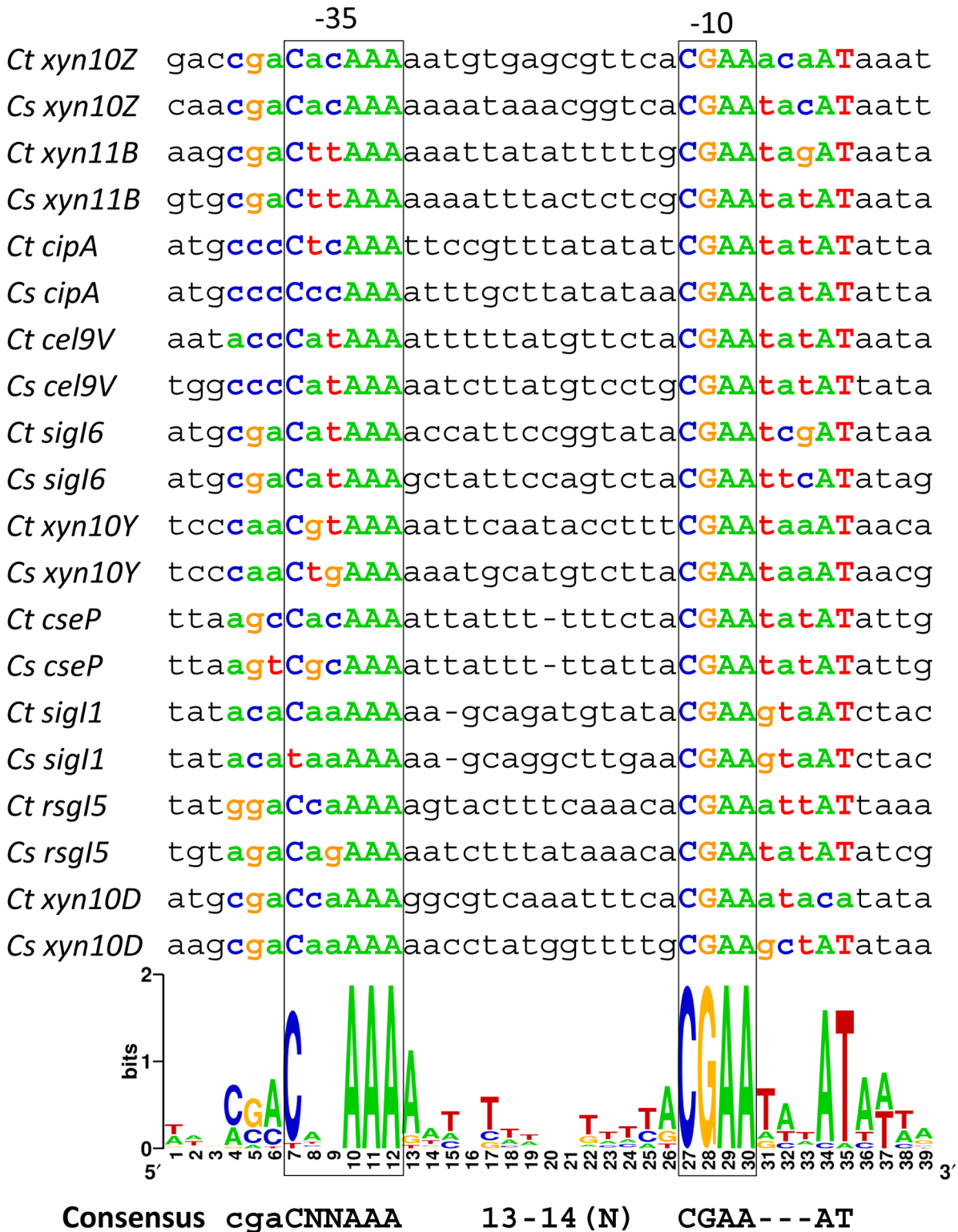


Fig 2. Identification of conserved elements of σ^{16} -dependent promoter sequences. WebLogo generated with σ^{16} -dependent promoter sequences of *C. thermocellum* and orthologous promoter sequences of *C. straminisolvens*.

doi:10.1371/journal.pone.0146316.g002

and 14 nucleotides (Fig 2). It is interesting to note that downstream of the -10 element there is a highly conserved AT sequence.

In the -35 element, next to the highly conserved C, a less conserved CGA sequence can be observed in the WebLogo generated with the 10 σ^{16} -dependent promoter sequences of *C. thermocellum* and *C. straminisolvens* (Fig 2). Interestingly, this CGA sequence is present in the promoter sequences of *sigI6*, *xyn10D*, and in the sequences of the two strongest promoters identified, *xyn10Z* and *xyn11B* (Fig 2 and Table 3). In order to analyze the importance of these bases for the recognition by *C. thermocellum* σ^{16} , we performed site-directed mutagenesis analysis using the promoter sequence of *xyn10Z*. Additionally, we evaluated the most conserved bases, which are suggested as the *C. thermocellum* σ^{16} -promoter consensus, including the highly conserved AT sequence downstream of the -10 element. The analysis is shown in Fig 3. It can be seen that individual changes in the nucleotides C and G of the CGA sequence (*xyn10Z mut1* and *mut2*) at the 5' of the -35 element reduced dramatically the LacZ activity. Interestingly, the mutation from A to T in the CGA sequence (*xyn10Z mut3*) increased the LacZ activity by 33%. This result shows that although the nucleotides C and G in the CGA sequence of the -35 element is less conserved, they play an important role in recognition by *C. thermocellum* σ^{16} .

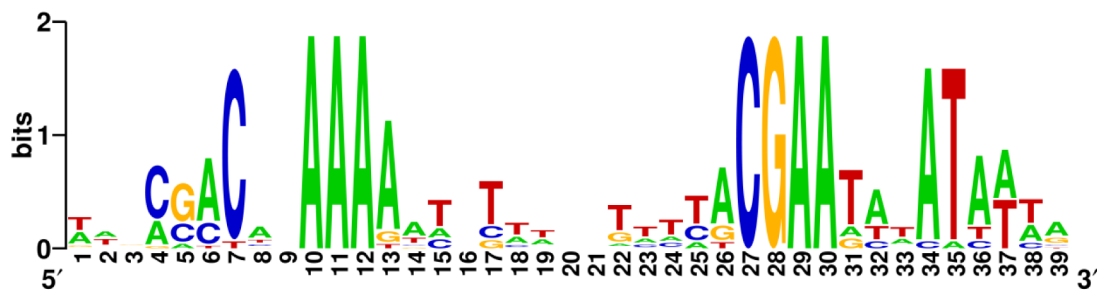
As expected, individual mutations in the highly conserved nucleotide C (*xyn10Z mut4*) and in the AAA sequence (*xyn10Z mut5*, *mut6* and *mut7*) of the -35 element abolished or reduced dramatically the LacZ activity. The least "sensitive" nucleotide was the middle A (*xyn10Z mut6*) in the AAA triplet of the -35 element. Regarding the -10 element, the most conserved bases were also highly sensitive to mutations. Individual changes in the highly conserved CGAA sequence showed that the first 3 nucleotides CGA (*xyn10Z mut10*, *mut11* and *mut13*) are more sensitive to changes than the last A at the 3' of the CGAA sequence (*xyn10Z mut14*). Finally, changes of the highly conserved AT sequence at the 3' of the -10 element showed a dramatic reduction of LacZ activity when the sequence was changed to TA (*xyn10Z mut12*). However, when the AT sequence was changed to CC, LacZ activity was not detected (*xyn10Z mut13*). All these results confirm the importance of the highly conserved sequences of the -35 and -10 elements, as well as the less conserved nucleotides C and G of the CGA sequence at the 5' of the -35 element.

Searching for σ^{13} -dependent promoter sequences of cellulosomal genes in *C. thermocellum*

In this work, we developed a new methodology that employs *B. subtilis* as a heterologous host to verify *C. thermocellum* σ^1 -dependent promoters. By exploiting this approach, we extended the promoter analysis to other *C. thermocellum* σ^1 factors. The above-mentioned 40 predicted σ^1 -dependent promoters (Tables 1 and 2) were also analyzed with *C. thermocellum* σ^{13} expressed in *B. subtilis*. Table 4 shows LacZ activity of the four predicted promoters that were recognized by *C. thermocellum* σ^{13} . These promoters were deduced upstream of the *C. thermocellum* genes *sigI3*, *pl11*, *ce12* and *cipA*. As expected, *C. thermocellum* σ^{13} was able to recognize its own promoter, again suggesting autoregulation in *C. thermocellum*. Two of the predicted promoters recognized by *C. thermocellum* σ^{13} belong to genes coding for pectin-degrading enzyme (*pl11* and *ce12*). Furthermore, during the quantification of σ^{13} -dependent promoter activities, the promoters of *pl11* and *ce12* showed the highest activities (Table 4). Interestingly, the previously verified σ^{16} -dependent promoter of *cipA* was also recognized by *C. thermocellum* σ^{13} (but to a lesser degree) suggesting a possible overlap between the regulons of *C. thermocellum* σ^{13} and σ^{16} (Tables 3 and 4).

In order to identify the important promoter nucleotides for σ^{13} recognition, we performed an alignment using the experimentally validated promoter sequences recognized by *C.*

	-35		-10	Relative activity (%)
<i>xyn10Z short</i>	cgaCacAAA	14 (N)	CGAAacaAT	100
<i>xyn10Z mut1</i>	ggaCacAAA	14 (N)	CGAAacaAT	11 ± 0.05
<i>xyn10Z mut2</i>	c <u>c</u> aCacAAA	14 (N)	CGAAacaAT	9 ± 3.1
<i>xyn10Z mut3</i>	c <u>g</u> tCacAAA	14 (N)	CGAAacaAT	133 ± 16
<i>xyn10Z mut4</i>	cga <u>G</u> acAAA	14 (N)	CGAAacaAT	ND
<i>xyn10Z mut5</i>	cgaCac <u>T</u> AA	14 (N)	CGAAacaAT	ND
<i>xyn10Z mut6</i>	cgaCacA <u>T</u> A	14 (N)	CGAAacaAT	6 ± 2.3
<i>xyn10Z mut7</i>	cgaCacAA <u>T</u>	14 (N)	CGAAacaAT	ND
<i>xyn10Z mut8</i>	cgaCacAAA	14 (N)	<u>G</u> GAAacaAT	ND
<i>xyn10Z mut9</i>	cgaCacAAA	14 (N)	<u>C</u> CAAacaAT	ND
<i>xyn10Z mut10</i>	cgaCacAAA	14 (N)	CG <u>T</u> AacaAT	ND
<i>xyn10Z mut11</i>	cgaCacAAA	14 (N)	CGA <u>T</u> acaAT	10 ± 3.4
<i>xyn10Z mut12</i>	cgaCacAAA	14 (N)	CGAAaca <u>T</u> A	11 ± 3.8
<i>xyn10Z mut13</i>	cgaCacAAA	14 (N)	CGAAaca <u>CC</u>	ND



σ^{I6} consensus cgaCNNAAA 14 - 13 (N) CGAA - - - AT

Fig 3. Evaluation of σ^{I6} promoter sequence validity by mutagenesis. The activities are shown as relative activities, with the control promoter *xyn10Zshort* without mutations set to 100%. ND means not detected.

doi:10.1371/journal.pone.0146316.g003

Table 4. Quantitative evaluation of possible promoters under the control of *C. thermocellum* σ^{13} .

Gene	Locus tags ^a	Gene product ^b	Ref.	Activity (U) ^d
<i>pl11</i>	Clo1313_1983, Cthe_0246	Cellulosomal pectinase:Doc-CBM35-RGL11	—	195 ± 13
<i>ce12</i>	Clo1313_0693, Cthe_3141	Cellulosomal pectinase:RGAE-Doc-CBM35-RGAE	—	62 ± 8
<i>cipA</i>	Clo1313_0627, Cthe_3077	Cellulosomal scaffoldin subunit:2(Coh)-CBM3-6(Coh)-X-Doc	[45]	58 ± 7
<i>sig3</i>	Clo1313_1911, Cthe_0315	Alternative σ^{13} factor	[17]	28 ± 13
<i>sig6^c</i>	Clo1313_2778, Cthe_2120	Alternative σ^{16} factor	[17]	ND

^a Clo1313 and Cthe are the locus tag prefixes of *C. thermocellum* strains DSM 1313 and ATCC 27405, respectively.

^b CE, carbohydrate esterase; CBM, carbohydrate binding module; Doc, dockerin; RGL, rhamnogalacturonan lyase; RGAE, rhamnogalacturonan acetyltransferase Coh, cohesin; X, CipA X-module; UNK, unknown domain.

^c The predicted promoter of *C. thermocellum sig6* was used as negative control.

^d The activity was measured using MUG as substrate. One unit of enzyme activity (U) was defined as the amount of β -galactosidase that releases 1 μ mol of 4-MU per minute. Numbers in parenthesis represent the standard deviation of at least three independent experiments. ND means not detected.

doi:10.1371/journal.pone.0146316.t004

thermocellum σ^{13} . To improve the analysis, the four experimentally validated σ^{13} -dependent promoter sequences were compared with orthologous promoter sequences of *C. straminisolvens*. The result is shown in Fig 4. It can be observed that σ^{13} -dependent promoters have two highly conserved sequences. The suggested *C. thermocellum* σ^{13} -promoter consensus is CCCYAAA for the -35 element and CGWA for the -10 element (where Y represents C or T, and W represents A or T). The spacing between the -35 and -10 elements is between 13 and 14 nucleotides, resembling the organization of the σ^{16} promoter (Fig 2).

Discussion

Since the original discovery of the cellulosome, numerous observations indicated that its production and composition is influenced by the nature of the carbon source present in the growth media [7,50]. However, until now, there are only a few reports in the literature regarding the regulation of cellulosomal genes [51–54]. Most of these studies were focused on key cellulosomal genes, such as *cipA* [51], *cel48S* [52], *cel8A* [53], *cel9D* [54] and *cel9F* [54]. In the majority of these works, the authors were able to identify σ^A -dependent promoters upstream of the analyzed genes [51–54]. Additionally, these genes appeared to be regulated by alternative σ factors [51–54]. Nonetheless, the assignment of alternative σ factors was problematic, and in the case of *cel9D* [54] and *cel9F* [54], the authors could not suggest a convincing alternative σ factor. This obstacle surfaced since basic knowledge about *C. thermocellum* was limited, and the DNA sequence upstream of the start sites did not contain homologies with described consensus promoters [54].

In an effort to gather knowledge about the regulation of cellulosomal genes, our research group discovered a set of eight alternative σ^1 factors [17] where six of them showed up-regulation by environmental polysaccharides [19]. This set of *C. thermocellum* σ^1 factors is homologous to the *B. subtilis* σ^1 factor [17,18]. Hence, in order to identify the important *C. thermocellum* σ^1 promoter elements, in the present work, we decided to compare the upstream regions of *C. thermocellum* σ^1 factor genes with the *Bacillales* σ^1 -dependent promoter sequences. The high conservation of the *Bacillales* σ^1 -dependent promoter sequences identified in this study indicated that the *C. thermocellum sigI-rsgI* operons may likely have similar -35 and -10 promoter elements. Additionally, it was interesting that the *Bacillales* σ^1 -dependent promoters have a C "enrichment" in their -35 elements (Fig 1A). This observation suggested an easy way to search for σ^1 -dependent promoter candidates in other species of the *Firmicutes* phylum, and notably, for our purposes, the cellulosome-producing bacteria, especially since

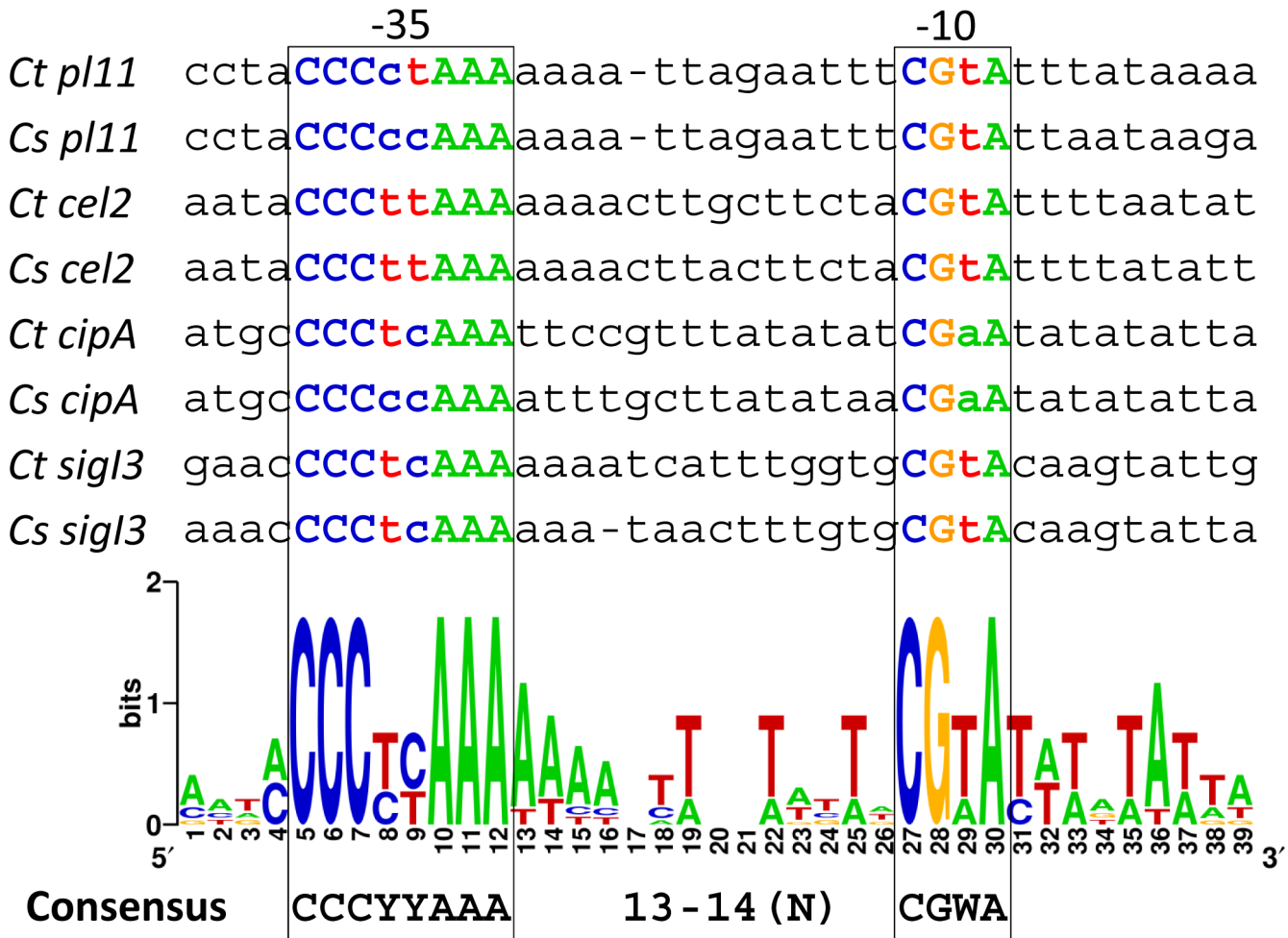


Fig 4. Identification of conserved elements of σ^{13} -dependent promoter sequences. WebLogo generated with σ^{13} -dependent promoter sequences of *C. thermocellum* and orthologous promoter sequences of *C. straminisolvans*.

doi:10.1371/journal.pone.0146316.g004

one striking characteristic of the *Firmicutes* phylum is the low G/C content of their genomes [55]. However, during the search of putative σ^I -dependent promoters in *C. thermocellum*, only *sigI2* and *sigI3* showed the anticipated high C-enrichment in the -35 element motif (Table 1). Subsequently, we compared the *C. thermocellum sigI2* and *sigI3* predicted promoter sequences with those of *C. thermocellum sigI1* and *sigI6* genes proposed by Nataf and co-workers [19]. This comparison revealed that the promoter regions of *sigI1*, *sigI2*, *sigI3* and *sigI6* have specific "signatures", such as an AAA triad at the -35 region and a CGWA tetrad at the -10 region (Table 1).

Using these specific "signatures", we were able to predict the putative promoters of *C. thermocellum sigI4*, *sigI7* and *sigI8*. Moreover, the conservation of these specific "signatures" within σ^I -dependent promoters was additionally supported by the orthologous *sigI*-promoter sequences of the closely related cellulosome-producing bacterium, *C. straminisolvans* JCM 21531 (Table 1). Furthermore, these specific "signatures" were corroborated with the predicted promoter sequences of *sigI*-like genes of *C. clariflavum*, *A. cellulolyticus* and *P. cellulosolvans* and the currently identified 33 σ^I -dependent promoters of *C. thermocellum* cellulosomal genes (Table 2 and S5 Table).

The predicted promoter sequences of the σ^I -dependent promoters of cellulosome-producing bacteria can be divided into three regions. Two regions are highly conserved and contain the proposed specific "signatures" of the σ^I -dependent promoters, namely the AAA triad and the CGWA tetrad sequences of the "extended -35" and -10 element, respectively (Tables 1 and 2, and Fig 1). The third region is highly divergent and is located in the 5'-terminal sequence of the -35 elements (corresponding to the *B. subtilis* ACCCC sequence, Tables 1 and 2 and Fig 1). We predict that, whereas the most conserved sequences of the "extended -35" and -10 elements are implicated in the "general" recognition of promoters by their cognate σ^I factors, the most divergent 5'-terminal sequence of the -35 elements is likely implicated in the specificity of the different σ^I factors. This phenomenon could reflect the general strategy followed by σ^I factors in order to recognize their target promoters in cellulosome-producing bacteria, whose genomes encode multiple σ^I factors. Our hypothesis was herein supported by experimental identification of the putative *C. thermocellum* σ^{16} - and σ^{13} -dependent promoters (Figs 2 and 3). Whereas the *C. thermocellum* σ^{16} -dependent promoters have a highly conserved C nucleotide upstream of the AAA in the -35 element, the *C. thermocellum* σ^{13} -dependent promoters have a highly conserved CCC triad (Figs 2 and 3). Additionally, the analysis of the *C. thermocellum* σ^{16} - and σ^{13} -dependent promoters indicated that, in addition, some nucleotides in the -10 element probably have an important role in the specificity of the different σ^I factors of cellulosome-producing bacteria. For example, all of the identified *C. thermocellum* σ^{16} -dependent promoters have the 5'-located CGAA sequence in their -10 elements, and most of the promoters have an AT sequence three bases downstream (Fig 2). Regarding the identified *C. thermocellum* σ^{13} -dependent promoters, they have a less conserved -10 element with the sequence CGWA, where W could be A or T (Fig 3).

Four of the 10 promoters that were activated by *C. thermocellum* σ^{16} belong to genes implicated in the hydrolysis of xylan (*xyn10Z*, *xyn11B-xyn11A* operon, *xyn10Y* and *xyn10D*). This observation is in accordance with previous experiments performed by Nataf and co-workers [19] which showed that when *C. thermocellum* was grown on cellulose, the expression of the *sigI6* gene was up-regulated 2.5-fold; and when the cells were grown on cellulose and xylan, *sigI6* was up-regulated at least 10-fold. Moreover, the anti- σ^{16} factor, RsgI6, bears an extracytoplasmic C-terminal sensing module that belongs to the glycoside hydrolase family 10 (GH10). Interestingly, the RsgI6 GH10-family module is highly similar to Xyn10D (it is 57% identical and has 79% similar residues in the 381-aa gapless alignment; data not shown). Bahari and co-workers [21] showed that the RsgI6 GH10-like domain binds to oat-spelt xylan and Avicel (cellulose). Furthermore, a recent study performed by Wei and co-workers [16], showed that the genes *xyn10Y* and *xyn10D* were up-regulated when *C. thermocellum* was grown in dilute acid-pretreated yellow poplar, containing 65% cellulose, 4% xylan and 31% lignin.

Regarding *C. thermocellum* σ^{13} , two of the four promoters that were activated by *C. thermocellum* σ^{13} belong to genes implicated in the solubilization of pectin (*pl11/Clo1313_1983* and *ce12/Clo1313_0693*). Interestingly, the anti- σ^{13} factor RsgI3, has an extracytoplasmic sensing module that is composed of two tandem PA14 superfamily motifs that were shown to bind pectin by Kahel-Raifer and co-workers [17]. Taken together, these results suggest that while σ^{16} likely plays a role in the regulation of xylan-degrading enzymes, σ^{13} likely plays a role in the regulation of pectin-degrading enzymes.

The *C. thermocellum* *sigI1* and *rsgI5* genes are part of the proposed genes involved in the regulation of cellulosomal genes in response to environmental polysaccharides [17]. In the present work, the two predicted σ^I -dependent promoters for *sigI1* and *rsgI5* were recognized by *C. thermocellum* σ^{16} (Table 3). Furthermore, the predicted σ^I -dependent promoter for *cipA* was recognized by both *C. thermocellum* σ^{16} and σ^{13} (Tables 3 and 4). These results suggest possible crosstalk between different *C. thermocellum* σ^I factors and a possible overlap of their respective

regulons. Interestingly, this phenomenon is common in ECF sigma factors. For example, Huang and co-workers [56] found in *B. subtilis* that σ^W recognizes a subset of promoters that are partially dependent on σ^X for expression. Additionally, Mascher and co-workers [57] found 7 ECF sigma factors in *B. subtilis* that regulate partially overlapping regulons related to cell envelope homeostasis and antibiotic resistance. To unwrap its preferred substrate, cellulose, that is covered with different types of polysaccharides in the plant cell wall, *C. thermocellum* would presumably produce an array of different hydrolytic cellulosomal components. Hence, to have partially overlapping regulons for the multiple σ^I factors could be advantageous, because expression of a variety of cellulosomal components is crucial for efficient solubilization of the different types of polysaccharides that conceal the cellulose fibers in their native state.

Our results and observations reveal several promising options to improve the performance of the industrially prominent bacterium *C. thermocellum*. First of all, by changing promoter designs by metabolic engineering, we may try to modify the expression of selected cellulosomal genes that might be crucial for production of natural forms of designer cellulosomes. Secondly, by using strong sigma-dependent promoters (e.g., those of *xyn10Z*, *xyn11B*, *pl11*, etc.) one can introduce additional, synthetic cellulosomal genes in *C. thermocellum* and use their products for improvement of either saccharolytic activity or, alternatively, ethanol production. Continued analysis and harnessing of the various σ and anti- σ factors in *C. thermocellum* will allow us to control and enhance the capacity of this ecologically prominent and industrially relevant bacterium for deconstruction of plant-derived polysaccharides *en route* to the production of biofuels.

Supporting Information

S1 Fig. Schematic depiction of plasmid pLOXErysIrgIBs.

(PDF)

S2 Fig. ClustalW alignment of the RNAP subunits sequences of *Bacillus subtilis* strain 168 and *Clostridium thermocellum* strain DSM 1313.

(PDF)

S1 Table. Primers used in the present work.

(PDF)

S2 Table. Plasmids constructed in the present work.

(PDF)

S3 Table. *Bacillus subtilis* strains constructed in the present work.

(PDF)

S4 Table. Alignment of experimentally confirmed and putative *sigI* promoters from different *Bacillales* species, including the experimentally confirmed *sigI*-dependent promoters of the *B. subtilis* *bcrC* and *mreBH* genes.

(PDF)

S5 Table. Alignments of predicted *sigI* promoters from *Clostridium clariflavum*, *Acetivibrio cellulolyticus*, *Pseudobacteroides cellulosolvens*, *Clostridium thermocellum* and *Clostridium straminisolvens*.

(PDF)

S6 Table. σ^{16} -dependent promoter sequence alignment of *xyn11B-xyn11A* operon and *xyn11A* of available sequences of *C. thermocellum* and *C. straminisolvens* JCM 21531.

(PDF)

Acknowledgments

Appreciation is expressed to Olga Zhivin (The Weizmann Institute of Science) and Andy Sand (Technion—Israel Institute of Technology) for helpful discussions and advice. Additionally, we want to thank Dr. Avigdor Eldar (Tel Aviv University) and the Bacillus Genetic Stock Center (BGSC) for providing *B. subtilis* strains and plasmid. 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.

Author Contributions

Conceived and designed the experiments: IB IMG LOO. Performed the experiments: IMG LOO YG IB. Analyzed the data: IMG LOO IB IRG RL. Contributed reagents/materials/analysis tools: RL IB EAB YS. Wrote the paper: IMG LOO IB. Made manuscript revisions: IMG LOO EAB RL IB.

References

1. Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS. Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol Mol Biol Rev.* 2002; 66: 506–577. doi: [10.1128/MMBR.66.3.506-577.2002](https://doi.org/10.1128/MMBR.66.3.506-577.2002) PMID: [12209002](https://pubmed.ncbi.nlm.nih.gov/12209002/)
2. Viljoen JA, Fred EB, Peterson WH. The fermentation of cellulose by thermophilic bacteria. *J Agric Sci.* 1926; 16: 1–17. doi: [10.1017/S0021859600088249](https://doi.org/10.1017/S0021859600088249)
3. Akinosho H, Yee K, Close D, Ragauskas A. The emergence of *Clostridium thermocellum* as a high utility candidate for consolidated bioprocessing applications. *Front Chem.* 2014; 2: 66. doi: [10.3389/fchem.2014.00066](https://doi.org/10.3389/fchem.2014.00066) PMID: [25207268](https://pubmed.ncbi.nlm.nih.gov/25207268/)
4. Lamed R, Setter E, Kenig R, Bayer EA. The cellulosome—a discrete cell surface organelle of *Clostridium thermocellum* which exhibits separate antigenic, cellulose-binding and various cellulolytic activities. *Biotechnol Bioeng Symp.* 1983; 13: 163–181.
5. Bayer EA, Belaich JP, Shoham Y, Lamed R. The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. *Annu Rev Microbiol.* 2004; 58: 521–554. doi: [10.1146/annurev.micro.57.030502.091022](https://doi.org/10.1146/annurev.micro.57.030502.091022) PMID: [15487947](https://pubmed.ncbi.nlm.nih.gov/15487947/)
6. Lamed R, Setter E, Bayer EA. Characterization of a cellulose-binding, cellulase-containing complex in *Clostridium thermocellum*. *J Bacteriol.* 1983; 156: 828–836. PMID: [6195146](https://pubmed.ncbi.nlm.nih.gov/6195146/)
7. Bayer EA, Kenig R, Lamed R. Adherence of *Clostridium thermocellum* to cellulose. *J Bacteriol.* 1983; 156: 818–827. PMID: [6630152](https://pubmed.ncbi.nlm.nih.gov/6630152/)
8. Salamitou S, Tokatlidis K, Béguin P, Aubert JP. Involvement of separate domains of the cellulosomal protein S1 of *Clostridium thermocellum* in binding to cellulose and in anchoring of catalytic subunits to the cellulosome. *FEBS Lett.* 1992; 304: 89–92. PMID: [1618304](https://pubmed.ncbi.nlm.nih.gov/1618304/)
9. Tokatlidis K, Salamitou S, Béguin P, Dhurjati P, Aubert JP. Interaction of the duplicated segment carried by *Clostridium thermocellum* cellulases with cellulosome components. *FEBS Lett.* 1991; 291: 185–188. doi: [10.1016/0014-5793\(91\)81279-H](https://doi.org/10.1016/0014-5793(91)81279-H) PMID: [1936262](https://pubmed.ncbi.nlm.nih.gov/1936262/)
10. Hong W, Zhang J, Feng Y, Mohr G, Lambowitz AM, Cui GZ, et al. The contribution of cellulosomal scaffoldins to cellulose hydrolysis by *Clostridium thermocellum* analyzed by using thermotargetrons. *Biotechnol Biofuels.* 2014; 7: 80. doi: [10.1186/1754-6834-7-80](https://doi.org/10.1186/1754-6834-7-80) PMID: [24955112](https://pubmed.ncbi.nlm.nih.gov/24955112/)
11. Yaniv O, Morag E, Borovok I, Bayer EA, Lamed R, Frolow F, et al. 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.* 2013; 69: 733–737. doi: [10.1107/S174430911301614X](https://doi.org/10.1107/S174430911301614X) PMID: [23832198](https://pubmed.ncbi.nlm.nih.gov/23832198/)
12. Zverlov VV, Kellermann J, Schwarz WH. Functional subgenomics of *Clostridium thermocellum* cellulosomal genes: identification of the major catalytic components in the extracellular complex and detection of three new enzymes. *Proteomics.* 2005; 5: 3646–3653. doi: [10.1002/pmic.200401199](https://doi.org/10.1002/pmic.200401199) PMID: [16127726](https://pubmed.ncbi.nlm.nih.gov/16127726/)
13. Zverlov VV, Fuchs KP, Schwarz WH. Chi18A, the endochitinase in the cellulosome of the thermophilic, cellulolytic bacterium *Clostridium thermocellum*. *Appl Environ Microbiol.* 2002; 68: 3176–3179. doi: [10.1128/AEM.68.6.3176-3179.2002](https://doi.org/10.1128/AEM.68.6.3176-3179.2002) PMID: [12039789](https://pubmed.ncbi.nlm.nih.gov/12039789/)

14. Raman B, Pan C, Hurst GB, Rodriguez M, McKeown CK, Lankford PK, et al. Impact of pretreated Switchgrass and biomass carbohydrates on *Clostridium thermocellum* ATCC 27405 cellulosome composition: a quantitative proteomic analysis. PLoS One. 2009; 4: e5271. doi: [10.1371/journal.pone.0005271](https://doi.org/10.1371/journal.pone.0005271) PMID: [19384422](https://pubmed.ncbi.nlm.nih.gov/19384422/)
15. Gold ND, Martin VJJ. Global view of the *Clostridium thermocellum* cellulosome revealed by quantitative proteomic analysis. J Bacteriol. 2007; 189: 6787–6795. doi: [10.1128/JB.00882-07](https://doi.org/10.1128/JB.00882-07) PMID: [17644599](https://pubmed.ncbi.nlm.nih.gov/17644599/)
16. Wei H, Fu Y, Magnusson L, Baker JO, Maness PC, Xu Q, et al. Comparison of transcriptional profiles of *Clostridium thermocellum* grown on cellobiose and pretreated yellow poplar using RNA-Seq. Front Microbiol. 2014; 5: 142. doi: [10.3389/fmicb.2014.00142](https://doi.org/10.3389/fmicb.2014.00142) PMID: [24782837](https://pubmed.ncbi.nlm.nih.gov/24782837/)
17. Kahel-Raifer H, Jindou S, Bahari L, Nataf Y, Shoham Y, Bayer EA, et al. 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. 2010; 308: 84–93. doi: [10.1111/j.1574-6968.2010.01997.x](https://doi.org/10.1111/j.1574-6968.2010.01997.x) PMID: [20487018](https://pubmed.ncbi.nlm.nih.gov/20487018/)
18. Asai K, Ootsuji T, Obata K, Matsumoto T, Fujita Y, Sadaie Y. Regulatory role of RsgI in *sigI* expression in *Bacillus subtilis*. Microbiology. 2007; 153: 92–101. doi: [10.1099/mic.0.29239-0](https://doi.org/10.1099/mic.0.29239-0) PMID: [17185538](https://pubmed.ncbi.nlm.nih.gov/17185538/)
19. Nataf Y, Bahari L, Kahel-Raifer H, Borovok I, Lamed R, Bayer EA, et al. *Clostridium thermocellum* celulosomal genes are regulated by extracytoplasmic polysaccharides via alternative sigma factors. Proc Natl Acad Sci U S A. 2010; 107: 18646–18651. doi: [10.1073/pnas.1012175107](https://doi.org/10.1073/pnas.1012175107) PMID: [20937888](https://pubmed.ncbi.nlm.nih.gov/20937888/)
20. Morag E, Bayer EA, Hazlewood GP, Gilbert HJ, Lamed R. Cellulase Ss (CelS) is synonymous with the major cellobiohydrolase (subunit S8) from the cellulosome of *Clostridium thermocellum*. Appl Biochem Biotechnol. 1993; 43: 147–151. doi: [10.1007/BF02916439](https://doi.org/10.1007/BF02916439) PMID: [8267404](https://pubmed.ncbi.nlm.nih.gov/8267404/)
21. Bahari L, Gilad Y, Borovok I, Kahel-Raifer H, Dassa B, Nataf Y, et al. Glycoside hydrolases as components of putative carbohydrate biosensor proteins in *Clostridium thermocellum*. J Ind Microbiol Biotechnol. 2011; 38: 825–832. doi: [10.1007/s10295-010-0848-9](https://doi.org/10.1007/s10295-010-0848-9) PMID: [20820855](https://pubmed.ncbi.nlm.nih.gov/20820855/)
22. Yaniv O, Fichman G, Borovok I, Shoham Y, Bayer EA, Lamed R, et al. Fine-structural variance of family 3 carbohydrate-binding modules as extracellular biomass-sensing components of *Clostridium thermocellum* anti- σ^I factors. Acta Crystallogr D Biol Crystallogr. 2014; 70: 522–534. doi: [10.1107/S139900471302926X](https://doi.org/10.1107/S139900471302926X) PMID: [24531486](https://pubmed.ncbi.nlm.nih.gov/24531486/)
23. Rigden DJ, Mello LV, Galperin MY. The PA14 domain, a conserved all- β domain in bacterial toxins, enzymes, adhesins and signaling molecules. Trends Biochem Sci. 2004; 29: 335–339. doi: [10.1016/j.tibs.2004.05.002](https://doi.org/10.1016/j.tibs.2004.05.002) PMID: [15236739](https://pubmed.ncbi.nlm.nih.gov/15236739/)
24. Helmann JD. The extracytoplasmic function (ECF) sigma factors. Adv Microb Physiol. 2002; 46: 47–110. doi: [10.1016/S0065-2911\(02\)46002-X](https://doi.org/10.1016/S0065-2911(02)46002-X) PMID: [12073657](https://pubmed.ncbi.nlm.nih.gov/12073657/)
25. Mascher T. Signaling diversity and evolution of extracytoplasmic function (ECF) σ factors. Curr Opin Microbiol. 2013; 16: 148–155. doi: [10.1016/j.mib.2013.02.001](https://doi.org/10.1016/j.mib.2013.02.001) PMID: [23466210](https://pubmed.ncbi.nlm.nih.gov/23466210/)
26. Paget M. Bacterial sigma factors and anti-sigma factors: structure, function and distribution. Biomolecules. 2015; 5: 1245–1265. doi: [10.3390/biom5031245](https://doi.org/10.3390/biom5031245) PMID: [26131973](https://pubmed.ncbi.nlm.nih.gov/26131973/)
27. Sand A, Holwerda EK, Ruppertsberger NM, Maloney M, Olson DG, Nataf Y, et al. Three celulosomal xylanase genes in *Clostridium thermocellum* are regulated by both vegetative SigA (σ^A) and alternative SigI6 (σ^I6) factors. FEBS Lett. 2015; 589: 3133–3140. doi: [10.1016/j.febslet.2015.08.026](https://doi.org/10.1016/j.febslet.2015.08.026) PMID: [26320414](https://pubmed.ncbi.nlm.nih.gov/26320414/)
28. Youngman P, Perkins JB, Losick R. Construction of a cloning site near one end of Tn917 into which foreign DNA may be inserted without affecting transposition in *Bacillus subtilis* or expression of the transposon-borne *erm* gene. Plasmid. 1984; 12: 1–9. doi: [10.1016/0147-619X\(84\)90061-1](https://doi.org/10.1016/0147-619X(84)90061-1) PMID: [6093169](https://pubmed.ncbi.nlm.nih.gov/6093169/)
29. Sambrook J, Rusell D. Molecular cloning a laboratory manual. 3rd ed. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press; 2001.
30. Härtl B, Wehrl W, Wiegert T, Homuth G, Schumann W. Development of a new integration site within the *Bacillus subtilis* chromosome and construction of compatible expression cassettes. J Bacteriol. 2001; 183: 2696–2699. doi: [10.1128/JB.183.8.2696-2699.2001](https://doi.org/10.1128/JB.183.8.2696-2699.2001) PMID: [11274134](https://pubmed.ncbi.nlm.nih.gov/11274134/)
31. Radeck J, Kraft K, Bartels J, Cikovic T, Dürr F, Emenegger J, et al. The *Bacillus* BioBrick Box: generation and evaluation of essential genetic building blocks for standardized work with *Bacillus subtilis*. J Biol Eng. 2013; 7: 29. doi: [10.1186/1754-1611-7-29](https://doi.org/10.1186/1754-1611-7-29) PMID: [24295448](https://pubmed.ncbi.nlm.nih.gov/24295448/)
32. Harwood CR, Cutting S. Molecular biological methods for bacillus. Harwood CR, Cutting S, editors. Chichester, UK: John Wiley & Sons, Ltd; 1990.
33. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. 1997; 25: 3389–3402. doi: [10.1093/nar/25.17.3389](https://doi.org/10.1093/nar/25.17.3389) PMID: [9254694](https://pubmed.ncbi.nlm.nih.gov/9254694/)
34. Higgins DG, Thompson JD, Gibson TJ. Using CLUSTAL for multiple sequence alignments. Methods Enzymol. 1996; 266: 383–402. doi: [10.1016/S0076-6879\(96\)66024-8](https://doi.org/10.1016/S0076-6879(96)66024-8) PMID: [8743695](https://pubmed.ncbi.nlm.nih.gov/8743695/)

35. Crooks G, Hon G, Chandonia J, Brenner S. NCBI GenBank FTP Site\nWebLogo: a sequence logo generator. *Genome Res.* 2004; 14: 1188–1190. doi: [10.1101/gr.849004](https://doi.org/10.1101/gr.849004) PMID: [15173120](https://pubmed.ncbi.nlm.nih.gov/15173120/)
36. Tseng CL, Shaw GC. Genetic evidence for the actin homolog gene *mreBH* and the bacitracin resistance gene *bcrC* as targets of the alternative sigma factor SigI of *Bacillus subtilis*. *J Bacteriol.* 2008; 190: 1561–1567. doi: [10.1128/JB.01497-07](https://doi.org/10.1128/JB.01497-07) PMID: [18156261](https://pubmed.ncbi.nlm.nih.gov/18156261/)
37. Kato S, Haruta S, Cui ZJ, Ishii M, Yokota A, Igarashi Y. *Clostridium straminisolvens* sp. nov., a moderately thermophilic, aerotolerant and cellulolytic bacterium isolated from a cellulose-degrading bacterial community. *Int J Syst Evol Microbiol.* 2004; 54: 2043–2047. doi: [10.1099/ijs.0.63148-0](https://doi.org/10.1099/ijs.0.63148-0) PMID: [15545431](https://pubmed.ncbi.nlm.nih.gov/15545431/)
38. Yuki M, Oshima K, Suda W, Sakamoto M, Kitamura K, Iida T, et al. Draft genome sequence of *Clostridium straminisolvens* strain JCM 21531T, isolated from a cellulose-degrading bacterial community. *Genome Announc.* 2014; 2: 5–6. doi: [10.1128/genomeA.00110-14](https://doi.org/10.1128/genomeA.00110-14)
39. Fang C, Stiegeler E, Cook GM, Mascher T, Gebhard S. *Bacillus subtilis* as a platform for molecular characterisation of regulatory mechanisms of *Enterococcus faecalis* resistance against cell wall antibiotics. *PLoS One.* 2014; 9: e93169. doi: [10.1371/journal.pone.0093169](https://doi.org/10.1371/journal.pone.0093169) PMID: [24676422](https://pubmed.ncbi.nlm.nih.gov/24676422/)
40. Gebhard S, Fang C, Shaaly A, Leslie DJ, Weimar MR, Kalamorz F, et al. Identification and characterization of a bacitracin resistance network in *Enterococcus faecalis*. *Antimicrob Agents Chemother.* 2014; 58: 1425–1433. doi: [10.1128/AAC.02111-13](https://doi.org/10.1128/AAC.02111-13) PMID: [24342648](https://pubmed.ncbi.nlm.nih.gov/24342648/)
41. Yano K, Inoue H, Mori H, Yee LM, Matsuoka S, Sadaie Y, et al. Heterologous expression of the *Oceanobacillus iheyensis* SigW and its anti-protein RsiW in *Bacillus subtilis*. *Biosci Biotechnol Biochem.* 2011; 75: 966–975. doi: [10.1271/bbb.110035](https://doi.org/10.1271/bbb.110035) PMID: [21597191](https://pubmed.ncbi.nlm.nih.gov/21597191/)
42. Ho TD, Ellermeier CD. PrsW is required for colonization, resistance to antimicrobial peptides, and expression of extracytoplasmic function σ factors in *Clostridium difficile*. *Infect Immun.* 2011; 79: 3229–3238. doi: [10.1128/IAI.00019-11](https://doi.org/10.1128/IAI.00019-11) PMID: [21628514](https://pubmed.ncbi.nlm.nih.gov/21628514/)
43. Grépinet O, Chebrou MC, Béguin P. Purification of *Clostridium thermocellum* xylanase Z expressed in *Escherichia coli* and identification of the corresponding product in the culture medium of *C. thermocellum*. *J Bacteriol.* 1988; 170: 4576–4581. PMID: [3139631](https://pubmed.ncbi.nlm.nih.gov/3139631/)
44. Hayashi H, Takehara M, Hattori T, Kimura T, Karita S, Sakka K, et al. Nucleotide sequences of two contiguous and highly homologous xylanase genes *xynA* and *xynB* and characterization of XynA from *Clostridium thermocellum*. *Appl Microbiol Biotechnol.* 1999; 51: 348–357. doi: [10.1007/s002530051401](https://doi.org/10.1007/s002530051401) PMID: [10222584](https://pubmed.ncbi.nlm.nih.gov/10222584/)
45. Gerngross UT, Romaniec MP, Kobayashi T, Huskisson NS, Demain AL. Sequencing of a *Clostridium thermocellum* gene (*cipA*) encoding the cellulosomal SL-protein reveals an unusual degree of internal homology. *Mol Microbiol.* 1993; 8: 325–334. doi: [10.1111/j.1365-2958.1993.tb01576.x](https://doi.org/10.1111/j.1365-2958.1993.tb01576.x) PMID: [8316083](https://pubmed.ncbi.nlm.nih.gov/8316083/)
46. Jindou S, Xu Q, Kenig R, Shulman M, Shoham Y, Bayer EA, et al. Novel architecture of family-9 glycoside hydrolases identified in cellulosomal enzymes of *Acetivibrio cellulolyticus* and *Clostridium thermocellum*. *FEMS Microbiol Lett.* 2006; 254: 308–316. doi: [10.1111/j.1574-6968.2005.00040.x](https://doi.org/10.1111/j.1574-6968.2005.00040.x) PMID: [16445761](https://pubmed.ncbi.nlm.nih.gov/16445761/)
47. Fontes CM, Hazlewood GP, Morag E, Hall J, Hirst BH, Gilbert HJ. Evidence for a general role for non-catalytic thermostabilizing domains in xylanases from thermophilic bacteria. *Biochem J.* 1995; 307: 151–158. PMID: [7717969](https://pubmed.ncbi.nlm.nih.gov/7717969/)
48. Zverlov VV, Velikodvorskaya GA, Schwarz WH. Two new cellulosome components encoded downstream of cell in the genome of *Clostridium thermocellum*: The non-processive endoglucanase CelN and the possibly structural protein CseP. *Microbiology.* 2003; 149: 515–524. doi: [10.1099/mic.0.25959-0](https://doi.org/10.1099/mic.0.25959-0) PMID: [12624213](https://pubmed.ncbi.nlm.nih.gov/12624213/)
49. Zverlov VV, Schantz N, Schmitt-Kopplin P, Schwarz WH. Two new major subunits in the cellulosome of *Clostridium thermocellum*: Xyloglucanase Xgh74A and endoxylanase Xyn10D. *Microbiology.* 2005; 151: 3395–3401. doi: [10.1099/mic.0.28206-0](https://doi.org/10.1099/mic.0.28206-0) PMID: [16207921](https://pubmed.ncbi.nlm.nih.gov/16207921/)
50. Bayer EA, Setter E, Lamed R. Organization and distribution of the cellulosome in *Clostridium thermocellum*. *J Bacteriol.* 1985; 163: 552–559. PMID: [4019409](https://pubmed.ncbi.nlm.nih.gov/4019409/)
51. Dror TW, Rolider A, Bayer EA, Lamed R, Shoham Y. Regulation of expression of scaffoldin-related genes in *Clostridium thermocellum*. *J Bacteriol.* 2003; 185: 5109–5116. doi: [10.1128/JB.185.17.5109-5116.2003](https://doi.org/10.1128/JB.185.17.5109-5116.2003) PMID: [12923083](https://pubmed.ncbi.nlm.nih.gov/12923083/)
52. Dror TW, Morag E, Rolider A, Bayer EA, Lamed R, Shoham Y. Regulation of the cellulosomal CelS (*cel48A*) gene of *Clostridium thermocellum* is growth rate dependent. *J Bacteriol.* 2003; 185: 3042–3048. doi: [10.1128/JB.185.10.3042-3048.2003](https://doi.org/10.1128/JB.185.10.3042-3048.2003) PMID: [12730163](https://pubmed.ncbi.nlm.nih.gov/12730163/)
53. Béguin P, Rocancourt M, Chebrou MC, Aubert JP. Mapping of mRNA encoding endoglucanase A from *Clostridium thermocellum*. *Mol Gen Genet.* 1986; 202: 251–254. doi: [10.1007/BF00331645](https://doi.org/10.1007/BF00331645) PMID: [3010048](https://pubmed.ncbi.nlm.nih.gov/3010048/)

54. Mishra S, Béguin P, Aubert JP. Transcription of *Clostridium thermocellum* endoglucanase genes *celF* and *celD*. *J Bacteriol.* 1991; 173: 80–85. PMID: [1987137](#)
55. Lightfield J, Fram NR, Ely B. Across bacterial phyla, distantly-related genomes with similar genomic GC content have similar patterns of amino acid usage. *PLoS One.* 2011; 6: e17677. doi: [10.1371/journal.pone.0017677](#) PMID: [21423704](#)
56. Huang X, Fredrick KL, Helmann JD. Promoter recognition by *Bacillus subtilis* σ^W : autoregulation and partial overlap with the σ^X regulon. *J Bacteriol.* 1998; 180: 3765–3770. PMID: [9683469](#)
57. Mascher T, Hachmann AB, Helmann JD. Regulatory overlap and functional redundancy among *Bacillus subtilis* extracytoplasmic function sigma factors. *J Bacteriol.* 2007; 189: 6919–69127. doi: [10.1128/JB.00904-07](#) PMID: [17675383](#)