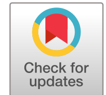
AMERICAN SOCIETY FOR MICROBIOLOGY | mBio®

RESEARCH ARTICLE



A Flexible Binding Site Architecture Provides New Insights into CcpA Global Regulation in Gram-Positive Bacteria

Yunpeng Yang,[a,b] Lu Zhang,[a,b] He Huang,[a] Chen Yang,[a] Sheng Yang,[a,c] Yang Gu,[a,d] Weihong Jiang[a,c]

Key Laboratory of Synthetic Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China[a]; University of Chinese Academy of Sciences, Beijing, China[b]; Jiangsu National Synergetic Innovation Center for Advanced Materials, SICAM, Nanjing, China[c]; Shanghai Collaborative Innovation Center for Biomanufacturing Technology, Shanghai, China[d]

ABSTRACT Catabolite control protein A (CcpA) is the master regulator in Gram-positive bacteria that mediates carbon catabolite repression (CCR) and carbon catabolite activation (CCA), two fundamental regulatory mechanisms that enable competitive advantages in carbon catabolism. It is generally regarded that CcpA exerts its regulatory role by binding to a typical 14- to 16-nucleotide (nt) consensus site that is called a catabolite response element (*cre*) within the target regions. However, here we report a previously unknown noncanonical flexible architecture of the CcpA-binding site in solventogenic clostridia, providing new mechanistic insights into catabolite regulation. This novel CcpA-binding site, named cre_{var}, has a unique architecture that consists of two inverted repeats and an intervening spacer, all of which are variable in nucleotide composition and length, except for a 6-bp core palindromic sequence (TGTAAA/TTTACA). It was found that the length of the intervening spacer of cre_{var} can affect CcpA binding affinity, and moreover, the core palindromic sequence of cre_{var} is the key structure for regulation. Such a variable architecture of cre_{var} shows potential importance for CcpA's diverse and fine regulation. A total of 103 potential cre_{var} sites were discovered in solventogenic *Clostridium acetobutylicum*, of which 42 sites were picked out for electrophoretic mobility shift assays (EMSAs), and 30 sites were confirmed to be bound by CcpA. These 30 cre_{var} sites are associated with 27 genes involved in many important pathways. Also of significance, the cre_{var} sites are found to be widespread and function in a great number of taxonomically different Gram-positive bacteria, including pathogens, suggesting their global role in Gram-positive bacteria.

IMPORTANCE In Gram-positive bacteria, the global regulator CcpA controls a large number of important physiological and metabolic processes. Although a typical consensus CcpA-binding site, *cre*, has been identified, it remains poorly explored for the diversity of CcpA-mediated catabolite regulation. Here, we discovered a novel flexible CcpA-binding site architecture (cre_{var}) that is highly variable in both length and base composition but follows certain principles, providing new insights into how CcpA can differentially recognize a variety of target genes to form a complicated regulatory network. A comprehensive search further revealed the wide distribution of cre_{var} sites in Gram-positive bacteria, indicating it may have a universal function. This finding is the first to characterize such a highly flexible transcription factor-binding site architecture, which would be valuable for deeper understanding of CcpA-mediated global catabolite regulation in bacteria.

Received 31 October 2016 **Accepted** 21 December 2016 **Published** 24 January 2017

Citation Yang Y, Zhang L, Huang H, Yang C, Yang S, Gu Y, Jiang W. 2017. A flexible binding site architecture provides new insights into CcpA global regulation in Gram-positive bacteria. mBio 8:e02004-16. https://doi.org/10.1128/mBio.02004-16.

Editor Sang Yup Lee, Korea Advanced Institute of Science and Technology

Copyright © 2017 Yang et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Yang Gu, ygu02@sibs.ac.cn, or Weihong Jiang, whjiang@sibs.ac.cn.

Carbon catabolite repression (CCR) and carbon catabolite activation (CCA) are two of the most fundamental regulatory mechanisms in microbes (1, 2), enabling them to adapt quickly to environmental changes. In Gram-positive bacteria, the master regu-

lator mediating CCR and CCA is catabolite control protein A (CcpA), a protein of the LacI-GalR family (3). CcpA is a pleiotropic regulator involved in many important cellular processes, including bacterial pathogenicity (4–6).

It is known that CcpA executes its regulation via binding to a so-called catabolite-responsive element (*cre*) within the promoter or protein-coding regions of the target genes (7). The consensus sequence of *cre* has been determined to be TGWAANCGNT NWCA in *Bacillus subtilis*, a model organism of Gram-positive bacteria, in which N represents any base and W represents A or T (8). Additional *cre*s identified later in *B. subtilis* also closely match this consensus sequence (7, 9). However, it has recently been found that, in some cases, CcpA employed two different binding motifs (one is a typical *cre* and the other one is an atypical *cre*) to regulate the central carbon metabolism (10). This indicates that the general understanding of CcpA activity is superficial and the mechanism by which CcpA exerts its regulation is more sophisticated than we know.

Solventogenic clostridia are of great interest because they are able to produce a series of bulk chemicals (11), among which *n*-butanol and ethanol are important liquid fuels. Our research group previously revealed a core *cre* consensus sequence, WTGW AAACGWTWWCAW (where W represents A or T) that is responsible for CcpA binding in *Clostridium acetobutylicum*, a typical species of solventogenic clostridia (12). This sequence is highly similar to that of *B. subtilis*, but meanwhile, a large number of genes that exhibited greatly altered transcription after *ccpA* inactivation do not contain the *cre* sites in their promoter or protein-coding region (12), thus raising the question as to whether there exist atypical binding sites recognized by CcpA.

Here, we identified a novel flexible architecture of binding sites recognized by CcpA. This binding motif was then shown to be widespread in Gram-positive bacteria, indicating its importance in CcpA-mediated regulation. Based on these results, we identify a large number of new target genes controlled by CcpA and thereby chart a more complete CcpA regulatory network in *C. acetobutylicum*.

RESULTS

Discovery of novel transcriptional binding sites for CcpA regulation. Based on our previously performed comparative transcriptomic analysis (12), we found that, among a total of 1,394 genes that showed greatly altered transcription after *ccpA* inactivation, only 154 genes contain the typical *cre* (WTGWAAACGWTWWCAW [W represents A or T]) sites within their promoter or coding region. Thus, the presence of noncanonical CcpA-binding sites within these genes is strongly suggested. To explore this possibility, we chose *sol* (CAP0162-0164), a key operon responsible for acid reassimilation and solvent formation in *C. acetobutylicum* (13), containing no typical *cre* sites but showing high binding affinity to CcpA (12), for a detailed examination. In the first step, a 663-bp promoter region (−663 to −1 bp relative to the translational start point) of *sol* was divided into three fragments (213, 350, and 100 bp) for electrophoretic mobility shift assays (EMSAs) (Fig. 1A). A strongly shifted band was observed for the 350-bp fragment (Fig. 1B), suggesting the existence of binding sites within this region. Next, this 350-bp fragment was further divided into three 170-bp segments (P_{sol}-170-1, P_{sol}-170-2, and P_{sol}-170-3), with 80 bp overlapping one by one, for EMSAs (Fig. 1C). Interestingly, a DNA bind shift was observed for all three segments (Fig. 1D), implying that there may exist more than one CcpA-binding site within the promoter region of *sol*.

To confirm this hypothesis, P_{sol}-170-1, which had the strongest shifted signal among these three segments, was gradually truncated, and the resulting three truncated fragments, namely, P_{sol}-170-1 minus 20, 40, and 60 bp, respectively, were examined (Fig. 2A). The results showed that the affinity of P_{sol}-170-1 for CcpA was almost completely abolished with a 40- or 60-bp deletion (Fig. 2B), suggesting a binding site overlapping or within the deleted region. Encouragingly, visual scanning of this 60-bp region identified a 41-nucleotide (nt) palindromic sequence (AAACTGCTAAATGTAAA-TTATACG-TTTACATTTAGCAGTTT) comprising two 17-nt inverted repeats separated by

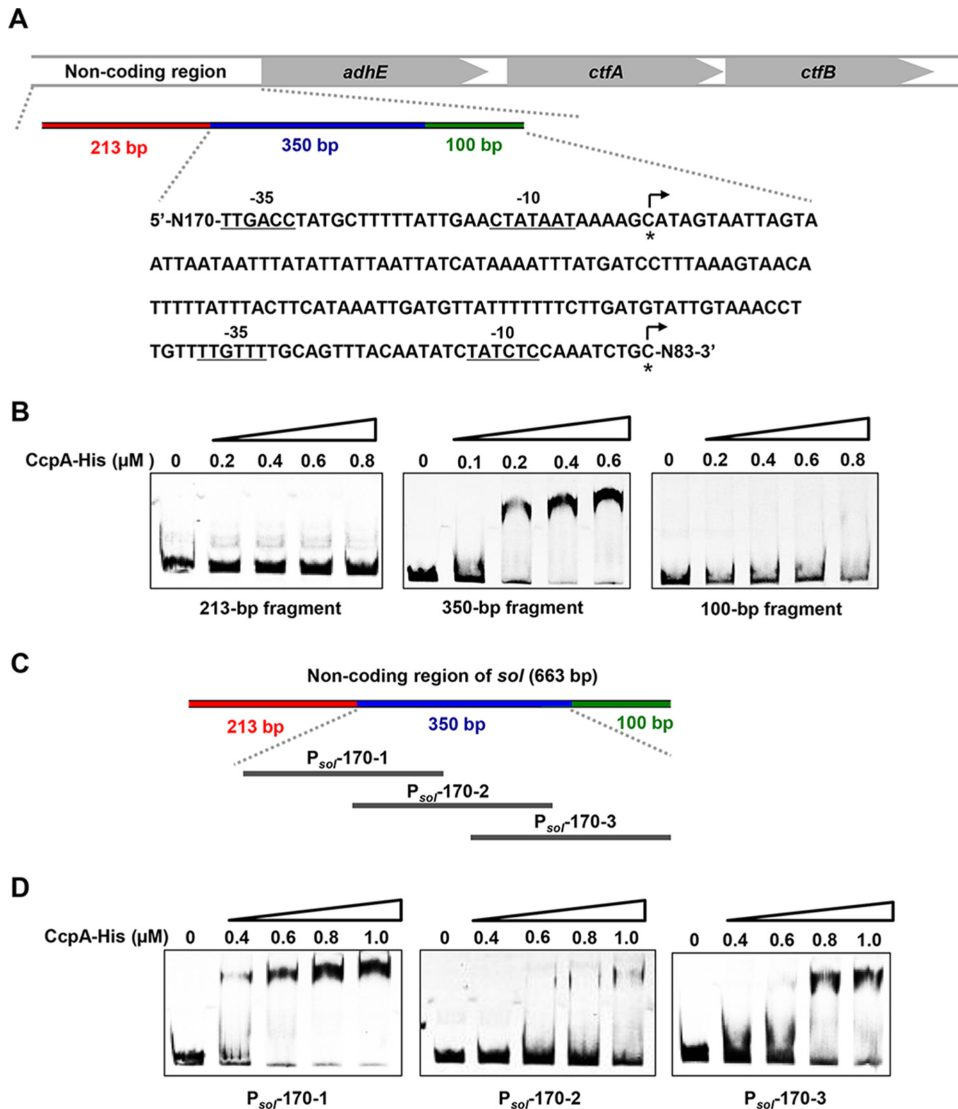


FIG 1 Functional analysis of the CcpA-binding sites in the noncoding region of the *sol* operon. (A) The noncoding region of the *sol* operon was divided into three fragments (213, 350, and 100 bp). The −10 region and −35 region are underlined. The two transcription start sites (13) are indicated by bent arrows. (B) EMSAs of His_6-CcpA binding to the 213-, 350-, and 100-bp fragments labeled with Cy5. (C) The 350-bp fragment of *sol* was further divided into three 170-bp fragments (P_{sol}-170-1, P_{sol}-170-2, and P_{sol}-170-3). (D) EMSAs of His_6-CcpA binding to P_{sol}-170-1, P_{sol}-170-2 and P_{sol}-170-3.

7 nt (Fig. 2C). According to the characteristics of this palindromic sequence, we further found two other similar palindromic sequences within the 350-bp fragment, which harbor 6- and 9-nt inverted repeats separated by 8 and 18 nt, respectively (Fig. 2D). A common feature of these three palindromic sequences (designated *sol*-41, *sol*-20, and *sol*-36, respectively) is the two repeats that contain the core palindromic sequence TGTAAA/TTTACA, as well as the intervening spacer region; the difference is the variable length of the two repeats and the intervening spacer region. Thus, this yielded the architecture N_xTGTAAA-Y_x-TTTACAM_x (where Y represents any base, N and M also represent any base but are complementary to each other, and *x* represents the base number) (Fig. 2E). Compared with the known *cre* consensus in bacteria such as *Bacillus* (14), *Lactobacillus* (15), and *Staphylococcus* species (16), this binding site architecture is quite distinct, which is an inverted TA-rich sequence separated by a variable (length and nucleotide) spacer region. This architecture is different from all known CcpA-binding *cre* motifs, which are normally 14 or 16 bp in length, including several highly

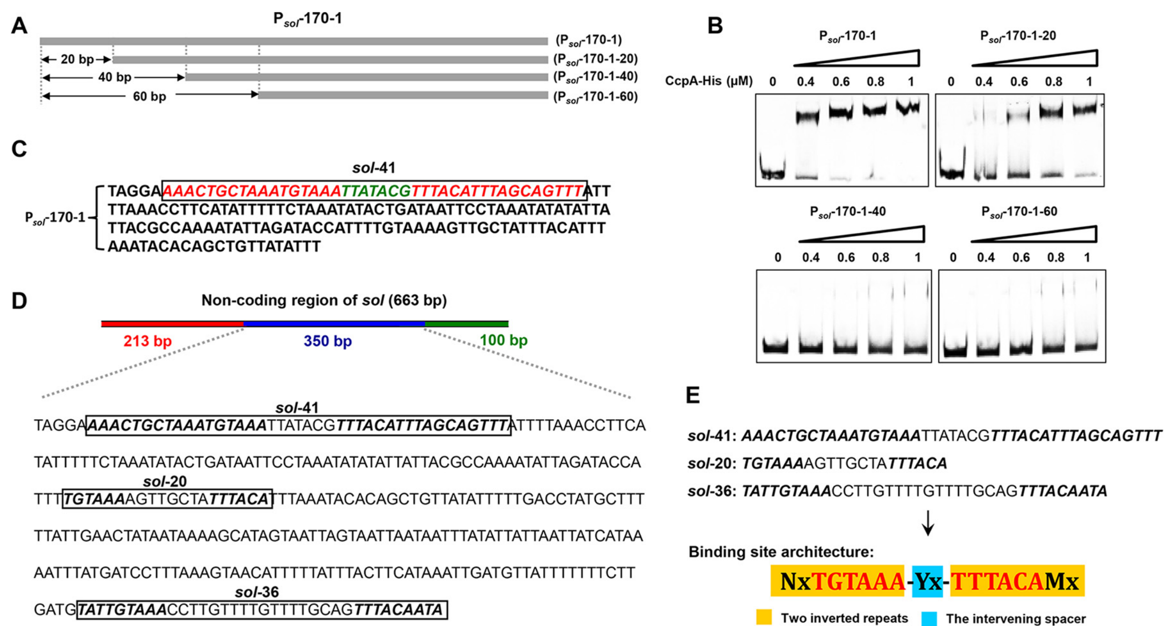


FIG 2 Identification of novel CcpA-binding sites in *C. acetobutylicum*. (A) Truncation of P_{sol}-170-1. Sequences 20, 40, and 60 bp in length were cut off from P_{sol}-170-1, yielding three truncated fragments (P_{sol}-170-1-20, P_{sol}-170-1-40, and P_{sol}-170-1-60). (B) EMSAs of His_6-CcpA binding to P_{sol}-170-1 and the three truncated fragments. His_6-CcpA concentrations of 0 to 1.0 μM were used. (C) Mining of a potential CcpA-binding site in P_{sol}-170-1 via visual analysis. The putative CcpA-binding site is represented by a box. The two inverted repeats and the intervening spacer of the bind site are colored red and green. (D) Three putative CcpA-binding sites (*sol*-41, *sol*-20, and *sol*-36) found in the 350-bp noncoding region of *sol*. These three sites share two conserved inverted repeats (TGTAAA/TTTACA), which are highlighted in boldface and italics. (E) A common sequence (N_xTGTAAA-Y_x-TTTACAM_x) was derived from these three CcpA-binding sites. N and M represent any bases complementary to each other, Y represents any base, and *x* represents the number of the bases included in the two inverted repeats and intervening spacer, which was variable.

conserved nucleotides (7, 17–20). We named this novel flexible CcpA binding site architecture "*cre*$_{var}$".

Wide occurrence of the novel binding site *cre*$_{var}$ in *C. acetobutylicum*. To explore the distribution of the *cre*$_{var}$ sites in *C. acetobutylicum*, we performed a genome-wide scan using the RegPredict web server (21), in which the two 6-nt repeats of *cre*$_{var}$ were fixed but the length of the intervening region was variable, covering 0 to 40 nt. The search result revealed 103 potential *cre*$_{var}$ sites that belong to 99 genes (see Table S1 in the supplemental material). Next, 42 (the associated genes showed $\geq$2-fold transcriptional changes after *ccpA* inactivation) (12) of these 103 potential *cre*$_{var}$ sites were picked out for EMSAs to examine the quality of the prediction result. Finally, 30 *cre*$_{var}$ sites were confirmed to be bound by CcpA, including the above-mentioned three *cre*$_{var}$ sites in the upstream region of *sol* (see Fig. 4C) and the other 27 *cre*$_{var}$ sites that are associated with 26 genes (see Fig. S1 in the supplemental material). Among these 30 *cre*$_{var}$ sites, 20 sites are located in promoter regions, whereas 10 sites are inside protein-coding regions (see Table S2 in the supplemental material). The majority of these *cre*$_{var}$-associated genes can be grouped into certain functional subsets (Fig. 3). Several genes are involved in important bioprocesses, including substance transport and metabolism, redox balancing, sporulation, and solvent production. Besides, by comparing the distribution of *cre* and *cre*$_{var}$ in the genome, *cre*$_{var}$ sites were found to be predominantly associated with genes belonging to certain function categories, i.e., sporulation, solvent production, and purine and pyrimidine metabolism (Fig. 3), which suggests that *cre*$_{var}$ may play more important roles in CcpA regulation of these genes, thus enabling a more comprehensive regulatory network of CcpA.

Characterization of the binding motif *cre*$_{var}$. Next, we attempted to assess the importance of the signature sequences of *cre*$_{var}$ for CcpA binding. The two inverted repeats and the intervening spacers of the three *cre*$_{var}$ sites (*sol*-41, *sol*-20, and *sol*-36) mentioned above were mutated (Fig. 4A, B, and D), and then the binding activities of

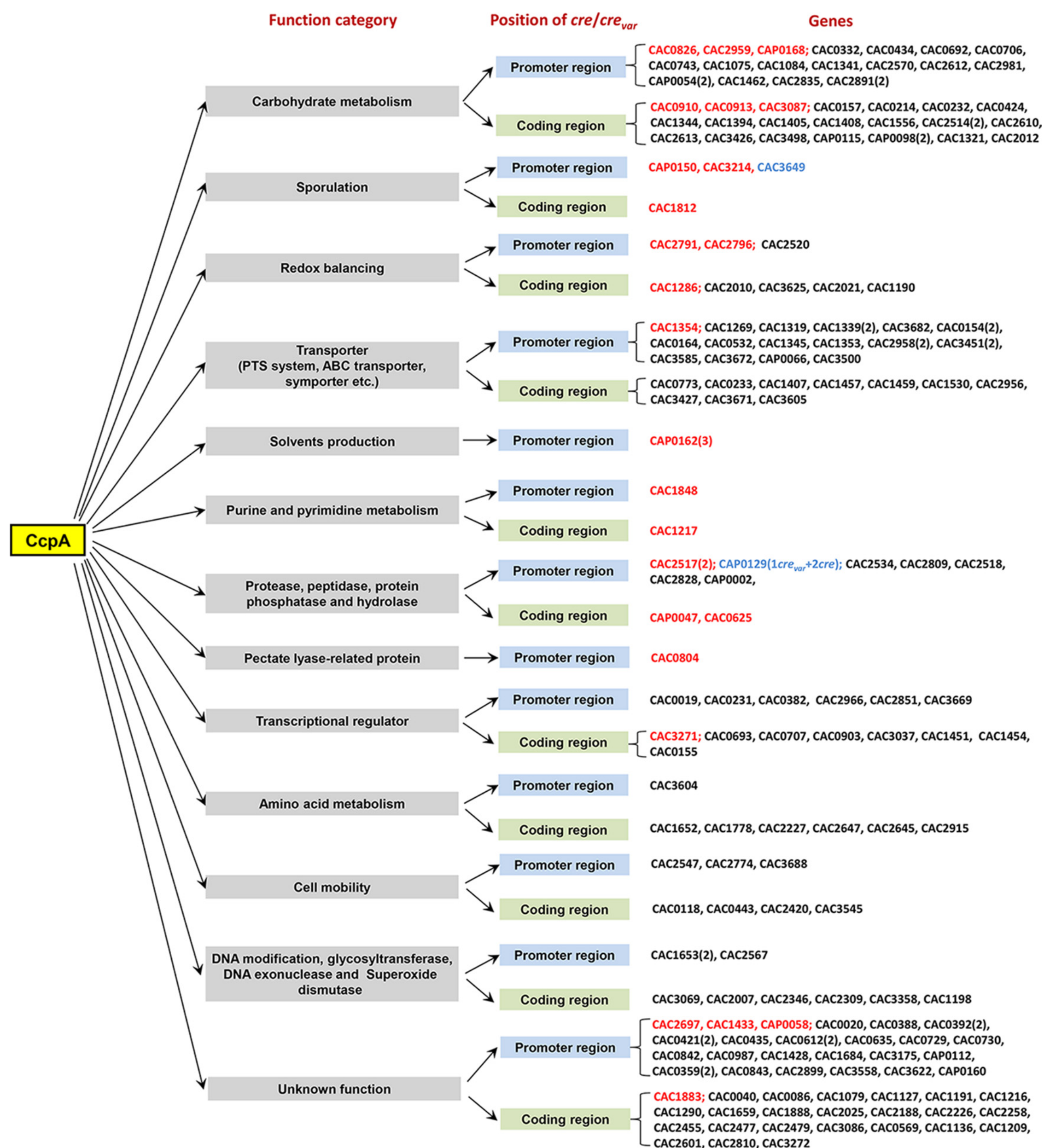


FIG 3 Overview of the genes affected by CcpA. The genes that have been confirmed to be directly regulated by CcpA (from EMSA data in Fig. S1) via *cre*$_{var}$ sites are shown in red; the *cre*-associated genes that were predicted to be significantly affected by CcpA (>2-fold transcriptional change after *ccpA* inactivation) according to our previous work (12) are shown in black; and the genes associated with both *cre* and *cre*$_{var}$ and also showing significant transcriptional changes (>2-fold) after *ccpA* inactivation are shown in blue. The genes associated with multiple *cre* or *cre*$_{var}$ sites are annotated with the number of sites given in parentheses. CAC, genes located in chromosome of *C. acetobutylicum*; CAP, genes located in the megaplasmid of *C. acetobutylicum*.

CcpA with the three mutated 120-bp sequences were determined. The results showed that the mutations at two repeats completely abolished the binding of CcpA to *sol*-41, whereas a light binding to *sol*-20 and *sol*-36 was maintained (Fig. 4C). In contrast, mutations in the intervening region weakened, to different extent, the binding affinities of CcpA to *sol*-41, *sol*-20, and *sol*-36 (Fig. 4E). These findings suggest that both the two inverted repeats and intervening spacer are crucial for CcpA-*cre*$_{var}$ binding.

In addition to *in vitro* experiments, we also examined CcpA binding to *cre*$_{var}$ sites *in vivo* by using a reporter gene. As shown in Fig. S2 in the supplemental material, the LacZ activity assay revealed that single mutation of either the *sol*-41 or *sol*-20 site and mutation of both the *sol*-41 and *sol*-20 sites resulted in significantly decreased strength

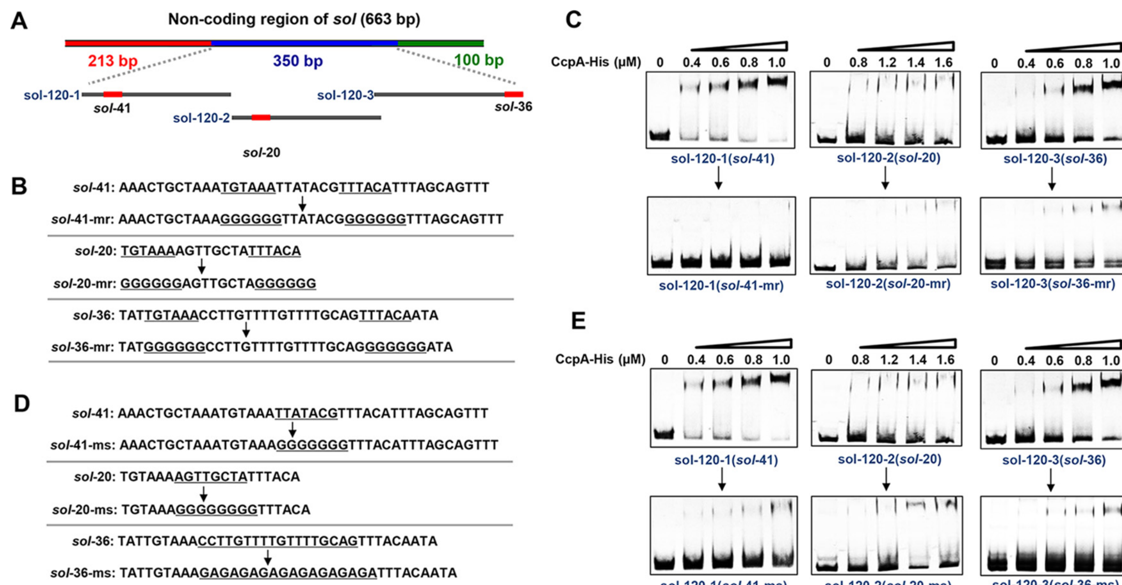


FIG 4 Mutational analysis of the cre_{var} sites in the noncoding region of *sol*. (A) The 350-bp sequence containing the three putative cre_{var} sites was divided into three 120-bp fragments (sol-120-1, sol-120-2, and sol-120-3), each of which harbored one cre_{var} site (*sol*-41, *sol*-20, and *sol*-36), which is represented by a red box. (B) Mutation of the two inverted repeats of the three cre_{var} sites. The two 6-nt repeats of *sol*-41, *sol*-20, and *sol*-36 were all mutated to GGGGGG, yielding *sol*-41-mr, *sol*-20-mr, and *sol*-36-mr, respectively. (C) EMSAs of His_6-CcpA binding to the three 120-bp fragments as well as their derived fragments mutated in the two repeats. (D) Mutation of the intervening spacer of each of the three cre_{var} sites (*sol*-41-ms, *sol*-20-ms, and *sol*-36-ms). (E) EMSAs of His_6-CcpA binding to the three 120-bp fragments as well as their derived fragments mutated in the intervening spacers.

of promoter P_{sol} in the wild-type strain, whereas no significant difference was observed in the 824*ccpA* strain (in which *ccpA* was disrupted). This further confirmed that *sol*-41 and *sol*-20 are the CcpA-binding sites.

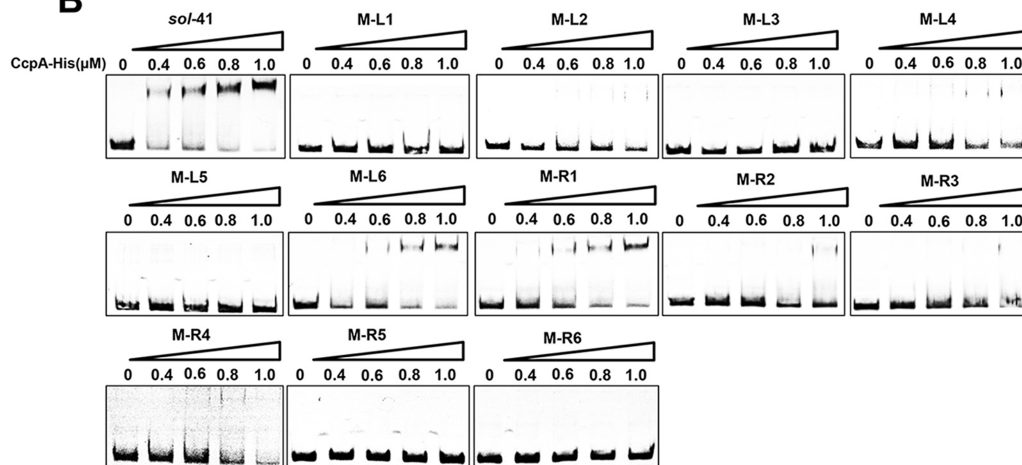
Specific to the two inverted 6-nt repeats, since they were important for CcpA-cre_{var} binding, we attempted to examine whether each nucleotide is essential. Thus, each nucleotide in the two 6-nt repeats of *sol*-41 was separately mutated, yielding 12 derivative probes for EMSA analysis (Fig. 5A). The EMSA results showed that single mutation of each one of the outer five nucleotides (L1, L2, L3, L4, and L5 or R2, R3, R4, R5, and R6) thoroughly eliminated CcpA–*sol*-41 binding, while mutation of L6 or R1 still retained a slight binding (Fig. 5B). Next, the *in vivo* experiments using a *lacZ* reporter were performed to see the strength variations between P_{sol} and its 12 derivatives. While no significant difference was observed in 824*ccpA* (the control with *ccpA* disruption), all 12 single mutations resulted in greatly decreased LacZ activity in the wild-type strain (Fig. 5C), which are consistent with the *in vitro* EMSA results. These findings demonstrate that each nucleotide in the two 6-nt arms of cre_{var} is important for CcpA binding.

The variable intervening region length of cre_{var} sites affects CcpA binding affinity. Since the intervening spacer of cre_{var} is variable, the question arose as to whether the spacer length affected CcpA-DNA binding. We explored this possibility by using promoter $P_{cac0804\text{-}15(wt)}$, which contains a 27-nt cre_{var} site with a 15-nt intervening spacer (Fig. 6A). Here, the major reason for choosing $P_{cac0804\text{-}15(wt)}$ for investigation is that, among the 15 genes identified to contain a sole cre_{var} site in their promoter regions (Fig. 3), the *cac0804* gene was the only one showing steady and significant upregulation (over 2-fold) in transcriptional level after *ccpA* overexpression (data not shown), indicating a high CcpA binding affinity to the cre_{var} site in $P_{cac0804\text{-}15(wt)}$. When this 15-nt spacer was truncated to 10 nt, a significantly altered CcpA-DNA binding affinity occurred (Fig. 6B and C); further truncated to 6 nt, no binding affinity changes were found between $P_{cac0804\text{-}6}$ and $P_{cac0804\text{-}15(wt)}$ (Fig. 6C). In contrast, for the *ccpA*-inactivated strain (used here as a control), no significant differences in LacZ expression were observed after truncation of the spacer (Fig. 6C). These results suggest that the intervening spacer length within cre_{var} sites can influence CcpA-cre_{var} binding affinity.

A

	L1	L2	L3	L4	L5	L6		R1	R2	R3	R4	R5	R6
sol-41	T	G	T	A	A	A	TTATACG	T	T	T	A	C	A
M-L1	G	G	T	A	A	A	TTATACG	T	T	T	A	C	A
M-L2	T	A	T	A	A	A	TTATACG	T	T	T	A	C	A
M-L3	T	G	G	A	A	A	TTATACG	T	T	T	A	C	A
M-L4	T	G	T	G	A	A	TTATACG	T	T	T	A	C	A
M-L5	T	G	T	A	G	A	TTATACG	T	T	T	A	C	A
M-L6	T	G	T	A	A	G	TTATACG	T	T	T	A	C	A
M-R1	T	G	T	A	A	A	TTATACG	G	T	T	A	C	A
M-R2	T	G	T	A	A	A	TTATACG	T	G	T	A	C	A
M-R3	T	G	T	A	A	A	TTATACG	T	T	G	A	C	A
M-R4	T	G	T	A	A	A	TTATACG	T	T	T	G	C	A
M-R5	T	G	T	A	A	A	TTATACG	T	T	T	A	A	A
M-R6	T	G	T	A	A	A	TTATACG	T	T	T	A	C	G

B



C

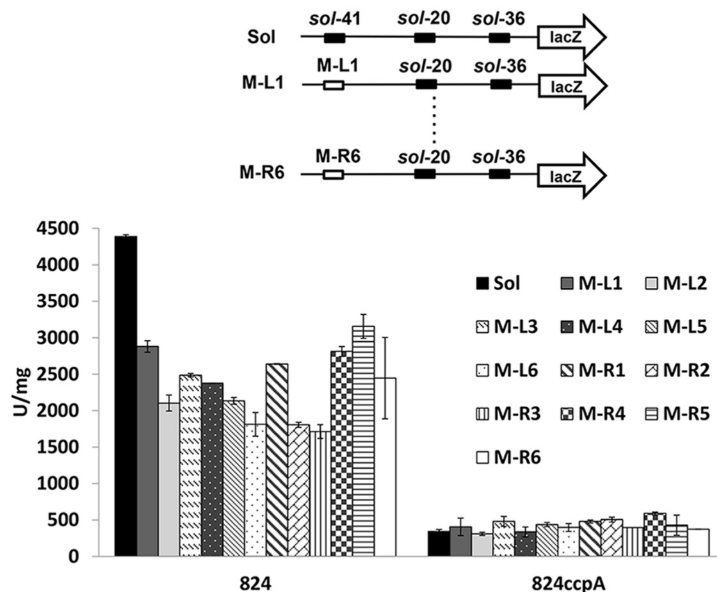


FIG 5 Characterization of the novel CcpA-binding sites in *C. acetobutylicum*. (A) Single point mutation of the inverted repeats of *sol*-41. The mutation site is marked in red. (B) EMSAs of His_6-CcpA binding to sol-120-1 containing *sol*-41 and its mutated derivatives. Concentrations of 0 to 1.0 μM of His_6-CcpA were used. (C) *In vivo* assay of P_{sol} and its derivatives in both the *C. acetobutylicum* wild-type and *ccpA*-inactivated strain. The data represent the average from two independent samples.

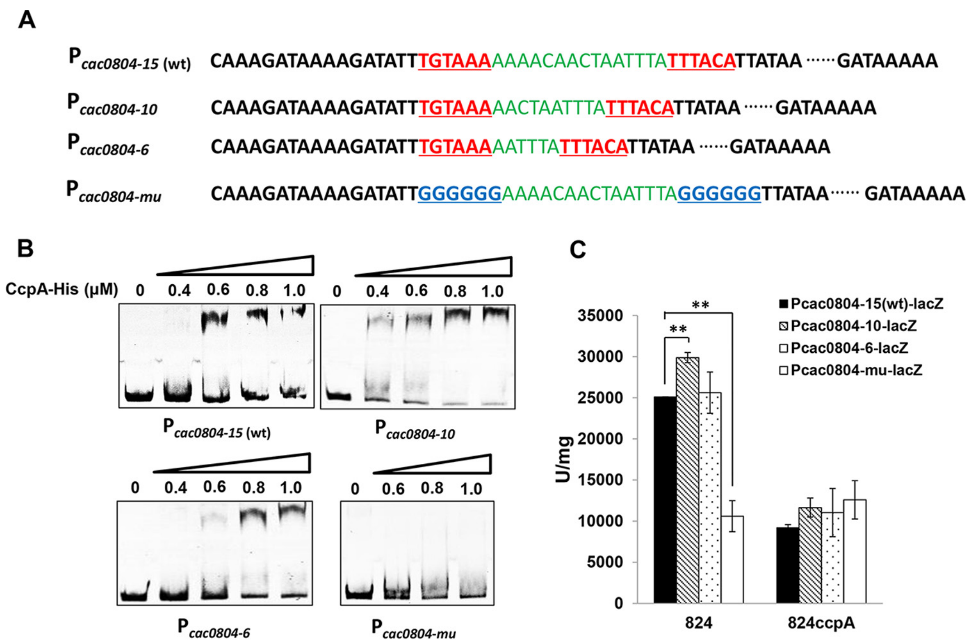


FIG 6 Influence of the intervening spacer length within *cre*$_{var}$ on CcpA regulation. (A) Construction of artificial promoters harboring the *cre*$_{var}$ sites with different spacer lengths. The two inverted repeats and the intervening spacer are highlighted with red and green, respectively. P$_{cac0804-15(wt)}$ is the natural promoter sequence of the *cac0804* gene. P$_{cac0804-mu}$ was constructed as a negative control by mutating the two inverted repeats of P$_{cac0804-15(wt)}$. (B) EMSAs of His$_6$-CcpA binding to P$_{cac0804-15(wt)}$, P$_{cac0804-10}$, P$_{cac0804-6}$, and P$_{cac0804-mu}$. (C) *In vivo* assay of CcpA regulation of P$_{cac0804-15(wt)}$ and its derivatives in both the *C. acetobutylicum* wild-type strain (824) and *ccpA*-inactivated strain (824*ccpA*). Means and standard deviations were from two independent biological replicates (**, $P \leq 0.01$, Student's *t* test).

The *cre*$_{var}$ sites are widely distributed in Gram-positive bacteria. Because the *cre*$_{var}$ sites occurred frequently in the *C. acetobutylicum* genome, we are curious whether this *cis* element is also present in other bacteria. To this end, we performed genome-wide searches in the classes *Clostridia* and *Bacilli*, two large groups in Gram-positive bacteria. Surprisingly, the *cre*$_{var}$ sites were found in the genome of several members of these two classes, including pathogens, and were especially abundant in *Clostridium* and *Bacillus* species, in which over 100 *cre*$_{var}$ sites were predicted to be present in *Clostridium acetobutylicum*, *Clostridium cellulolyticum*, *Clostridium difficile*, and *Bacillus cereus* (see Table S3 in the supplemental material). Importantly, like those identified in *C. acetobutylicum*, the *cre*$_{var}$ sites present in these species also exhibited high diversity in the two inverted repeats and intervening spacer regions (Table S3). To our knowledge, only very few proteins have been found capable of recognizing DNA sequence separated by a variable spacer (22–25); however, *cre*$_{var}$-like binding motifs that contain such a highly flexible spacer region have not been reported.

We chose five potential *cre*$_{var}$ sites from both *B. subtilis* (BSU10020, BSU14580, BSU22720, BSU27620, and BSU35080) and *C. perfringens* (CPF0042, CPF0484, CPF0526, CPF0580, and CPF1663) for EMSA verification. *B. subtilis* CcpA and *C. perfringens* CcpA were purified and used for functional analysis of the *cre*$_{var}$ sites in *B. subtilis* and *C. perfringens*, respectively. Encouragingly, among these candidates, a substantial DNA band shift was observed for six *cre*$_{var}$ sites (BSU22720, BSU27620, BSU35080, CPF0526, CPF0580, and CPF1663) (Fig. 7), indicating a high reliability of the predicted *cre*$_{var}$ sites in the classes *Clostridia* and *Bacilli*.

DISCUSSION

As an important regulator in Gram-positive bacteria, CcpA has remained little understood with respect to its pleiotropic regulatory function. This study has expanded CcpA's target genes to a broader range in clostridia as well as some other Gram-positive bacteria, thereby providing new insights into CcpA regulation. Importantly, identifica-

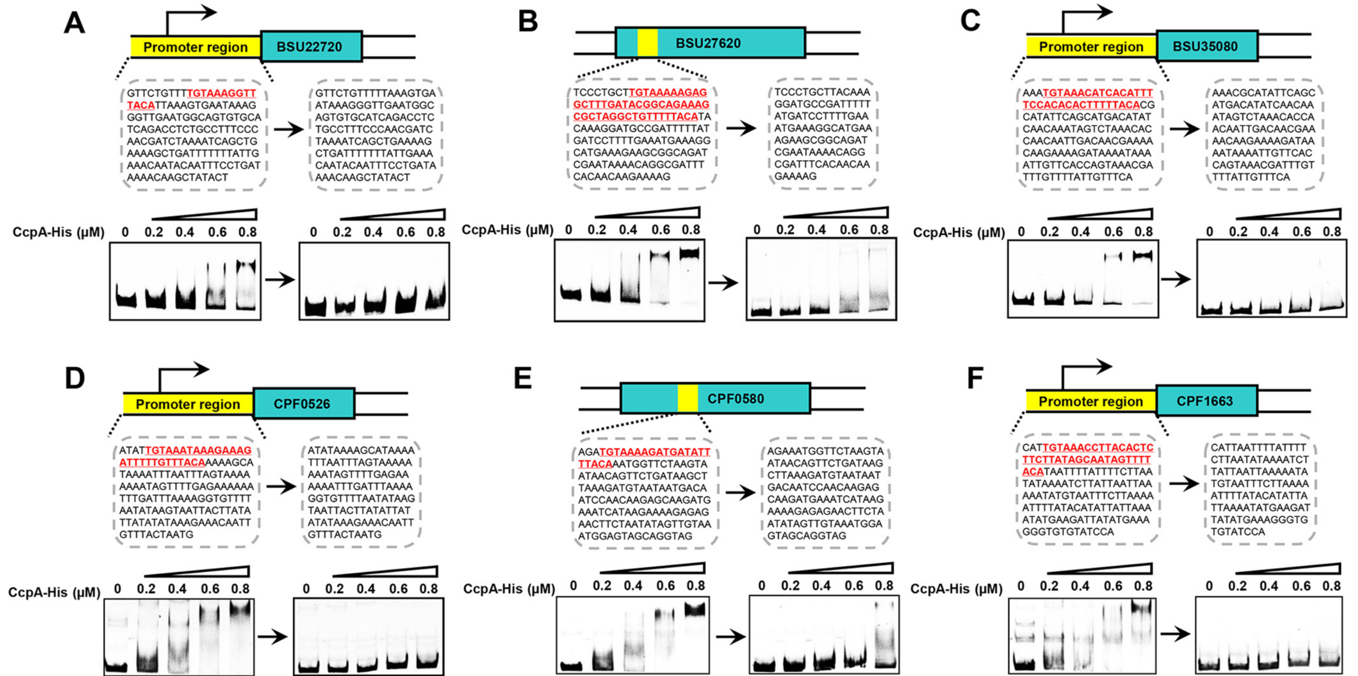


FIG 7 EMSA verification of CcpA binding to the putative *cre*$_{var}$ sites in *B. subtilis* and *C. perfringens*. The sequences highlighted with red indicate the putative *cre*$_{var}$ sites. *B. subtilis* CcpA and *C. perfringens* CcpA were purified and used for functional analysis of the *cre*$_{var}$ sites in *B. subtilis* and *C. perfringens*, respectively.

tion of the novel *cre*$_{var}$ sites revealed a flexible binding site architecture used by CcpA to regulate its target genes. The variation in both the intervening spacer region and two inverted repeats of this *cre*$_{var}$ motif, as well as its widespread occurrence in Gram-positive bacteria, suggests a more complex CcpA regulation than was previously understood.

To date, only very few proteins have been found capable of recognizing repeats separated by a variable spacer. As an example, the *Escherichia coli* cyclic AMP (cAMP) receptor protein (CRP)-binding sites contain a 6- or 8-bp spacer (22); additionally, the *E. coli* CytR repressor, with the assistance of the CRP, can recognize two inverted repeats separated by 10 to 13 bp or direct repeats separated by 1 bp (23). A latest example is the *E. coli* transcription factor HipB, which can recognize palindromic sequences with variable intervening spacer regions (24); moreover, the crystal structures of the HipB-HipA-*hipBA* promoter complex showed that HipBA binding to DNA with a long spacer can be achieved by DNA extrusion (25).

Here, the variation in the *cre*$_{var}$ sites suggests diversity in the CcpA-binding sites for both repressed and activated target genes. The most distinct feature of *cre*$_{var}$ is its intervening spacer region, which is flexible in both length (0- to 40-nt span) and base composition. This feature makes *cre*$_{var}$ quite different from all known CcpA-binding *cre* motifs, which are normally 14 or 16 bp in length, including several highly conserved nucleotides (7, 17–20). For typical *cre* sites, the base variations may cause them to display different bend angles during CcpA binding; CcpA is also able to adjust its conformation to meet the changes in target DNA (7). However, such changes in binding angle appeared insufficient to affect the affinity of the DNA for CcpA (7). In contrast, for the atypical *cre*$_{var}$, the spacer within its motif is variable in both length and base composition, which may cause greater changes in CcpA conformation during its binding to the targets.

Given the wide variation of the *cre*$_{var}$ sites as well as the coexistence of *cre*$_{var}$ and *cre*, we propose that this variability may be an effective mechanism for the diverse regulation of CcpA in Gram-positive bacteria. First, the variable spacer might affect the binding affinity of CcpA for its targets, which would enable CcpA to produce diverse regulatory outputs. For example, the regulation of *E. coli* CytR, a regulator belonging to

the LacI family, was affected by artificially altering the half-site spacing in its binding sites, and the maximum changes in CytR regulation occurred in the short spacing variants (26). Second, the coexistence of *cre*$_{var}$ and *cre* suggests a complementary or independent role of *cre*$_{var}$ relative to *cre* in CcpA regulation, which would confer more choices to CcpA during its regulation. At least in some cases, CcpA may require more than one binding site to exert sophisticated gene regulation. For example, it has been found that CcpA employed two different binding motifs (WWGAAARCGYTTTCWW and TTTTYHWDHHWWTTTY) to regulate the central carbon metabolism in *Streptococcus suis* (10); besides, *cre*$_{var}$ sites were predominantly found to be related to genes of certain function categories (Fig. 3), indicating a more important role of *cre*$_{var}$, rather than *cre*, in CcpA regulation of these genes.

Another interesting finding here is that the widespread occurrence of the *cre*$_{var}$ sites in Gram-positive bacteria, especially classes *Clostridia* and *Bacilli*. Using *B. subtilis* as an example, many essential genes involved in core metabolism were shown to be controlled by CcpA via *cre*$_{var}$ (Table S3), although the predicted *cre*$_{var}$ sites appeared to be fewer than the typical *cre* sites, which were previously estimated to number over 100 in *B. subtilis* (9, 27). The validation experiments in *B. subtilis* also supported this finding (Fig. 7). More importantly, the *cre*$_{var}$ sites were found to be associated with several essential genes in the pathogenic bacteria, such as genes responsible for the phosphotransferase (PTS) system, cell motility and division, DNA replication and mismatch repair, and sporulation (Table S3). Of note, the *cre*$_{var}$ sites were also present in the promoter or protein-coding regions of certain toxin or virulence genes, e.g., the *texT* gene in *Clostridium tetani* (28) and a possible virulence factor gene (SE0184) in *Staphylococcus epidermidis* (29) (Table S3). All of these findings further suggest the potential importance of *cre*$_{var}$ as a *cis* element.

It should be noted for the two 6-nt-sequence core region that although this sequence appears to be conserved in the *cre*$_{var}$ motif, we cannot exclude the possibility that the sequence is also changeable to a certain extent without impacting CcpA recognition of the targets. In this study, we used the common sequence (TGTAAA-Y_x-TTTACA) that was extracted from the three binding sites upstream of the *sol* genes as a template, in which the two inverted repeats were fixed. Apparently, using such a template to search for more CcpA-binding sites has limited the 6-nt core region of the repeats; thus, the yielded binding sites do not reflect all the potential variations in this region. To determine the occurrence frequency of each base at each location of this 6-bp inverted repeat, the strategy such as chromatin immunoprecipitation followed by high-throughput sequencing (chromatin immunoprecipitation sequencing [ChIP-seq]) should be useful. This study is under way.

In summary, we have discovered an unrealized highly flexible architecture of CcpA-binding sites. The motif *cre*$_{var}$, which is variable in both the two repeats and the intervening spacer region, provides new insight into the structure of CcpA recognition sites in Gram-positive bacteria. Such a variation of *cre*$_{var}$ may provide an effective means to CcpA for fine-tuning the regulatory network. Given the wide distribution of the *cre*$_{var}$ in Gram-positive bacteria, it is conceivable that this flexible motif plays an important role in CcpA-mediated regulation of cellular properties.

MATERIALS AND METHODS

Strains and plasmid construction. The strains and plasmids used in this work are listed in Table S4 in the supplemental material. To express the CcpA protein of *C. acetobutylicum*, *ccpA* (CAC3037) was PCR amplified and cloned into pET-28a (Novagen, Madison, WI), yielding the plasmid pET-28a-ccpAcac. Similarly, pET-28a-ccpAbsu and pET-28a-ccpAcpf were constructed to express CcpA from *Bacillus subtilis* and *Clostridium perfringens*. pET-28a-HPrK and pGEX4T1-HPr were used for HPr kinase (HPrK) and HPr expression (30). P_{sol} and $P_{cac0804}$ and their derivatives were PCR amplified and cloned into pIMP1-lacZ (31) for β-galactosidase assays.

Media and cultivation conditions. *Escherichia coli* was grown in Luria-Bertani (LB) medium at 37°C with the addition of chloramphenicol (Chloromycetin [25 μg/ml]), kanamycin (50 μg/ml), ampicillin (100 μg/ml), and spectinomycin (50 μg/ml) when needed. *C. acetobutylicum* was cultivated in CGM medium (32) and P2 medium (33). Thiamphenicol (8 μg/ml) were added to the P2 medium when needed.

Identification of CcpA-binding sites. The RegPredict web server (21) was used to search all potential *cre*$_{var}$ sites on the genome of *C. acetobutylicum* based on the architecture of the template

TGTAAA-Y_x-TTTACA (Yx ranged from 0 to 40 nt). The search regions cover nucleotide positions from −500 to +2000 relative to the translational start sites of all the genes. The results were further artificially analyzed to eliminate the redundant data.

Protein overexpression and purification. The His_6-tagged CcpA and HPrK and glutathione *S*-transferase (GST)-tagged HPr were expressed and purified as described previously (12, 30). The purified proteins were checked by SDS-PAGE.

EMSAs. The DNA probes used in EMSAs were generated as follows. First, the unlabeled DNA fragments were amplified from the genome using specific primer pairs containing a universal sequence (5′ AGCCAGTGGCGATAAG 3′) at the 5′ terminal. Second, the DNA fragments were Cy5 labeled by PCR using the universal primer 5′ AGCCAGTGGCGATAAG 3′, with Cy5 labeled at the 5′ end. Finally, the resulting Cy5-labeled probes were recovered by agarose gel electrophoresis.

The EMSAs with *C. acetobutylicum* CcpA were performed as described previously (30). The EMSAs with *B. subtilis* CcpA and *C. perfringens* CcpA were performed similarly, except that the phosphorylated Hpr was not used.

***β*-Galactosidase assays.** The *C. acetobutylicum* strains harboring the plasmids pIMP1-P_{sol}-*lacZ*, pIMP1-$P_{cac0804}$-*lacZ*, and their derivatives (listed in Table S4) were grown in P2 medium containing 60 g/liter glucose as the sole carbon source. $CaCO_3$ was added at 0.5% (wt/vol) to the medium to control pH. The cell pellets were harvested by centrifugation (5,000 × *g*, 4°C, 10 min), dissolved in B-PER reagent (Thermo Scientific Pierce), and vortexed for 1 min for cell lysis. The cell lysate was then heat treated at 60°C for 30 min to remove the heat-unstable proteins. Finally, the cell lysate was centrifuged at 12,000 × *g* for 30 min, and the supernatant was used for *β*-galactosidase assays as previously reported (34).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/mBio.02004-16.

FIG S1, DOCX file, 0.4 MB.
FIG S2, DOCX file, 0.1 MB.
TABLE S1, DOCX file, 0.03 MB.
TABLE S2, DOCX file, 0.02 MB.
TABLE S3, DOCX file, 0.1 MB.
TABLE S4, DOCX file, 0.03 MB.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (31630003, 31570043, and 31421061), National High-Tech Research and Development Program of China (2015AA020202), Youth Innovation Promotion Association CAS, Natural Science Foundation of Shanghai (15ZR1446000), and Synthetic Biology China-United Kingdom Partnering Award.

REFERENCES

1. Görke B, Stülke J. 2008. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. Nat Rev Microbiol 6:613–624. https://doi.org/10.1038/nrmicro1932.
2. Lorca GL, Chung YJ, Barabote RD, Weyler W, Schilling CH, Saier MH, Jr. 2005. Catabolite repression and activation in *Bacillus subtilis*: dependency on CcpA, HPr, and HprK. J Bacteriol 187:7826–7839. https://doi.org/10.1128/JB.187.22.7826-7839.2005.
3. Swint-Kruse L, Matthews KS. 2009. Allostery in the LacI/GalR family: variations on a theme. Curr Opin Microbiol 12:129–137. https://doi.org/10.1016/j.mib.2009.01.009.
4. Chiang C, Bongiorni C, Perego M. 2011. Glucose-dependent activation of *Bacillus anthracis* toxin gene expression and virulence requires the carbon catabolite protein CcpA. J Bacteriol 193:52–62. https://doi.org/10.1128/JB.01656-09.
5. Shelburne SA, Keith D, Horstmann N, Sumby P, Davenport MT, Graviss EA, Brennan RG, Musser JM. 2008. A direct link between carbohydrate utilization and virulence in the major human pathogen group A *Streptococcus*. Proc Natl Acad Sci U S A 105:1698–1703. https://doi.org/10.1073/pnas.0711767105.
6. Iyer R, Baliga NS, Camilli A. 2005. Catabolite control protein A (CcpA) contributes to virulence and regulation of sugar metabolism in *Streptococcus pneumoniae*. J Bacteriol 187:8340–8349. https://doi.org/10.1128/JB.187.24.8340-8349.2005.
7. Schumacher MA, Sprehe M, Bartholomae M, Hillen W, Brennan RG. 2011. Structures of carbon catabolite protein A-(HPr-Ser46-P) bound to diverse catabolite response element sites reveal the basis for high-affinity binding to degenerate DNA operators. Nucleic Acids Res 39:2931–2942. https://doi.org/10.1093/nar/gkq1177.
8. Weickert MJ, Chambliss GH. 1990. Site-directed mutagenesis of a catabolite repression operator sequence in *Bacillus subtilis*. Proc Natl Acad Sci U S A 87:6238–6242. https://doi.org/10.1073/pnas.87.16.6238.
9. Miwa Y, Nakata A, Ogiwara A, Yamamoto M, Fujita Y. 2000. Evaluation and characterization of catabolite-responsive elements (*cre*) of *Bacillus subtilis*. Nucleic Acids Res 28:1206–1210. https://doi.org/10.1093/nar/28.5.1206.
10. Willenborg J, de Greeff A, Jarek M, Valentin-Weigand P, Goethe R. 2014. The CcpA regulon of *Streptococcus suis* reveals novel insights into the regulation of the streptococcal central carbon metabolism by binding of CcpA to two distinct binding motifs. Mol Microbiol 92:61–83. https://doi.org/10.1111/mmi.12537.
11. Tracy BP, Jones SW, Fast AG, Indurthi DC, Papoutsakis ET. 2012. Clostridia: the importance of their exceptional substrate and metabolite diversity for biofuel and biorefinery applications. Curr Opin Biotechnol 23:364–381. https://doi.org/10.1016/j.copbio.2011.10.008.
12. Ren C, Gu Y, Wu Y, Zhang W, Yang C, Yang S, Jiang W. 2012. Pleiotropic functions of catabolite control protein CcpA in butanol-producing *Clostridium acetobutylicum*. BMC Genomics 13:349. https://doi.org/10.1186/1471-2164-13-349.
13. Fischer RJ, Helms J, Dürre P. 1993. Cloning, sequencing, and molecular analysis of the *sol* operon of *Clostridium acetobutylicum*, a chromosomal

locus involved in solventogenesis. J Bacteriol 175:6959–6969. https://doi.org/10.1128/jb.175.21.6959-6969.1993.
14. Moreno MS, Schneider BL, Maile RR, Weyler W, Saier MH, Jr. 2001. Catabolite repression mediated by the CcpA protein in *Bacillus subtilis*: novel modes of regulation revealed by whole-genome analyses. Mol Microbiol 39:1366–1381. https://doi.org/10.1111/j.1365-2958.2001.02328.x.
15. Castaldo C, Siciliano RA, Muscariello L, Marasco R, Sacco M. 2006. CcpA affects expression of the *groESL* and *dnaK* operons in *Lactobacillus plantarum*. Microb Cell Fact 5:35. https://doi.org/10.1186/1475-2859-5-35.
16. Crooke AK, Fuller JR, Obrist MW, Tomkovich SE, Vitko NP, Richardson AR. 2013. CcpA-independent glucose regulation of lactate dehydrogenase 1 in *Staphylococcus aureus*. PLoS One 8:e54293. https://doi.org/10.1371/journal.pone.0054293.
17. Almengor AC, Kinkel TL, Day SJ, McIver KS. 2007. The catabolite control protein CcpA binds to P*mga* and influences expression of the virulence regulator Mga in the group A streptococcus. J Bacteriol 189:8405–8416. https://doi.org/10.1128/JB.01038-07.
18. Kim JH, Yang YK, Chambliss GH. 2005. Evidence that *Bacillus* catabolite control protein CcpA interacts with RNA polymerase to inhibit transcription. Mol Microbiol 56:155–162. https://doi.org/10.1111/j.1365-2958.2005.04496.x.
19. Marciniak BC, Pabijaniak M, de Jong A, Dűhring R, Seidel G, Hillen W, Kuipers OP. 2012. High- and low-affinity cre boxes for CcpA binding in *Bacillus subtilis* revealed by genome-wide analysis. BMC Genomics 13:401. https://doi.org/10.1186/1471-2164-13-401.
20. Antunes A, Camiade E, Monot M, Courtois E, Barbut F, Sernova NV, Rodionov DA, Martin-Verstraete I, Dupuy B. 2012. Global transcriptional control by glucose and carbon regulator CcpA in *Clostridium difficile*. Nucleic Acids Res 40:10701–10718. https://doi.org/10.1093/nar/gks864.
21. Novichkov PS, Rodionov DA, Stavrovskaya ED, Novichkova ES, Kazakov AE, Gelfand MS, Arkin AP, Mironov AA, Dubchak I. 2010. RegPredict: an integrated system for regulon inference in prokaryotes by comparative genomics approach. Nucleic Acids Res 38:W299–W307. https://doi.org/10.1093/nar/gkq531.
22. Barber AM, Zhurkin VB, Adhya S. 1993. CRP-binding sites: evidence for two structural classes with 6-bp and 8-bp spacers. Gene 130:1–8. https://doi.org/10.1016/0378-1119(93)90339-5.
23. Kallipolitis BH, Valentin-Hansen P. 2004. A role for the interdomain linker region of the *Escherichia coli* CytR regulator in repression complex formation. J Mol Biol 342:1–7. https://doi.org/10.1016/j.jmb.2004.05.067.
24. Lin CY, Awano N, Masuda H, Park JH, Inouye M. 2013. Transcriptional repressor HipB regulates the multiple promoters in *Escherichia coli*. J Mol Microbiol Biotechnol 23:440–447. https://doi.org/10.1159/000354311.
25. Schumacher MA, Balani P, Min JK, Chinnam NB, Hansen S, Vulić M, Lewis K, Brennan RG. 2015. HipBA-promoter structures reveal the basis of heritable multidrug tolerance. Nature 524:59–64. https://doi.org/10.1038/nature14662.
26. Jørgensen CI, Kallipolitis BH, Valentin-Hansen P. 1998. DNA-binding characteristics of the *Escherichia coli* CytR regulator: a relaxed spacing requirement between operator half-sites is provided by a flexible, unstructured interdomain linker. Mol Microbiol 27:41–50. https://doi.org/10.1046/j.1365-2958.1998.00655.x.
27. Lulko AT, Buist G, Kok J, Kuipers OP. 2007. Transcriptome analysis of temporal regulation of carbon metabolism by CcpA in *Bacillus subtilis* reveals additional target genes. J Mol Microbiol Biotechnol 12:82–95. https://doi.org/10.1159/000096463.
28. Brüggemann H, Gottschalk G. 2004. Insights in metabolism and toxin production from the complete genome sequence of *Clostridium tetani*. Anaerobe 10:53–68. https://doi.org/10.1016/j.anaerobe.2003.08.001.
29. Dubin G, Chmiel D, Mak P, Rakwalska M, Rzychon M, Dubin A. 2001. Molecular cloning and biochemical characterisation of proteases from *Staphylococcus epidermidis*. Biol Chem 382:1575–1582. https://doi.org/10.1515/BC.2001.192.
30. Wu Y, Yang Y, Ren C, Yang C, Yang S, Gu Y, Jiang W. 2015. Molecular modulation of pleiotropic regulator CcpA for glucose and xylose coutilization by solvent-producing *Clostridium acetobutylicum*. Metab Eng 28:169–179. https://doi.org/10.1016/j.ymben.2015.01.006.
31. Xiao H, Li Z, Jiang Y, Yang Y, Jiang W, Gu Y, Yang S. 2012. Metabolic engineering of D-xylose pathway in *Clostridium beijerinckii* to optimize solvent production from xylose mother liquid. Metab Eng 14:569–578. https://doi.org/10.1016/j.ymben.2012.05.003.
32. Wiesenborn DP, Rudolph FB, Papoutsakis ET. 1988. Thiolase from *Clostridium acetobutylicum* ATCC 824 and its role in the synthesis of acids and solvents. Appl Environ Microbiol 54:2717–2722.
33. Baer SH, Blaschek HP, Smith TL. 1987. Effect of butanol challenge and temperature on lipid composition and membrane fluidity of butanol-tolerant *Clostridium acetobutylicum*. Appl Environ Microbiol 53:2854–2861.
34. Tummala SB, Welker NE, Papoutsakis ET. 1999. Development and characterization of a gene expression reporter system for *Clostridium acetobutylicum* ATCC 824. Appl Environ Microbiol 65:3793–3799.