

The functions of DNA methylation by CcrM in *Caulobacter crescentus*: a global approach

Diego Gonzalez¹, Jennifer B. Kozdon^{2,3}, Harley H. McAdams², Lucy Shapiro² and Justine Collier^{1,*}

¹Department of Fundamental Microbiology, University of Lausanne, Lausanne, CH 1015, Switzerland,

²Department of Developmental Biology, Stanford University, CA 94305, USA and ³Department of Chemistry, Stanford University, CA 94305, USA

Received October 7, 2013; Revised December 3, 2013; Accepted December 5, 2013

ABSTRACT

DNA methylation is involved in a diversity of processes in bacteria, including maintenance of genome integrity and regulation of gene expression. Here, using *Caulobacter crescentus* as a model, we exploit genome-wide experimental methods to uncover the functions of CcrM, a DNA methyltransferase conserved in most *Alphaproteobacteria*. Using single molecule sequencing, we provide evidence that most CcrM target motifs (GANTC) switch from a fully methylated to a hemi-methylated state when they are replicated, and back to a fully methylated state at the onset of cell division. We show that DNA methylation by CcrM is not required for the control of the initiation of chromosome replication or for DNA mismatch repair. By contrast, our transcriptome analysis shows that >10% of the genes are misexpressed in cells lacking or constitutively over-expressing CcrM. Strikingly, GANTC methylation is needed for the efficient transcription of dozens of genes that are essential for cell cycle progression, in particular for DNA metabolism and cell division. Many of them are controlled by promoters methylated by CcrM and co-regulated by other global cell cycle regulators, demonstrating an extensive cross talk between DNA methylation and the complex regulatory network that controls the cell cycle of *C. crescentus* and, presumably, of many other *Alphaproteobacteria*.

INTRODUCTION

Methylated bases can be found in the genomes of organisms from all three domains of life as well as in the genome

of viruses (1–4). Aberrant cytosine methylation is associated with a variety of human diseases, including neuropsychiatric disorders and cancers, exemplifying the critical functions of cytosine methylation, largely through its effects on the regulation of gene expression, in higher metazoans (5). The functions of DNA methylation have been poorly investigated for most of the bacterial kingdom (6). Bacterial DNA methyltransferases are mostly associated with endonucleases in restriction-modification systems, which are generally considered as a defense mechanism that bacteria use to identify and destroy differentially methylated foreign DNA (7–9). A number of ‘orphan’ DNA methyltransferases that are not associated with a cognate endonuclease have also been identified (7,10–12). The best studied examples are the DNA adenine methyltransferases Dam and CcrM. Dam methylates 5'-GATC-3' (hereafter called GATC) motifs in the genomes of a subset of *Gammaproteobacteria* (13), whereas CcrM methylates 5'-GANTC-3' (hereafter called GANTC) motifs in the genomes of many *Alphaproteobacteria*.

Dam has pleiotropic functions in *Gammaproteobacteria*: it is involved in the control of the initiation of chromosome replication (14–16), in the DNA mismatch repair (MMR) process (17,18) and in the regulation of gene expression (7,10). DNA microarrays were used to compare global transcription patterns in wild type, Δdam and Dam-over-expressing *Escherichia coli* cells (19–22). These studies revealed that hundreds of genes were misregulated in these two mutant strains. Similar effects were observed in *Salmonella enterica* (23). In addition, detailed studies in *E. coli* and *S. enterica* have led to the identification of a number of genes that show a Dam-dependent transcriptional phase variation, creating heterogeneity of expression rates within the bacterial population (11,24). Such mechanisms of regulation often involve transcription factors that compete with Dam for DNA binding and whose binding affinity is sensitive to the methylation

*To whom correspondence should be addressed. Tel: +41 21 692 56 10; Fax: +41 21 692 56 05; Email: justine.collier@unil.ch

state of the DNA at or near their binding site. In some cases, the presence of one or more under-methylated GATC motifs in a promoter is an indication that it is regulated by such a bistable epigenetic switch, although there also exists under-methylated GATC motifs that are not part of epigenetic mechanisms regulating gene expression (25,26). Dam is to date the only DNA adenine methyltransferase whose global effects on gene expression have been studied in bacteria.

The cell cycle-regulated methyltransferase CcrM was first identified in *Caulobacter crescentus*, an *Alphaproteobacterium* that divides asymmetrically into two morphologically different progeny cells: the daughter stalked cell that immediately initiates DNA replication and cell division and the daughter swarmer cell that first differentiates into a stalked cell before replication and cell division (Figure 1) (27,28). This bacterium initiates the replication of its chromosome only once per cell cycle and specifically in the stalked progeny or during the swarmer-to-stalked cell transition (29,30). The expression of CcrM is restricted to the late predivisive stage near the end of chromosome replication, which implies that GATC motifs on the chromosome are supposedly fully methylated (methylated on both DNA strands) in swarmer cells and then become hemi-methylated (methylated on one strand only) on DNA replication in stalked cells, and stay so until the end of the replication of the chromosome in pre-divisive cells (Figure 1) (30–32). Maintaining the chromosome constitutively fully methylated by over-expressing CcrM leads to morphological and cell cycle-related defects in *C. crescentus*, suggesting that the temporal regulation of the transition from the fully methylated to the hemi-methylated state has biological importance (31,33). In *C. crescentus*, the methylation state of GATC motifs has a direct effect on the transcription of at least four genes: the *ctrA* and *dnaA* genes encoding master regulators of the cell cycle (34,35), the *ftsZ* and *mipZ* genes required for cell division (36) and, most probably, the *ccrM* gene itself (37). The transition from the fully methylated to the hemi-methylated state of a GATC motif located in *cis* is sufficient to cause a small but significant change in the transcriptional activity of the *ctrA*, *dnaA* and *ftsZ* promoters. This supports a model according to which the change in GATC methylation state during DNA replication might function to co-ordinate gene expression changes with the progression of the replication fork, setting the mechanical basis for the regulation of the timing of multiple cell cycle events (32,34,35). We recently showed that an increase in *ftsZ* transcription is sufficient to restore the viability of $\Delta ccrM$ cells in rich medium, demonstrating that the direct activation of *ftsZ* transcription by DNA methylation is one of the most important functions of CcrM in *C. crescentus* (36). These rescued cells still exhibited a residual phenotype. For example, cells were elongated and straighter than wild-type cells and had shorter stalks. These observations suggested that more genes are probably regulated by CcrM-dependent methylation in *C. crescentus*. Interestingly, the *ccrM* gene is not essential anymore for the viability of *C. crescentus* when cells are cultivated in

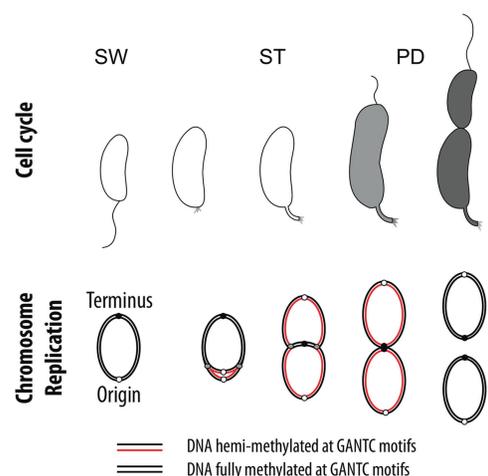


Figure 1. CcrM is a cell cycle-regulated DNA methyltransferase in *C. crescentus*. Upper panel: schematic of the *C. crescentus* cell cycle, showing the differentiation of a swarmer cell (SW) into a stalked cell (ST), which then turns into a pre-divisive cell (PD). Lower panel: schematic showing the changes in the methylation state of GATC motifs on the chromosome as a function of chromosome replication and cell cycle progression.

slow-growing conditions, such as minimal medium: these cells are nevertheless slightly elongated and grow slightly more slowly than wild-type cells (36). The physiological importance of GATC methylation is probably not restricted to *C. crescentus*, as the homolog of *ccrM* was shown to be essential for viability in fast-growing cultures of each of the species where this was tested (33,36,38–40).

Here, we assess the global influence of GATC methylation on transcription profiles in *C. crescentus* by analyzing the transcriptome of a $\Delta ccrM$ strain and of a strain in which CcrM is constitutively expressed so that the chromosome remains in the fully methylated state throughout the cell cycle. We show that hundreds of genes are differentially expressed in both mutant strains compared with the wild type. Genes whose transcription could directly be affected by GATC methylation are predicted on the basis of their differential expression in the mutant strains and through bioinformatic analysis of GATC conservation in promoters. Our findings reveal a strong link between CcrM-dependent methylation and cell cycle control in *C. crescentus* that is also likely to exist in many other *Alphaproteobacteria*.

MATERIALS AND METHODS

Phylogenetic analysis

NCBI whole bacterial proteomes were searched for proteins with homology to the CcrM protein from *C. crescentus* strain CB15 (blastp e-value < 1e-10) (41); one strain was selected per genus and the protein with the lowest blastp e-value was used for each individual strain. On the basis of a MUSCLE alignment of the sequences (42,43), the subtype β N6-adenine-methyltransferases were selected, using the motif IV [DPPY \neq SPP(Y/F)] and the motif X [T(Q/E/D)KP \neq AXFP] to discriminate between

N6-methyltransferases and N4-methyltransferases (44). A tree was calculated with FastTree (45,46) using the WAG matrix for amino-acid substitutions to get a general idea of the topology (Supplementary Figure S1B). Our final set comprised all sequences of CcrM-homologs, which had a conserved C-terminal domain, i.e. sequences from *Alphaproteobacteria*, *Helicobacter* and some *Chloroflexi*, with the addition of *Haemophilus influenzae* HinI [from REBASE (8)]; on the basis of the FastTree tree, two sequences were added as outgroups (from *Streptococcus mitis* and *Stigmatella aurantiaca*). A new alignment was made with MUSCLE and curated with GBLOCKS 0.91b (47) with default options and gaps allowed at all positions. The final trees were calculated on the basis of the curated alignment using RAxML (48) or MrBayes version 3.2.1 with 200 000 iterations and 2 × 4 chains (49,50) with the WAG matrix for amino-acid substitution and the Gamma model of substitution rate heterogeneity. Branches with support values <0.5 were left unresolved (Supplementary Figure S1C).

Distribution of pentanucleotides in the genomes of *Alphaproteobacteria*

Bacterial genome sequences were downloaded from the NCBI repository. GANTC motifs and other pentanucleotides were counted in the complete genome sequences (.fna file), in the protein coding sequences (.ffn file) and in the RNA coding sequences (.frn file). The number of GANTC motifs and other pentanucleotides in intergenic sequences was calculated as follows: number in the genome – (number in protein coding sequences + number in RNA coding sequences). The frequency of each nucleotide in each complete genome sequence was calculated as the ratio between the number of each nucleotide and the total number of nucleotides of the whole genome. The expected number of a given pentanucleotide with a central variable N in a genome was calculated as $\text{freq}(\text{A}) \cdot \text{freq}(\text{C}) \cdot \text{freq}(\text{G}) \cdot \text{freq}(\text{A}) \cdot \text{freq}(\text{N})$ (genome length), where $\text{freq}(\text{A})$ is the frequency of As in the genome, for example, and ‘genome length’ is the total number of nucleotides in the considered genome. The expected number of GANTC motifs and other pentanucleotides in protein coding, RNA coding and non-coding sequences was calculated as (subsequence length/genome sequence length) × (total pentanucleotide number), where ‘subsequence length’ is the length of the protein coding, RNA coding and non-coding sequences, respectively, and the total pentanucleotide number is the total number of a given pentanucleotide found in the considered genome. As ‘Other *Proteobacteria*’ we used the genomes of the following: *Neisseria gonorrhoeae* FA1090, *Burkholderia mallei* ATCC 23344, *Nitrosomonas europaea* ATCC 19718, *Myxococcus xanthus* DK1622, *Bdellovibrio bacteriovorus* HD100, *Desulfovibrio vulgaris* DP4, *Helicobacter pylori* F32, *Nautilia profundicola* AmH, *Sulfurospirillum deleyianum* DSM 6946, *E. coli* str. K-12 substr. DH10B, *Pseudomonas aeruginosa* PAO1, *Vibrio cholerae* O1 biovar El Tor str. N16961.

Growth conditions and bacterial strains

Three *C. crescentus* strains were used in this study: (i) Strain NA1000 (CB15N) is a synchronizable derivative of the wild-type CB15 strain (51,52); (ii) Strain JC362 (53) is a derivative of the NA1000 strain expressing a second copy of the *ccrM* gene under the control of the endogenous *lacZ* promoter (*Plac::ccrM*), which is constitutively active (31); and (iii) Strain JC1149 (36) is a derivative of the NA1000 strain with a deletion of the *ccrM* gene (ΔccrM). Cells were cultivated in M2G minimal medium at 28°C. Experiments were all performed using unsynchronized cell populations because well-synchronized JC1149 populations could not be isolated, probably due to the slightly abnormal morphology of the ΔccrM cells (36).

Single molecule real time sequencing analysis

Genomic DNA was extracted from cells cultivated to exponential phase using the Qiagen Puregene Yeast/Bact DNA extraction kit with a second ethanol precipitation step after resuspension; the DNA was finally dissolved in 10 mM Tris-HCl, pH 8. For each DNA sample, Single Molecule Real Time (SMRT) Bell libraries with a 2-kb insert size were sequenced in 6 SMRT cells and the data were analyzed with tools provided in the Pacific Biosciences SMRT portal using the NC_011916 NA1000 genome (NCBI) as a reference. Mean depth coverage was 182 for strain NA1000, 187 for strain JC362 and 37 for strain JC1149. SMRT confidence scores, defined as $-10 \log(\text{p-value})$, for methylated adenines were typically high in GANTC motifs for the NA1000 and JC362 strains (median is 115 and 145, respectively) and low for the JC1149 strain (median is 3). For adenines in all GANTC motifs in the chromosome, the inter-pulse duration (IPD) ratio was calculated by dividing the average measured IPD for the nucleotide by the IPD value predicted in an *in silico* model for non-methylated DNA. IPD ratios were calculated for all adenines in GANTC contexts on both DNA strands. We used the average IPD ratio for adenines at GANTC motifs in the chromosome of strain JC362 to approximate the value for adenines fully methylated in chromosomes in the whole population (= 7.68); the theoretical average IPD ratio for adenines hemi-methylated in chromosomes in the whole population would then be $(7.68 + 1)/2 = 4.34$. An IPD of 2.4, corresponding to an adenine fully methylated in chromosomes from 20% of the population or hemi-methylated in 40% of the population was chosen as a threshold to consider that a GANTC motif was under-methylated.

Preparation of RNA samples for transcriptome or qRT-PCR analysis

All strains were cultivated in triplicates and independent cultures (NA1000 versus JC1149) or independent biological samples (NA1000 versus JC362) were used for subsequent transcriptome and quantitative real-time polymerase chain reaction (qRT-PCR) analysis. Two milliliters of cell culture aliquots were pelleted, frozen in liquid nitrogen and stored at –80°C. Pellets were

homogenized in 1 ml of Trizol reagent and incubated at 65°C for 10 min. Two hundred microliters of chloroform were added and, after 3 min at room temperature, the extracts were centrifuged for 15 min at maximum speed. Five hundred fifty microliters of 100% ethanol were added to 550 µl of aqueous phase from the Trizol treatment and RNA was purified using the Ambion PureLink RNA mini-kit (standard protocol without the lysis step). RNA was eluted in 40 µl of H₂O, treated in a 60-µl reaction with 10 µl of Invitrogen DNase I Amp Grade for 30 min at room temperature.

Transcriptome analysis

Custom microarrays were designed by Nimblegen using *C. crescentus* NA1000 protein and small RNA coding sequences. When possible, three different probes of ~60 bp were designed per gene. Microarrays had a 4-plex format (four sub-arrays per chip); each probe was printed in randomly positioned triplicates in each of the sub-arrays. One sub-array was used per biological sample. RNA samples were prepared as described above and 6 µl of EDTA 25 mM (Invitrogen) were added for a 10-min incubation at 65°C to inactivate the DNase I. Sixty microliters of lysis buffer from the PureLink RNA kit with 1% 2-mercaptoethanol and 60 µl of 100% ethanol were added to the reaction and the RNA was purified again using the PureLink standard protocol (without the lysis step). The RNA was resuspended in 32 µl, from which 2 µl were used for quality check via Bioanalyzer and quantification with Nanodrop. cDNA was synthesized from 5 µg of RNA using the Roche double-stranded cDNA synthesis system (cat. 11'117'831'001) and the protocol described in the 'Roche cDNA Synthesis System for use with NimbleGen Gene Expression Microarrays' technical note. Promega random hexamers at the same concentration were used instead of the oligo(dT)₁₅. The double-stranded cDNA was precipitated with isopropanol, purified and resuspended in 40 µl according to the protocol. The whole purified double-stranded cDNA (0.5–1 µg cDNA) was used as input for the standard Klenow-based NimbleGen Cy3 labelling protocol. Two micrograms of labelled cDNA were hybridized 16 h in one sub-array using the standard Nimblegen protocol for assembly, hybridization and washing. Arrays were scanned in an Agilent High-Resolution Microarray scanner at 2 µm resolution. Normalized intensities were extracted from the pictures with the NimbleScan 2.6 software using the standard workflow and default options. The Robust Multi-array Average (RMA) module of the software was used to normalize the intensity values for all chips. The statistical analysis 'mutant versus wild type' was performed with the R limma package on the basis of the intensity values of all biological replicates for the mutant and wild-type strains. Genes with an adjusted $P < 0.01$ were considered significantly up- or downregulated. The logFC value given by the limma function was used as the log-ratio shown in Tables.

Flow cytometry analysis, methods to determine spontaneous mutation rates and methods for qRT-PCR analysis are described in Supplementary Information.

RESULTS

CcrM homologs are conserved in most Alphaproteobacteria

The functions of the CcrM family of N6-adenine methyltransferases have been studied in a limited number of *Alphaproteobacteria*, including *C. crescentus*, *Agrobacterium tumefaciens*, *Rhizobium meliloti* and *Brucella abortus* (33,38–40). To assess the phylogenetic conservation of CcrM homologs beyond these species and more thoroughly than before (54), we searched for proteins with a significant similarity to the *C. crescentus* CcrM protein in all available bacterial proteomes (Supplementary Table S1). We found that all *Alphaproteobacteria* except *Rickettsiales* and *Magnetococcales* (55) had a CcrM homolog with a high similarity score (Figure 2A). As the target recognition domain of CcrM is conserved in each of these proteins, these enzymes most likely also methylate the adenine in GANTC motifs. The only exceptions were the bacteria belonging to the SAR11 clade (e.g. *Pelagibacter ubique* and related species) that do have a CcrM homolog although they have sometimes been classified as *Rickettsiales* (56,57), and the cicada endosymbiont *Candidatus Hodgkinia cicadicola* Dsem (*Rhizobiales*) whose genome does not encode a CcrM homolog. Outside *Alphaproteobacteria*, the genomes of sporadic species and clades across the phylogeny encode a protein closely related to CcrM. *ccrM* homologs were only found in close proximity to an endonuclease gene on a chromosome in several *Epsilonproteobacteria* or *Gammaproteobacteria* (Supplementary Table S2), suggesting that CcrM homologs are orphan N6-adenine methyltransferases in all *Alphaproteobacteria*.

A phylogenetic tree was built using the sequences of proteins with a high similarity to the *C. crescentus* CcrM protein, belonging to the β-class of N6-adenine methyltransferases and comprising a conserved additional C-terminal domain specific to the CcrM family (Supplementary Figure S1). The most parsimonious scenario compatible with the phylogenetic tree is that the gene encoding the ancestor of the CcrM homolog entered the genome of an ancestral *Alphaproteobacterium* after the divergence of *Magnetococcales* and *Rickettsiales* and that it was thereafter transmitted vertically to the descendent species (Figure 2A). The pervasiveness of vertical transmission suggests that the CcrM homolog is part of the core genome and plays a critical physiological role in the large and diverse terminal clade of the alphaproteobacterial tree.

GANTC motifs are under-represented in the genomes of Alphaproteobacteria but over-represented in intergenic regions

Previous studies indicated that GANTC motifs are less frequent in the genome of *C. crescentus* than expected in a random sequence of nucleotides and that these motifs are over-represented in intergenic regions; these biases might be connected to the biological role that DNA methylation plays in controlling transcription or other

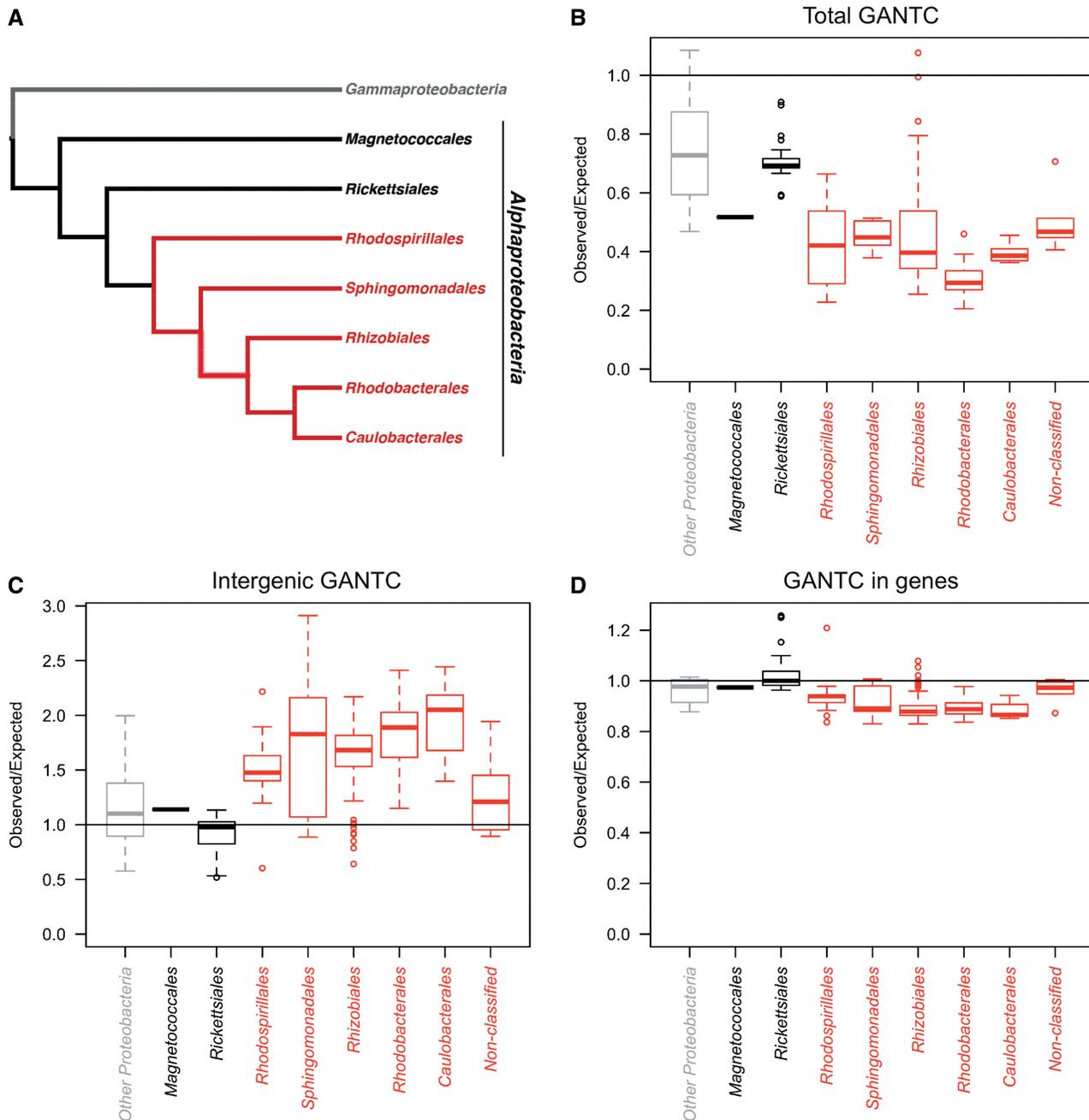


Figure 2. Conservation of CcrM in *Alphaproteobacteria* and distribution of GANTC motifs in their genomes. (A) Schematic showing the conservation of CcrM in different orders of *Alphaproteobacteria*. Panels (B), (C) and (D) show the observed-over-expected GANTC ratios using the whole genome (B), the intergenic sequences (C) and the coding sequences (D) of sequenced bacterial species from different orders of *Alphaproteobacteria*. 'Other Proteobacteria' correspond to a control of 10 mixed other *Proteobacteria*. The expected number of motifs was calculated for each genome taking the nucleotide composition into account. The limits of the box represent the first and third quartiles and the bold line the median of the distribution. In the whole figure, orders of bacterial species containing a CcrM homolog are indicated in red.

cellular processes in *C. crescentus* (35,58,59). To explore this possibility more systematically, we analyzed the frequency and distribution of GANTC motifs in all available genomes of *Alphaproteobacteria* and of a control set of other *Proteobacteria*. We found that GANTC motifs were on average at least 2-fold less frequent than expected in the genomes of all orders of *Alphaproteobacteria*, including *Magnetococcales* but excluding *Rickettsiales* (Figure 2B). In the group of other *Proteobacteria*, this

bias was largely attenuated. Notably, we also observed that GANTC motifs showed a distribution bias between protein coding and intergenic sequences in each order comprising bacterial species encoding a CcrM homolog (Figure 2C and D), being on average >1.5-fold over-represented in intergenic regions and slightly under-represented in coding regions. These biases were largely attenuated in the group of other *Proteobacteria* lacking a CcrM homolog. Overall, these findings suggest that

selective pressure limits the overall frequency of GANTC motifs in protein coding sequences on the genomes of bacteria encoding a CcrM homolog, whereas they are tolerated or favoured in intergenic sequences.

To check whether these biases were specific to GANTC motifs, we also analyzed the frequency and distribution of the 24 pentanucleotides of the same structure as GANTC (each of the four bases present one time, one central variable N) (Supplementary Figure S2). Surprisingly, another one of these pentanucleotides, CTNAG, presented similar strong frequency biases, not only in *Alphaproteobacteria* encoding CcrM homologs, but also in the control group of other *Proteobacteria* (Supplementary Figure S3). Thus, a bias in the frequency and distribution of pentanucleotides of that structure can be found in bacterial genomes independently of their capacity to be methylated by CcrM homologs, indicating that the biased distribution of GANTC motifs in the genomes of CcrM-encoding *Alphaproteobacteria* might also be explained by other factors.

The methylation of GANTC motifs in the genome of *C. crescentus* is strictly dependent on CcrM

The temporal regulation of the methylation of GANTC motifs seems important in a variety of *Alphaproteobacteria*, since cells that constitutively express homologs of CcrM often appear abnormal (31,34,39,40,60). A system of DNA methylation probes distributed at a dozen positions along the *C. crescentus* chromosome previously suggested that newly replicated GANTC motifs stay hemi-methylated until the pre-divisional stage of the *C. crescentus* cell cycle, independently of the moment when they got replicated, due to the strict temporal regulation of CcrM concentration during the cell cycle (30,31). According to this model, GANTC motifs located close to the chromosomal origin should stay hemi-methylated for a longer period during the cell cycle than motifs located close to the chromosomal terminus, as they are replicated earlier during the cell cycle. To confirm this model for each of the GANTC motifs on the *C. crescentus* chromosome, we made use of the recently developed SMRT sequencing technology of PacBio on genomic DNA extracted from wild-type, $\Delta ccrM$ (36) or CcrM-over-expressing (31) cells cultivated to exponential phase. This technology measures the time needed for a DNA polymerase to incorporate a nucleotide on a template DNA (IPD), which varies depending on the methylation state of the nucleotide found on the template DNA (61). This method is powerful at detecting N6-methyladenines on bacterial chromosomes (62,63) and we used it to estimate the proportion of cells in the bacterial population in which a given GANTC motif is methylated on either strand. We observed a positive correlation between the proximity of a GANTC motif to the chromosomal origin and its probability to be hemi-methylated in a mixed population of the wild-type strain (Figure 3A). We estimated that motifs located close to the chromosomal origin were, on average, hemi-methylated in ~80% of the cells, while sites located close to the chromosomal terminus were hemi-methylated in <10% of the

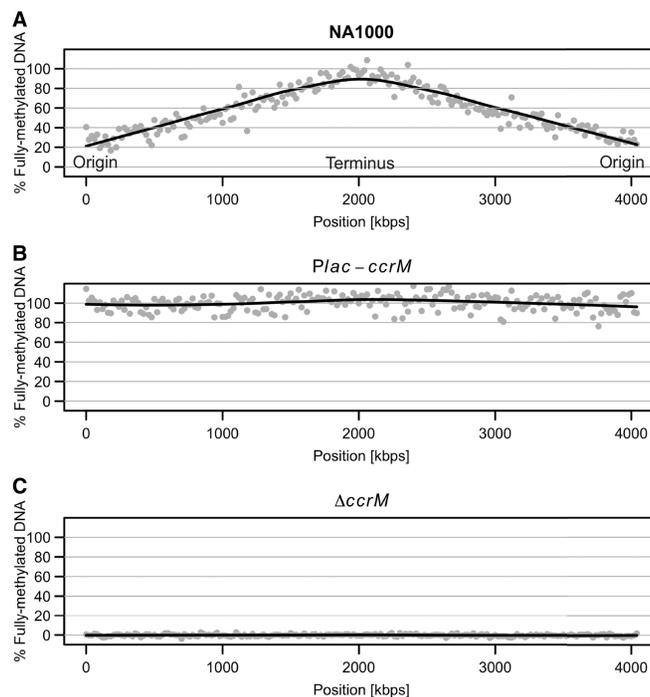


Figure 3. CcrM-dependent methylation of GANTC motifs in the genome of *C. crescentus*. Predicted methylation state of the DNA, as % of the chromosomal population being fully methylated at a particular locus, along the chromosome of *C. crescentus* (A) NA1000, (B) *Plac::ccrM* [JC362], (C) $\Delta ccrM$ [JC1149]). We used the SMRT portal to detect methylated motifs in the chromosome of *C. crescentus* and extracted the IPD ratios for adenines in all GANTC motifs (both strands). We then calculated the average IPD ratio in 20 kb windows along the chromosome and plotted the values according to the chromosomal co-ordinates (the origin of replication is situated at the junction between 4000 kb and 1 kb). The average IPD ratio for strain JC362 was equalled to 100% and the theoretical value for hemi-methylated DNA [(value for fully methylated DNA + 1)/2] as 0%; the curve was fitted with the R loess function.

cells. In the $\Delta ccrM$ population, we found that none of the GANTC motifs were fully methylated (Figure 3C), confirming that the methylation of each GANTC motif on the *C. crescentus* chromosome is dependent on CcrM. In the CcrM-over-expressing population, we observed that nearly all GANTC motifs were fully methylated in the whole population (Figure 3B), showing that the fully methylated to hemi-methylated switch that takes place for most of the GANTC motifs on the *C. crescentus* chromosome is dependent on the precise temporal regulation of the concentration of CcrM during the cell cycle. These results confirmed, at the genomic scale and in a quantitative manner, the model previously proposed by Zweiger *et al.* for most GANTC motifs (31).

Interestingly, through our SMRT sequencing analysis, we found 35 exceptions to this rule: 24 GANTC motifs were strongly under-methylated and 11 additional GANTC motifs were asymmetrically methylated (one strand being more often under-methylated than the other one) in a population of unsynchronized wild-type *C. crescentus* cells (Supplementary Table S3). A recent study using a synchronized population of *C. crescentus* cells demonstrated that most of these motifs remain

under-methylated throughout the whole cell cycle (64). Here, we show that only three of these motifs remained under-methylated in cells over-expressing CcrM, while the other 32 GANTC motifs were efficiently methylated on both DNA strands.

Overall, these results demonstrate that the $\Delta ccrM$ and CcrM-over-expressing strains are accurate tools for exploring the functions of GANTC methylation in *C. crescentus* and the functions of the methylation switch that occurs at most GANTC motifs over the course of the cell cycle of *C. crescentus*.

Control of chromosome replication and spontaneous mutation rate in the absence of GANTC methylation in *C. crescentus*

The orphan adenine methyltransferase Dam was shown to be required for the replication of one of the two chromosomes in *V. cholerae* (15,16) and to be involved in the control of the initiation of chromosome replication in *E. coli* (14,65). Several GANTC motifs can be found in the origin of replication of *C. crescentus* and of other closely related species (66) and it was previously shown that cells over-expressing CcrM accumulate additional chromosomal copies (31). We thus hypothesized that CcrM might be required for the regulation of chromosome replication in *C. crescentus*, ensuring that it occurs only once per cell cycle (30,31). To test this hypothesis, we used flow cytometry to compare the DNA content of cells in wild-type and $\Delta ccrM$ populations treated with rifampicin. Rifampicin blocks the initiation of DNA replication, but not the elongation of DNA replication, so that wild-type *C. crescentus* cells accumulate either one or two complete chromosomes after rifampicin treatment (Supplementary Figure S4A). We observed that $\Delta ccrM$ populations cultivated in minimal medium had a DNA content profile similar to wild-type populations (Supplementary Figure S4B). We concluded that CcrM is neither required for the control of the initiation of chromosome replication, nor for the correct segregation of replicating chromosomes before cell division, unlike Dam in several enterobacteria (7,14).

In addition to its regulatory roles, Dam plays an important role in DNA MMR in a subset of *Gammaproteobacteria* (17,18): the hemi-methylated state of GATC motifs behind the replication fork provides critical information to identify the newly replicated strand, more likely to contain errors that need to be repaired (18). This raised the possibility that the methylation of GANTC motifs by CcrM may similarly be used for DNA repair in *C. crescentus*. We therefore compared the spontaneous mutation rates of the wild-type and $\Delta ccrM$ strains by calculating the frequency of rifampicin resistant mutants arising in each population and found that they were similar (Supplementary Figure S5). This result shows that methylation of GANTC motifs by CcrM is not needed for the MMR mechanism in *C. crescentus*.

Cumulatively, these observations indicate that CcrM in *Alphaproteobacteria* does not share the most conserved functions of Dam in *Gammaproteobacteria*. This is consistent with the different evolutionary histories of the two

proteins. Furthermore, we provide evidence below that the impact of CcrM-dependent methylation on the regulation of gene expression in *C. crescentus* also differs from the known impact of Dam-dependent methylation in *E. coli* (19–21).

Alterations in gene expression in the absence of GANTC methylation in *C. crescentus*

To evaluate the impact of CcrM-dependent DNA methylation on global gene expression profiles in *C. crescentus*, we used custom-made DNA microarrays to compare the transcription levels in wild-type and $\Delta ccrM$ populations cultivated to exponential phase in minimal medium. We found that out of the 3932 genes in the *C. crescentus* NA1000 genome for which probes were designed, the expression levels of 388 genes were significantly (corrected Student's test $P < 0.01$) changed in the $\Delta ccrM$ strain compared with the wild-type strain, with 152 of the 388 genes being affected >2 -fold (Supplementary Table S4). For 17 genes, the effects were verified by qRT-PCR and were essentially consistent with the microarray results (Supplementary Figure S6 and Table 1). We classified each gene strongly misregulated in $\Delta ccrM$ cells according to the known or predicted function of the protein that it encodes (NCBI single letter COG) (Figure 4A). We found that genes encoding proteins involved in DNA replication, recombination or repair were significantly enriched among the genes that were upregulated in the absence of CcrM, while genes encoding proteins involved in cell motility were significantly enriched among the genes that were downregulated in the absence of CcrM (Figure 4A).

A subset of the transcriptional effects observed in the $\Delta ccrM$ strain was probably directly due to the absence of methylation at promoter regions containing GANTC motifs (140 misregulated genes contained minimum one GANTC motif in their promoter region). To assess whether the transcriptome analysis was likely to reveal such direct effects of GANTC methylation on transcription, we calculated the probability that significantly misregulated genes contained a GANTC motif in their promoter region. Because many cell cycle-regulated genes are controlled by relatively long promoter regions in *C. crescentus* (67), we looked for the presence of GANTC motifs up to 200 bp upstream of the annotated translational start codon of each gene. We found that the probability to find a GANTC motif was ~ 1.8 -fold higher than in a random set of genes (Fisher's exact test $P < 0.05$) (Figure 5A). This value rose to ~ 2.5 -fold when considering only the genes misregulated >2 -fold (Fisher's exact test $P < 0.05$). There was a clear enrichment in genes containing a GANTC motif in their promoter region among the genes that were the most significantly up- or downregulated in the $\Delta ccrM$ strain (Figure 5B and C). The conservation of a GANTC motif in the promoter regions of homologous genes in closely related bacteria has been previously used as an indicator that the methylation of the motif may play a role in regulating the activity of the promoter (36). We thus asked whether genes strongly misregulated in the $\Delta ccrM$ strain had a higher probability of containing a 'conserved' GANTC

Table 1. Essential genes containing GANTC motifs in their promoter region and significantly misregulated in $\Delta ccrM$ cells compared with wild-type *C. crescentus* cells

| Gene name | Known or predicted function | Log-ratio MA | Log-ratio qRT-PCR | CtrA regulon | DnaA regulon | GcrA regulon | GANTC conservation |
|--|---|--------------------|-------------------|--------------|--------------|--------------|--------------------|
| Essential genes repressed in $\Delta ccrM$ | | | | | | | |
| <i>CCNA_00390</i> | ADP-heptose—LPS heptosyltransferase | -0.85 | -0.72 | No | No | No | 4 |
| <i>CCNA_01104</i> | Glycosyltransferase | -0.77 | ND | No | No | No | 2 |
| <i>CCNA_01275</i> | Methylenetetrahydrofolate dehydrogenase (NADP+) | -0.84 | ND | No | No | No | 1 |
| <i>CCNA_01450</i> | Single-stranded DNA-specific exonuclease RecJ | -1.03 | -1.20 | No | No | Yes | 4 |
| <i>CCNA_01535</i> | Single-stranded DNA binding protein | -0.56 | ND | No | Yes | No | 2 |
| <i>CCNA_01590</i> | NAD-dependent DNA ligase | -1.23 | ND | No | No | No | 3 |
| <i>CCNA_01651</i> | DNA gyrase subunit A | -1.14 | -1.01 | No | No | Yes | 5 |
| <i>CCNA_01776</i> | Ribonuclease D | -0.68 | -0.92 | No | No | No | 5 |
| <i>CCNA_02052</i> | Topoisomerase IV subunit B | -1.97 | ND | No | No | Yes | 3 |
| <i>CCNA_02086</i> | Cell division protein FtsN | -1.09 | ND | No | No | Yes | 5 |
| <i>CCNA_02127</i> | Oxacillin resistance-associated protein FmtC | -1.93 | ND | No | No | Yes | 2 |
| <i>CCNA_02208</i> | Thymidylate synthase | -1.03 | ND | No | No | No | 3 |
| <i>CCNA_02246</i> | Division plane positioning ATPase MipZ | -0.52 ^a | -2.54 | Yes | Yes | Yes | 5 |
| <i>CCNA_02256</i> | Mg2+ transporter MgtE | -0.83 | -1.14 | No | No | No | 5 |
| <i>CCNA_02389</i> | Glucosamine-1-phosphate acetyltransferase/UDP-N-acetylglucosamine pyrophosphorylase | -0.87 | -0.69 | No | No | No | 5 |
| <i>CCNA_02623</i> | Cell division protein FtsZ | -1.44 | -1.75 | Yes | Yes | No | 5 |
| <i>CCNA_02679</i> | Hypothetical protein | -1.07 | ND | No | No | No | 3 |
| <i>CCNA_03607</i> | Ribonucleoside-diphosphate reductase subunit alpha NrdA | -0.47 | -0.77 | No | Yes | No | 5 |
| Essential genes activated in $\Delta ccrM$ | | | | | | | |
| <i>CCNA_00379</i> | Thiol:disulfide interchange protein DsbA | 0.69 | 0.53 | No | No | No | 5 |
| <i>CCNA_00845</i> | Antitoxin protein RelB-1 | 0.82 | ND | No | No | No | 1 |
| <i>CCNA_00893</i> | Transcriptional regulatory protein | 0.53 | ND | No | No | No | 0 |
| <i>CCNA_01946</i> | Tyrosyl-tRNA synthetase | 0.89 | 0.28 | No | No | No | 4 |
| <i>CCNA_02635</i> | Cell division protein FtsW | 0.65 | ND | Yes | No | No | 2 |
| <i>CCNA_03130</i> | Cell cycle response regulator CtrA | 0.46 | 0.34 | Yes | No | Yes | 5 |
| <i>CCNA_03471</i> | 4-hydroxy-3-methylbut-2-enyl diphosphate reductase | 0.74 | 0.23 | No | No | No | 4 |
| <i>CCNA_03766</i> | 16S rRNA processing protein RimM | 1.11 | ND | No | No | No | 3 |
| <i>CCNA_03820</i> | Outer-membrane lipoproteins carrier protein | 1.09 | ND | No | No | No | 3 |

Genes were considered essential according to (67). The log-ratios indicate the expression change observed using microarrays (MA) or qRT-PCR comparing wild-type and $\Delta ccrM$ cells. ND indicates that the log-ratio was not measured by qRT-PCR for that gene.

^aIndicates that unspecific low-signal probes were unfortunately designed to detect *mipZ* by microarrays. The CtrA and DnaA direct regulons were described in (68) and (69), respectively. The GcrA regulon was described in (70). The 'GANTC conservation' column indicates the number of species in which a GANTC motif is found in the promoter region (200 bp upstream of their translational start codon) of homologous genes among five bacterial species closely related to *C. crescentus* (*C. crescentus*, *C. segnis*, *C. K31*, *Phenylobacter zincinum*, *Brevundimonas subvibrioides*).

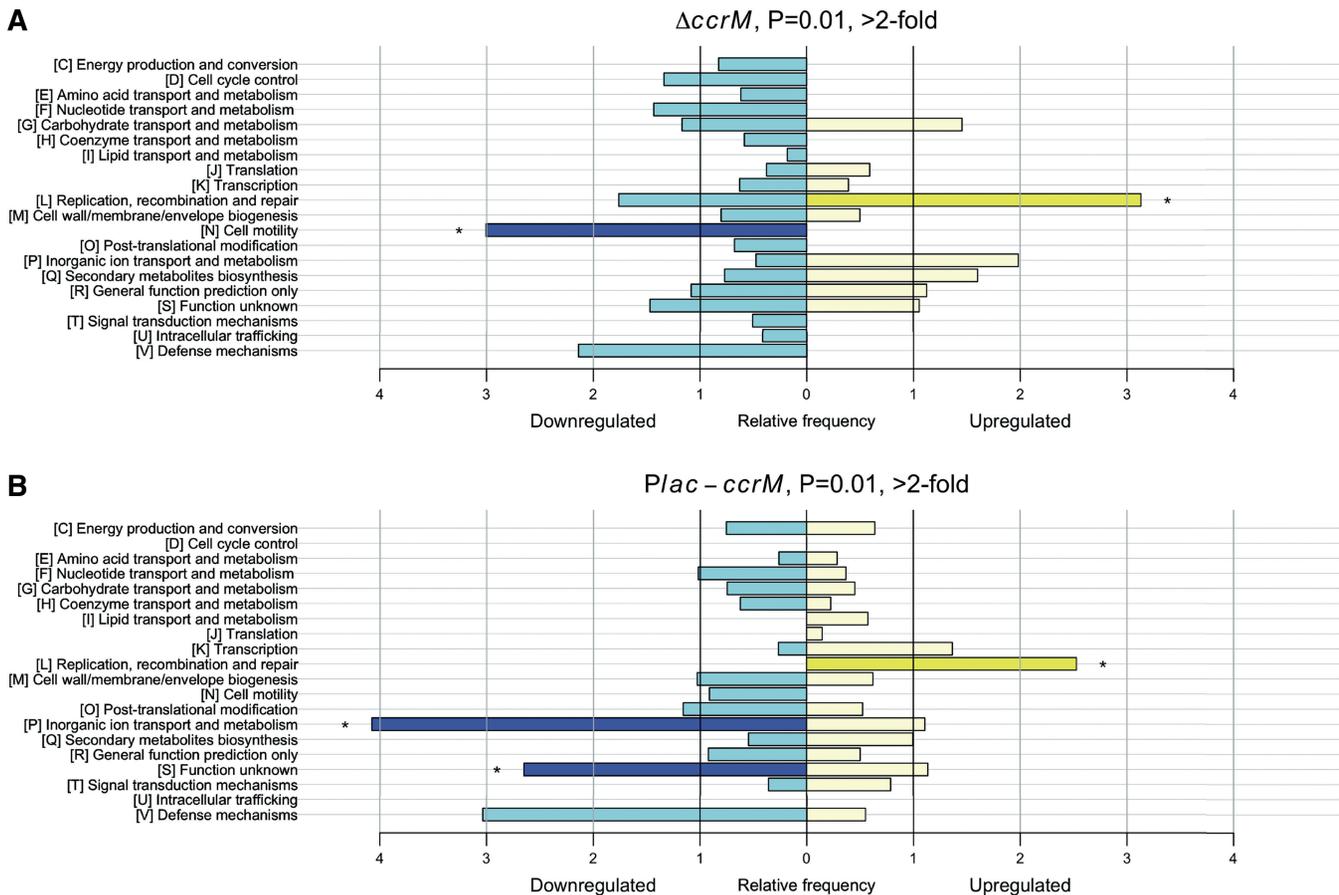


Figure 4. Relative frequency of functional categories among genes found as strongly misregulated during the transcriptome analysis. Genes whose expression changed > 2 -fold in the $\Delta ccrM$ (A) and the CcrM-over-expressing (B) strains were selected. Blue and yellow bars represent the set of genes whose expression is downregulated and upregulated, respectively, in the corresponding strain. Stars, dark blue and bright yellow colours indicate a significant over-representation ($P < 0.05$, Fisher's exact test).

motif, compared with a random set of genes. We found that genes containing a conserved GANTC motif were more frequent among the misregulated genes than expected randomly. Genes containing a GANTC motif that was conserved in seven closely related species were, for example, five times more frequent among the most downregulated genes than expected randomly (Fisher's exact test $P < 0.05$) (Figure 5D). This result indicates that promoters that contain conserved GANTC motifs are more likely to be affected by the absence of CcrM. Thus, we predict that genes that are significantly misregulated in the $\Delta ccrM$ strain and that are under the control of a promoter region that contains a conserved GANTC motif have a high probability of being directly regulated by the methylation state of their promoter region (Supplementary Table S4). The *ctrA*, *ftsZ* and *mipZ* genes, which were shown to be directly regulated by DNA methylation (35,36), fit these criteria (Table 1). Additional interesting candidates from our study include, for example, *CCNA_01450* (*recJ*), *CCNA_01651* (*gyrA*), *CCNA_02283* (annotated as an endonuclease), *CCNA_02389* (UDP-N-acetylglucosamine pyrophosphorylase) and *CCNA_02726* (annotated as an acetyltransferase) (Supplementary Table S4).

To evaluate the major physiological consequences of the transcriptional changes observed in the absence of CcrM-mediated methylation, we first examined the set of genes that were reported to be essential for the viability of *C. crescentus* (67) and that were significantly (corrected Student's test $P < 0.01$) downregulated: we found that 31 essential genes (Supplementary Table S4), including four genes encoding cell division proteins (*FtsZ*, *MipZ*, *FtsN* and *FtsE*; *ftsX*, encoding *FtsX*, is co-transcribed with *ftsE* and misregulated at $P < 0.05$) and seven genes encoding proteins involved in DNA replication or repair or in the maintenance of chromosome topology (*Ssb*, *RecJ*, ligase, *GyrA*, *Topo IV*, thymidylate synthase and *NrdA*). Each of these 18 essential genes contained a relatively conserved GANTC motif in their promoter region (Table 1), suggesting that their expression might be directly activated by CcrM-mediated DNA methylation. We also observed that genes belonging to the GcrA (70), DnaA (69) and CtrA (68) regulons were strongly enriched among the most significantly downregulated genes in the $\Delta ccrM$ strain (Figures 6A–C), with the GcrA regulon being the most striking case (~ 10 -fold enrichment when considering genes whose transcription is changed at least 2-fold in the $\Delta ccrM$ strain). Genes whose expression is temporally

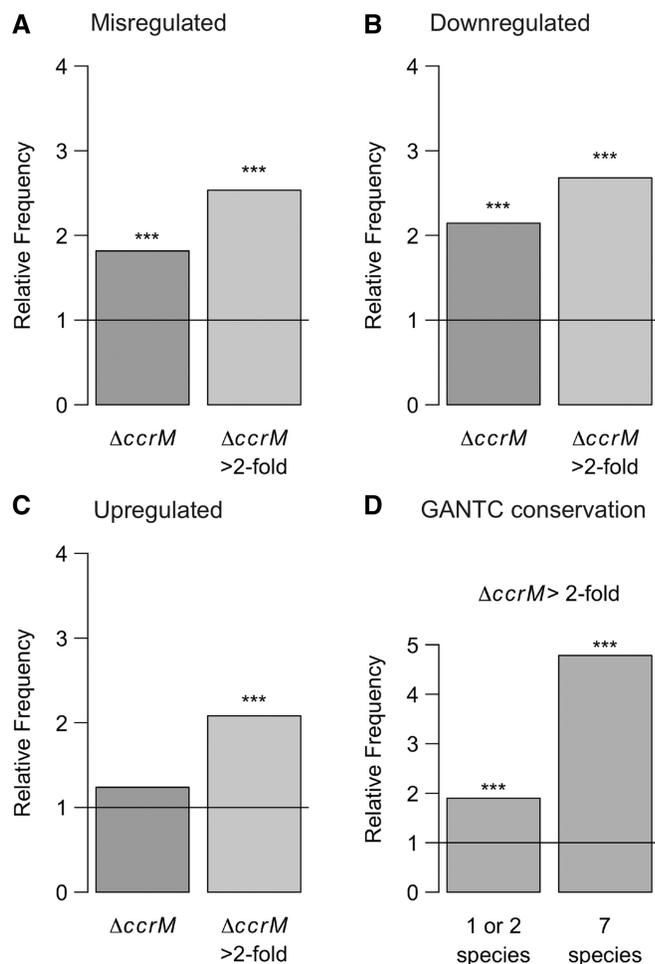


Figure 5. Conserved GANTC motifs are enriched upstream of genes significantly misregulated in the $\Delta ccrM$ strain. Frequency, relative to the entire genome, of genes whose promoter region (here, 200 bp upstream of their translational start codon) contains a GANTC motif among genes significantly misregulated (upregulated or downregulated) (A), significantly downregulated (B) and significantly upregulated (C) in the $\Delta ccrM$ strain compared with the wild-type strain. (D) Frequency, relative to the entire genome, of genes whose promoter region (200 bp upstream the translational start codon) contains a GANTC in one or two or in seven bacterial species closely related to *C. crescentus* (*C. crescentus*, *Caulobacter segnis*, *Caulobacter K31*, *Phenylobacter zucineum*, *Brevundimonas subvibrioides*, *Asticcacaulis excentricus*, *Maricaulis maris*) among genes significantly downregulated in the $\Delta ccrM$ strain. Stars indicate that the bias is significant ($P < 0.05$, Fisher's exact test).

regulated during the cell cycle (71) were also significantly enriched (Figure 6D). Altogether, these observations indicate a strong connection between DNA methylation by CcrM and cell cycle control, considering that GcrA, DnaA and CtrA are essential master transcriptional regulators of the *C. crescentus* cell cycle and that many essential genes might be directly regulated by DNA methylation (Table 1).

Effect of the constitutive over-expression of CcrM on global transcription profiles in *C. crescentus*

To evaluate the global impact of CcrM-dependent methylation on gene expression in *C. crescentus*, we also

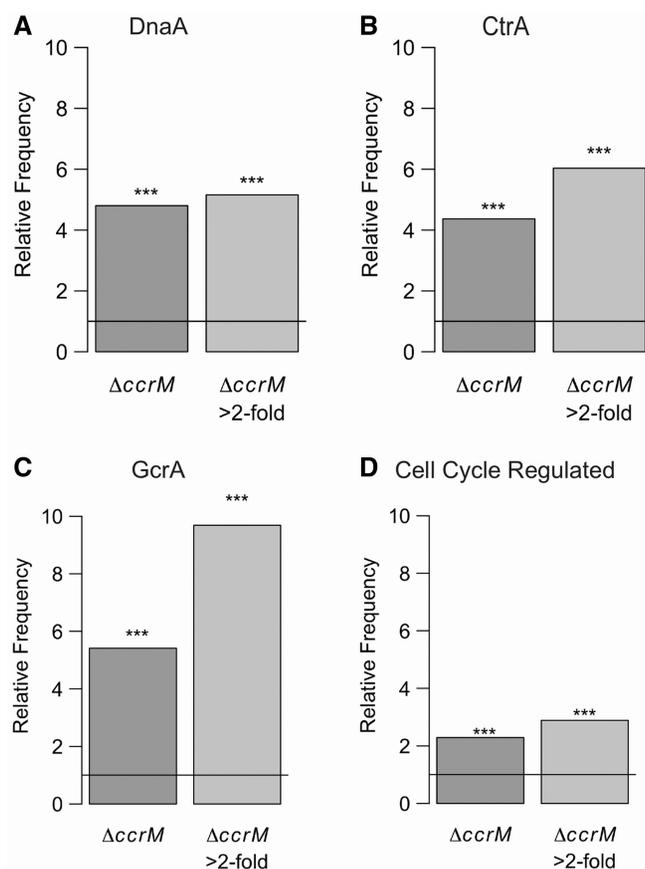


Figure 6. Involvement of CcrM in the regulation of the *C. crescentus* cell cycle. Frequency, relative to the entire genome, of genes belonging to the direct DnaA regulon (69) (A), the direct CtrA regulon (68) (B), the GcrA regulon (70) (C) or whose expression levels vary during the cell cycle (71) (D) among genes significantly downregulated or strongly downregulated (>2-fold change) in the $\Delta ccrM$ strain. Stars indicate a significant bias ($P < 0.05$, Fisher's exact test). A comparison between genes significantly misregulated in the $\Delta ccrM$ strain and possible direct targets of GcrA, as determined by a ChIP-seq experiment (72), is shown in Supplementary Figure S10.

analyzed the transcriptional profiles of bacterial populations in which CcrM is expressed at all times in the cell cycle. When CcrM is over-expressed, the period of post-replicative hemi-methylation of GANTC motifs that is characteristic of the wild-type strain is abolished having potential consequences on gene expression (Figure 3). We found that the expression levels of 546 genes were significantly changed in the strain with constitutive over-expression of CcrM compared with the wild-type strain, with 214 of these genes being affected >2-fold (Supplementary Table S5). For 15 genes, the effects were verified by qRT-PCR and were essentially consistent with the microarray results (Supplementary Figure S6).

To evaluate whether these effects may result from direct effects of GANTC methylation on promoter activity, we calculated the probability that significantly misregulated genes contained a GANTC motif in their promoter region and compared this value with the probability obtained for a random set of genes. We found no significant

(Fisher's test $P < 0.05$) difference, even when considering the most misregulated genes (Supplementary Figure S7), which suggests that most of the changes in gene expression observed in cells over-expressing CcrM result from indirect, rather than direct, effects of GANTC methylation on the activity of promoters.

Only a few genes involved in the regulation of the cell cycle were found among the genes that were significantly misregulated in the CcrM over-expressing strain. Instead, genes encoding proteins involved in inorganic ion transport and metabolism were over-represented among the most downregulated genes in this mutant strain, while genes encoding proteins involved in DNA replication, recombination and repair were over-represented among the most upregulated genes (Figure 4B). Notably, regulons involved in the resistance to different forms of environmental stress, such as the Fur (73), SigT (74) and FixKL (75) regulons, were over-represented among misregulated genes (Figure 7A–C). Similarly, genes upregulated under heavy-metal stress (76) were also over-represented (Figure 7D). These observations suggest that cells with over-expressed CcrM throughout the cell cycle exhibit a general and unspecific stress response.

DISCUSSION

In the present study, we analyzed the methylation state of all GANTC motifs on the chromosome of three *C. crescentus* strains expressing different levels of the CcrM DNA methyltransferase (Figure 3) and determined the global transcriptional changes associated with the differences in the methylation state. We observed that hundreds of genes were misregulated in cells lacking CcrM or in cells in which CcrM is constitutively over-expressed throughout the cell cycle, compared with wild-type cells (Supplementary Tables S4 and S5). A significant proportion of the gene expression changes detected in the CcrM-deficient strain appeared to be due to direct methylation-mediated gene regulation, since conserved GANTC motifs were more frequently found in the promoter regions of genes misregulated in this strain than in random promoter regions (Figure 5 and Supplementary Table S4). Considering that 80 genes with a GANTC motif in their promoter region are found among genes strongly misregulated in the $\Delta ccrM$ strain and that this number is 2-fold higher than expected by chance, we estimate that ~ 40 of these genes could be directly regulated by CcrM (some essential candidates are shown in Table 1 and Figure 8). In contrast, gene expression changes detected in the CcrM over-expressing strain appeared to be mostly due to secondary effects of the constitutive methylation of the chromosome that yields a stress phenotype (Figure 7 and Supplementary Figure S7). Consistent with these findings, we found no negative correlation between the changes in gene expression observed in $\Delta ccrM$ cells and in CcrM-over-expressing cells for most genes (Supplementary Figure S8), as would have been expected if the absence of methylation and constitutive-methylation had detectable opposite effects on the activity

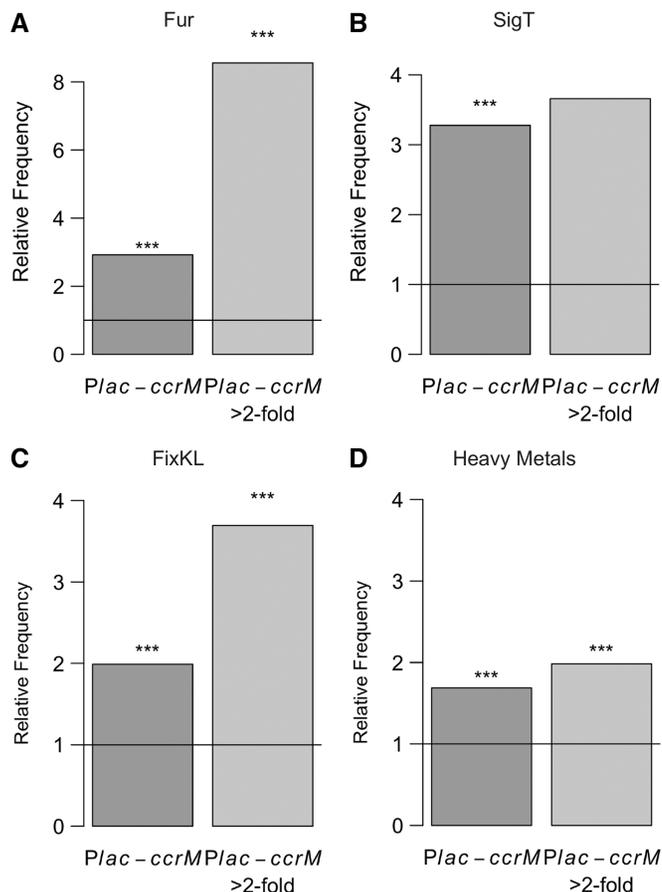


Figure 7. Constitutive over-expression of CcrM yields a stress phenotype. Frequency, relative to the entire genome, of genes belonging to the Fur regulon (A) or the SigT regulon (B), among genes significantly misregulated (upregulated or downregulated) or strongly misregulated (>2-fold change) in the *Plac::ccrM* strain. Frequency, relative to the entire genome, of genes belonging to the FixKL regulon (C) or the set of genes induced under heavy metal stress (D), among genes significantly misregulated or strongly misregulated (>2-fold change) in the *Plac::ccrM* strain. Stars indicate a significant bias ($P < 0.05$, Fisher's exact test).

of many promoters. We also observed that genes misregulated in each strain generally encoded proteins involved in different processes (Figure 4).

In the absence of CcrM-dependent methylation, multiple genes involved in processes that are essential to the viability of *C. crescentus* were misregulated (Table 1 and Supplementary Tables S4). Figure 8 shows 17 essential genes that are good candidates for being directly activated by CcrM-dependent methylation, as they have a relatively conserved GANTC motif in their promoter region. These genes include many genes encoding proteins required for cell division, such as FtsZ, FtsN and MipZ, and proteins involved in DNA replication, repair and topology, such as the ligase, the gyrase and Ssb (Figure 8, Table 1 and Supplementary Table S4). We previously showed that the artificial expression of FtsZ can restore the viability of $\Delta ccrM$ cells in fast-growing conditions, but that these cells were still longer, straighter and with shorter stalks than wild-type cells (36).

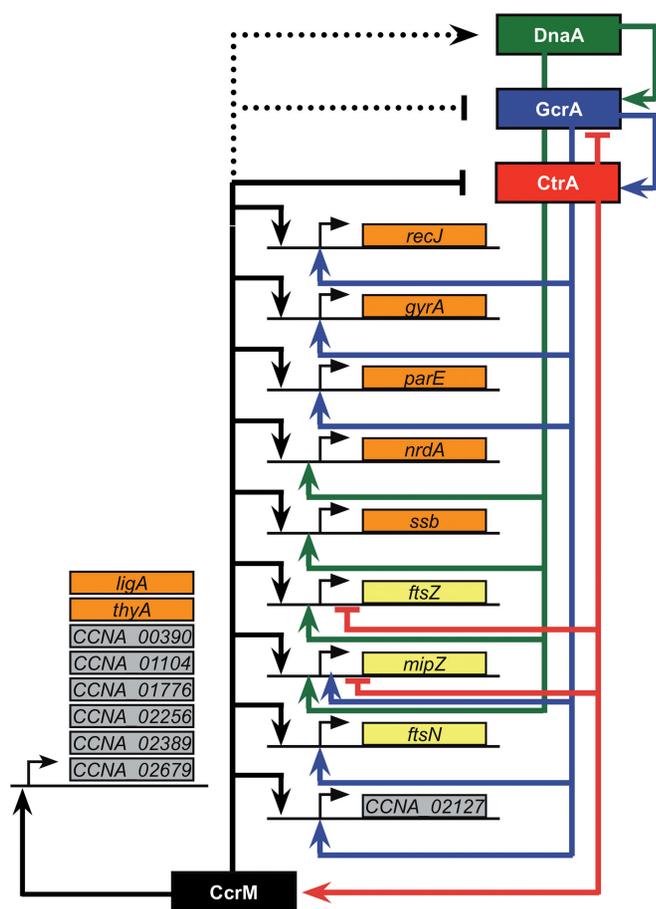


Figure 8. Model for the CcrM-dependent activation of essential genes containing conserved GANTC motifs in their promoter. Genes included in this schematic contain a GANTC motif in their promoter region (200 bp upstream of their translational start codon), which is conserved in a minimum of two out of five bacterial species closely related to *C. crescentus* (Table 1). Genes coloured in orange encode proteins involved in DNA replication, repair or topology. Genes coloured in yellow encode proteins involved in cell division. Genes coloured in grey encode proteins involved in other functions. Solid black arrows indicate regulatory pathways identified as significant during the transcriptome analysis using the $\Delta ccrM$ strain in this study. Dashed black arrows indicate regulatory effects previously identified (34,77), but not found as significantly affected using the $\Delta ccrM$ strain in this study. Solid blue, red and green arrows indicate the GcrA-, CtrA- and DnaA-dependent regulatory pathways identified in (70), (68) and (69), respectively. This schematic suggests that many, but not all, essential genes activated by CcrM are also co-regulated by DnaA, CtrA or GcrA master regulators, showing the strong interconnection between CcrM-dependent DNA methylation and the *C. crescentus* cell cycle regulatory network.

Our transcriptome analysis suggests that these cells might be longer due to insufficient quantities of FtsN or to excessive quantities of FtsW (Supplementary Table S4: the *ftsN* and *ftsW* messengers were 2.1 times less and 1.6 times more abundant in $\Delta ccrM$ than in wild-type cells, respectively). Interestingly, $\Delta ccrM$ cells still seemed to replicate their chromosome efficiently and had a normal spontaneous mutation rate (Supplementary Figures S4 and S5), despite the downregulation of several genes involved in DNA replication or repair and the over-expression of the CtrA inhibitor of replication (Figure 8

and Table 1 and Supplementary Table S4). This result suggests that proteins like Ssb or the ligase are not in limiting quantities for DNA replication and that the post-transcriptional regulation of CtrA maintains intracellular levels of active CtrA low enough for DNA replication to initiate (29). Consistent with our results, it was previously shown that a *C. crescentus* strain that carries a unique copy of the *ctrA* gene under the control of a mutant un-methylatable *ctrA* promoter does not have an abnormal phenotype (35). In addition to these many essential genes, several genes involved in processes that are required for the development of *C. crescentus* were also misregulated in the absence of CcrM-dependent methylation. For example, the *staR* and *creS* non-essential genes, involved in stalk biogenesis (78) and in cell curvature (79), respectively, were significantly downregulated in $\Delta ccrM$ cells (Supplementary Table S4: the *staR* and *creS* messengers were 2.9 and 1.6 times less abundant in $\Delta ccrM$ than in wild-type cells, respectively) and this might provide an explanation for the additional phenotypes that were observed.

Another remarkable finding was that genes regulated by the three global transcriptional regulators of the *C. crescentus* cell cycle, GcrA, CtrA and DnaA, were strongly over-represented in the set of genes that were the most misregulated in the absence of CcrM-mediated methylation (Figure 6 and Supplementary Figure S9 and S10). These three regulators are essential to the viability of *C. crescentus*, and they regulate the expression of hundreds of genes involved in various events of the cell cycle (27). Our observation suggests that *C. crescentus* genes encoding proteins that are critical for cell cycle progression are often regulated by one or more global regulators and by DNA methylation (Figures 6 and 8 and Table 1), showing the strong interconnection between CcrM-dependent DNA methylation and the *C. crescentus* cell cycle regulatory network. A possible explanation is that the binding or the activity of one or more of these global regulators is sensitive to the methylation state of the DNA at or next to their binding site. The consensus binding sites of CtrA and DnaA do not contain a sequence that resembles a GANTC motif (68,69). As for GcrA, its consensus binding site is not yet clearly characterized, but a recent study demonstrated that GcrA activates the transcription of several genes by promoting the recruitment of the RNA polymerase to promoter regions containing methylated GANTC motifs that overlapped a GcrA binding site (72). Interestingly, the homologs of GcrA have the same conservation pattern as CcrM homologs (Figure 2A). The expression of many other members of the GcrA direct regulon (68,72) was nevertheless not significantly affected in the CcrM-deficient strain (Supplementary Table S4 and Supplementary Figure S9), suggesting that the activity of GcrA might not always be influenced by CcrM-dependent methylation. Further studies will be required to test whether the activities of the CtrA and DnaA global regulators may also be influenced by the methylation state of some promoter regions or if the over-representation of the CtrA and DnaA regulons among genes significantly misregulated

in the $\Delta ccrM$ strain is only a consequence of regulation by more than one regulator.

The chromosomes of cells over-expressing CcrM are maintained in the fully methylated state throughout the cell cycle (Figure 3); these cells are slightly elongated and tend to accumulate additional copies of the chromosome (31). The microarray analyses revealed that the transcription levels of many genes are changed in this mutant (Supplementary Table S5) and that cells exhibit a stress response (Figure 7). Transcriptional effects directly due to the elimination of the fully methylated to hemi-methylated switch characteristic of the wild-type strain were, however, either rare or so weak that the method was not sensitive enough to detect them (Supplementary Figure S7). A possible link between the over-replication phenotype of the strain and the transcriptome can be found in the downregulation of the gene encoding the PopZ protein (Supplementary Table S5: the *popZ* messenger was 1.4 times less abundant in CcrM-over-expressing cells than in wild-type cells), which is required for the anchoring of the two chromosomal origins at opposite cell poles after the initiation of chromosome replication (80,81). Insufficient intracellular levels of PopZ may contribute to the abnormal DNA content in cells over-expressing CcrM, as was previously described for PopZ-depleted cells (80,81). Other phenotypes may be attributable to a global stress response that seems to take place in these cells (Figure 7).

Our SMRT sequencing analysis showed that 35 GANTC motifs along the chromosome were frequently under-methylated on minimum one strand in wild-type cells (Supplementary Table S3). We found that only three of these motifs remained under-methylated in cells over-expressing CcrM. To address whether the under-methylation of the other 32 GANTC motifs in wild-type cells may affect gene expression, we compared the mRNA levels of nearby genes in the wild-type and in the CcrM-over-expressing cells. We observed a significant difference in expression of two genes: these were the *CCNA_01150* gene of unknown function and the *CCNA_03248* gene encoding a TonB-dependent receptor (Supplementary Tables S3 and S5): the *CCNA_01150* and the *CCNA_03248* messengers were five times more and 1.3 times less abundant in CcrM-over-expressing than in wild-type cells, respectively). The difference in expression might be linked to the difference in the methylation state of GANTC motifs in the two strains. Similarly, under-methylated GATC sites were observed on the chromosomes of Dam-expressing bacteria and they are sometimes involved in epigenetic switches of gene expression (7,11,24–26,82). These sites often overlap the binding sites of the Lrp, OxyR and Fur global regulators, which were shown to compete with Dam for the DNA and to be sometimes sensitive to its methylation state. Interestingly, we found that the predicted Fur regulon was over-represented among the genes that were misregulated when the CcrM enzyme was over-produced in *C. crescentus* (Figure 7). We nevertheless did not find Fur-regulated promoters carrying frequently under-methylated GANTC motifs on the *C. crescentus* chromosome. This observation suggests that the *C. crescentus* Fur

protein does not often protect the promoter that it controls from CcrM-mediated methylation, at least when cells were cultivated in iron-rich medium, contrarily to what was previously observed at the *sciI* promoter on the *E. coli* chromosome (82). It is still possible that CcrM-dependent methylation affects the activity of the *C. crescentus* Fur protein using a mechanism different from the one previously described in *E. coli* (82). An excellent candidate gene that may be regulated by an epigenetic switch mediated by Fur and DNA methylation in *C. crescentus* is the *CCNA_02275* gene, as its promoter contains two GANTC sites that overlap a Fur binding site and since its messenger was 2.2 times less abundant in CcrM-over-expressing than in wild-type cells (Supplementary Table S5). This gene encodes a trans-membrane protein of unknown function and is repressed by Fur (83). These results suggest that Fur may have a higher affinity for the *CCNA_02275* promoter when it is in a fully methylated, rather than in a hemi-methylated, state. Additional detailed studies will be required to understand how this gene might be epigenetically regulated in *C. crescentus*.

Overall, our study demonstrates that the overlap between the roles of Dam-dependent methylation in enterobacteria and the roles of CcrM-dependent methylation in *Alphaproteobacteria* is restricted. In both cases, the methylation of adenines has pleiotropic effects on gene expression, but the genes and functions represented in the Dam and CcrM regulons are different. In *C. crescentus*, but not *E. coli*, the methylation of GANTC motifs may mediate the integration of the cell cycle regulatory circuit with chromosome replication. In addition, we found that CcrM-dependent methylation is not required to control the initiation of chromosome replication (Supplementary Figure S4) or to correct DNA mismatches during the MMR process (Supplementary Figure S5), as it is often the case for Dam-dependent methylation (11,12,14). Dam is generally co-conserved with MutH, which is the protein that recognizes the newly synthesized non-methylated DNA strand that needs to be repaired during methyl-directed MMR (18). Most bacteria lack Dam and MutH, like *C. crescentus* and *Bacillus subtilis* (13,84,85). These bacteria probably use a methylation-independent strand recognition mechanism during MMR, which might be spatially coupled with the DNA polymerase complex (85). This study exemplifies the need to explore the multifaceted use of DNA methylation in a variety of bacterial species.

ACCESSION NUMBERS

The microarray data are publicly available at the GEO database (GEO accession numbers GSE52722 and GSE52375, <https://www.ncbi.nlm.nih.gov/geo>).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful to Keith Harshman, Johann Weber and Hannes Richter from the Lausanne Genomic Technologies Facility and to Jonas Korfach and Tyson Clark from the Pacific Biosciences Company (Menlo Park, CA, USA) for helping us during the transcriptome, qRT-PCR and SMRT sequencing analysis. We thank members of the Collier laboratory for helpful discussions.

FUNDING

University of Lausanne; Swiss National Science Fellowships [3100A0_122541 and 31003A_140758 to J.C.]; National Institute of Health Grant [RO1-GM051426 to L.S.]. Funding for open access charge: Swiss National Science Fellowship [31003A_140758 to J.C.].

Conflict of interest statement. None declared.

REFERENCES

- Schlagman,S.L., Miner,Z., Feher,Z. and Hattman,S. (1988) The DNA [adenine-N6]methyltransferase (Dam) of bacteriophage T4. *Gene*, **73**, 517–530.
- Hoelzer,K., Shackelton,L.A. and Parrish,C.R. (2008) Presence and role of cytosine methylation in DNA viruses of animals. *Nucleic Acids Res.*, **36**, 2825–2837.
- Bochow,S., Elliman,J. and Owens,L. (2012) Bacteriophage adenine methyltransferase: a life cycle regulator? Modelled using *Vibrio harveyi* myovirus like. *J. Appl. Microbiol.*, **113**, 1001–1013.
- Jeltsch,A. (2002) Beyond watson and crick: DNA methylation and molecular enzymology of DNA methyltransferases. *Chembiochem*, **3**, 274–293.
- Ooi,S.K., O'Donnell,A.H. and Bestor,T.H. (2009) Mammalian cytosine methylation at a glance. *J. Cell Sci.*, **122**, 2787–2791.
- Davis,B.M., Chao,M.C. and Waldor,M.K. (2013) Entering the era of bacterial epigenomics with single molecule real time DNA sequencing. *Curr. Opin. Microbiol.*, **16**, 192–198.
- Wion,D. and Casadesus,J. (2006) N6-methyl-adenine: an epigenetic signal for DNA-protein interactions. *Nat. Rev. Microbiol.*, **4**, 183–192.
- Roberts,R.J., Vincze,T., Posfai,J. and Macelis,D. (2010) REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.*, **38**, D234–D236.
- Jeltsch,A. (2003) Maintenance of species identity and controlling speciation of bacteria: a new function for restriction/modification systems? *Gene*, **317**, 13–16.
- Marinus,M.G. and Casadesus,J. (2009) Roles of DNA adenine methylation in host-pathogen interactions: mismatch repair, transcriptional regulation, and more. *FEMS Microbiol. Rev.*, **33**, 488–503.
- Low,D.A. and Casadesus,J. (2008) Clocks and switches: bacterial gene regulation by DNA adenine methylation. *Curr. Opin. Microbiol.*, **11**, 106–112.
- Casadesus,J. and Low,D. (2006) Epigenetic gene regulation in the bacterial world. *Microbiol. Mol. Biol. Rev.*, **70**, 830–856.
- Brezellec,P., Hoebeke,M., Hiet,M.S., Pasek,S. and Ferat,J.L. (2006) DomainSieve: a protein domain-based screen that led to the identification of dam-associated genes with potential link to DNA maintenance. *Bioinformatics*, **22**, 1935–1941.
- Collier,J. (2009) Epigenetic regulation of the bacterial cell cycle. *Curr. Opin. Microbiol.*, **12**, 722–729.
- Demarre,G. and Chattoraj,D.K. (2010) DNA adenine methylation is required to replicate both *Vibrio cholerae* chromosomes once per cell cycle. *PLoS Genet.*, **6**, e1000939.
- Val,M.E., Skovgaard,O., Ducos-Galand,M., Bland,M.J. and Mazel,D. (2012) Genome engineering in *Vibrio cholerae*: a feasible approach to address biological issues. *PLoS Genet.*, **8**, e1002472.
- Pukkila,P.J., Peterson,J., Herman,G., Modrich,P. and Meselson,M. (1983) Effects of high levels of DNA adenine methylation on methyl-directed mismatch repair in *Escherichia coli*. *Genetics*, **104**, 571–582.
- Iyer,R.R., Pluciennik,A., Burdett,V. and Modrich,P.L. (2006) DNA mismatch repair: functions and mechanisms. *Chem. Rev.*, **106**, 302–323.
- Oshima,T., Wada,C., Kawagoe,Y., Ara,T., Maeda,M., Masuda,Y., Hiraga,S. and Mori,H. (2002) Genome-wide analysis of deoxyadenosine methyltransferase-mediated control of gene expression in *Escherichia coli*. *Mol. Microbiol.*, **45**, 673–695.
- Lobner-Olesen,A., Marinus,M.G. and Hansen,F.G. (2003) Role of SeqA and Dam in *Escherichia coli* gene expression: a global/microarray analysis. *Proc. Natl Acad. Sci. USA*, **100**, 4672–4677.
- Robbins-Manke,J.L., Zdraveski,Z.Z., Marinus,M. and Essigmann,J.M. (2005) Analysis of global gene expression and double-strand-break formation in DNA adenine methyltransferase- and mismatch repair-deficient *Escherichia coli*. *J. Bacteriol.*, **187**, 7027–7037.
- Seshasayee,A.S. (2007) An assessment of the role of DNA adenine methyltransferase on gene expression regulation in *E. coli*. *PLoS One*, **2**, e273.
- Balbontin,R., Rowley,G., Pucciarelli,M.G., Lopez-Garrido,J., Wormstone,Y., Lucchini,S., Garcia-Del Portillo,F., Hinton,J.C. and Casadesus,J. (2006) DNA adenine methylation regulates virulence gene expression in *Salmonella enterica* serovar Typhimurium. *J. Bacteriol.*, **188**, 8160–8168.
- van der Woude,M.W. (2011) Phase variation: how to create and coordinate population diversity. *Curr. Opin. Microbiol.*, **14**, 205–211.
- van der Woude,M., Hale,W.B. and Low,D.A. (1998) Formation of DNA methylation patterns: nonmethylated GATC sequences in *gut* and *pap* operons. *J. Bacteriol.*, **180**, 5913–5920.
- Dalia,A.B., Lazinski,D.W. and Camilli,A. (2013) Characterization of undermethylated sites in *Vibrio cholerae*. *J. Bacteriol.*, **195**, 2389–2399.
- Collier,J. and Shapiro,L. (2007) Spatial complexity and control of a bacterial cell cycle. *Curr. Opin. Biotechnol.*, **18**, 333–340.
- Curtis,P.D. and Brun,Y.V. (2010) Getting in the loop: regulation of development in *Caulobacter crescentus*. *Microbiol. Mol. Biol. Rev.*, **74**, 13–41.
- Collier,J. (2012) Regulation of chromosomal replication in *Caulobacter crescentus*. *Plasmid*, **67**, 76–87.
- Marczynski,G.T. (1999) Chromosome methylation and measurement of faithful, once and only once per cell cycle chromosome replication in *Caulobacter crescentus*. *J. Bacteriol.*, **181**, 1984–1993.
- Zweiger,G., Marczynski,G. and Shapiro,L. (1994) A *Caulobacter* DNA methyltransferase that functions only in the predivisive cell. *J. Mol. Biol.*, **235**, 472–485.
- Reisenauer,A., Kahng,L.S., McCollum,S. and Shapiro,L. (1999) Bacterial DNA methylation: a cell cycle regulator? *J. Bacteriol.*, **181**, 5135–5139.
- Stephens,C., Reisenauer,A., Wright,R. and Shapiro,L. (1996) A cell cycle-regulated bacterial DNA methyltransferase is essential for viability. *Proc. Natl Acad. Sci. USA*, **93**, 1210–1214.
- Collier,J., McAdams,H.H. and Shapiro,L. (2007) A DNA methylation ratchet governs progression through a bacterial cell cycle. *Proc. Natl Acad. Sci. USA*, **104**, 17111–17116.
- Reisenauer,A. and Shapiro,L. (2002) DNA methylation affects the cell cycle transcription of the CtrA global regulator in *Caulobacter*. *EMBO J.*, **21**, 4969–4977.
- Gonzalez,D. and Collier,J. (2013) DNA methylation by CcrM activates the transcription of two genes required for the division of *Caulobacter crescentus*. *Mol. Microbiol.*, **88**, 203–218.
- Stephens,C.M., Zweiger,G. and Shapiro,L. (1995) Coordinate cell cycle control of a *Caulobacter* DNA methyltransferase and the flagellar genetic hierarchy. *J. Bacteriol.*, **177**, 1662–1669.
- Wright,R., Stephens,C. and Shapiro,L. (1997) The CcrM DNA methyltransferase is widespread in the alpha subdivision of *proteobacteria*, and its essential functions are conserved in *Rhizobium meliloti* and *Caulobacter crescentus*. *J. Bacteriol.*, **179**, 5869–5877.

39. Kahng, L.S. and Shapiro, L. (2001) The CcrM DNA methyltransferase of *Agrobacterium tumefaciens* is essential, and its activity is cell cycle regulated. *J. Bacteriol.*, **183**, 3065–3075.
40. Robertson, G.T., Reisenauer, A., Wright, R., Jensen, R.B., Jensen, A., Shapiro, L. and Roop, R.M. II (2000) The *Brucella abortus* CcrM DNA methyltransferase is essential for viability, and its overexpression attenuates intracellular replication in murine macrophages. *J. Bacteriol.*, **182**, 3482–3489.
41. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
42. Edgar, R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, **5**, 113.
43. Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, **32**, 1792–1797.
44. Malone, T., Blumenthal, R.M. and Cheng, X. (1995) Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. *J. Mol. Biol.*, **253**, 618–632.
45. Price, M.N., Dehal, P.S. and Arkin, A.P. (2010) FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One*, **5**, e9490.
46. Price, M.N., Dehal, P.S. and Arkin, A.P. (2009) FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.*, **26**, 1641–1650.
47. Castresana, J. (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.*, **17**, 540–552.
48. Stamatakis, A., Hoover, P. and Rougemont, J. (2008) A rapid bootstrap algorithm for the RAxML web servers. *Syst. Biol.*, **57**, 758–771.
49. Huelsenbeck, J.P. and Ronquist, F. (2001) MRBAYES: bayesian inference of phylogenetic trees. *Bioinformatics*, **17**, 754–755.
50. Ronquist, F. and Huelsenbeck, J.P. (2003) MrBayes 3: bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.
51. Evinger, M. and Agabian, N. (1977) Envelope-associated nucleoid from *Caulobacter crescentus* stalked and swarmer cells. *J. Bacteriol.*, **132**, 294–301.
52. Marks, M.E., Castro-Rojas, C.M., Teiling, C., Du, L., Kapratl, V., Walunas, T.L. and Crosson, S. (2010) The genetic basis of laboratory adaptation in *Caulobacter crescentus*. *J. Bacteriol.*, **192**, 3678–3688.
53. Collier, J. and Shapiro, L. (2009) Feedback control of DnaA-mediated replication initiation by replisome-associated HdaA protein in *Caulobacter*. *J. Bacteriol.*, **191**, 5706–5716.
54. Brill, M., Fondi, M., Fani, R., Mengoni, A., Ferri, L., Bazzicalupo, M. and Biondi, E.G. (2010) The diversity and evolution of cell cycle regulation in *alpha*-proteobacteria: a comparative genomic analysis. *BMC Syst. Biol.*, **4**, 52.
55. Bazylinski, D.A., Williams, T.J., Lefevre, C.T., Berg, R.J., Zhang, C.L., Bowser, S.S., Dean, A.J. and Beveridge, T.J. (2013) *Magnetococcus marinus* gen. nov., sp. nov., a marine, magnetotactic bacterium that represents a novel lineage (Magnetococcaceae fam. nov., Magnetococcales ord. nov.) at the base of the *Alphaproteobacteria*. *Int. J. Syst. Evol. Microbiol.*, **63**, 801–808.
56. Viklund, J., Ettema, T.J. and Andersson, S.G. (2012) Independent genome reduction and phylogenetic reclassification of the oceanic SAR11 clade. *Mol. Biol. Evol.*, **29**, 599–615.
57. Gupta, R.S. and Mok, A. (2007) Phylogenomics and signature proteins for the *alpha* proteobacteria and its main groups. *BMC Microbiol.*, **7**, 106.
58. Nierman, W.C., Feldblyum, T.V., Laub, M.T., Paulsen, I.T., Nelson, K.E., Eisen, J.A., Heidelberg, J.F., Alley, M.R., Ohta, N., Maddock, J.R. et al. (2001) Complete genome sequence of *Caulobacter crescentus*. *Proc. Natl Acad. Sci. USA*, **98**, 4136–4141.
59. Paulino, L.C., de Mello, M.P. and Ottoboni, L.M. (2002) Differential gene expression in response to copper in *Acidithiobacillus ferrooxidans* analyzed by RNA arbitrarily primed polymerase chain reaction. *Electrophoresis*, **23**, 520–527.
60. Wright, R., Stephens, C., Zweiger, G., Shapiro, L. and Alley, M.R. (1996) *Caulobacter* Lon protease has a critical role in cell-cycle control of DNA methylation. *Genes Dev.*, **10**, 1532–1542.
61. Flusberg, B.A., Webster, D.R., Lee, J.H., Travers, K.J., Olivares, E.C., Clark, T.A., Korlach, J. and Turner, S.W. (2010) Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods*, **7**, 461–465.
62. Clark, T.A., Murray, I.A., Morgan, R.D., Kislyuk, A.O., Spittle, K.E., Boitano, M., Fomenkov, A., Roberts, R.J. and Korlach, J. (2012) Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. *Nucleic Acids Res.*, **40**, e29.
63. Murray, I.A., Clark, T.A., Morgan, R.D., Boitano, M., Anton, B.P., Luong, K., Fomenkov, A., Turner, S.W., Korlach, J. and Roberts, R.J. (2012) The methylomes of six bacteria. *Nucleic Acids Res.*, **40**, 11450–11462.
64. Kozdon, J.B., Melfi, M.D., Luong, K., Clark, T.A., Boitano, M., Wang, S., Zhou, B., Gonzalez, D., Collier, J., Turner, S.W. et al. (2013) Global methylation state at base-pair resolution of the *Caulobacter* genome throughout the cell cycle. *Proc. Natl Acad. Sci. USA*, **110**, E4658–E4667.
65. Nievera, C., Torgue, J.J., Grimwade, J.E. and Leonard, A.C. (2006) SeqA blocking of DnaA-oriC interactions ensures staged assembly of the *E. coli* pre-RC. *Mol. Cell*, **24**, 581–592.
66. Shaheen, S.M., Ouimet, M.C. and Marczyński, G.T. (2009) Comparative analysis of *Caulobacter* chromosome replication origins. *Microbiology*, **155**, 1215–1225.
67. Christen, B., Abeliuk, E., Collier, J.M., Kalogeraki, V.S., Passarelli, B., Collier, J.A., Fero, M.J., McAdams, H.H. and Shapiro, L. (2011) The essential genome of a bacterium. *Mol. Syst. Biol.*, **7**, 528.
68. Laub, M.T., Chen, S.L., Shapiro, L. and McAdams, H.H. (2002) Genes directly controlled by CtrA, a master regulator of the *Caulobacter* cell cycle. *Proc. Natl Acad. Sci. USA*, **99**, 4632–4637.
69. Hottes, A.K., Shapiro, L. and McAdams, H.H. (2005) DnaA coordinates replication initiation and cell cycle transcription in *Caulobacter crescentus*. *Mol. Microbiol.*, **58**, 1340–1353.
70. Holtzendorff, J., Hung, D., Brende, P., Reisenauer, A., Viollier, P.H., McAdams, H.H. and Shapiro, L. (2004) Oscillating global regulators control the genetic circuit driving a bacterial cell cycle. *Science*, **304**, 983–987.
71. Laub, M.T., McAdams, H.H., Feldblyum, T., Fraser, C.M. and Shapiro, L. (2000) Global analysis of the genetic network controlling a bacterial cell cycle. *Science*, **290**, 2144–2148.
72. Fioravanti, A., Fumeaux, C., Mohapatra, S.S., Bompard, C., Brill, M., Frandi, A., Castric, V., Villeret, V., Viollier, P.H. and Biondi, E.G. (2013) DNA binding of the cell cycle transcriptional regulator GcrA depends on N6-adenosine methylation in *Caulobacter crescentus* and other *Alphaproteobacteria*. *PLoS Genet.*, **9**, e1003541.
73. da Silva Neto, J.F., Braz, V.S., Italiani, V.C. and Marques, M.V. (2009) Fur controls iron homeostasis and oxidative stress defense in the oligotrophic *alpha*-proteobacterium *Caulobacter crescentus*. *Nucleic Acids Res.*, **37**, 4812–4825.
74. Alvarez-Martinez, C.E., Lourenco, R.F., Baldini, R.L., Laub, M.T. and Gomes, S.L. (2007) The ECF sigma factor sigma(T) is involved in osmotic and oxidative stress responses in *Caulobacter crescentus*. *Mol. Microbiol.*, **66**, 1240–1255.
75. Crosson, S., McGrath, P.T., Stephens, C., McAdams, H.H. and Shapiro, L. (2005) Conserved modular design of an oxygen sensory/signaling network with species-specific output. *Proc. Natl Acad. Sci. USA*, **102**, 8018–8023.
76. Hu, P., Brodie, E.L., Suzuki, Y., McAdams, H.H. and Andersen, G.L. (2005) Whole-genome transcriptional analysis of heavy metal stresses in *Caulobacter crescentus*. *J. Bacteriol.*, **187**, 8437–8449.
77. Collier, J., Murray, S.R. and Shapiro, L. (2006) DnaA couples DNA replication and the expression of two cell cycle master regulators. *EMBO J.*, **25**, 346–356.
78. Biondi, E.G., Skerker, J.M., Arif, M., Prasol, M.S., Perchuk, B.S. and Laub, M.T. (2006) A phosphorelay system controls stalk biogenesis during cell cycle progression in *Caulobacter crescentus*. *Mol. Microbiol.*, **59**, 386–401.

79. Ausmees,N., Kuhn,J.R. and Jacobs-Wagner,C. (2003) The bacterial cytoskeleton: an intermediate filament-like function in cell shape. *Cell*, **115**, 705–713.
80. Bowman,G.R., Comolli,L.R., Zhu,J., Eckart,M., Koenig,M., Downing,K.H., Moerner,W.E., Earnest,T. and Shapiro,L. (2008) A polymeric protein anchors the chromosomal origin/ParB complex at a bacterial cell pole. *Cell*, **134**, 945–955.
81. Ebersbach,G., Briegel,A., Jensen,G.J. and Jacobs-Wagner,C. (2008) A self-associating protein critical for chromosome attachment, division, and polar organization in caulobacter. *Cell*, **134**, 956–968.
82. Brunet,Y.R., Bernard,C.S., Gavioli,M., Lloubes,R. and Cascales,E. (2011) An epigenetic switch involving overlapping fur and DNA methylation optimizes expression of a type VI secretion gene cluster. *PLoS Genet.*, **7**, e1002205.
83. da Silva Neto,J.F., Lourenco,R.F. and Marques,M.V. (2013) Global transcriptional response of *Caulobacter crescentus* to iron availability. *BMC Genomics*, **14**, 549.
84. Martins-Pinheiro,M., Marques,R.C. and Menck,C.F. (2007) Genome analysis of DNA repair genes in the *alpha* proteobacterium *Caulobacter crescentus*. *BMC Microbiol.*, **7**, 17.
85. Simmons,L.A., Davies,B.W., Grossman,A.D. and Walker,G.C. (2008) Beta clamp directs localization of mismatch repair in *Bacillus subtilis*. *Mol. Cell*, **29**, 291–301.