

SCIENTIFIC REPORTS



OPEN

Regulation of biomass degradation by alternative σ factors in cellulolytic clostridia

Lizett Ortiz de Ora¹, Raphael Lamed¹, Ya-Jun Liu², Jian Xu², Qiu Cui², Yingang Feng², Yuval Shoham³, Edward A. Bayer⁴ & Iván Muñoz-Gutiérrez^{4,5}

Bacteria can adjust their genetic programs via alternative σ factors to face new environmental pressures. Here, we analyzed a unique set of paralogous alternative σ factors, termed σ 's, which fine-tune the regulation of one of the most intricate cellulolytic systems in nature, the bacterial cellulosome, that is involved in degradation of environmental polysaccharides. We combined bioinformatics with experiments to decipher the regulatory networks of five σ 's in *Clostridium thermocellum*, the epitome of cellulolytic microorganisms, and one σ in *Pseudobacteroides cellulosolvens* which produces the cellulosomal system with the greatest known complexity. Despite high homology between different σ 's, our data suggest limited cross-talk among them. Remarkably, the major cross-talk occurs within the main cellulosomal genes which harbor the same σ '-dependent promoter elements, suggesting a promoter-based mechanism to guarantee the expression of relevant genes. Our findings provide insights into the mechanisms used by σ 's to differentiate among their corresponding regulons, representing a comprehensive overview of the regulation of the cellulosome to date. Finally, we show the advantage of using a heterologous host system for analysis of multiple σ 's, since information generated by their analysis in their natural host can be misinterpreted owing to a cascade of interactions among the different σ 's.

Bacteria can sense the extracellular environment and transmit information intracellularly by using different types of signal transduction mechanisms¹. After sensing the environment, one type of response is regulation of genes at the level of transcription initiation by alternative sigma (σ) factors allowing bacteria to adjust their transcriptional programs to environmental changes^{2,3}. σ factors are the key component of RNA polymerase, since they provide promoter specificity. All bacteria harbor one primary σ factor (known as the housekeeping σ factor, σ^{70} or σ^A) that is responsible for basal expression level of most genes. When the environmental conditions change, the housekeeping σ factor is substituted by the alternative σ factors, thereby redirecting the RNA polymerase to alternative promoters of genes that will help the bacterium deal with the new environmental conditions⁴.

Cellulolytic clostridia are anaerobic bacteria that use plant cell-wall polysaccharides as carbon sources. The structural complexity and composition of their substrate, together with the anoxic conditions of their ecosystem, have generated selective pressures for the evolution of extracellular multi-enzyme nanomachines called cellulosomes for efficient degradation of plant cell-wall polysaccharides⁵. During the process of cellulolytic biomass breakdown, cellulolytic clostridia control the type(s) of enzymatic subunits present in the cellulosome complex to suit the type(s) of polysaccharide(s) that are exposed during the degradation process^{6,7}. In the cellulosome-producing bacteria *Clostridium (Ruminiclostridium) thermocellum*, the enzymatic composition of the cellulosome is probably regulated by a group of at least 6 paralogous alternative σ^1 factors that are related to the *Bacillus subtilis* σ^1 factor^{8–12}. However, the regulons of this set of σ 's are poorly understood.

The features of the σ 's are consistent with almost all of the characteristics of the ECF (extracytoplasmic function) σ factors². Both σ 's and ECF σ factors share the following features. (i) They usually autoregulate their own

¹Department of Molecular Microbiology and Biotechnology, Tel Aviv University, Tel Aviv, Israel. ²CAS Key Laboratory of Biofuels and Shandong Provincial Key Laboratory of Energy Genetics, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, Shandong, China. ³Department of Biotechnology and Food Engineering, Technion-IIT, Haifa, Israel. ⁴Department of Biomolecular Sciences, The Weizmann Institute of Science, Rehovot, Israel. ⁵Present address: Outreach Research Training and Minority Science Programs, Francisco Ayala School of Biological Sciences, University of California, Irvine, California, USA. Correspondence and requests for materials should be addressed to I.M.-G. (email: ivan.munoz.gutierrez@gmail.com)

expression. (ii) They are usually located in an operon with an anti- σ factor gene. In the case of σ^1 factors, the latter gene is termed *rsgI* (i.e., regulation of *sigI*) that controls the activity of its cognate σ factor. (iii) The anti- σ factor is composed of three parts: an extracytoplasmic sensory module(s), a transmembrane domain, and an intracellular domain that sequesters the σ factor. (iv) The σ factor is activated by inhibiting the activity of the anti- σ factor.

The main difference between σ^1 and ECF σ factors resides in their protein structure. Alternative σ^1 factors harbor only the σ_2 domain of the σ^{70} family that is involved in the recognition of the -10 promoter element, and the σ_4 domain is substituted by a C-terminal domain termed σ_{1-C} that is likely involved in the recognition of the -35 promoter element⁸. Taken together, the above-mentioned characteristics render σ^1 factors unique members of the σ^{70} family.

Some cellulosome-producing clostridia are characterized by multiple σ^1 factors. The σ^1 factors of *C. thermocellum* are highly homologous with identities between 36 and 45% among them. This observation raises the question of how a set of highly homologous alternative σ^1 factors can avoid crosstalk among each other. This is of particular interest in a bacterium like *C. thermocellum*, which contains 8 paralogous σ^1 s, as well as other cellulosome-producing clostridia [notably, *Clostridium clariflavum*, *Clostridium straminosolvens*, *Clostridium* sp. Bc-iso-3, *Acetivibrio cellulolyticus*, and *Bacteroides* (*Pseudobacteroides*) *cellulosolvens*] which also produce between 8 and 16 highly homologous alternative σ^1 factors that are presumably involved in similar regulatory networks. Moreover, there are reports where a set of ECF σ factors exhibits high crosstalk^{13–15}, wherein the latter set shows less conservation compared to that of the *C. thermocellum* σ^1 factors. For example, the *B. subtilis* ECF σ^M , σ^W and σ^X factors share identities of only 25 to 32% but present high regulatory overlap¹⁴.

It should be noted that there is continued controversy regarding the classification of the cellulosome-producing clostridia¹⁶, and, in particular, for those species that produce complex or multiplex cellulosomes, characterized by a multiplicity of scaffoldin genes^{17,18}. In this context, *Acetivibrio cellulolyticus* and *Bacteroides cellulosolvens* were clearly misclassified in the original works^{19,20}, and both species were later determined to be members of the greater clostridial assemblage²¹. These latter discrepancies were not fully resolved by the recent attempt to reclassify of *Bacteroides cellulosolvens* as *Pseudoaeroides cellulosolvens*²².

In the present work, we demonstrate how a collection of five alternative σ^1 factors in *C. thermocellum*, namely σ^{11} , σ^{12} , σ^{13} , σ^{14} and σ^{16} , regulate the expression of 17 genes encoding different cellulosomal components. This analysis shows for the first time a sophisticated regulatory network of several alternative σ factors, which control the enzymatic composition of the cellulosome. Furthermore, our results show that σ^1 factors from cellulosome-producing bacteria use highly conserved promoter sequences to delimit the genes that are under control of a given σ^1 . Our findings indicate that the -35 promoter element, proposed to be recognized by the novel domain σ_{1-C} , is critical for the specificity of each σ^1 factor. This promoter element can be divided into two regions: a highly-conserved homopolymeric A-tract motif in the 3' region, that we herein propose as a general motif for σ^1 -dependent promoter recognition, and a more divergent region upstream of the A-tract motif that provides specificity to each σ^1 factor. By using this information, we identified the regulons of one σ^1 factor in *P. cellulosolvens*, a bacterium that produces the most complex cellulosomal system described until now²³. Our results provide a better view into the mechanisms used by multiple alternative σ^1 s to differentiate their corresponding regulons. This information is crucial for future efforts to predict regulons of multiple σ^1 factors in cellulolytic clostridia.

Results

Deciphering the regulatory networks of σ^1 factors. In a previous work, we predicted 40 putative σ^1 -dependent promoters in *C. thermocellum* by bioinformatic analysis¹⁰. To overcome the lack of convenient genetic tools to work directly in *C. thermocellum*²⁴, we analyzed the recognition of the predicted promoters by *C. thermocellum* σ^{13} and σ^{16} in a heterologous *B. subtilis* host system¹⁰. This analysis revealed that the main types of enzymatic genes regulated by *C. thermocellum* σ^{13} and σ^{16} are pectin-degrading enzymes and xylanases, respectively¹⁰. In order to confirm the predicted σ^1 -dependent promoters of the spectrum of *C. thermocellum* σ^1 factors that are proposed to be involved in the regulation of genes coding saccharolytic enzymes (σ^{11} to σ^{16})^{9,10}, we herein mapped the transcriptional start sites (TSSs) of the *sigI2-rsgI2*, *sigI3-rsgI3*, *sigI4-rsgI4* and *sigI5-rsgI5* operons by the rapid amplification of cDNA ends (5'-RACE) technique (Supplementary Fig. S1). The TSSs of the *sigI1-rsgI1* and *sigI6-rsgI6* operons were mapped in a previous work⁹. During this analysis, however, we failed to identify a TSS of the *sigI4-rsgI4* and *sigI5-rsgI5* operons using this technique. The σ^1 -dependent promoter sequences associated with the TSSs mapped in the present work are shown in Fig. 1A.

Subsequently, in order to decipher the regulatory networks of *C. thermocellum* σ^1 factors, we tested the recognition of the previously predicted σ^1 -dependent promoters¹⁰ with each of the *C. thermocellum* σ^1 -factors (i.e., σ^{11} – σ^6) in a heterologous *B. subtilis* host system (Fig. 1B). The members of the promoter library were previously fused to a lacZ reporter gene¹⁰. In the present work, the activation of the reporter was tested in 24-well cell culture plates with Spizizen's minimal medium using 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside (X-gal) as substrate (Fig. 1B). During the analysis, *C. thermocellum* σ^{15} did not recognize any of the predicted promoters. As shown in Fig. 1B, with few exceptions, the majority of the σ^1 -dependent promoters are specific for their own σ^1 factor. Interestingly, the major crosstalk between σ^1 s occurs in the two most important cellulosomal genes, *cipA* and *cel48S* (Fig. 1B). These genes encode the primary scaffoldin of the cellulosome, CipA^{11,18,25}, and the most abundant enzyme in the cellulosome, Cel48S^{6,26,27}, respectively.

σ^1 -dependent promoters are structured into three motifs. According to the consensus promoter sequence, generated with the σ^1 -dependent promoters that were experimentally confirmed (Fig. 1), we divided the σ^1 -dependent promoters into three regions: (i) a highly conserved CGAA tetrad in the -10 element, (ii) a homopolymeric AAAA tetrad, herein termed the “A-tract motif” at the 3' end of the -35 element, and (iii) a divergent region upstream of the A-tract motif. The most highly conserved promoter motifs, i.e., the A-tract in

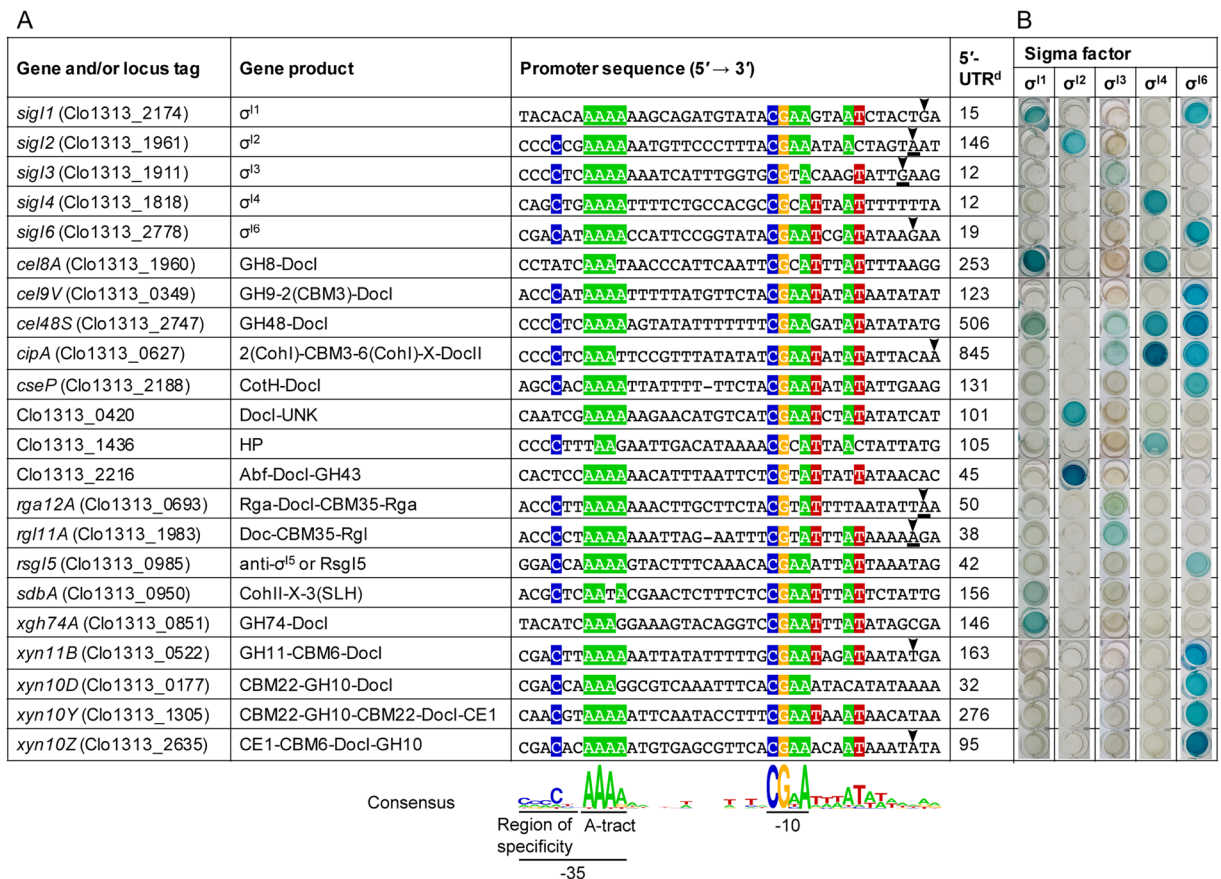


Figure 1. Regulons of *C. thermocellum* alternative σ^1 factors. **(A)** Alignment of σ^1 -dependent promoters that were experimentally confirmed. The most conserved nucleotides are highlighted in color. TSSs are indicated by arrowheads and the TSSs identified in the present work (Supplementary Fig. S1) are underlined. The TSS of *sig11*, *sig16*, *cipA*, *xyn11B* and *xyn10Z* were identified in previous works^{9,11,43}. Distances between the promoter sequences and the first codon of corresponding genes are shown in the column labeled 5'-UTR (5'-untranslated region). GH, glycoside hydrolase; Doc, dockerin; CBM, carbohydrate-binding module; Coh, cohesin; X, X-module (module of unknown function); CotH, spore coat protein H; UNK, unknown sequence; HP, hypothetical protein; Abf, Alpha-L-arabinofuranosidase; Rga, rhamnolacturan acylesterase; Rgl, rhamnolacturan lyase; SLH, S-layer homology domain; CE, carbohydrate esterase. **(B)** Recognition of σ^1 -dependent promoters by *C. thermocellum* σ^1 factors. The respective promoters were fused to a LacZ reporter gene and their recognition by the different *C. thermocellum* σ^1 factors was tested in a *B. subtilis* heterologous host, grown in 24-well cell culture plates with Spizizen's minimal medium and X-gal. The development of blue color indicated the activation of the σ^1 -dependent promoters by a given *C. thermocellum* σ^1 factor.

the -35 element and the CGAA tetrad in the -10 element, are proposed to be implicated in the “general” recognition of promoters by σ^1 s in cellulosome-producing bacteria (Fig. 1A). We suggest that the least conserved region, upstream of the A-tract motif in the -35 element, is implicated in the specificity of the different σ^1 factors. Hence, we termed this divergent region as “region of specificity” (Fig. 1A). The alignments generated with the promoter sequences of each σ^1 regulon support this observation (Fig. 2). A comparison of the region of specificity shows how each σ^1 factor has different preferences. For example, whereas *C. thermocellum* σ^{11} favors promoters with a CTC triad immediately upstream of the A-tract motif, *C. thermocellum* σ^{13} prefers a CCC triad two nucleotides upstream of the A-tract motif (Fig. 2).

Site-directed mutagenesis analysis confirmed the three motifs of σ^1 promoters. The high stringency of promoter recognition by alternative σ^1 factors was analyzed using *C. thermocellum* σ^{13} , because it recognizes highly conserve promoters with a characteristic C triad upstream of the A-tract motif in the -35 promoter element (Fig. 2). First, in order to verify the σ^{13} -dependent promoters, we mapped the TSSs of *rgl11A* and *rga12A* by the 5'-RACE technique (Supplementary Fig. S1). Next, the crucial nucleotides for recognition of *C. thermocellum* σ^1 -dependent promoters were analyzed by site-directed mutagenesis experiments using the σ^{13} -dependent promoter of *rgl11A*. The σ^{13} -dependent promoter of *rgl11A* was selected for the analysis because previous work showed that this promoter presented the highest activation by the *C. thermocellum* σ^{13} factor¹⁰.

A promoter library with single transversion mutations (A to T or T to A, and C to G or G to C) in both -35 and -10 promoter elements was created. Additional transversion mutations were created upstream and

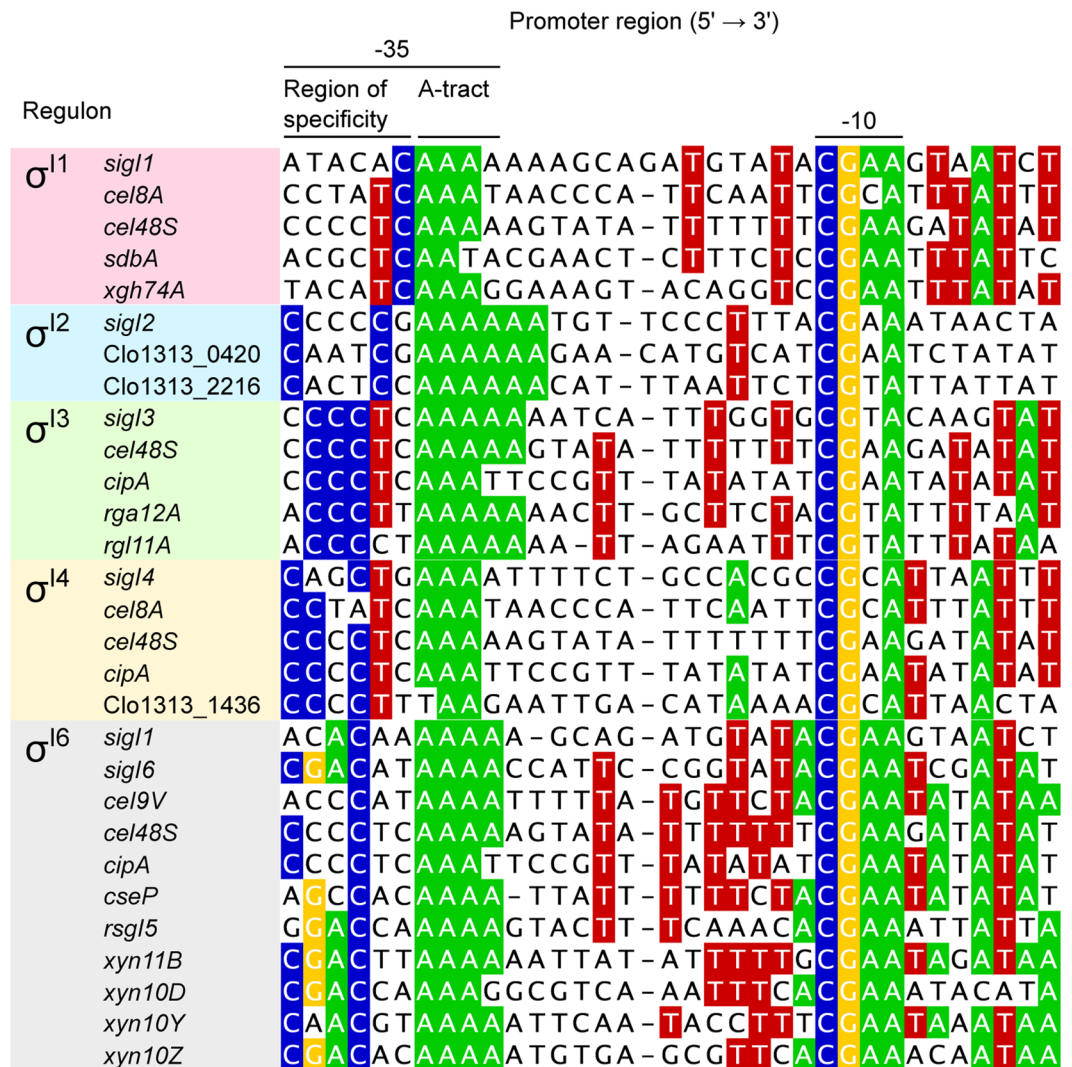


Figure 2. Dissection of *C. thermocellum* σ^1 -dependent promoters. The conservation of nucleotides shown in the alignment is divided into σ^{11} -, σ^{12} -, σ^{13} -, σ^{14} -, and σ^{16} -dependent promoters. The promoter sequences of each σ^1 regulon were taken from Fig. 1A. The most conserved nucleotides of each σ^1 regulon are highlighted.

downstream of both -35 and -10 promoter elements, replacing less conserved nucleotides. Subsequently, the promoter library was fused to a *lacZ-gfp* reporter operon. The promoter activities were thus studied in the heterologous *B. subtilis* host system. Quantification of promoter activities was performed by measuring GFP fluorescence, and the fluorescence of the different promoter variants was compared to that of the wild-type *rgl11A* promoter.

As shown in Fig. 3, mutations in the most conserved nucleotides in both -35 and -10 promoter elements have a negative effect, abolishing the detection of the GFP fluorescence. Mutations in less conserved nucleotides have a negligible effect, showing promoter activities at the same level as that of the wild-type promoter. Interestingly, the mutation from T to A between the highly conserved CCC and AAA triads of the -35 promoter element (*rgl11A*-Mut8; sequence underlined in CCCCTAAA) increased the fluorescence by 53%. As can be observed in the WebLogo shown in Fig. 3, *C. thermocellum* σ^{13} apparently prefers promoter sequences enriched in adenines downstream of the -35 promoter element. This observation can explain the rise in activity identified in the promoter version of *rgl11A*-Mut8.

σ^1 factor recognizes highly stringent promoter sequences. To confirm the implication of the region of specificity in the stringency of σ^1 -dependent promoters, we searched for σ^1 -dependent promoters in the genome of *P. cellulosolvens* that resemble those of the *C. thermocellum* σ^{13} -dependent promoters. *P. cellulosolvens* was chosen for further comparison, because it produces the most complex cellulosome known to date²³, and a previous report showed that as the cellulosome becomes more elaborate, the bacterium harbors more σ^1 factors¹¹. This observation was confirmed by our analysis which revealed that *P. cellulosolvens* genome encodes 16 σ^1 paralogues (Supplementary Table S4), making this bacterium an excellent model for the study of multiple σ^1 factors.

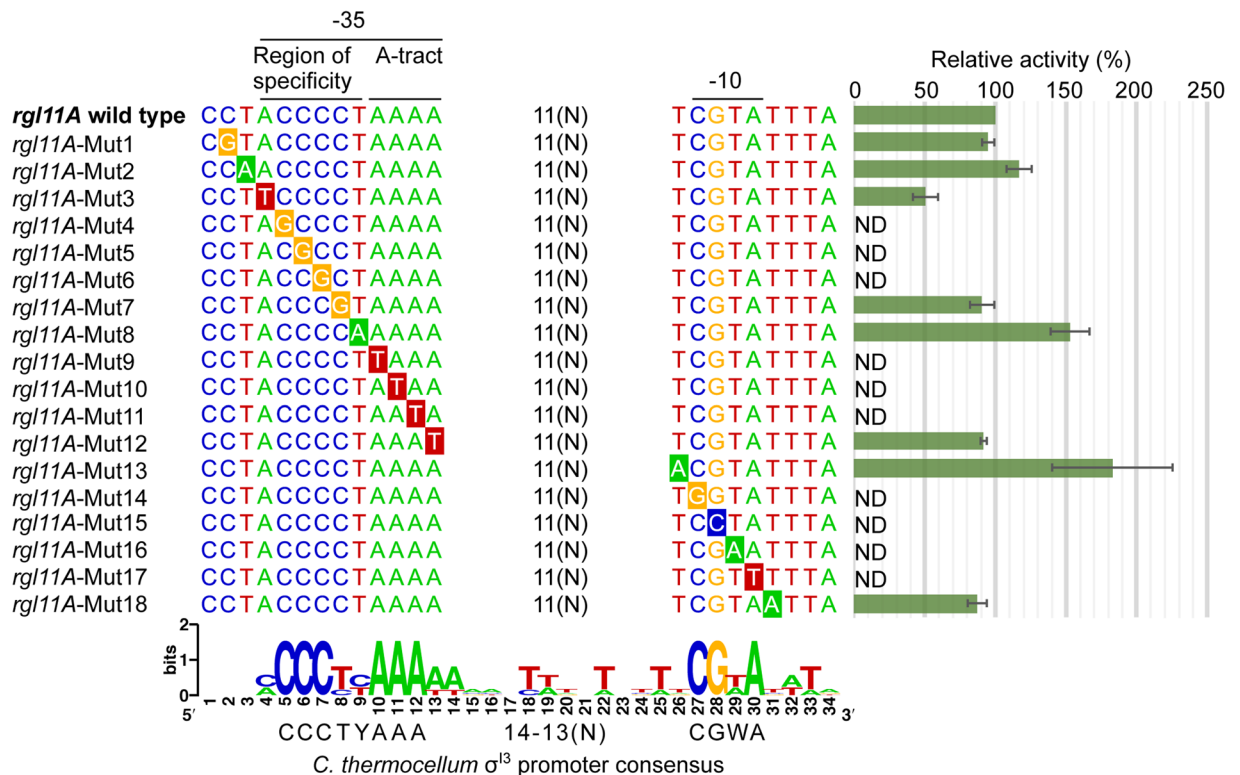


Figure 3. Evaluation of the validity of the σ^1 promoter motifs by mutagenesis. The *rgl11A* σ^{13} -dependent promoter was selected for this analysis, because it presented the highest activation by the *C. thermocellum* σ^{13} factor¹⁰. The WebLogo was generated with σ^{13} -dependent promoter sequences (Fig. 1A) and is shown to illustrate the effect of mutations in the most conserved nucleotides of the three motifs: (i) the -10 promoter element, (ii) the A-tract motif and (iii) the region of specificity. Mutations are indicated by highlighted nucleotides, and the effect of each mutation is shown as relative activity, compared to that of the wild-type control promoter *rgl11A*, defined as 100%. The relative activity shown is the average of three independent experiments. Predicted -35 and -10 promoter elements are indicated by lines above the *rgl11A* promoter sequences. The nucleotide code Y represents C or T, W represents A or T, and N represents any nucleotide. ND means not detected.

For the identification of putative σ^1 -dependent promoters in *P. cellulosolvens*, we first analyzed the upstream regions of all predicted σ^1 genes (Supplementary Table S4). The search for promoters upstream of *P. cellulosolvens* σ^1 genes was performed taking into account that the AAA triad and the CG dyad in the -35 and -10 promoter regions, respectively, are the most conserved nucleotides of the σ^1 -dependent promoters of cellulosome-producing bacteria (Fig. 1). The predicted putative σ^1 -dependent promoters of *P. cellulosolvens* σ^1 genes are shown in Supplementary Table S4.

In order to predict putative σ^1 -dependent promoters, the genome sequence of *P. cellulosolvens* was analyzed using the same promoter motifs that were employed during the analysis of the upstream region of σ^1 genes (AAA in the -35 region and CG in the -10 region). To delimit the promoter search, we included A, T or C downstream of the CG dyad in the -10 region (CGHH, where H represents A, T or C), because the predicted promoters of σ^1 genes contain these nucleotides (Supplementary Table S4). The spacing between the -10 and -35 promoter elements was allowed to be between 12 to 15 bases. During this analysis, a collection of 140 σ^1 -dependent promoters were predicted (Supplementary Table S5).

Subsequently, in order to find putative *P. cellulosolvens* σ^1 -dependent promoter sequences resembling those of *C. thermocellum* σ^{13} -dependent promoters, the collection of predicted σ^1 -dependent promoters of *P. cellulosolvens* was analyzed by searching for cytosine enrichments in the region of specificity in the -35 promoter element. This analysis allowed the identification of one putative promoter upstream of *P. cellulosolvens* *sig111* and five putative promoters upstream of genes encoding saccharolytic enzymes (Bccel_3806, Bccel_5179, Bccel_5541, Bccel_5619 and Bccel_5627). The alignment of the *P. cellulosolvens* predicted promoters is shown in Fig. 4. The highly conserved CCC triad can be observed in the region of specificity immediately upstream of the A-tract motif in the -35 promoter element. Additionally, the highly conserved CGCAT pentad in the -10 promoter element can also be observed. The majority of the *P. cellulosolvens* predicted promoters correspond to genes encoding catalytic modules which are probably involved in pectin degradation, such as pectate lyase, pectin esterase, Rga and Rgl [Fig. 4; ref.^{28–30}].

Finally, to test the ability of *P. cellulosolvens* σ^{111} to recognize the predicted promoters shown in Fig. 4, we fused the predicted promoters to a *gfp-lacZ* reporter operon. To overcome the lack of genetic tools in *P. cellulosolvens*, the

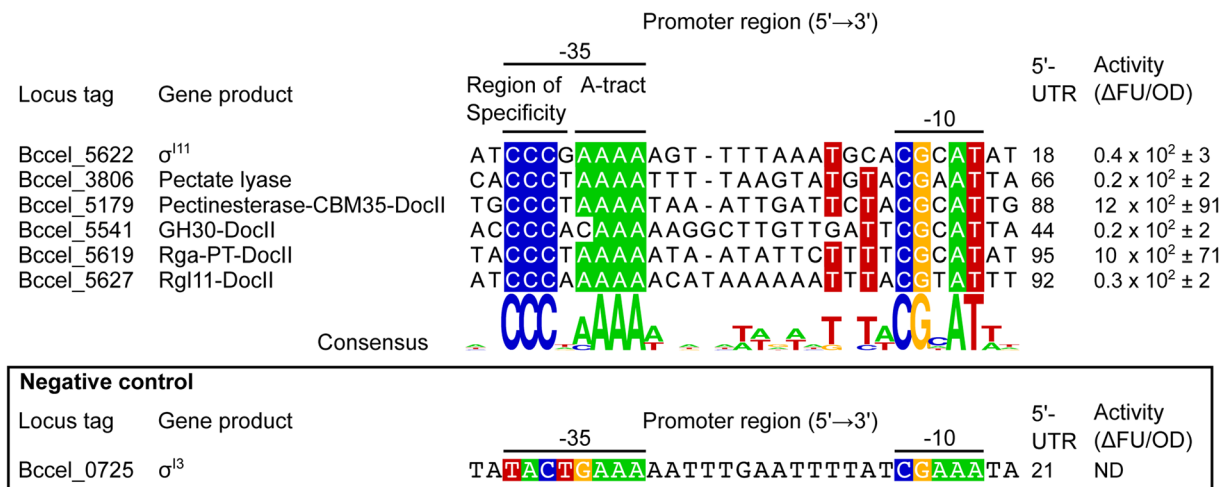


Figure 4. Alignment of *P. cellulosolvens* σ^{11} -dependent promoter sequences and quantitative evaluation of their recognition by the *P. cellulosolvens* σ^{11} factor. The WebLogo was generated with the sequence shown in the alignment. Distances between the promoter region sequences used for the alignment and the first codon of corresponding genes are shown in the 5'-UTR (5'-untranslated region) column. Promoter activity was measured by quantifying the fluorescence (average of three independent experiments). Fluorescence units (FU) were calculated when cells reached an optical density (600 nm) of 1. ΔFU represents the activity of the induced promoter after subtracting the fluorescence obtained under uninduced conditions. ND (not detected) was assigned where ΔFU was negative, zero or the standard deviation exceeded ΔFU. As a negative control, the predicted σ^1 -dependent promoter of the *P. cellulosolvens* σ^{13} gene was used. CBM, carbohydrate-binding module; Doc, dockerin; GH, glycoside hydrolase; Rga, rhamnogalacturan acetylerase; PT, Pro-Thr repeat (pfam04886); Rgl, rhamnogalacturonan lyase.

promoter activities were also studied in a heterologous *B. subtilis* host system¹⁰. The recognition of the predicted promoters by *P. cellulosolvens* σ^{11} was analyzed by measuring GFP fluorescence. As shown in Fig. 4, all the predicted promoters were recognized by *P. cellulosolvens* σ^{11} , suggesting that this alternative σ factor is likely involved in the regulation of genes encoding pectin-degrading enzymes, similar to its orthologous *C. thermocellum* σ^{13} (Fig. 1B).

Discussion

Cellulolytic clostridia produce one of the most efficient systems to degrade plant-cell wall polysaccharides — the multi-enzyme cellulosome complex³¹. The genomes of this group of bacteria contain dozens of genes, which encode different types of carbohydrate-active hydrolyzing enzymes and structural scaffoldin subunits, whose regulation is fine-tuned during the hydrolysis of biomass³². Although the cellulolytic capacities of cellulosome-producing bacteria have been the subject of study for many years^{33–35}, the regulatory mechanisms that govern these processes are poorly known. Here, we show how a collection of five alternative σ^1 factors in *C. thermocellum*, namely σ^{11} , σ^{12} , σ^{13} , σ^{14} and σ^{16} , regulate the expression of 17 genes encoding different cellulosomal components (Fig. 1). This analysis reveals for the first time a sophisticated regulatory network of several alternative σ factors, which together control the enzymatic composition of the cellulosome (Fig. 1). Our results show how each σ^1 factor has a particular regulon (Fig. 2) that correlates with its corresponding anti- σ^1 factor that selectively senses a target polysaccharide^{8,36,37}. Furthermore, with the results collected in *C. thermocellum*, we were able to predict the putative regulon of the *P. cellulosolvens* σ^{11} factor (Fig. 4) and provide experimental evidence, for the first time in *P. cellulosolvens*, of the regulation of genes encoding cellulosomal components by an alternative σ factor (Fig. 4).

Alternative σ^1 factors are a subfamily of the σ^{70} family which are unique, because not all σ^1 factors are involved in the regulation of genes encoding saccharolytic enzymes or cellulosomal components. *B. subtilis* has only one σ^1 factor that is induced by heat-shock¹² and is involved in the maintenance of cell envelope integrity and homeostasis³⁸. These observations show how σ^1 s are malleable alternative σ factors that have been adapted in different bacteria to perform various functions. However, during the course of evolution, when σ^1 genes were extensively duplicated in the genome of cellulolytic clostridia to regulate the enzymatic composition of the cellulosome, their dependent promoter sequences were specialized in order to avoid regulatory overlap.

In the current work, three important observations support the hypothesis that alternative σ^1 factors recognize stringent promoter sequences to discriminate among dozens of promoters that share similar motifs. First, each σ^1 factor involved in the regulation of genes encoding saccharolytic enzymes and cellulosomal components in *C. thermocellum* has specific regulons with characteristic promoter sequences (Fig. 2). Second, mutagenesis analysis performed with the σ^{13} -dependent promoter of *C. thermocellum* *rgl11A* revealed a drastic reduction in promoter activity when the most conserved nucleotides were mutated (Fig. 3). Finally, the *C. thermocellum* σ^{13} promoter consensus sequence allowed the identification of genes encoding putative pectin-degrading enzymes in another cellulolytic clostridium, *P. cellulosolvens* (Fig. 4).

Our results demonstrate that the region located upstream of the A-tract motif in the -35 promoter element has a crucial role for the discrimination of σ^1 -dependent promoters by the different σ^1 factors (Figs 1 and 2). In

general, all σ^1 factors recognize promoters with a CGAA tetrad and an A-tract motif in the -10 and -35 promoter elements, respectively (Figs 1 and 2). However, the region upstream of the A-tract motif is characteristic to each σ^1 regulon (Fig. 2). Indeed, this region, herein named region of specificity, was used to identify σ^1 -dependent promoters of genes involved in pectin degradation in the genome of *P. cellulosolvens* (Fig. 4).

The key role of the -35 promoter element for the recognition of σ^1 factors is also supported by the conservation of homopolymeric A-tract motifs (Fig. 1). Intriguingly, in contrast to the A-tract motifs of σ^1 -dependent promoters, a recent report demonstrated in *B. subtilis* that a homopolymeric T-tract motif contributes to the activation of several ECF σ factors¹³. The latter T-tract motif is located downstream of the highly conserved AAC triad in the -35 promoter element of ECF σ factors. Here, the relevance of the A-tract motif was demonstrated during the mutagenesis analysis of the σ^{13} -dependent promoter of *C. thermocellum* *rgl11A*. We demonstrated how single mutations in the AAA triad of the -35 promoter element destroyed the recognition of the promoter by *C. thermocellum* σ^{13} (Fig. 3). Furthermore, the addition of an extra A, by mutating the T located immediately upstream of the A-tract motif (i.e., *rgl11A*-Mut8; sequence underlined in CCCCTAAA) increased the strength of the promoter, providing promoter activity 53% higher than that of the wild-type promoter (Fig. 3).

It is also interesting that both T-tract and A-tract motifs are located in the same position in the -35 promoter region of ECF σ - and σ^1 -dependent promoters, respectively [ref.¹³ and Fig. 1, respectively]. Hence, we propose that this motif may help avoid crosstalk between ECF σ s and σ^1 s. In this context, the respective -10 promoter elements of ECF σ - and σ^1 -dependent promoters are highly similar [ref.¹³ and Fig. 1, respectively]. The discrimination of promoters is critical in cellulolytic clostridia that harbor multiple σ^1 factors, not only to avoid crosstalk between the different σ^1 s, but also to discriminate between the promoters of other types of alternative σ factors, such as the ECF σ s. According to the Microbial Signal Transduction database (MiST_{2.2}, <http://mistdb.com/>), the genome of *C. thermocellum* encodes 7 ECF σ s, and our analysis revealed that the *P. cellulosolvens* genome encodes at least 22 ECF σ s (Supplementary Table S6). Therefore, the possibility of having ECF σ - and σ^1 -dependent promoters with similar -35 regions is very high. Consequently, the A-tract motif of σ^1 -dependent promoters may also serve to avoid their recognition by ECF σ factors. This idea is further supported by the observation that none of the σ^1 -dependent promoters, which have been tested in our laboratory by using the *B. subtilis* heterologous host cell system (that is devoid of its native σ^1 /RsgI system), were activated by the resident *B. subtilis* σ factors^{10,11}.

It is worth mentioning that genomic context may also play a defined role in promoter selectivity. It has been reported that some important sequences which reside outside of the classic -35 and -10 promoter elements can be implicated in the recognition of the promoter^{39,40}. For example, most of the promoters that are dependent on the *Escherichia coli* and *Salmonella enterica* σ^E , an ECF σ factor, require a sequence upstream of the -35 promoter element (UP-element) to increase their strength³⁹. Likewise, promoters that are dependent on σ^1 factors may also require UP-elements. The presence of motifs that reside outside of the classic -35 and -10 promoter elements that help to increase the strength of the promoter, or compensate for a poor -35 or -10 promoter element that deviates from consensus, is a phenomenon known as “mix and match”^{40–42}. Future identification of σ^1 -dependent promoters may be improved by taking into account mix-and-matching as a promoter recognition mechanism, thereby generating a better understanding of the biological function of each σ^1 factor.

It is also important to note that, although the *C. thermocellum* σ^1 factors have a defined regulon with little crosstalk, the major regulatory overlap between σ^1 s occurs in the two most important cellulosomal genes, *cipA*^{11,18,25} and *cel48S*^{6,26,27} (Fig. 1B). Close inspection of the σ^1 -dependent promoter of *C. thermocellum* *cipA* and *cel48S* reveals that both -35 and -10 promoter regions are identical, with a CCCCTAAA nonad and a CGAA tetrad, respectively (Fig. 1A). Additionally, the σ^1 -dependent promoter of *C. thermocellum* *cipA* and *cel48S* have a conserved AT dyad, three nucleotides downstream of the CGAA tetrad in the -10 promoter element that is present in nearly all of the σ^1 -dependent promoters shown in Fig. 1A. These observations suggest that the σ^1 -dependent promoter of *C. thermocellum* *cipA* and *cel48S* may represent a type of “universal” promoter that is used by the bacterium to assure expression of relevant genes.

Previous works have shown that both *CipA* and *Cel48S* are fundamental components of the cellulosome^{25–27}. Hence their expression should be assured in the presence of a wide number of polysaccharide substrates and conditions^{6,27}. In this sense, it would seem logical to use more than one alternative σ factor to regulate their expression. However, the recognition of specific promoter sequences by each of the regulators would be the consequence of a long and complex evolutionary process. A more practical approach to avoid this process is the utilization of a single “universal promoter” such as the one presented in this work. From an evolutionary point of view, it is very likely that the common ancestor of these σ^1 factors recognized a similar promoter sequence to that used by *cipA* and *cel48S*. Later, each of the duplicated genes evolved to encode specialized versions of the σ^1 s capable of recognizing unique promoters, thus reducing unnecessary crosstalk and limiting their regulons while maintaining regulatory overlap of these critical components.

Interestingly, the σ^1 -dependent promoters of *C. thermocellum* *sig13* harbor almost the same promoter elements present in *cipA* and *cel48S* σ^1 -dependent promoters. In contrast, the σ^1 -dependent promoters of *C. thermocellum* *sig13* are only recognized by σ^{13} (Fig. 1B). The difference is present in the -10 promoter element. Whereas the σ^1 -dependent promoter of *cipA* and *cel48S* have the CGAA tetrad in the -10 element, the σ^1 -dependent promoter of *sig13* has the CGTA tetrad. Additionally, the conserved AT dyad, downstream of the -10 element that is present in the σ^1 -dependent promoter of *cipA* and *cel48S* (Fig. 1A), is less conserved in the promoter of *sig13*, since it has a GT dyad in the same position (Fig. 1A). Thus, these small changes in the -10 promoter element can also help to improve specificity and avoid regulatory overlap.

Direct analysis of the multiple σ^1 factors in *C. thermocellum* represents a serious challenge, owing to the regulatory overlap between the different σ^1 factors with some of the important genes, such as *cipA* and *cel48S* (Fig. 1). Moreover, the genome of *C. thermocellum* may harbor additional as-yet-unidentified σ^1 -dependent promoters that may be activated by more than one σ^1 factor. Therefore, during future studies of a particular σ^1 factor directly in *C. thermocellum*, we could expect a cascade of interactions among the different σ^1 s. Additionally, cellulosomal component genes can also have σ^A -dependent promoters^{11,43} and other regulatory proteins⁴⁴, making the analysis

of σ^1 s directly in *C. thermocellum* more complex. If these observations are not taken into account during the direct analysis of σ^1 s in *C. thermocellum*, the results obtained can be misinterpreted. These observations apply also for other cellulolytic clostridia with multiple σ^1 s, such as *P. cellulosolvens* (Supplementary Table S4). Hence, as shown in the present report, the application of the heterologous *B. subtilis* host system for analysis of the multiple σ^1 factors of cellulolytic clostridia is advantageous. The predicted σ^1 -dependent promoters can be experimentally tested in *B. subtilis* and corroborated *in vivo* by mapping the TSSs of their associated genes (Supplementary Fig. S1).

In conclusion, in the present work, we show how the employment of classical microbiology genetic tools, such as the *LacZ* reporter system⁴⁵, together with the well-known Gram-positive bacterium *B. subtilis* as heterologous host¹⁰, enabled us to decipher the regulatory networks of the multiple alternative σ^1 factors, of one of the most efficient and most intricate cellulolytic systems in nature – the bacterial cellulosome.

Methods

Bacterial strains, growth media and culture conditions. *C. thermocellum* DSM 1313 and *P. cellulosolvens* DSM2933 were obtained from the DSMZ (German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany). The recognition of predicted σ^1 -dependent promoters by *C. thermocellum* and *P. cellulosolvens* σ^1 factors was analyzed in a heterologous *B. subtilis* host system that was developed in a previous work¹⁰. The *B. subtilis* strains constructed used in this work (Supplementary Table S1) are isogenic derivatives of the *B. subtilis* strain CO02 that is devoid of its *sigI-rsgI* operon¹⁰.

C. thermocellum and *P. cellulosolvens* were grown using the media and condition described by the ref.⁹ and the DSMZ, respectively. *B. subtilis* and *E. coli* were cultivated routinely on solid LB Broth (Lennox, Difco, BD Diagnostics, Maryland, USA) or in liquid LB Broth (at 250 rpm) at 37 °C. The expression of genes under the P_{xyIA} promoter was induced with D-xylose using a final concentration of 10 g/L (Sigma-Aldrich). When required the following antibiotics (all from Sigma-Aldrich) were added at the indicated final concentration: ampicillin (100 μ g/mL, Amp), kanamycin (50 μ g/mL, Kan), chloramphenicol (5 μ g/mL, Cam) or erythromycin (3 μ g/mL, Erm).

DNA manipulation techniques and construction of plasmids. The primers and plasmids used in the present work are listed in Tables S2 and S3, respectively. Plasmids were constructed by standard molecular cloning techniques using restriction enzymes and ligase, or by the ligase-independent cloning technique based on the In-Fusion HD Cloning Kit (Clontech Laboratories, Inc., California, USA).

The pAX01 integration vector was used to express the *C. thermocellum* and *P. cellulosolvens* σ^1 factors in *B. subtilis*⁴⁶. This plasmid has the xylose-inducible promoter P_{xyIA} to control the expression of a gene of interest, integrates at *B. subtilis lacA* locus and harbors an *erm* cassette as a selectable marker. First, pAX01 was linearized with the restriction enzyme BamHI. Later, *P. cellulosolvens sigI11* was PCR-amplified using primers P1 and P2. Finally, the PCR product was cloned using the In-Fusion HD Cloning Kit into the linearized pAX01 vector, thereby obtaining the pAX01-Bc-SigI11 plasmid. To express the *C. thermocellum* σ^{11} , σ^{12} , σ^{13} , σ^{14} , σ^{15} and σ^{16} factors in *B. subtilis* we used the pAX01 derived plasmids, pAX01-SigI1, pAX01-SigI2, pAX01-SigI3, pAX01-SigI4, pAX01-SigI5 and pAX01-SigI6, that were constructed in previous works^{10,11}.

The regulons of *C. thermocellum* σ^{11} , σ^{12} , σ^{13} , σ^{14} , σ^{15} and σ^{16} were analyzed by using the 40 putative σ^1 -dependent promoters that were predicted in a previous work¹⁰. With the exception of the putative predicted σ^1 -dependent promoters of *C. thermocellum cel48S*, we used a library of the previously predicted promoters that was fused to the *LacZ* reporter gene of the pBS1C-*LacZ* integration vector in a previous work¹⁰. pBS1C-*LacZ* integrates at the *B. subtilis amyE* locus and carries a *cam* resistance cassette as a selectable marker⁴⁷. In the case of the putative predicted σ^1 -dependent promoters of *C. thermocellum cel48S*, its sequence was PCR-amplified using primers P3 and P4. Subsequently, the PCR product was digested with the restriction enzymes EcoRI and BamHI and fused to the *LacZ* reporter gene of pBS1C-*LacZ* that was previously cut with the same restriction enzymes, thereby obtaining the pProm-Ct-Cel48S derived plasmid.

In order to study the important nucleotides for the recognition of σ^1 -dependent promoters, we used the pBS1C-GFP-*LacZ* integration vector which harbors a promoterless *gfp-lacZ* reporter operon¹¹. pBS1C-GFP-*LacZ* integrates at the *B. subtilis amyE* locus and carries a *cam* resistance cassette as a selectable marker¹¹. The analysis of promoter recognition was performed with mutant versions of the σ^{13} -dependent promoter of *C. thermocellum rgl11A* that were created by site-directed mutagenesis and fused to the *gfp-lacZ* reporter operon of pBS1C-GFP-*LacZ*. To introduce individual mutations in both –35 and –10 promoter regions, the reverse primers from P5 to P22, which contain the mutated nucleotide, were used with the forward primer P23. In order to compare the mutant versions, a wild type version of the *rgl11A* σ^{13} -dependent promoter was PCR-amplified using the primer pair P23-P24. After PCR amplification of the promoter mutant versions, the PCR products were digested with the restriction enzymes EcoRI and BamHI. Finally, each digested PCR product was cloned into pBS1C-GFP-*LacZ* that was digested previously with the same restriction enzymes, thereby obtaining the pBS1C-GFP-*LacZ* derived plasmids listed in Supplementary Table S3.

The recognition of σ^1 -dependent promoters by *P. cellulosolvens* σ^{111} was analyzed with the pBS1C-GFP-*LacZ* integration vector. The predicted σ^1 -dependent promoters of Bccel_5622 (*P. cellulosolvens* σ^{111} gene), Bccel_3856, Bccel_5179, Bccel_5541, Bccel_5619 and Bccel_5627 were PCR-amplified using the primer pairs P25-P26, P27-P28, P29-P30, P31-P32, P33-P34 and P35-P36, respectively. Later, each PCR product was digested with the restriction enzymes EcoRI and BamHI, except the promoter of Bccel_5541 that was PCR-amplified for cloning with the In-Fusion system. Finally, each digested PCR product was cloned into the pBS1C-GFP-*LacZ* plasmid that was digested previously with the same restriction enzymes, thereby obtaining the pBS1C-GFP-*LacZ* derived plasmids listed in Supplementary Table S3 (plasmid #26 to #30). In the case of the predicted promoter of Bccel_5541, its DNA sequence was cloned into pBS1C-GFP-*LacZ* (previously linearized with EcoRI and BamHI) using the In-Fusion HD Cloning Kit (Supplementary Table S3, plasmid #31).

Mapping of the TSSs. In order to map the TSSs we purified total RNA of *C. thermocellum* following the protocols described in a previous publication¹¹. The mRNA 5'-ends were mapped using the 5'-RACE technique with the SMARTer[®] RACE 5'/3' kit (Clontech) according to supplier protocols. Briefly, total RNA was subjected to RT-PCR with random primers and the SMARTer II oligonucleotide. Subsequently the 5'-RACE-Ready cDNA was submitted to a PCR amplification using the Universal Primer A Mix (a combination of Universal Primer Long and Universal Primer Short) and the gene specific primer (P40, P41, P42, P43 or P44 for *sigI2*, *sigI3*, *sigI4*, *rgl11A* and *rga12A*, respectively; Supplementary Table S2). Then, this PCR-product was subjected to a second PCR with the Universal Primer Short and the nested gene specific primer (P45, P46, P47, P48 or P49 for *sigI2*, *sigI3*, *sigI4*, *rgl11A* and *rga12A*, respectively; Supplementary Table S2). Finally, the PCR-products were gel purified, cloned into pRACE using the In-Fusion[®] HD Cloning kit and sequenced.

Construction of *B. subtilis* strains and analysis of promoter activities. *B. subtilis* was transformed by using the natural competence method⁴⁸. Chromosomal integration of plasmids by a double-crossover event was confirmed by colony PCR using the primers listed in Tables S2. To analyze the LacZ reporter system, *B. subtilis* strain samples were taken from the -80°C glycerol stock, inoculated in LB broth (3 mL) with Cam and grown overnight. The next day, the cells were centrifuged 15 min at $3000 \times g$ and resuspended in 1 mL of Spizizen's minimal media⁴⁸ with fructose (1.8 g/L final) as the carbon source. Finally, the cells were incubated in 24-well cell culture plates at 150 rpm. To observe the LacZ activity, the Spizizen's minimal media was supplemented with X-gal (40 mg/L final). To measure the fluorescence associated to the GFP reporter system, we followed the protocol described elsewhere¹¹. Fluorescence units (FU) were calculated when the cells reached an optical density at 600 nm of 1 and represent the activity of the induced promoter after subtracting the values obtained under uninduced conditions, i.e. $\Delta\text{FU} = \text{FU}_{\text{induced cells}} - \text{FU}_{\text{uninduced cells}}$.

Bioinformatics. Promoter motifs searches were carried out with the Pattern Locator program⁴⁹. The analysis of the promoter motif sequences was performed with the Jalview software⁵⁰. Multiple sequence alignment (MSA) of promoter sequences was performed using the T-Coffee algorithm⁵¹ implemented by Jalview. DNA sequence logos were generated with the program WebLogo⁵².

Data availability. All data generated or analysed during this study are included in this published article (and its Supplementary Information files).

References

- Mascher, T. Signaling diversity and evolution of extracytoplasmic function (ECF) σ factors. *Curr. Opin. Microbiol.* **16**, 148–55 (2013).
- Staron, A. *et al.* The third pillar of bacterial signal transduction: classification of the extracytoplasmic function (ECF) σ factor protein family. *Mol. Microbiol.* **74**, 557–581 (2009).
- Browning, D. F. & Busby, S. J. W. Local and global regulation of transcription initiation in bacteria. *Nat. Rev. Microbiol.* **14**, 638–650 (2016).
- Helmann, J. D. The extracytoplasmic function (ECF) sigma factors. *Adv. Microb. Physiol.* **46**, 47–110 (2002).
- Bayer, E. A., Belaich, J.-P., Shoham, Y. & Lamed, R. The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. *Annu. Rev. Microbiol.* **58**, 521–554 (2004).
- Raman, B. *et al.* Impact of pretreated Switchgrass and biomass carbohydrates on *Clostridium thermocellum* ATCC 27405 cellulosome composition: a quantitative proteomic analysis. *PLoS One* **4**, e5271 (2009).
- Blouzard, J. C. *et al.* Modulation of cellulosome composition in *Clostridium cellulolyticum*: adaptation to the polysaccharide environment revealed by proteomic and carbohydrate-active enzyme analyses. *Proteomics* **10**, 541–554 (2010).
- Kahel-Raifer, H. *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.* **308**, 84–93 (2010).
- Nataf, Y. *et al.* *Clostridium thermocellum* cellulosomal genes are regulated by extracytoplasmic polysaccharides via alternative sigma factors. *Proc. Natl. Acad. Sci.* **107**, 18646–18651 (2010).
- Muñoz-Gutiérrez, I. *et al.* Decoding biomass-sensing regulons of *Clostridium thermocellum* alternative Sigma-I factors in a heterologous *Bacillus subtilis* host system. *PLoS One* **11**, e0146316 (2016).
- Ortiz de Ora, L. *et al.* Revisiting the regulation of the primary scaffoldin gene in *Clostridium thermocellum*. *Appl. Environ. Microbiol.* **83**, e03088–16 (2017).
- Zuber, U., Drzewiecki, K. & Hecker, M. Putative sigma factor SigI (YkoZ) of *Bacillus subtilis* is induced by heat shock. *J. Bacteriol.* **183**, 1472–1475 (2001).
- Gaballa, A. *et al.* Modulation of extracytoplasmic function (ECF) sigma factor promoter selectivity by spacer region sequence. *Nucleic Acids Res.* **46**, 134–145 (2018).
- Mascher, T., Hachmann, A.-B. & Helmann, J. D. Regulatory overlap and functional redundancy among *Bacillus subtilis* extracytoplasmic function σ factors. *J. Bacteriol.* **189**, 6919–6927 (2007).
- Huang, X., Fredrick, K. L. & Helmann, J. D. Promoter recognition by *Bacillus subtilis* σ^W : autoregulation and partial overlap with the σ^X regulon. *J. Bacteriol.* **180**, 3765–3770 (1998).
- Yutin, N. & Galperin, M. Y. A genomic update on clostridial phylogeny: Gram-negative spore formers and other misplaced clostridia. *Environ. Microbiol.* **15**, 2631–2641 (2013).
- Bayer, E. A., Shoham, Y. & Lamed, R. In *The Prokaryotes: Prokaryotic Physiology and Biochemistry*. (ed Rosenberg, E.), 215–266 (Springer-Verlag, Berlin, 2013).
- Xu, Q. *et al.* Dramatic performance of *Clostridium thermocellum* explained by its wide range of cellulase modalities. *Sci. Adv.* **2**, e1501254 (2016).
- Giuliano, C. & Khan, A. W. Cellulase and sugar formation by *Bacteroides cellulosolvens*, a newly isolated cellulolytic anaerobe. *Appl. Environ. Microbiol.* **48**, 446–448 (1984).
- Saddler, J. N. & Khan, A. W. Cellulase production by *Acetivibrio cellulolyticus*. *Can. J. Microbiol.* **26**, 760–765 (1980).
- Lin, C., Urbance, J. W. & Stahl, D. A. *Acetivibrio cellulolyticus* and *Bacteroides cellulosolvens* are members of the greater clostridial assemblage. *FEMS Microbiol. Lett.* **124**, 151–155 (1994).
- Horino, H., Fujita, T. & Tonouchi, A. Description of *Anaerobacterium chartisolvens* gen. nov., sp. nov., an obligately anaerobic bacterium from *Clostridium* rRNA cluster III isolated from soil of a Japanese rice field, and reclassification of *Bacteroides cellulosolvens* Murray *et al.* 1984 as *Pseudobacteroides cellulosolvens* gen. nov., comb. nov. *Int. J. Syst. Evol. Microbiol.* **64**, 1296–1303 (2014).

23. Zhivin, O. *et al.* Unique organization and unprecedented diversity of the *Bacteroides* (*Pseudobacteroides*) *cellulosolvens* cellulosome system. *Biotechnol. Biofuels* **10**, 211 (2017).
24. Guss, A. M., Olson, D. G., Caiazza, N. C. & Lynd, L. R. Dcm methylation is detrimental to plasmid transformation in *Clostridium thermocellum*. *Biotechnol. Biofuels* **5**, 30 (2012).
25. Olson, D. G., Giannone, R. J., Hettich, R. L. & Lynd, L. R. Role of the CipA scaffolding protein in cellulose solubilization, as determined by targeted gene deletion and complementation in *Clostridium thermocellum*. *J. Bacteriol.* **195**, 733–739 (2013).
26. Morag, E., Bayer, E. A., Hazlewood, G. P., Gilbert, H. J. & Lamed, R. Cellulase S_8 (CelS) is synonymous with the major cellobiohydrolase (subunit S8) from the cellulosome of *Clostridium thermocellum*. *Appl. Biochem. Biotechnol.* **43**, 147–151 (1993).
27. Yoav, S. *et al.* How does cellulosome composition influence deconstruction of lignocellulosic substrates in *Clostridium* (*Ruminiclostridium*) *thermocellum* DSM 1313? *Biotechnol. Biofuels* **10**, 1–16 (2017).
28. McDonough, M. A., Kadirvelraj, R., Harris, P., Poulsen, J. C. N. & Larsen, S. Rhamnogalacturonan lyase reveals a unique three-domain modular structure for polysaccharide lyase family 4. *FEBS Lett.* **565**, 188–194 (2004).
29. Mølgaard, A., Kauppinen, S. & Larsen, S. Rhamnogalacturonan acetyltransferase elucidates the structure and function of a new family of hydrolases. *Structure* **8**, 373–383 (2000).
30. Jayani, R. S., Saxena, S. & Gupta, R. Microbial pectinolytic enzymes: a review. *Process Biochem.* **40**, 2931–2944 (2005).
31. Artzi, L., Bayer, E. A. & Morais, S. Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides. *Nat. Rev. Microbiol.* **15**, 83–95 (2017).
32. Zverlov, V. V., Kellermann, J. & Schwarz, W. H. Functional subgenomics of *Clostridium thermocellum* cellulosomal genes: identification of the major catalytic components in the extracellular complex and detection of three new enzymes. *Proteomics* **5**, 3646–3653 (2005).
33. Lamed, R., Setter, E., Kenig, R. & Bayer, E. A. The cellulosome: a discrete cell surface organelle of *Clostridium thermocellum* which exhibits separate antigenic, cellulose-binding and various cellulolytic activities. *Biotechnol. Bioeng. Symp.* **13**, 163–181 (1983).
34. Bayer, E. A., Kenig, R. & Lamed, R. Adherence of *Clostridium thermocellum* to cellulose. *J. Bacteriol.* **156**, 818–827 (1983).
35. Lamed, R., Setter, E. & Bayer, E. A. Characterization of a cellulose-binding, cellulose-containing complex in *Clostridium thermocellum*. *J. Bacteriol.* **156**, 828–836 (1983).
36. Bahari, L. *et al.* Glycoside hydrolases as components of putative carbohydrate biosensor proteins in *Clostridium thermocellum*. *J. Ind. Microbiol. Biotechnol.* **38**, 825–832 (2011).
37. Yaniv, O. *et al.* Fine-structural variance of family 3 carbohydrate-binding modules as extracellular biomass-sensing components of *Clostridium thermocellum* anti- σ^1 factors. *Acta Crystallogr. D Biol. Cryst.* **70**, 522–534 (2014).
38. Tseng, C. L. & Shaw, G. C. 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.* **190**, 1561–1567 (2008).
39. Mutalik, V. K., Nonaka, G., Ades, S. E., Rhodius, V. A. & Gross, C. A. Promoter strength properties of the complete Sigma E regulon of *Escherichia coli* and *Salmonella enterica*. *J. Bacteriol.* **191**, 7279–7287 (2009).
40. Hook-Barnard, I. G. & Hinton, D. M. Transcription initiation by mix and match elements: flexibility for polymerase binding to bacterial promoters. *Gene Regul. Syst. Biol.* **1**, 275–293 (2007).
41. Guzina, J. & Djordjevic, M. Promoter recognition by extracytoplasmic function σ factors: analyzing DNA and protein interaction motifs. *J. Bacteriol.* **198**, 1927–1938 (2016).
42. Guzina, J. & Djordjevic, M. Mix-and-matching as a promoter recognition mechanism by ECF σ factors. *BMC Evol. Biol.* **17**, 12 (2017).
43. Sand, A. *et al.* Three cellulosomal xylanase genes in *Clostridium thermocellum* are regulated by both vegetative SigA (σ^A) and alternative SigI6 (σ^{I6}) factors. *FEBS Lett.* **589**, 3133–3140 (2015).
44. Wilson, C. M. *et al.* LacI transcriptional regulatory networks in *Clostridium thermocellum* DSM1313. *Appl. Environ. Microbiol.* **83**, e02751–16 (2017).
45. Zuber, P. & Losick, R. Use of a *lacZ* fusion to study the role of the *spoO* genes of *Bacillus subtilis* in developmental regulation. *Cell* **35**, 275–283 (1983).
46. Hartl, 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.* **183**, 2696–2699 (2001).
47. Radeck, J. *et al.* The Bacillus BioBrick Box: generation and evaluation of essential genetic building blocks for standardized work with *Bacillus subtilis*. *J. Biol. Eng.* **7**, 29 (2013).
48. Cutting, S. & Vander Horn, P. In *Molecular biological methods for Bacillus* (eds Harwood, C. & Cutting, S.) 27–74 (John Wiley and Sons Ltd, 1990).
49. Mrázek, J. & Xie, S. Pattern locator: A new tool for finding local sequence patterns in genomic DNA sequences. *Bioinformatics* **22**, 3099–3100 (2006).
50. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
51. Notredame, C., Higgins, D. G. & Heringa, J. T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217 (2000).
52. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).

Acknowledgements

The authors appreciate Dr. Ilya Borovok (Tel Aviv University) for his helpful advice during the course of this work and Dr. Milana Voronov-Goldman (Tel Aviv University) for technical support. This research was supported by the Israel Science Foundation (24/11 to R.L., 1349/13 to E.A.B. and 177/14 to Y.S.). Additional support was issued to R.L. by The Sidney E. Frank Foundation through the ISF. This research was funded in part by the joint Israel-China ISF-NSFC Research Grant (2566/160 to E.A.B. and Y.F.) – National Natural Science Foundation of China (31661143023 to Y.F.). I.M.-G. is grateful for the award of a Martin Kushner Schnur Post-Doctoral Fellowship at the Weizmann Institute. L.O.O. was supported by the “Consejo Nacional de Ciencia y Tecnología - México” with a PhD scholarship (440354). Y.S. holds the Erwin and Rosl Pollak Chair in Biotechnology at the Technion. E.A.B. is the incumbent of The Maynard I. and Elaine Wishner Chair of Bio-organic Chemistry at the Weizmann Institute of Science.

Author Contributions

L.O.O. and I.M.-G. contributed to conception and design; L.O.O. acquired data; L.O.O., R.L., E.B. and I.M.-G. analyzed and interpreted data; Y.-J.L., J.X., Q.C., Y.F. and Y.S. contributed reagents, materials, analysis and tools; L.O.O., R.L., E.B. and I.M.-G. prepared figures and drafted the article; Y.-J.L., J.X., Q.C., Y.F. and Y.S. critically revised the article; L.O.O., R.L., E.B. and I.M.-G. approved the final version to be published; R.L., E.B. and I.M.-G. directed and supervised the research.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-29245-5>.

Competing Interests: The authors declare no competing interests.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018