BMC
Genomics

High- and low-affinity *cre* boxes for CcpA binding in *Bacillus subtilis* revealed by genome-wide analysis

Bogumiła C Marciniak[1], Monika Pabijaniak[1], Anne de Jong[1], Robert Dűhring[2], Gerald Seidel[2],
Wolfgang Hillen[2]^ and Oscar P Kuipers[1,3*]

Abstract

Background: In *Bacillus subtilis* and its relatives carbon catabolite control, a mechanism enabling to reach maximal efficiency of carbon and energy sources metabolism, is achieved by the global regulator CcpA (carbon catabolite protein A). CcpA in a complex with HPr-Ser-P (seryl-phosphorylated form of histidine-containing protein, HPr) binds to operator sites called catabolite responsive elements, *cre*. Depending on the *cre* box position relative to the promoter, the CcpA/HPr-Ser-P complex can either act as a positive or a negative regulator. The *cre* boxes are highly degenerate semi-palindromes with a lowly conserved consensus sequence. So far, studies aimed at revealing how CcpA can bind such diverse sites were focused on the analysis of single *cre* boxes. In this study, a genome-wide analysis of *cre* sites was performed in order to identify differences in *cre* sequence and position, which determine their binding affinity.

Results: The transcriptomes of *B. subtilis* cultures with three different CcpA expression levels were compared. The higher the amount of CcpA in the cells, the more operons possessing *cre* sites were differentially regulated. The *cre* boxes that mediated regulation at low CcpA levels were designated as strong (high affinity) and those which responded only to high amounts of CcpA, as weak (low affinity). Differences in the sequence and position in relation to the transcription start site between strong and weak *cre* boxes were revealed.

Conclusions: Certain residues at specific positions in the *cre* box as well as, to a certain extent, a more palindromic nature of *cre* sequences and the location of *cre* in close vicinity to the transcription start site contribute to the strength of CcpA-dependent regulation. The main factors contributing to *cre* regulatory efficiencies, enabling subtle differential control of various subregulons of the CcpA regulon, are identified.

Keywords: CcpA, Catabolite responsive elements (*cre*) affinity, *Cre* box motif

Background

A well-known phenomenon among bacteria is the sole utilization of the most favored carbon source (e.g., glucose, fructose or malate) over other sugars present in the environment. The regulatory mechanism coordinating the metabolism of carbon and energy sources in order to maximize the metabolic efficiency is called carbon catabolite control, i.e., carbon catabolite repression (CCR) and carbon catabolite activation (CCA). Carbon catabolite control in *Bacillus subtilis* and other low-GC Gram-positive bacteria is exerted by the CcpA protein (catabolite control protein A) [1]. CcpA is a member of the LacI/GalR family of transcriptional regulators [2] and it can act either as a positive or negative regulator of genes that are in most cases involved in carbon acquisition or metabolism [3]. CcpA is synthesized constitutively, regardless to the availability of preferred carbon sources [4], it forms a dimer [5] and its activity is modulated by a complex interaction with either one of the

* Correspondence: o.p.kuipers@rug.nl
^Deceased
[1]Department of Molecular Genetics, Groningen Biomolecular Sciences and Biotechnology Institute, Centrum voor Levenswetenschappen, University of Groningen, Nijenborgh 7, 9747 AG, Groningen, The Netherlands
[3]Kluyver Center for Genomics of Industrial Fermentation, Delft/Groningen, The Netherlands
Full list of author information is available at the end of the article

corepressors, HPr or Crh [4-8]. In the presence of glucose or other rapidly metabolized carbon sources, the histidine-containing protein (HPr), and an HPr-like protein (Crh), are phosphorylated on a conserved Ser-46 residue by HPr kinase [9,10]. Binding of the seryl-phosphorylated HPr (HPr-Ser-P) or Crh (Crh-Ser-P) to CcpA stimulates the activity of CcpA [6-8,10]. During growth on carbohydrates there is much more HPr than Crh in the cell [11]. Notably, the Crh-specific function in the regulation of expression during growth on substrates other than carbohydrates was recently revealed [1]. Hence, Crh seems to play a secondary role in CCR. Next to HPr and Crh, low-molecular-weight molecules like NADP, glucose-6-phosphate (G6P), and fructose-1,6-bisphosphate (FBP) modulate CcpA activity by either stimulation of HPr kinase activity (FBP) [12,13], enhancement of CcpA affinity for HPr-Ser-P (FBP) [14], triggering cooperative CcpA binding to DNA (G6P) [15], or enhancing the CcpA interaction with the transcription machinery (NADP/NADPH) [16].

CcpA binds to DNA at *cis*-acting sequences called catabolite responsive elements (*cre*) located in the promoter region or within open reading frames of the regulated genes and operons. So far more than 50 *cre* sites were identified in the *B. subtilis* genome [1]. A general rule was deduced, stating that genes with *cre* boxes located upstream of −35 sequences of the promoter are subject to activation by the CcpA complex, as shown for *ackA* [17], *pta* [18] *ilvB* [19,20]. However, *ackA* is co-operatively activated by CcpA and CodY [21,22] and full activation of *ackA* requires also an additional conserved sequence present upstream of the *cre* box [23]. Moreover, the *lev* operon is subject to CcpA repression, although the *lev cre* site is located upstream of the promoter. However, regulation of the *lev* operon involves also the LevR transcriptional activator: binding of CcpA to the *lev cre* site prevents a productive interaction between LevR and RNA polymerase [24]. Binding of CcpA to *cre* boxes overlapping the promoter leads to transcriptional repression by interfering with the transcription machinery binding, as for *amyE*, *bglP*, *cccA*, *dctP*, *glpF*, *phoP*, *acuA* [25-31]. The binding of the protein complex to *cre* boxes that are located downstream of the transcription start site blocks transcription elongation, as is the case for most of the genes and operons regulated by CcpA [1,7].

Cre boxes are highly degenerate pseudo-palindromes with the consensus sequence WT<u>G</u>NNAR<u>C</u>GNWWW CAW, where the strongly conserved residues are underlined [32-34]. Little is known about how CcpA can bind to such diverse *cre* sequences. Our hypothesis was that CcpA can bind with different affinities to *cre* boxes with particular sequence and/or position in relation to the transcription start site (TSS). In order to identify *cre*

boxes with different affinities, CcpA expression was induced to three different levels using a tetracycline-dependent gene regulation system [35] and genome wide analysis of *cre* boxes was performed using transcriptome analyses combined with bioinformatics tools. High- and low-affinity *cre* boxes with subtle differences in their sequence and/or position in relation to the TSS are revealed.

Results

Tight regulation of CcpA production level

In order to enable very tight control of the CcpA expression level in *B. subtilis*, strain MP902 (P*tet-ccpA*, P*xyl-tetR*) was constructed. Strain MP902 carries the *ccpA* gene under control of the tetracycline-inducible promoter, P*tet*, integrated in the native promoter locus and the P*tet* repressor, *tetR*, under control of the xylose-inducible promoter, P*xyl*, located on the plasmid pWH119 [35]. To show tight regulation of the CcpA expression level, the MP902 strain was grown in rich TY medium [36] supplemented with 1 % glucose, 0.2 % xylose and a wide range of concentrations (0.1 – 20 nM) of P*tet* inducer, anhydrotetracycline (ATc) which is a non-bacteriostatic tetracycline analog. As demonstrated in Figure 1, the system allows obtaining several distinct expression levels of CcpA.

In order to test the influence of the different CcpA amounts in the cells on the CcpA regulon, three representative CcpA expression levels (hereafter referred to as low, medium and high) were chosen and the cultures were used for microarray experiments. For transcriptome analyses, the MP902 strain was grown in rich TY medium [36], since most likely it contains inducers for secondary regulators which could hide CCR in minimal medium and the samples were taken during exponential growth because CCR is expected to be strongest during maximal cell growth. The strain was grown in the presence of 0.2 % xylose to induce TetR expression and a
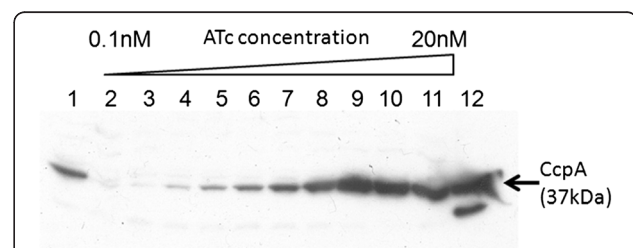
Figure 1 Tight regulation of the CcpA expression level in *B. subtilis* strain MP902 (P*tet-ccpA*, P*xyl-tetR*). Lane 1, wild type strain 168; lanes 2 – 11, MP902 grown in presence of 0.2 % xylose and increasing concentration of anhydrotetracycline (ATc): 0.1, 0.2, 04, 0.7, 1, 2, 4, 8, 10 and 20 nM, respectively; lane 12, 200 ng of purified CcpA. The representative graph of three reproducible experiments is shown.

high concentration of glucose (1 %) in order to ensure sufficient production of CcpA cofactors like HPr-Ser-P, NADP, glucose-6-phosphate (G6P) or fructose-1,6-bisphosphate (FBP) and optimal activity of CcpA. The medium was supplemented with different concentrations of ATc, exerting different CcpA production levels in the different cultures: 0.1 nM ATc (low CcpA induction level), 2 nM ATc (medium CcpA induction level) and 20 nM ATc (high CcpA induction level). The control culture was grown without ATc leading to no or only residual CcpA production. The CcpA production levels of the different MP902 cultures used for microarray experiments were assessed by Western blotting (Figure 2).

Effect of different CcpA amounts on gene regulation

The transcriptional profiles of exponentially growing cells of *B. subtilis* MP902 (P*tet-ccpA*, P*xyl-tetR*) grown in rich medium supplemented with glucose and xylose and expressing CcpA at low, medium and high levels (Figure 2) due to the presence of different concentrations of the P*tet* inducer, ATc, were compared to the transcriptional profile of MP902 cells grown in the corresponding medium but without ATc (no CcpA expression induction). Our first observation was that the more CcpA present in the cells the more genes were found to be significantly regulated (Table 1). Genes were considered to be regulated if they were at least 1.8 fold up- or downregulated. When CcpA was expressed at low, medium and high levels, 128, 343 and 408 genes were found to be differentially expressed, respectively. CcpA is known to act, depending on the *cre* box position in relation to the transcriptional start site (TSS), as a repressor or activator [37-39], but many more cases of repression than of activation are known [40]. Consistently, most of the regulated genes found in the microarray analyses with different CcpA induction levels were downregulated. For the list of expression fold changes of all the genes in the *B. subtilis* genome in all the three microarray experiments see Additional file 1 in the Supplementary Material.

The first genes of operons known from the literature to possess *cre* boxes (DataBase of Transcriptional Regulation in *Bacillus subtilis*, DBTBS [41] and reviewed by Fujita [1]) and which were differentially expressed at least under the high CcpA production level were extracted from the microarray data. Since it is estimated that the CcpA regulon includes more members than known so far [1], as also
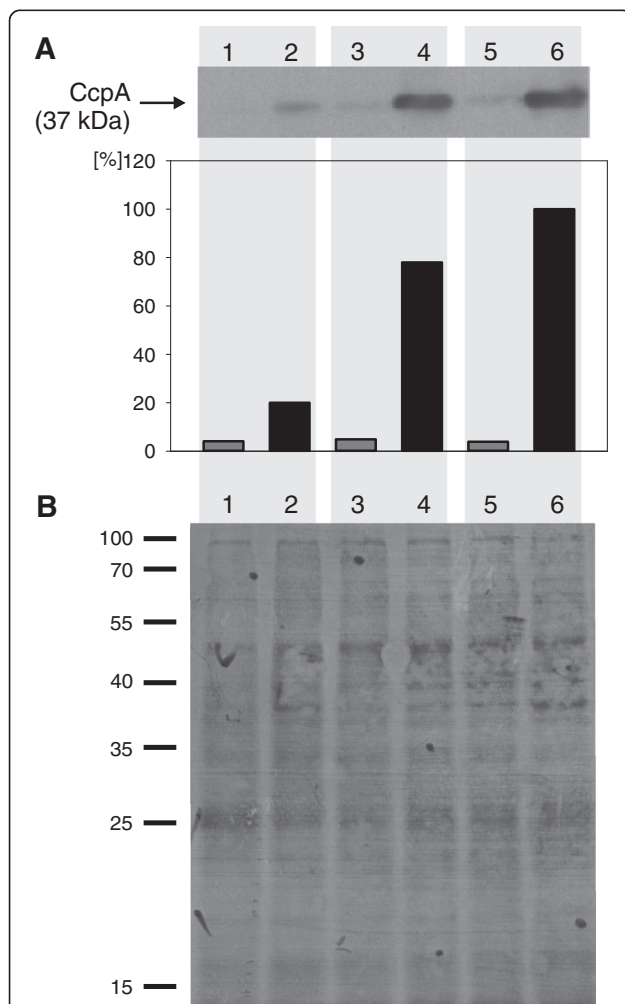
Figure 2 CcpA expression levels in *B. subtilis* cultures used for DNA microarray experiments. (**A**) *upper panel*, CcpA detection using anti-CcpA antibody; *lower panel*, signal quantification with ImageJ. Four CcpA expression levels were achieved by growing *B. subtilis* strain MP902 (P*tet-ccpA*, P*xyl-tetR*) in absence (lanes 1, 3 and 5) and in presence of 0.1, 2 and 20 nM ATc (lanes 2, 4 and 6), respectively. All cultures were grown in presence of 0.2 % xylose and 1 % glucose. Shadows in the background of the picture indicate culture pairs used in microarray experiments (**B**) Ponceau S control membrane staining for protein load verification. Lane numbers correspond to lane numbers in panel A. The representative graphs of three reproducible experiments are shown.

Table 1 Number of analyzed predicted *cre* boxes and regulated genes in response to different CcpA expression levels

	Level of the CcpA expression		
	Low (0.1 nM ATc)	Medium (2 nM ATc)	High (20 nM ATc)
N° of all genes	4106	4106	4106
N° of regulated genes	128	343	408
N° all predicted *cre* boxes	418	418	418
N° *cre* boxes of operons [a]	161	161	161
N° regulated operons with *cre* box [a]	30	58	67

[a] *cre* boxes within −500 and +100 nucleotides from start codon of first genes of operons.

shown recently [42], a prediction of putative *cre* boxes was performed. Using Genome2D [43] and a list of described *cre* boxes in the literature (reviewed by Fujita [1]) a Weight Matrix of *cre* boxes was generated: $T_1G_2A_3A_4A_5R_6C_7G_8Y_9T_{10}W_{11}W_{12}C_{13}A_{14}$. This *cre* motif was used to search the whole *B. subtilis* genome for putative *cre* boxes. As a result, 418 putative *cre* boxes were found: 200 in the upper and 218 in the lower strand (Table 1 and, for the complete list of found *cre*, Additional file 2). Most of the predicted *cre* boxes may not be functional taking into account their large distance from the promoter. Therefore, *cre* boxes located within −500

and +100 nucleotides relative to the start codon of the first gene of an operon were extracted. There were 161 genes possessing *cre* boxes that met these criteria (Table 1 and, for the complete list, Additional file 3). Since the search did not entirely cover the list of the known *cre* sites (for review see [1]), *cre* sites known from literature were also added to the analyzed *cre* sites. In total, there were 30, 58 and 67 operons possessing (known and predicted) *cre* sites and which were significantly downregulated under low, medium or high CcpA induction level, respectively. Three operons with known and predicted *cre* sites were activated under all these conditions (Table 2

Table 2 High- and low-affinity *cre* boxes of the first genes of operons

	Gene	Strand	*cre* sequence	Gene expression fold change [a]			*cre* box to TSS distance [b]
				Low CcpA induction	Medium CcpA induction	High CcpA induction	
High affinity cre boxes							
1	*acoR*	upper	TGAAAGCGCTTTAT	**−4.8**	**−18.7**	**−21.7**	−27
2	*acsA*	lower	TGAAAGCGTTACCA	**−2.3**	**−2.5**	**−2.7**	+44
3	*acuA*	upper	TGAAACGCTTTAT	**−2.2**	**−4.6**	**−7.7**	−26
4	*amyE*	upper	TGTAAGCGTTAACA	**−2.6**	**−10.7**	**−12.9**	+4
5	*bglP*	lower	TGAAAGCGTTGACA	**−2.5**	**−4.7**	**−4.6**	−36
6	*cccA*	lower	TGTAAGCGTATACA	**−2.2**	**−1.8**	**−2.8**	−29
7	*citM*	upper	TGTAAGCGGATTCA	**−2.6**	**−2.7**	**−2.9**	+46
8	*cstA*	lower	TGAATGCGGTTACA	**−2.2**	**−1.9**	**−2.4**	+32
9	*dctP*	upper	TGAAACGCTATCA	**−7.4**	**−12.3**	**−16.6**	−14
10	*glpF*	upper	TGACACCGCTTTCA	**−4.3**	**−21.9**	**−35.6**	−27
11	*gmuB*	upper	TGTAAGCGTTTTAA	**−3.0**	**−15.6**	**−35.8**	+6
12	*iolA-1*	lower	TGAAAGCGTTTAAT	**−1.8**	**−1.9**	**−2.1**	+93
13	*iolA-2 (within iolB)*	lower	TGAAACGTTGTCA	**−2.2**	**−2.5**	**−2.4**	+2404
14	*manR*	upper	TGTAAACGGTTTCT	**−2.0**	**−3.7**	**−8.0**	0
15	*msmX*	lower	AGAAAGCGTTTACA	**−2.0**	**−2.6**	**−3.1**	−15
16	*rbsR*	upper	TGTAAACGGTTACA	**−6.7**	**−15.2**	**−23.1**	+6
17	*rocG*	lower	TTAAAGCGCTTACA	**−2.6**	**−3.5**	**−3.1**	+43
18	*sacP*	lower	CGAAAACGCTATCA	**−2.1**	**−7.9**	**−8.1**	−19
19	*sucC*	upper	TGAAAGCGCAGTCT	**−2.0**	**−5.8**	**−3.4**	0
20	*treP*	upper	TGAAAACGCTTGCA	**−3.2**	**−13.0**	**−17.5**	+372
21	*uxaC*	upper	TGAAAGCGTTATCA	**−2.5**	**−3.7**	**−8.9**	+1237
22	*xsa*	lower	TAAAAGCGCTTACA	**−1.9**	**−1.8**	**−2.6**	+7
23	*xylA*	upper	TGGAAGCGCAAACA	**−2.4**	**−11.9**	**−11.1**	+144
24	*xynP*	upper	TGAAAGCGCTTTTA	**−4.0**	**−11.0**	**−17.9**	+230
25	*yisS*	upper	AGAAACGCTTTCT	**−1.9**	**−3.5**	**−3.7**	+74
26	*yjmD*	upper	TGAAAGCGGTTCAA	**−2.2**	**−2.4**	**−8.8**	ND
27	*ykoM*	upper	TGCAAGGGCTTTCA	**−2.0**	**−3.4**	**−3.5**	+150
28	*yrpD*	upper	TGATAGCGTTTTCT	**−1.9**	**−8.0**	**−6.8**	+127
29	*ytkA*	lower	TGTAAGCGTTTGCT	**−1.9**	**−6.4**	**−6.8**	ND
30	*yulD*	lower	TGAAAGCGCTATCT	**−2.3**	**−4.9**	**−5.3**	ND
31	*yvfK*	lower	TTAAAGCGCTTTCA	**−4.0**	**−6.1**	**−10.6**	+5

Table 2 High- and low-affinity *cre* boxes of the first genes of operons *(Continued)*

Low affinity cre boxes							
1	abnA	lower	TGTAAGCGCTTTCT	−1.8	−1.7	**−2.5**	+85
2	acoA	lower	TGTAAGCGTTTGCT	−1.1	−1.0	**−1.8**	+462
3	citZ	lower	TGTAAGCATTTTCT	−1.5	**−1.8**	**−2.1**	+88
4	csbX	lower	TGAAACGGTGCCA	−1.4	**−2.8**	**−2.1**	−401
5	cydA	lower	TGAAATGAATCGTT	1.6	1.0	**−2.7**	−21
6	drm	lower	TGAAACGGTTTAT	−1.3	**−3.6**	**−3.2**	−16
7	gntR-1	upper	TGAAAGTGTTTGCA	−1.3	**−2.8**	**−3.2**	−41
8	gntR-2	upper	TGAAAGCGGTACCA	−1.3	**−2.8**	**−3.2**	+148
9	hutP	upper	TGAAACCGCTTCCA	−1.3	**−1.9**	**−2.6**	+209
10	lcfA	lower	TGAAAACGTTATCA	−1.4	**−2.6**	**−2.6**	+450
11	levD	lower	TGAAAACGCTTAAC	−1.5	−1.2	**−2.2**	−45
12	malA	upper	TGTAAACGTTATCA	−1.7	**−2.0**	**−2.6**	+6
13	mleN	lower	TGAAAGCGTTTTAG	−1.5	**−3.5**	**−2.4**	+21
14	msmR	upper	TGTAACCGCTTACT	−1.7	**−4.2**	**−12.2**	−28
15	mtlR	upper	TGAAAGCGTTTTAT	−1.5	**−2.7**	**−2.5**	−16
16	odhA	lower	TGGAAGCGTTTTTA	−1.6	**−6.6**	**−3.4**	+21
17	pbuG	upper	TGAAAACGTTTTTT	−1.1	−1.5	**−1.9**	+245
18	pta	lower	TGAAAGCGCTATAA	1.3	**−3.2**	**−2.7**	−55
19	resA	lower	TAAAAACGCTTTCT	−1.1	**−1.9**	**−1.9**	−72
20	sigL	lower	GGAAAACGCTTTCA	−1.1	**−3.1**	**−3.3**	ND
21	wprA	upper	TGTAAGCGGTATCT	−1.6	**−5.5**	**−4.2**	+43
22	yckB	lower	TGAAAACGCGATCA	−1.4	**−3.5**	**−2.1**	−48
23	ycsA	upper	AGAAAGCGCTTACG	−1.7	**−6.0**	**−10.3**	+67
24	ydzA	lower	TGAAAACGTGTCCA	−1.3	**−6.4**	**−6.4**	+9
25	yesL	upper	TGAAAGCGTTTTCC	−1.3	−1.6	**−2.0**	+125
26	yfiG	upper	AGAAAGCGGTTACA	−1.6	**−2.7**	**−4.6**	+38
27	yncC	upper	TGTAAACGGTTACA	−1.3	**−2.4**	**−3.8**	+84
28	yojA	lower	TGAAAGCGCTTTCT	1.1	−1.5	**−1.8**	+57
29	yqgW	upper	TGAAAACGCTATCG	−1.1	**−4.5**	**−4.2**	−39
30	yqgY	upper	TGAAAATGTTTACA	−1.4	**−5.4**	**−4.1**	−38
31	ysbA	lower	TGTAAGCGCTTTAT	1.0	**−3.8**	**−7.6**	ND
32	ysfC	upper	TGAAAGCGTTTTTT	−1.5	−1.5	**−2.0**	+196
33	yugN	lower	TGAATGCGCTTTCT	−1.7	**−2.4**	**−2.3**	ND
34	yuxG	lower	TGAAAACGGATACA	−1.2	**−4.2**	**−6.1**	0
35	yvdG	lower	TGTAACCGCTTTCT	−1.4	−1.5	**−2.1**	−28
36	yxlH	upper	TTGAAACGCTTTCA	−1.4	**−2.0**	**−2.3**	+260
37	yydK	upper	TGTAAGCGGTTTAT	−1.5	**−3.2**	**−2.4**	−21
38	yyzE	lower	TGAAAGCGTAACCA	−1.2	**−3.0**	**−2.1**	0
Activating cre boxes							
1	ilvB	lower	TGAAAGCGTATACA	**3.0**	**6.2**	**2.7**	+88
2	opuE	lower	TGAAAGCGTTTTAT	**2.3**	**2.5**	**2.3**	−103
3	ycbP	lower	TGAAAGCGCTCGCT	**2.5**	**3.3**	**2.6**	+30

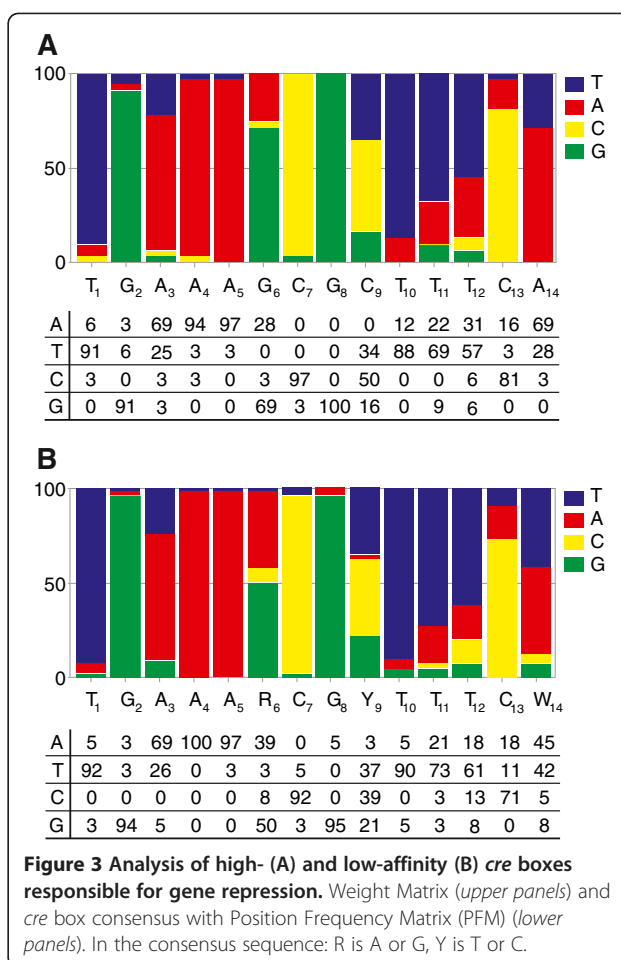*ª* In bold – genes significantly regulated (1.8 < fold < −1.8).
ᵇ cre box distance to transcriptional start site calculated from the conserved G residue in the middle CpG of the *cre* box.

and, in more detail, Additional file 4). For the sequences of the regions between −500 and +100 nucleotides from the start codon of first genes of operons possessing *cre* boxes that were analyzed in this study, see Additional file 5. The increase in amount of CcpA-regulated operons upon increasing amounts of CcpA indicates the presence of high-affinity *cre* boxes titrating away CcpA from the weaker *cre* boxes, which can trigger regulation of additional genes only when more functional CcpA is present in the cell. Therefore, the 31 *cre* boxes of the 30 operons (*iol operon* possesses two *cre* boxes: within *iolA* and *iolB*) repressed when CcpA was present in low amounts were designated as strong (high affinity to CcpA) and the other 38 *cre* sites of 37 operons (*gntR* possesses two *cre* sites), which were repressed only in the presence of higher amounts of CcpA in the cells (medium and high CcpA induction levels), were designated as weak (low affinity to CcpA) (Table 2). The high- and low-affinity, and the three activating *cre* boxes (Table 2) were analyzed with respect to their sequence and their position relative to the TSS. The term 'affinity' in this study is contractual, as direct binding assays were not performed in this study, and it is used to denote hierarchy in CcpA target genes regulation. From other (mutational) studies it is however apparent that strong regulation commonly coincides with high affinity and vice versa, so the term affinity appears to be adequate to describe differences in strong or weak regulation.

Analysis of *cre* box affinities in relation to their sequence

In order to detect differences within the sequence between different *cre* boxes, which putatively determine the *cre* box affinity, separate Weight Matrices for high- and low-affinity *cre* boxes that are responsible for gene repression were generated using Genome2D [43] (Figure 3). The resulting consensus sequences are $T_1G_2A_3A_4A_5G_6$ $C_7G_8C_9T_{10}T_{11}T_{12}C_{13}A_{14}$ and $T_1G_2A_3A_4A_5R_6C_7G_8Y_9T_{10}$ $T_{11}T_{12}C_{13}W_{14}$, for strong and weak *cre* boxes, respectively. *Cre* boxes from both groups have very conserved G_2, C_7 and G_8 residues, as in *cre* motifs proposed before [32-34]. Although the differences between high- and low-affinity *cre* are not very pronounced, the *cre* boxes with high affinity to CcpA seem to have a more conserved sequence around the middle CpG (conserved GCpGC instead of RCpGY) and at the C_{13} and A_{14} positions (Figure 3). To analyze the differences in the *cre* sequences in more detail, the high- and low-affinity *cre* boxes were aligned. The alignments show that the strong *cre* boxes (Table 3) have, on average, more palindromic residues than the weak *cre* boxes (Table 4) particularly at the external residues and in the middle CpG.

The *cre* sites of the genes that were activated in this study (*ilvB*, *opuE* and *ycbP*) were not included in the Weight Matrix generation nor *cre* alignment as *cre*

Figure 3 Analysis of high- (A) and low-affinity (B) *cre* boxes responsible for gene repression. Weight Matrix (*upper panels*) and *cre* box consensus with Position Frequency Matrix (PFM) (*lower panels*). In the consensus sequence: R is A or G, Y is T or C.

A

	T_1	G_2	A_3	A_4	A_5	G_6	C_7	G_8	C_9	T_{10}	T_{11}	T_{12}	C_{13}	A_{14}
A	6	3	69	94	97	28	0	0	0	12	22	31	16	69
T	91	6	25	3	3	0	0	0	34	88	69	57	3	28
C	3	0	3	3	0	3	97	0	50	0	0	6	81	3
G	0	91	3	0	0	69	3	100	16	0	9	6	0	0

B

	T_1	G_2	A_3	A_4	A_5	R_6	C_7	G_8	Y_9	T_{10}	T_{11}	T_{12}	C_{13}	W_{14}
A	5	3	69	100	97	39	0	5	3	5	21	18	18	45
T	92	3	26	0	3	3	5	0	37	90	73	61	11	42
C	0	0	0	0	0	8	92	0	39	0	3	13	71	5
G	3	94	5	0	0	50	3	95	21	5	3	8	0	8

boxes that are responsible for gene expression activation may need additional (upstream) sequences, as shown for instance for *ackA* [23]. Moreover, their sequence might putatively differ from the repressing *cre* sites, but the population of activating *cre* sites is too small to perform statistically significant analysis. However, taking into account the fact that all three genes that were activated in the microarray experiments in this study are regulated already in presence of low amounts of CcpA in the cell, the activating *cre* sites seem to take a higher place in the hierarchy of the genes regulated by CcpA. Additionally, the *cre* sites of *ilvB* and *ycbP* appear to match the consensus of the high-affinity *cre* boxes better compared to the consensus of low-affinity *cre* boxes (Table 2).

Analysis of the influence of relative *cre* box position on regulation

To find out whether the *cre* box position in relation to the promoter plays a role in determining the affinity to CcpA, the distance between *cre* boxes and the corresponding transcription start sites (TSS) was analyzed. The TSS of the regulated genes possessing a *cre* box

Table 3 Analysis of *cre* boxes with apparent high affinity to CcpA

Gene	Cre sequence [a]														Score [b]
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
acoR	T	G	A	A	A	G	C	G	C	T	T	T	A	T	5
acsA	T	G	A	A	A	G	C	G	T	T	A	C	C	A	4
acuA	T	G	A	A	A	A	C	G	C	T	T	T	A	T	4
amyE	T	G	T	A	A	G	C	G	T	T	A	A	C	A	5
bglP	T	G	A	A	A	G	C	G	T	T	G	A	C	A	4
cccA	T	G	T	A	A	G	C	G	T	A	T	A	C	A	5
citM	T	G	T	A	A	G	C	G	G	A	T	T	C	A	4
cstA	T	G	A	A	T	G	C	G	G	T	T	A	C	A	4
dctP	T	G	A	A	A	A	C	G	C	T	A	T	C	A	5
glpF	T	G	A	C	A	C	C	G	C	T	T	T	C	A	5
gmuB	T	G	T	A	A	G	C	G	T	T	T	T	A	A	4
iolA-1	T	G	A	A	A	G	C	G	T	T	T	A	A	T	3
iolA-2	T	G	A	A	A	A	C	G	T	T	G	T	C	A	6
manR	T	G	T	A	A	A	C	G	G	T	T	T	C	T	4
msmX	A	G	A	A	A	G	C	G	T	T	T	A	C	A	4
rbsR	T	G	T	A	A	A	C	G	G	T	T	A	C	A	6
rocG	T	T	A	A	A	G	C	G	C	T	T	A	C	A	5
sacP	C	G	A	A	A	A	C	G	C	T	A	T	C	A	4
sucC	T	G	A	A	A	G	C	G	C	A	G	T	C	T	4
treP	T	G	A	A	A	A	C	G	C	T	T	G	C	A	5
uxaC	T	G	A	A	A	G	C	G	T	T	A	T	C	A	5
xsa	T	A	A	A	A	G	C	G	C	T	T	A	C	A	5
xylA	T	G	G	A	A	G	C	G	C	A	A	A	C	A	4
xynP	T	G	A	A	A	G	C	G	C	T	T	T	T	A	6
yisS	A	G	A	A	A	A	C	G	C	T	T	T	C	T	6
yjmD	T	G	A	A	A	G	C	G	G	T	T	C	A	A	4
ykoM	T	G	C	A	A	G	G	G	C	T	T	T	C	A	5
yrpD	T	G	A	T	A	G	C	G	T	T	T	T	C	T	4
ytkA	T	G	T	A	A	G	C	G	T	T	T	G	C	T	4
yulD	T	G	A	A	A	G	C	G	C	T	A	T	C	T	5
yvfK	T	T	A	A	A	G	C	G	C	T	T	T	C	A	6
palindrome % [c]	68	71	52	61	84	32	97	97	32	84	61	52	71	68	Average score = 4.6

[a] In bold – palindromic residues.
[b] Score - a number of palindromic pairs.
[c] Occurrence of palindromic residue at each position.
Cre boxes of repressed genes are aligned.

were extracted from the literature or, when this information was lacking, predicted in this study (Table 2 and Additional file 5). The calculated *cre* to TSS distance (counting from the conserved G residue in the middle of the *cre* boxes to the TSS) was plotted against expression level fold change of the regulated genes under high levels of CcpA, separately for the genes with either high (Figure 4A) and low affinity *cre* boxes (Figure 4B). The majority of high affinity *cre* boxes are localized in close vicinity to the TSS (*cre*-TSS distance from 0 to +7, that is a TSS within the *cre* box) and around positions –27, -14 and +44. Repression of the genes with *cre* sites located with increment of approximately 10–11 nt (full helix turn) was significantly stronger, such as found for *cre* boxes of *acoR, glpF, dctP, gmuB, xynP, treP*, which are localized at positions –27, -27, -14, +6, +230, +372,

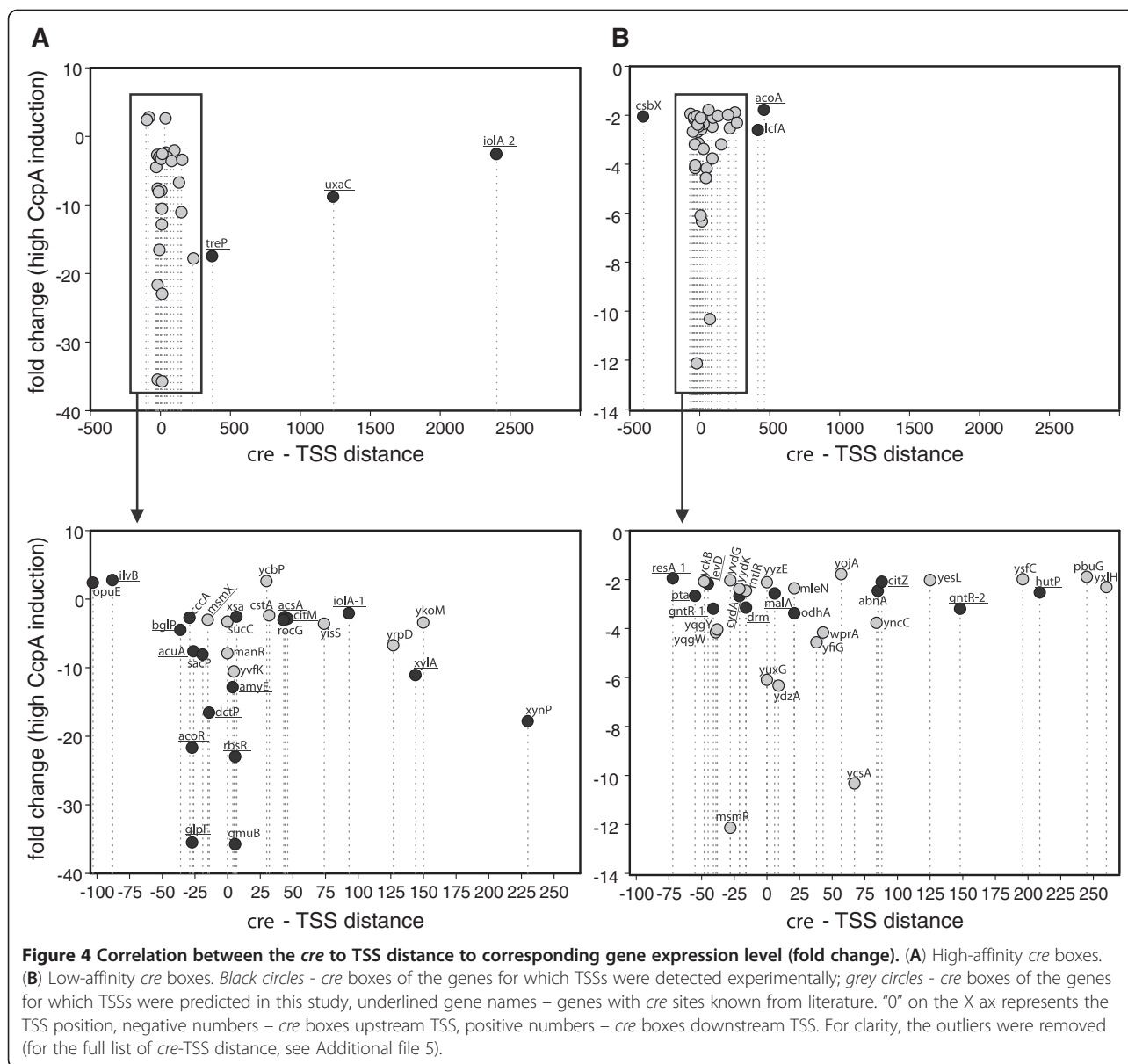**Table 4 Analysis of *cre* boxes with apparent low affinity to CcpA**

Gene	Cre sequence [a]														Score [b]
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
abnA	T	**G**	T	**A**	**A**	**G**	**C**	**G**	**C**	**T**	**T**	T	**C**	T	5
acoA	T	**G**	T	**A**	**A**	G	**C**	**G**	T	**T**	**T**	G	**C**	T	4
citZ	T	**G**	T	**A**	**A**	G	C	A	T	**T**	**T**	T	**C**	T	3
csbX	**T**	G	A	A	**A**	A	**C**	**G**	G	**T**	G	C	**C**	**A**	4
cydA	T	G	A	A	**A**	**T**	G	A	**A**	**T**	C	G	T	T	2
drm	T	G	**A**	**A**	**A**	A	**C**	**G**	G	**T**	**T**	**T**	A	T	4
gntR-1	**T**	**G**	A	A	**A**	G	T	G	T	**T**	**T**	G	**C**	**A**	4
gntR-2	**T**	**G**	A	A	**A**	G	**C**	**G**	G	**T**	A	C	**C**	**A**	4
hutP	**T**	**G**	A	**A**	**A**	C	**C**	**G**	C	**T**	**T**	C	**C**	**A**	5
lcfA	**T**	**G**	**A**	A	**A**	**A**	**C**	**G**	T	**T**	A	**T**	**C**	**A**	6
levD	T	G	A	**A**	**A**	A	**C**	**G**	C	**T**	**T**	A	A	C	3
malA	**T**	**G**	T	A	**A**	**A**	**C**	**G**	T	**T**	A	T	**C**	**A**	5
mleN	T	G	**A**	**A**	**A**	G	**C**	**G**	T	**T**	**T**	**T**	A	G	4
msmR	T	**G**	**T**	**A**	**A**	C	**C**	**G**	C	**T**	**T**	**A**	**C**	T	5
mtlR	T	G	**A**	**A**	**A**	G	**C**	**G**	T	**T**	**T**	**T**	A	T	4
odhA	**T**	G	G	**A**	**A**	G	**C**	**G**	T	**T**	**T**	T	T	**A**	4
pbuG	T	G	**A**	**A**	**A**	**A**	**C**	**G**	**T**	**T**	**T**	**T**	T	T	5
pta	**T**	G	**A**	A	**A**	**G**	**C**	**G**	**C**	**T**	A	**T**	A	**A**	5
resA	T	A	**A**	**A**	**A**	A	**C**	**G**	C	**T**	**T**	**T**	C	T	4
sigL	G	**G**	**A**	**A**	**A**	A	**C**	**G**	C	**T**	**T**	**T**	**C**	A	5
wprA	T	**G**	T	A	**A**	G	**C**	**G**	G	**T**	A	T	**C**	T	3
yckB	**T**	**G**	**A**	A	A	A	**C**	**G**	C	G	A	**T**	**C**	**A**	4
ycsA	A	**G**	A	**A**	**A**	**G**	**C**	**G**	**C**	**T**	**T**	A	**C**	G	5
ydzA	**T**	**G**	A	**A**	A	**A**	**C**	**G**	**T**	G	**T**	C	**C**	**A**	5
yesL	T	**G**	**A**	**A**	**A**	G	**C**	**G**	T	**T**	**T**	**T**	**C**	C	4
yfiG	A	**G**	A	**A**	**A**	G	**C**	**G**	G	**T**	**T**	A	**C**	A	4
yncC	**T**	**G**	**T**	**A**	**A**	A	**C**	**G**	G	**T**	**T**	**A**	**C**	**A**	6
yojA	T	**G**	**A**	**A**	**A**	**G**	**C**	**G**	**C**	**T**	**T**	**T**	**C**	T	6
yqgW	T	**G**	**A**	A	**A**	A	**C**	**G**	C	**T**	A	**T**	**C**	G	4
yqgY	**T**	**G**	A	**A**	**A**	**A**	T	G	**T**	**T**	**T**	A	**C**	**A**	5
ysbA	T	G	T	**A**	**A**	**G**	**C**	**G**	**C**	**T**	**T**	T	A	T	4
ysfC	T	G	**A**	**A**	**A**	G	**C**	**G**	T	**T**	**T**	**T**	T	T	4
yugN	T	**G**	**A**	**A**	T	**G**	**C**	**G**	**C**	T	**T**	**T**	**C**	T	5
yuxG	**T**	**G**	A	**A**	A	A	**C**	**G**	G	A	**T**	A	**C**	**A**	4
yvdG	T	**G**	T	**A**	**A**	C	**C**	**G**	C	**T**	**T**	T	**C**	T	5
yxlH	**T**	T	G	**A**	**A**	A	**C**	**G**	C	**T**	**T**	T	C	**A**	4
yydK	T	G	T	**A**	**A**	G	**C**	**G**	G	**T**	**T**	T	A	T	3
yyzE	**T**	**G**	A	A	**A**	G	**C**	**G**	T	A	A	C	**C**	**A**	3
palindrome % [c]	40	66	42	74	87	32	90	90	32	87	74	42	66	40	Average score = 4.3

[a] In bold – palindromic residues.
[b] Score - a number of palindromic pairs.
[c] Occurrence of palindromic residue at each position.
Cre boxes of repressed genes are aligned.

Figure 4 Correlation between the *cre* to TSS distance to corresponding gene expression level (fold change). (A) High-affinity *cre* boxes.
(**B**) Low-affinity *cre* boxes. *Black circles* - *cre* boxes of the genes for which TSSs were detected experimentally; *grey circles* - *cre* boxes of the genes for which TSSs were predicted in this study, underlined gene names – genes with *cre* sites known from literature. "0" on the X ax represents the TSS position, negative numbers – *cre* boxes upstream TSS, positive numbers – *cre* boxes downstream TSS. For clarity, the outliers were removed (for the full list of *cre*-TSS distance, see Additional file 5).

respectively. Further downstream from the TSS, there are more low affinity *cre* boxes than high affinity ones.

Discussion

CcpA is a global regulator of carbon catabolism [3] controlling expression of genes by binding to cognate operator sequences, *cre*, which is characterized by a low-conserved consensus sequence [32-34]. Hence, it seems possible that CcpA binds some *cre* sites with higher affinity than others. So far, the global studies of CcpA-dependent carbon catabolite repression were focused on identification of the members of the CcpA regulon [40,42,44], while the analysis of *cre* boxes in respect to their sequences, position and affinities in CcpA binding have been focused only on single examples [7,17,33,34,45]. A broader comparison of

32 *cre* boxes sequences and function was published by Miwa Y. *et al.* and it was deduced that a lower mismatching of *cre* sequences to the query sequence in the same direction as that of transcription of the target genes and a more palindromic sequence of *cre* boxes are desirable for their better function [33]. The goal of our study was to perform a genome-wide analysis of *cre* boxes in order to reveal *cre* boxes with high and low binding affinities by comparing the CcpA regulon under three distinct conditions, where different amounts of CcpA were present in the cells and to identify *cre* features that determine this affinity.

Using a tetracycline-dependent gene regulation system [35] we achieved a tightly-controlled *ccpA* expression, leading to a wide range of CcpA amounts in the cells.

B. subtilis cultures with relative low, medium or high amounts of CcpA in the cells were subjected to transcriptome analyses. The cells were grown in the presence of glucose to ensure sufficient production of low-molecular-weight modulators of CcpA activity (NADP, glucose-6-phosphate, fructose-1,6-bisphosphate). As expected, higher levels of CcpA protein lead to more genes significantly up- or downregulated. Most of the regulated genes, however, were affected indirectly, as they were lacking a *cre* site. Genes regulated indirectly in a CcpA-dependent manner (no *cre* or unfunctional *cre*) were already observed before and were proposed to be grouped in class II, next to class I that includes genes regulated by CcpA directly [40,46,47]. In our analysis, only genes belonging to class I were taken into account as the subject of this study was the nature of discriminating *cre* boxes. Many repressed genes are σA-dependent and do not need another inducing protein for their expression. However, expression of some genes is regulated by more than one regulator. In those rare cases of multiple regulation, the full extent of regulation would not be observed in our transcriptome analysis, but this does not affect our analysis since we are looking at the relative strength of repression at different CcpA concentrations.

The search for putative *cre* boxes in the *B. subtilis* genome, using a *cre* motif generated from the *cre* boxes known from DBTBS [41], $T_1G_2A_3A_4A_5R_6C_7G_8Y_9T_{10}W_{11}W_{12}C_{13}A_{14}$, resulted in 418 putative *cre* boxes. The majority of the predicted *cre* boxes were within ORFs far away from promoters and, although functional *cre* boxes located within coding sequences are present in the *B. subtilis* genome, a lot of the predicted *cre* sites seemed to be at a too large distance from the promoter to possibly be able to play a role in regulation of gene expression. Therefore, *cre* boxes located within −500 and +100 nucleotides from the first nucleotide of a start codon of the first genes of an operon were extracted. Also *cre* boxes triggering gene regulation that are known from the literature, but not predicted by our method, were included in our analysis. The genes differentially expressed at least at a high CcpA production level and possessing *cre* box(es) known from literature [1,41] and/or predicted in this study were selected. Among the selected genes, 30 were downregulated and 3 were upregulated at a low CcpA induction level, while the other 37 genes were downregulated only when CcpA was produced at higher levels (medium and high CcpA induction levels). For all these genes, expression fold changes were calculated as ratios of the amounts of transcripts downstream of *cre* boxes as the microarray chip probes were synthesized upstream from them. Of the regulated first genes of operons possessing known and/or predicted *cre* box, chip probes of only *kdgR* and *resA* were upstream from *kdgR-cre* and second *cre* of *resA* (located 1709 bp downstream from TSS). Therefore, these *cre*

boxes were not included in the sequence and position analysis of *cre* boxes. Since regulation depends on CcpA-*cre* binding, *cre* boxes causing significant regulation of downstream operons already when a small amount of CcpA is available are supposed to have a high affinity to CcpA and titrate CcpA away from low-affinity *cre* sites, which are able to exert regulation of other operons only when more CcpA is present. Notably, over a dozen of known *cre's* fell out of our data set, because the corresponding genes were not significantly regulated in any of the three microarray experiments. Despite of the fact that they could be considered as very low-affinity sites, they were not included in the analysis as lack of the differential expression might have been a false negative result due to, e.g., high background signal, bad spot quality on the microarray slides, mRNA degradation, growth conditions, more complex regulation or yet unidentified factors. Moreover, it should be noted that division of *cre* boxes to two affinity groups is a simplification necessary for this analysis. Very likely a gradient distribution of *cre* site affinities occurs in nature, which would be difficult to assess.

The detailed analysis of the sequences of high- and low-affinity *cre* boxes, led to a few interesting observations. The G_2 and middle C_7 and G_8 residues (Figure 3), known as highly conserved residues [32-34] are conserved in both high- and low-affinity *cre* boxes. Interestingly, the high-affinity *cre* boxes have more conserved G_6 and C_9 surrounding the middle CpG and C_{13} (palindromic to the conserved G_2) and A_{14} (palindromic to T_1) and their sequences are significantly more palindromic overall. It was observed before that a more palindromic sequence of *cre* sites contributes to a better function [33]. The more palindromic nature of the high-affinity *cre* sites (in comparison low-affinity *cre* sites) might create a more symmetric DNA conformation, preferred for CcpA binding. Although the bases at positions 4 and 11 are more often palindromic to each other in the weak *cre* boxes, this is obviously less important for the *cre* strength. In a previous study [34] it was shown that CcpA binds with similar affinities to different *cre* boxes, which explains well the role of CcpA as a global regulator. However, the three *cre* boxes tested in that work differ very little around the middle CpG and in their symmetry (palindromic sequence) and they did not differ at the residues corresponding to our C_{13} nor A_{14}.

Comparison of the high- and low-affinity *cre* boxes location in relation to the TSS also shows some trends. While the low-affinity *cre* sites can be located at any position from the TSS, the high-affinity *cre* sites cluster around the TSS, 14 and 27 base pairs upstream from TSS and 44 base pairs downstream from TSS. Simultaneously, the strongest repression by CcpA was observed for the genes with *cre* sites located around the TSS (*amyE, rbsR, gmuB*) and at positions −27 (*acoR, glpF*), -14 (*dctP*), +230

(*xynP*) and +372 (*treP*) base pairs from the TSS, which are separated by approximately 10 - 11-nt increments (corresponding with a full helical turn). This observation is in agreement with previous findings that activation or repression by CcpA binding to *cre* boxes is helix-face-dependent [17,45]. Also in *Lactococcus lactis* the strongest repression by CcpA was shown to occur when the center of *cre* box was located −39, -26, -16, +5 and +15 from the TSS [48].

It was shown before that genes with *cre* boxes located further upstream from −35 sequences of the promoter are subject to activation by the CcpA complex as in case of *ackA* [17], *pta* [18] and *ilvB* [19,20]. In our work however, under the tested conditions, only three genes were activated: *ilvB*, *opuE* and *ycbP* (the two latter genes with *cre* sites predicted in this study). We did not observe activation of *ackA* in this study. This is probably due to the very low basal expression of CcpA from the TetR repressed promoter that might be high enough for binding of CcpA to the *ackA cre* box and for full activation of the *ackA* promoter. In this case, a further increase of CcpA does not result in an additional increase of *ackA* expression. Surprisingly, *pta* was downregulated. However, in this study both test and control cultures were grown in medium supplemented with glucose. The mechanism of *pta* regulation in this case is thus different from low glucose-dependent CCA. Based on our criteria, the *cre* boxes of all three activated genes are of the high affinity type. Although the *ycbP cre* box appears to be downstream to the TSS (+30), both the *cre* box and the TSS in this case are not experimentally confirmed.

Some genes and operons possess multiple *cre* boxes. Since DNA microarray technology was used in this study to assess expression fold changes of genes and operons in the presence of different amounts of CcpA, we were not always able to judge whether the effect is due to one *cre* box (and which one) or more. In our set (Table 2) there were only two operons with two *cre* boxes (the first genes of these operons are: *iolA* and *gntR*). *gntR* was weakly regulated (low-affinity *cre* box), suggesting that the regulatory effects of the two *cre* boxes do not add up to exert strong regulation. In case of the *iolA* operon, each of the two *cre* boxes is located within another gene of the operon (*cre*-1 within *iolA* and *cre*-2 within the second gene of the operon, *iolB*). In this case, the regulatory effects of these *cre* boxes could be assessed independently. Based on the fold changes of *iolA* (*cre*-1) and *iolB* (*cre*-2), both *cre*-1 and *cre*-2 seem to be of high affinity. Multiple *cre* boxes could serve for fine tuning of CcpA-regulated genes and operons.

For the genes with *cre* boxes located close to the TSS and downstream, distinct repression mechanisms were proposed. Elongation blockage (roadblock) was shown for *xyl*, *ara* and *gnt* operons, as well as *sigL* and *acsA*

[49-53]. Prevention of binding RNAP to the promoter sequence was demonstrated for the *acuABC* and *bglPH* operons possessing *cre* partially overlapping with the promoter region [54,55]. Transcription inhibition by direct interaction of CcpA with the σ-subunit of RNAP already bound to the promoter was shown in case of the *amyE* gene and *xyl* operon [45]. The presence of a high-affinity *cre* box in close vicinity to the TSS shown in this study, suggests that repression by inhibition of RNAP binding is one of the most effective mechanism of negative regulation by CcpA.

Conclusions

In conclusion, we propose that besides the strongly conserved G_2 residue and the middle CpG, the residues G_6 and C_9 (surrounding the middle CpG), C_{13} and A_{14} and, to a certain extent a more palindromic sequence and a location of *cre* in close vicinity to the TSS, contribute to the high affinity of CcpA for certain *cre* boxes. This finding contributes to further understanding how CcpA binding to *cre* boxes is modulated and how subregulons can be formed. However, not all the *cre* boxes behave strictly according to this rule, suggesting that *cre* affinity is possibly determined in an even more complicated way. The *cre* sequence and position may play a role simultaneously and/or more factors may be involved, for instance additional conserved sequences as shown for *ackA* [23] or sequences flanking *cre* sites as in case of *acsA* [53].

It will be interesting to use these predictions for other G-positive organisms employing CcpA, like other *Bacilli*, lactic acid bacteria, or pathogenic *Streptococci* and *Staphylococci*.

Methods
Bacterial strains and growth conditions

Bacillus subtilis strain MP902 (*trpC2*, P*tet-ccpA*, pWH119, Km^R, Em^R) was grown in rich TY medium [36] in the dark at 37 °C with shaking. The medium was supplemented with 15 μg/ml kanamycin, 2.5 μg/ml erythromycin, 1 % glucose, 0.2 % xylose and anhydrotetracycline (ATc) at different concentrations. For inoculation, synchronized stocks were used. Synchronized stocks were prepared by growing the strain in TY medium with a corresponding composition as described before [44]. At OD_{600} = 0.8, the cells were collected for determination of the CcpA production level with Western blot and for RNA isolation to be used for microarray analysis.

Construction of the MP902 strain

All primers used in this work are listed in Table 5. To replace the *ccpA* promoter by a tetracycline-inducible promoter at the natural locus on the chromosome, the

Table 5 Primers used in this study

Primer	Sequence (5'- 3')
ccpAmut1	ATAATATCTAGAACCAAGTATACGTTTTCATC
Accout	ATAATAATAGGTACCGCTTCGAGTCCGGAATC
KmkfwR	ATAATAATATGTACAGATAAACCCAGCGAACCA
KmkbwR	AATAATAATAATAGTATACTATAAAACATCAGAGTATGGA
tetPccpAfw	ATAATAATATGTACAGCATGGTCCTAATTTTTGTT
tetPccpAbw	TACTGGATACACTTATCCTTCTGCAGGCATGCAAGCTA

integration vector pWH849 was constructed as follows. A *ccpA* fragment truncated at the 3' end was amplified from plasmid pWH1533 [56] using primers ccpAmut1 and Accout, restricted with *Bsr*GI and *Kpn*I and cloned into vector pWH618 [56]. The resulting vector was named pWH700 and contains the terminal 246 bases of *aroA*, the intergenic region between *aroA* and *ccpA* and 689 bases of *ccpA*. Next, a kanamycin resistance cassette was amplified from plasmid pDG792 [57], using primers KmkfwR and KmkbwR, and inserted in the intergenic region between *aroA* and *ccpA* via the restriction sites *Bsr*GI and *Acc*I. The resulting vector was named pWH800. The tetracycline-inducible promoter, P*tet* was amplified from the plasmid pWH1935-2 [58] with primers tetPccpAfw and tetPccpAbw. The resulting PCR fragment was used as a primer together with the primer Accout in order to fuse the tetracycline-inducible promoter P*tet* with *ccpA* at the intergenic region between *aroA* and *ccpA* in an overlap PCR with pWH800 as a template. The resulting PCR fragment was restricted with *Bsr*GI and *Kpn*I and cloned into vector pWH800, resulting in pWH849. *B. subtilis* 168 [59] was transformed with pWH849, linearized with *Sca*I, to replace the *ccpA* promoter on the chromosome by the tet inducible promoter via double homologous recombination. Positive candidates were selected on TY plates with kanamycin and verified by PCR screening. The resulting strain was named MP901. Strain MP901 was transformed with pWH119 plasmid [35] carrying tetracycline repressor gene, *tetR*, under control of xylose-inducible promoter, P*xyl* (P*xyl-tetR*), resulting in MP902 strain.

Quantification of the CcpA production level with Sodium dodecyl sulfate - polyacrylamide gel electrophoresis (SDS-PAGE) and Western blotting

B. subtilis MP902 cells were grown in LB medium with 0.2 % xylose and 0.1 to 20 nM ATc after one overnight culture with the respective xylose and ATc concentrations. In the mid log phase, 0.5 OD_{600} equivalents of the cells were sedimented and resuspended in SBT buffer (50 mM TrisHCl, 200 mM NaCl, 10 mM β-mercaptoethanol pH 7.5). After sonication, 0.05 OD_{600} equivalents of the crude protein extracts and 200 ng wild-type CcpA purified as described previously [56]

were subjected to SDS PAGE on a 10 % polyacrylamide gel. Proteins were then transferred to a PVDF membrane by electroblotting. After blocking, the membrane was incubated with a 1:10,000 dilution of rabbit polyclonal anti-CcpA antibodies [60]. For detection of CcpA on an X-ray film the membrane was incubated with anti-rabbit horseradish peroxidase conjugate and a luminol containing reagent mixture from an ECL + kit (GE Healthcare, Munich, Germany) according to manufacturer's instructions.

To analyze the CcpA production level in the cultures used for microarray experiments, the cells were collected at an optical density of $OD_{600} = 0.8$ (simultaneously with collection of the cells for total RNA isolation for microarray experiments). The signal on Western blot was quantified using ImageJ gel analyzer (http://rsb.info.nih. gov/ij/). For gel loading verification, the control blots were stained with 0.1 % Ponceau S dissolved in 5 % acetic acid. Images of Ponceau S – stained membranes were obtained using GS-800 calibrated densitometer (Bio-Rad).

DNA Microarray Analysis

16 ml of a culture was harvested at optical density of $OD_{600} = 0.8$ by centrifugation at 8,000 × g for 2 min. The pellet was rapidly frozen in liquid nitrogen and stored at −80 °C until RNA isolation. DNA microarray experiments were performed in general as described before [44]. Total RNA was isolated using High Pure RNA Isolation Kit (Roche) according to the manufacturer's protocol. RNA quantity and quality were tested with a ND-1000 spectrophotometer (NanoDrop Technologies) and an Agilent Bioanalyzer 2100 with RNA 6000 Lab-Chips (Agilent Technologies Netherlands BV), respectively. The amino allyl modified cDNA was synthesized with the Superscript III Reverse Transcriptase kit (Invitrogen), purified with Cyscribe GFX purification kit (Amersham Biosciences), labeled with Cy3 and Cy5 dyes and purified again. The labeled cDNA was hybridized to oligonucleotide microarrays in Ambion Slidehyb #1 buffer (Ambion Europe Ltd). Slides were washed, dried by centrifugation and scanned with a GeneTac LS V confocal laser scanner (Genomic Solutions Ltd). Scans were analyzed with ArrayPro 4.5 (Media Cybernetics Inc., Silver Spring, Md., USA). The resulting expression levels were normalized with Micro-Prep [61] and subjected to a t-test using the Cyber-T tool [62]. All microarray experiments were performed in three biological replicates. The complete microarray data is available at the GEO repository (http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE35154.

The sequences of the *cre* boxes known from DBTBS [41] were used to generate a weight matrix in Genome2D [43]. The resulting Weight Matrix was fed into

the Genome2D [43] to find the potential *cre* boxes in the whole genome of *B. subtilis*. In this search, a cut-off of 8.96 was used. The promoters (−35 and −10 boxes) and transcriptional start sites (TSS) were predicted using PePPER (Prediction of Prokaryote Promoter Elements and Regulons) tool [63] and sequence analysis. For the annotation, GenBank file NC000964.gbk last modified on the 19th of October 2011 available at http://www.ncbi.nlm.nih.gov/guide/ was used. The operons from DBTBS database [41] were confirmed with experimental evidence from microarray results obtained in this study (clustered up- or downregulation of genes belonging to one operon).

Additional files

Additional file 1: Gene expression fold changes "CcpA induction" vs "no CcpA induction". Expression fold changes in three microarrays experiments with low, medium and high CcpA expression induction. Activated genes - positive values, repressed genes - negative values.

Additional file 2: *Cre* boxes predicted in the *Bacillus subtilis* genome using Genome2D software and 8.96 calculated cutoff.

Additional file 3: *Cre* boxes predicted with Genome2D software using cutoff 8.96 and found in upper and lower strand within −500 and +100 nucleotides from the start codon of the first gene of operon.

Additional file 4: List of significantly regulated first genes of operons (1.8 < fold change < −1.8) with *cre* boxes within −500 and +100 nucleotides from start codon. Significant fold changes are shown in bold. *cre* box to start codon and transcriptional start site (TSS) distances are calculated from conserved G residue in the middle of *cre* box to first nucleotide of start codon and to TSS, respectively. For *cre* box affinity determination criteria see main text.

Additional file 5: First genes of operons with *cre* sites. Start codon and further coding sequences are shown in lowercase and intergenic regions in uppercase. Underlined – predicted *cre* boxes, bold – *cre* boxes known from the literature, solid box – promoter known from the literature, dotted box – predicted promoter, dark grey shadow - transcriptional start site (TSS) known from the literature, light grey shadow – predicted TSS.

Abbreviations
cre: Catabolite responsive elements; ATc: Anhydrotetracycline; TSS: Transcriptional start site; G6P: Glucose-6-phosphate; FBP: Fructose-1, 6-bisphosphate; RNAP: RNA polymerase.

Competing interests
The authors declare that they have no competing interests.

Authors' contribution
BCM, WH and OPK contributed to the design of the study. OPK and WH conceived the study. GS, RD and MP constructed the MP901 and MP902 strains. MP and RD performed Western blot experiments. BCM, MP and AdJ were involved in analysis of the data. BCM and OPK were involved in interpretation of data and writing the manuscript. All authors read and approved the final manuscript. OPK gave final approval of the version to be published.

Authors' information
The authors wish to dedicate this work to the memory of Prof. Wolfgang Hillen who died unexpectedly on 17 October 2010.

Author details
[1]Department of Molecular Genetics, Groningen Biomolecular Sciences and Biotechnology Institute, Centrum voor Levenswetenschappen, University of Groningen, Nijenborgh 7, 9747 AG, Groningen, The Netherlands. [2]Lehrstuhl für Microbiologie, Institut für Biologie der Friedrich-Alexander-Universität Erlangen-Nürnberg, Staudtstr. 5, 91058, Erlangen, Germany. [3]Kluyver Center for Genomics of Industrial Fermentation, Delft/Groningen, The Netherlands.

References
1. Fujita Y: **Carbon catabolite control of the metabolic network in *Bacillus subtilis*.** *Biosci Biotechnol Biochem* 2009, **73**:245–259.
2. Henkin TM, Grundy FJ, Nicholson WL, Chambliss GH: **Catabolite repression of α amylase gene expression in *Bacillus subtilis* involves a *trans*-acting gene product homologous to the *Escherichia coli lacI* and *galR* repressors.** *Mol Microbiol* 1991, **5**:575–584.
3. Sonenshein AL: **Control of key metabolic intersections in *Bacillus subtilis*.** *Nat Rev Microbiol* 2007, **5**:917–927.
4. Miwa Y, Saikawa M, Fujita Y: **Possible function and some properties of the CcpA protein of *Bacillus subtilis*.** *Microbiology* 1994, **140**:2567–2575.
5. Schumacher MA, Allen GS, Diel M, Seidel G, Hillen W, Brennan RG: **Structural basis for allosteric control of the transcription regulator CcpA by the phosphoprotein HPr-Ser46-P.** *Cell* 2004, **118**:731–741.
6. Deutscher J, Reizer J, Fischer C, Galinier A, Saier MH, Steinmetz M: **Loss of protein kinase-catalyzed phosphorylation of HPr, a phosphocarrier protein of the phosphotransferase system, by mutation of the *ptsH* gene confers catabolite repression resistance to several catabolic genes of *Bacillus subtilis*.** *J Bacteriol* 1994, **176**:3336–3344.
7. Fujita Y, Miwa Y, Galinier A, Deutscher J: **Specific recognition of the *Bacillus subtilis gnt cis*-acting catabolite-responsive element by a protein complex formed between CcpA and seryl-phosphorylated HPr.** *Mol Microbiol* 1995, **17**:953–960.
8. Galinier A, Haiech J, Kilhoffer M-C, Jaquinod M, Stülke J, Deutscher J, Martin-Verstraete I: **The Bacillus subtilis crh gene encodes a HPr-like protein involved in carbon catabolite repression.** *Proc Natl Acad Sci U S A* 1997, **94**:8439–8444.
9. Singh KD, Schmalisch MH, Stülke J, Görke B: **Carbon catabolite repression in *Bacillus subtilis*: quantitative analysis of repression exerted by different carbon sources.** *J Bacteriol* 2008, **190**:7275–7284.
10. Landmann JJ, Werner S, Hillen W, Stülke J, Görke B: **Carbon source control of the phosphorylation state of the *Bacillus subtilis* carbon-flux regulator Crh *in vivo*.** *FEMS Microbiol Lett* 2012, **327**:47–53.
11. Görke B, Fraysse L, Galinier A: **Drastic differences in Crh and HPr synthesis levels reflect their different impacts on catabolite repression in *Bacillus subtilis*.** *J Bacteriol* 2004, **186**:2992–2995.
12. Jault J-M, Fieulaine S, Nessler S, Gonzalo P, Di Pietro A, Deutscher J, Galinier A: **The HPr kinase from *Bacillus subtilis* is a homo-oligomeric enzyme which exhibits strong positive cooperativity for nucleotide and fructose 1,6-bisphosphate binding.** *J Biol Chem* 2000, **275**:1773–1780.
13. Nessler S, Fieulaine S, Poncet S, Galinier A, Deutscher J, Janin J: **HPr kinase/phosphorylase, the sensor enzyme of catabolite repression in Gram-positive bacteria: structural aspects of the enzyme and the complex with its protein substrate.** *J Bacteriol* 2003, **185**:4003–4010.
14. Deutscher J, Küster E, Bergstedt U, Charrier V, Hillen W: **Protein kinase-dependent HPr/CcpA interaction links glycolytic activity to carbon catabolite repression in Gram-positive bacteria.** *Mol Microbiol* 1995, **15**:1049–1053.
15. Gösseringer R, Küster E, Galinier A, Deutscher J, Hillen W: **Cooperative and non-cooperative DNA binding modes of catabolite control protein CcpA from *Bacillus megaterium* result from sensing two different signals.** *J Mol Biol* 1997, **266**:665–676.

16. Kim J-H, Voskuil MI, Chambliss GH: **NADP, corepressor for the** *Bacillus* catabolite control protein CcpA. *Proc Natl Acad Sci USA* 1998, **95**:9590–9595.

17. Turinsky AJ, Grundy FJ, Kim J-H, Chambliss GH, Henkin TM: **Transcriptional activation of the** *Bacillus subtilis* **ackA gene requires sequences upstream of the promoter.** *J Bacteriol* 1998, **180**:5961–5967.

18. Presecan-Siedel E, Galinier A, Longin R, Deutscher J, Danchin A, Glaser P, Martin-Verstraete I: **Catabolite regulation of the** *pta* **gene as part of carbon flow pathways in** *Bacillus subtilis*. *J Bacteriol* 1999, **181**:6889–6897.

19. Tojo S, Satomura T, Morisaki K, Deutscher J, Hirooka K, Fujita Y: **Elaborate transcription regulation of the** *Bacillus subtilis* **ilv-leu operon involved in the biosynthesis of branched-chain amino acids through global regulators of CcpA, CodY and TnrA.** *Mol Microbiol* 2005, **56**:1560–1573.

20. Shivers RP, Sonenshein AL: *Bacillus subtilis* **ilvB operon: an intersection of global regulons.** *Mol Microbiol* 2005, **56**:1549–1559.

21. Shivers RP, Dineen SS, Sonenshein AL: **Positive regulation of** *Bacillus subtilis* **ackA by CodY and CcpA: establishing a potential hierarchy in carbon flow.** *Mol Microbiol* 2006, **62**:811–822.

22. Wünsche A, Hammer E, Bartholomae M, Völker U, Burkovski A, Seidel G, Hillen W: **CcpA forms complexes with CodY and RpoA in** *Bacillus subtilis*. *FEBS J* 2012, **279**:2201–2214.

23. Moir-Blais TR, Grundy FJ, Henkin TM: **Transcriptional activation of the** *Bacillus subtilis* **ackA promoter requires sequences upstream of the CcpA binding site.** *J Bacteriol* 2001, **183**:2389–2393.

24. Martin-Verstraete I, Stülke J, Klier A, Rapoport G: **Two different mechanisms mediate catabolite repression of the** *Bacillus subtilis* **levanase operon.** *J Bacteriol* 1995, **177**:6919–6927.

25. Nicholson WL, Park YK, Henkin TM, Won M, Weickert MJ, Gaskell JA, Chambliss GH: **Catabolite repression-resistant mutations of the** *Bacillus subtilis* **α-amylase promoter affect transcription levels and are in an operator-like sequence.** *J Mol Biol* 1987, **198**:609–618.

26. Krüger S, Gertz S, Hecker M: **Transcriptional analysis of** *bglPH* **expression in** *Bacillus subtilis*: **evidence for two distinct pathways mediating carbon catabolite repression.** *J Bacteriol* 1996, **178**:2637–2644.

27. Monedero V, Boël G, Deutscher J: **Catabolite regulation of the cytochrome c550-encoding** *Bacillus subtilis* **cccA gene.** *J Mol Microbiol Biotechnol* 2001, **3**:433–438.

28. Asai K, Baik SH, Kasahara Y, Moriya S, Ogasawara N: **Regulation of the transport system for C4-dicarboxylic acids in Bacillus subtilis.** *Microbiology (Reading, Engl.)* 2000, **146**(Pt 2):263–271.

29. Darbon E, Servant P, Poncet S, Deutscher J: **Antitermination by GlpP, catabolite repression via CcpA and inducer exclusion triggered by P∼GlpK dephosphorylation control Bacillus subtilis glpFK expression.** *Mol Microbiol* 2002, **43**:1039–1052.

30. Puri-Taneja A, Paul S, Chen Y, Hulett FM: **CcpA causes repression of the** *phoPR* **promoter through a novel transcription start site, PA6.** *J Bacteriol* 2006, **188**:1266–1278.

31. Grundy FJ, Waters DA, Takova TY, Henkin TM: **Identification of genes involved in utilization of acetate and acetoin in** *Bacillus subtilis*. *Mol Microbiol* 1993, **10**:259–271.

32. Weickert MJ, Chambliss GH: **Site-directed mutagenesis of a catabolite repression operator sequence in** *Bacillus subtilis*. *Proc Natl Acad Sci U S A* 1990, **87**:6238–6242.

33. Miwa Y, Nakata A, Ogiwara A, Yamamoto M, Fujita Y: **Evaluation and characterization of catabolite-responsive elements (cre) of** *Bacillus subtilis*. *Nucleic Acids Res* 2000, **28**:1206–1210.

34. Schumacher MA, Sprehe M, Bartholomae M, Hillen W, Brennan RG: **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* 2010, **39**:2931–2942.

35. Kamionka A, Bertram R, Hillen W: **Tetracycline-dependent conditional gene knockout in** *Bacillus subtilis*. *Appl Environ Microbiol* 2005, **71**:728–733.

36. Killmann H, Herrmann C, Torun A, Jung G, Braun V: **TonB of** *Escherichia coli* **activates FhuA through interaction with the β-barrel.** *Microbiology* 2002, **148**:3497–3509.

37. Miwa Y, Nagura K, Eguchi S, Fukuda H, Deutscher J, Fujita Y: **Catabolite repression of the Bacillus subtilis gnt operon exerted by two catabolite-responsive elements.** *Mol Microbiol* 1997, **23**:1203–1213.

38. Henkin TM: **The role of CcpA transcriptional regulator in carbon metabolism in** *Bacillus subtilis*. *FEMS Microbiol Lett* 1996, **135**:9–15.

39. Galinier A, Deutscher J, Martin-Verstraete I: **Phosphorylation of either Crh or HPr mediates binding of CcpA to the** *Bacillus subtilis* **xyn cre and catabolite repression of the xyn operon.** *J Mol Biol* 1999, **286**:307–314.

40. Moreno MS, Schneider BL, Maile RR, Weyler W, Saier MH: **Catabolite repression mediated by the CcpA protein in** *Bacillus subtilis*: **novel modes of regulation revealed by whole-genome analyses.** *Mol Microbiol* 2001, **39**:1366–1381.

41. Makita Y, Nakao M, Ogasawara N, Nakai K: **DBTBS: database of transcriptional regulation in** *Bacillus subtilis* **and its contribution to comparative genomics.** *Nucleic Acids Res* 2004, **32**:D75–D77.

42. Buescher JM, Liebermeister W, Jules M, Uhr M, Muntel J, Botella E, Hessling B, Kleijn RJ, Le Chat L, Lecointe F, Mäder U, Nicolas P, Piersma S, Rügheimer F, Becher D, Bessieres P, Bidnenko E, Denham EL, Dervyn E, Devine KM, Doherty G, Drulhe S, Felicori L, Fogg MJ, Goelzer A, Hansen A, Harwood CR, Hecker M, Hubner S, Hultschig C, et al: **Global network reorganization during dynamic adaptations of** *Bacillus subtilis* **metabolism.** *Science* 2012, **335**:1099–1103.

43. Baerends RJS, Smits WK, de Jong A, Hamoen LW, Kok J, Kuipers OP: **Genome2D: a visualization tool for the rapid analysis of bacterial transcriptome data.** *Genome Biol* 2004, **5**:R37.

44. Lulko AT, Buist G, Kok J, Kuipers OP: **Transcriptome analysis of temporal regulation of carbon metabolism by CcpA in** *Bacillus subtilis* **reveals additional target genes.** *J Mol Microbiol Biotechnol* 2007, **12**:82–95.

45. Kim J-H, Yang Y-K, Chambliss GH: **Evidence that** *Bacillus* **catabolite control protein CcpA interacts with RNA polymerase to inhibit transcription.** *Mol Microbiol* 2005, **56**:155–162.

46. Blencke H-M, Homuth G, Ludwig H, Mäder U, Hecker M, Stülke J: **Transcriptional profiling of gene expression in response to glucose in** *Bacillus subtilis*: **regulation of the central metabolic pathways.** *Metab Eng* 2003, **5**:133–149.

47. Ludwig H, Rebhan N, Blencke H-M, Merzbacher M, Stülke J: **Control of the glycolytic gapA operon by the catabolite control protein A in** *Bacillus subtilis*: **a novel mechanism of CcpA-mediated regulation.** *Mol Microbiol* 2002, **45**:543–553.

48. Zomer AL, Buist G, Larsen R, Kok J, Kuipers OP: **Time-resolved determination of the CcpA regulon of Lactococcus lactis subsp. cremoris MG1363.** *J Bacteriol* 2007, **189**:1366–1381.

49. Jacob S, Allmansberger R, Gärtner D, Hillen W: **Catabolite repression of the operon for xylose utilization from** *Bacillus subtilis* **W23 is mediated at the level of transcription and depends on a cis site in the xylA reading frame.** *Mol Gen Genet* 1991, **229**:189–196.

50. Zalieckas JM, Wray J, Ferson AE, Fisher SH: **Transcription–repair coupling factor is involved in carbon catabolite repression of the** *Bacillus subtilis* **hut and gnt operons.** *Mol Microbiol* 1998, **27**:1031–1038.

51. Inácio JM, Costa C, de Sá-Nogueira I: **Distinct molecular mechanisms involved in carbon catabolite repression of the arabinose regulon in** *Bacillus subtilis*. *Microbiology* 2003, **149**:2345–2355.

52. Choi S-K, Saier MH: **Regulation of** *sigL* **expression by the catabolite control protein CcpA involves a roadblock mechanism in** *Bacillus subtilis*: **potential connection between carbon and nitrogen metabolism.** *J Bacteriol* 2005, **187**:6856–6861.

53. Zalieckas JM, Wray LV, Fisher SH: **Expression of the** *Bacillus subtilis* **acsA gene: position and sequence context affect cre-mediated carbon catabolite repression.** *J Bacteriol* 1998, **180**:6649–6654.

54. Grundy FJ, Turinsky AJ, Henkin TM: **Catabolite regulation of** *Bacillus subtilis* **acetate and acetoin utilization genes by CcpA.** *J Bacteriol* 1994, **176**:4527–4533.

55. Krüger S, Hecker M: **Regulation of the putative bglPH operon for aryl-β-glucoside utilization in** *Bacillus subtilis*. *J Bacteriol* 1995, **177**:5590–5597.

56. Seidel G, Diel M, Fuchsbauer N, Hillen W: **Quantitative interdependence of coeffectors, CcpA and cre in carbon catabolite regulation of** *Bacillus subtilis*. *FEBS J* 2005, **272**:2566–2577.

57. Guérout-Fleury AM, Shazand K, Frandsen N, Stragier P: **Antibiotic-resistance cassettes for** *Bacillus subtilis*. *Gene* 1995, **167**:335–336.

58. Bertram R, Kostner M, Muller J, Vazquez Ramos J, Hillen W: **Integrative elements for Bacillus subtilis yielding tetracycline-dependent growth phenotypes.** *Nucleic Acids Res* 2005, **33**:e153.

59. Kunst F, Ogasawara N, Moszer I, Albertini AM, Alloni G, Azevedo V, Bertero MG, Bessieres P, Bolotin A, Borchert S, Borriss R, Boursier L, Brans A, Braun M, Brignell SC, Bron S, Brouillet S, Bruschi CV, Caldwell B, Capuano V, Carter NM, Choi SK, Codani JJ, Connerton IF, Danchin A, et al: **The complete**

genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature* 1997, **390**:249–256.

60. Küster E, Luesink EJ, de Vos WM, Hillen W: **Immunological crossreactivity to the catabolite control protein CcpA *Bacillus megaterium* is found in many Gram-positive bacteria.** *FEMS Microbiol Lett* 1996, **139**:109–115.

61. van Hijum SAFT, García de la Nava J, Trelles O, Kok J, Kuipers OP: **MicroPreP: a cDNA microarray data pre-processing framework.** *Appl Bioinformatics* 2003, **2**:241–244.

62. Long AD, Mangalam HJ, Chan BY, Tolleri L, Hatfield GW, Baldi P: **Improved statistical inference from DNA microarray data using analysis of variance and a Bayesian statistical framework. Analysis of global gene expression in Escherichia coli K12.** *J Biol Chem* 2001, **276**:19937–19944.

63. de Jong A, Pietersma H, Cordes M, Kuipers OP, Kok J: **PePPER: a webserver for prediction of prokaryote promoter elements and regulons.** *BMC Genomics* 2012, **13**:299.