Global methylation state at base-pair resolution of the *Caulobacter* genome throughout the cell cycle

Jennifer B. Kozdon[a,b,1], Michael D. Melfi[a,b,1], Khai Luong[c,1], Tyson A. Clark[c], Matthew Boitano[c], Susana Wang[c], Bo Zhou[a], Diego Gonzalez[d], Justine Collier[d], Stephen W. Turner[c], Jonas Korlach[c], Lucy Shapiro[a,2], and Harley H. McAdams[a,2]

[a]Department of Developmental Biology, Stanford University, Stanford, CA 94305; [b]Department of Chemistry, Stanford University, Stanford, CA 94305; [c]Pacific Biosciences, Menlo Park, CA 94025; and [d]Department of Fundamental Microbiology, University of Lausanne, CH-1015 Lausanne, Switzerland

Contributed by Lucy Shapiro, October 22, 2013 (sent for review August 15, 2013)

The *Caulobacter* DNA methyltransferase CcrM is one of five master cell-cycle regulators. CcrM is transiently present near the end of DNA replication when it rapidly methylates the adenine in hemimethylated GANTC sequences. The timing of transcription of two master regulator genes and two cell division genes is controlled by the methylation state of GANTC sites in their promoters. To explore the global extent of this regulatory mechanism, we determined the methylation state of the entire chromosome at every base pair at five time points in the cell cycle using single-molecule, real-time sequencing. The methylation state of 4,515 GANTC sites, preferentially positioned in intergenic regions, changed progressively from full to hemimethylation as the replication forks advanced. However, 27 GANTC sites remained unmethylated throughout the cell cycle, suggesting that these protected sites could participate in epigenetic regulatory functions. An analysis of the time of activation of every cell-cycle regulatory transcription start site, coupled to both the position of a GANTC site in their promoter regions and the time in the cell cycle when the GANTC site transitions from full to hemimethylation, allowed the identification of 59 genes as candidates for epigenetic regulation. In addition, we identified two previously unidentified N^6-methyladenine motifs and showed that they maintained a constant methylation state throughout the cell cycle. The cognate methyltransferase was identified for one of these motifs as well as for one of two 5-methylcytosine motifs.

DNA methylation | SMRT sequencing | methylome

DNA methylation involves the addition of a methyl group to either adenine or cytosine by a site-specific DNA methyltransferase. The role of this epigenetic mechanism in the regulation of multiple bacterial processes is described in several reviews (1–5). Two adenine methyltransferases, Dam in the γ-proteobacterium *Escherichia coli* and CcrM in the α-proteobacterium *Caulobacter crescentus*, have regulatory functions and are not part of a restriction-modification system. Notably, the synthesis of CcrM, but not Dam, is cell cycle-regulated. CcrM plays an essential role in *Caulobacter* cell-cycle control (6–8) and in *Agrobacterium tumifaciens* (9), and it is involved in infective processes in *Brucella abortus* (10) and in the plant-microbe symbiotic relationship in *Mesorhizobium meliloti* (11), whereas Dam is involved in virulence of numerous bacterial species (1).

In α-proteobacteria, CcrM methylates adenines in GANTC sites (12). In *Caulobacter*, CcrM is expressed for only a brief time late in chromosome replication (12), and the Lon protease then rapidly degrades CcrM, with a half-life of about 10 min (13). Before the initiation of DNA replication, the chromosome is in the fully methylated state at GANTC sites. Upon bidirectional progression of the replication forks, newly replicated DNA becomes hemimethylated; each copy contains one parental methylated strand and one daughter unmethylated strand. It was predicted that newly replicated DNA therefore remains hemimethylated at these GANTC sites for an extended interval until the late transient expression of CcrM. In contrast, in γ-proteobacteria where the Dam DNA methyltransferase is continuously present (14), most of the Dam GATC recognition sites are only transiently in the hemimethylated state after passage of the replication fork, with the exception of regulatory GATC sites that are protected from Dam methylation by specific DNA-binding proteins (1, 4).

In *Caulobacter*, CcrM is one of five master regulators controlling cell-cycle progression. The remaining four master regulators are transcription factors (15, 16). Three of these transcription factors—DnaA, GcrA, and CtrA, together with the CcrM DNA methyltransferase—create a cyclical genetic circuit (Fig. 1*A*) that controls the expression of about 200 genes (6, 17–19). SciP, the fourth master transcription factor (not included in Fig. 1), adds feed-forward and feedback pathways that enhance the robustness of the performance of the cell-cycle control system (20).

Caulobacter has 4,542 GANTC sites, about 40% as many as expected by random occurrence (21), and 23% of the GANTC sites are in intergenic regions that together compose only 9.1% of the *Caulobacter* genome. The methylation state of GANTC sites in the intergenic promoter regions can affect the activity of these promoters. CcrM has been shown to methylate two GANTC sites in the *dnaA* promoter (6), one site in the *ctrA* P1 promoter (8) and sites in the *ftsZ* and *mipZ* promoters (22) modulating the activity of their respective genes. The *dnaA* promoter is preferentially active when in the fully methylated state.

Significance

***Caulobacter crescentus*, a bacterium with an inherent asymmetric cell division, uses dynamic changes in chromosome methylation state to synchronize chromosome replication with cell-cycle regulation. We identified the N^6-methyladenine and 5-methylcytosine methylation state of every base pair at five times in the cell cycle to show that 4,515 GANTC sites, recognized by the CcrM methyltransferase, change from full- to hemimethylation upon passage of the replication fork. Significantly, 27 of the GANTC sites are protected from methylation at all times. We also identified four previously unknown methylation motifs and the cognate methyltransferase for two of these motifs. The ability to track the state of the methylome in exquisite temporal detail will be invaluable to investigations of microbial epigenetic regulation.**

Author contributions: J.B.K., M.D.M., K.L., and T.A.C. designed research; J.B.K., M.D.M., K.L., T.A.C., M.B., S.W., B.Z., and D.G. performed research; K.L., T.A.C., M.B., S.W., B.Z., S.W.T., and J.K. contributed new reagents/analytic tools; J.B.K., M.D.M., K.L., T.A.C., M.B., S.W., B.Z., D.G., J.C., S.W.T., J.K., L.S., and H.H.M. analyzed data; and J.B.K., M.D.M., L.S., and H.H.M. wrote the paper.

L.S. is a non-executive member of the Pacific Biosciences board of directors. K.L., T.A.C., M.B., S.W., S.W.T., and J.K. are full-time employees at Pacific Biosciences, a company commercializing single molecule, real-time sequencing technologies.

[1]J.B.K., M.D.M., and K.L. contributed equally to this work.

[2]To whom correspondence may be addressed. E-mail: hmcadams@stanford.edu or shapiro@stanford.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1319315110/-/DCSupplemental.

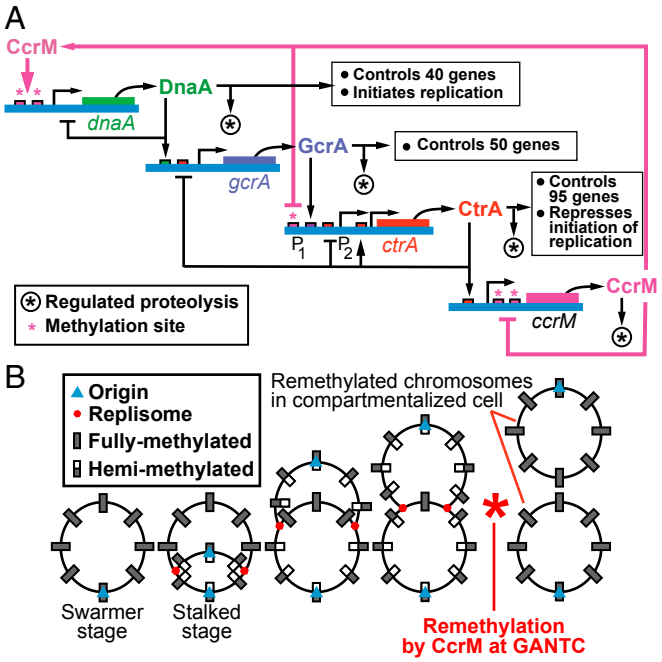


Fig. 1. (*A*) Four genes organized in a cyclical genetic circuit provide the core engine that drives the *Caulobacter* cell cycle (6, 15, 45). CcrM-catalyzed DNA methylation at GANTC sites is indicated in the promoter regions of *dnaA* and *ctrA*. The promoter for the origin-proximal *dnaA* gene is activated when the GANTC site is in the fully methylated state, and the more origin-distal P1 promoter for the *ctrA* gene is activated when in the hemimethylated state. (*B*) Model for the 6mA GANTC methylation state during chromosome replication. Chromosomes are drawn as circles, and the changing methylation state of GANTC sites versus position on the genome is indicated by the shaded rectangles. As chromosome replication proceeds bidirectionally from the origin, we expected that the chromosome would become increasingly hemimethylated due to the semiconservative nature of DNA replication and due to the absence of the DNA methyltransferase CcrM until the fully methylated state is restored near the end of replication when *ccrM* is expressed. Results reported in the text demonstrate the progressive wave of hemimethylation at GANTC sites with the passage of the replication fork.

The *dnaA* gene is near the origin region of the chromosome, and its promoter is in the fully methylated state and activated before replication initiation. Soon after replication initiation, the *dnaA* promoter becomes hemimethylated and less active until CcrM is expressed and methylates both new chromosomes near the end of DNA replication (Fig. 1*B*) (6). In contrast, the *ctrA*P1 promoter, which is farther from the origin than the *dnaA* promoter, is preferentially activated when its GANTC site becomes hemimethylated, presumably after the passage of the replication fork (8).

The circular *Caulobacter* chromosome is replicated once and only once per cell cycle (23). At the beginning of the cell cycle, in nonreplicating swarmer cells, the chromosome is fully methylated at GANTC sites (12). After replication initiation at the swarmer-to-stalked cell transition, the two replication forks proceed bidirectionally from the origin of replication (Fig. 1*B*) (24). Because the CcrM DNA methyltransferase is not present in the cell until near the end of replication, we expected that the two new DNA double helices would each remain in the hemimethylated state until the time of CcrM expression at the end of the cell cycle. (Fig. 1*B*). However, this dynamic cell-cycle pattern of GANTC site methylation has previously been demonstrated for only seven chromosomal loci (12, 23, 25). Cell cultures have also been monitored at 11 regions of the chromosome to show that the degree of hemimethylation at each locus correlates with chromosomal position (23).

To determine if progressive hemimethylation occurs throughout the chromosome coincident with the movement of the replication fork and to identify the methylation state of all GANTC sites in promoter regions of chromosomal ORFs as a function of the cell cycle, we sequenced the *Caulobacter* genome at five time points over the course of the cell cycle using single-molecule, real-time (SMRT) DNA sequencing (26). In SMRT sequencing, the rate of incorporation of each nucleotide is monitored, and two parameters, the pulse width (PW) and the interpulse duration (IPD), are used to describe the kinetics of this base incorporation. PW is the time that it takes to incorporate a single nucleotide, and IPD is the time between incorporation of two nucleotides. Because nucleotide incorporation is slowed when the DNA polymerase of SMRT sequencing encounters a modified base, the kinetics of nucleotide incorporation at each base position can be used to identify the locations of methylated bases. The ratio of the average IPD value of modified DNA at a given genomic position to the average IPD value of unmodified DNA at bases of similar sequence context is one useful metric for describing the extent of modification. Larger IPD ratios at a site indicate a modification at that site, whereas smaller IPD ratios indicate no modification. Different DNA modifications have distinct kinetic signatures, enabling not only the identification of the position of base modification, but also the type. SMRT sequencing for base modification detection has previously been used to identify methyltransferase motifs on plasmids and to identify the methylation status of entire bacterial genomes (27–30); however, no previous studies have investigated the cell cycle dynamics of the genome methylation state. We observed the progressive transformation of 4,515 dynamically methylated GANTC sites from the fully to hemimethylated state coincident with bidirectional progression of chromosome replication and identified the complete *Caulobacter* N^6-methyladenine (6mA) and 5-methylcytosine (5mC) methylome.

Results

Methylated Bases and Recognition Motifs in the *Caulobacter* Genome. SMRT sequencing was used to determine the genome-wide distribution of methylated bases in *Caulobacter* swarmer cells before the start of chromosome replication. Analysis of the sequencing kinetics indicated the presence of both 6mA and 5mC (*SI Appendix*, Table S1 and Fig. S1). Sequence context analysis of methylated bases resulted in the identification of three putative adenine methyltransferase recognition motifs (Fig. 2 *A*–*C*). In addition to the previously identified sequence context 5′-GANTC-3′ (bold face type marks the methylated adenine, and underlining marks the position of the methylated adenine in the complementary strand), resulting from the activity of the CcrM DNA methyltransferase (12), we observed two additional adenine methylation motifs: an asymmetric motif of 5′-CGACCAG-3′ and a characteristic Type I motif with the sequence context of 5′-CGAC(N7)TRGG-3′. Over 98% of these three motifs were determined to be highly methylated (*Materials and Methods*; *SI Appendix*, Table S1).

The locations of these motifs were mapped to the *Caulobacter* genome and categorized based on whether they were in an ORF or an intergenic region and whether in an essential region of the genome (31) (Table 1). In this classification, we required the methylated adenines on both strands of the motif to be fully contained in the essential region or ORF. Only the GANTC motif showed a bias toward location in intergenic regions: ~23% of the 4,542 GANTC sites are located in the 9.1% of the genome that is between ORFs. The two adenine motifs had only 2% and 5% of their sites, respectively, located within intergenic regions, which is reasonable because these methylation sites can be categorized as part of a restriction-modification system.

To obtain a robust SMRT sequencing kinetic signal for 5mC, the DNA was treated by Tet1 oxidation before sequencing (32). Sequencing of a Tet-treated sample allowed us to identify two 5mC sequence motifs, 5′-GGCGCC-3′ and 5′-YGCCGGCR-3′

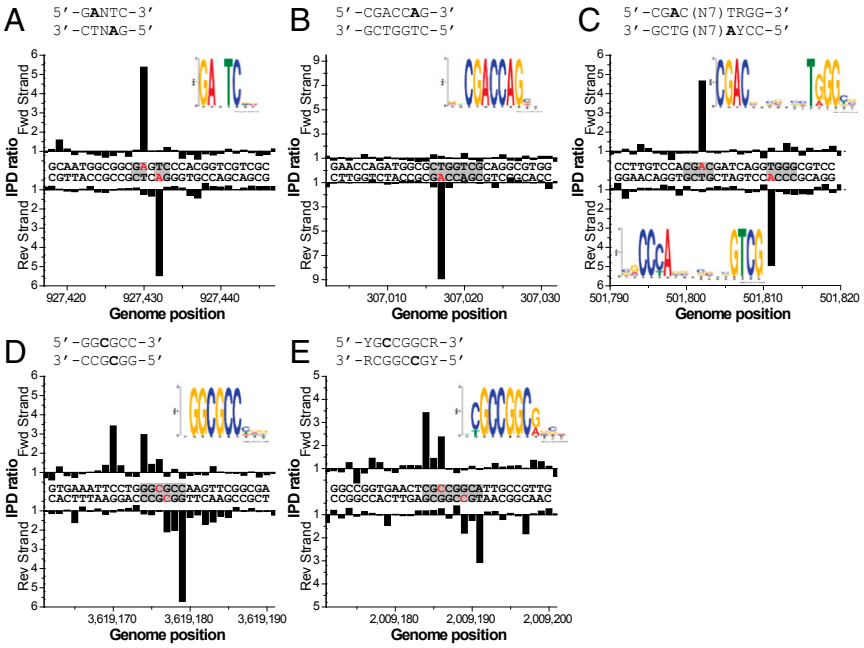


Fig. 2. DNA methylation motifs identified by SMRT sequencing. The methylated base within each motif is indicated by the large IPD ratios and the red letter in the DNA sequences. Motifs and representative IPD ratios: (*A*) GANTC sites. (*B* and *C*) Two 6mA motifs. (*D* and *E*) Two 5mC motifs.

(Y = C or T; R = A or G) (Fig. 2 *D* and *E*; *SI Appendix*, Table S1). The detection of the methylation state of the cytosine motifs yielded 54.4% methylated for GGCGCC and 75.6% for YGCCGGCR, as opposed to more than 98–99% for the detection of the methylation state of the 6mA motifs (*SI Appendix*, Table S1).

Adenine Methylome Dynamics During the Cell Cycle. The capability to obtain genome-wide detection of methylated adenine sites allowed us to observe changes in the DNA methylation state during the progression of chromosome replication at five different time points during the cell cycle as shown in Fig. 3*A*. The progression of chromosome replication as a function of the cell cycle is shown in Fig. 3*B*. Genome segments that have already been replicated have an increased amount of DNA in the sample, and thus these segments have a significantly increased coverage in the SMRT sequencing. Thus, we can estimate the position of the replication fork in samples from a synchronized cell population from the depth of sequencing coverage across the genome. As expected, the average sequencing coverage depth was uniform across the genome for the swarmer cell (SW) (Fig. 3*B*). However, after replication had initiated in the stalked cell (ST), the sequencing coverage was highest for the regions closest to the origin of replication. The samples from the later stages of the cell cycle, early predivisional 1 and 2 (EPD1, EPD2) and late predivisional (LPD), show the continued progression of the increased sequencing coverage as replication proceeds. The shape of the coverage transition region between the replicated and nonreplicated genome portions reflects the dispersion in synchrony of the cells in the sample.

Fig. 3*C* shows the IPD ratio for GANTC adenines at their various genomic positions for the five sampled cell-cycle time points. The average IPD ratio of sites in the hemimethylated state was much less than the average IPD ratio of those in the fully methylated state. In the swarmer cell stage at the top, both As in the GANTC sites were fully methylated, and the average IPD ratios were high across the entire genome. After passage of the replication fork, only the As on the parent strand of each double-stranded copy of GANTC motifs are methylated, reducing the IPD ratio at these sites until the As are remethylated later in the cell cycle. In stalked cells, after replication had started, the IPD ratio measured for adenines in GANTC sites near the origin decreased, whereas in regions not yet replicated, the IPD ratios maintained the higher value observed in the swarmer cell stage. An increasing length of the DNA becomes hemimethylated at the GANTC sites as chromosome replication continues along the genome through the EPD1 and EPD2

Table 1. Adenine motif location

Motif*	No.	Motifs in ORFs†	No.	Intergenic motifs†	No.
5′-G**A**NTC-3′	4,542	Essential	437	Essential	111
3′-CTN**A**G-5′		Nonessential	3,046	Nonessential	948
		Total	3,483	Total	1,059
5′-CGACC**A**G-3′	3,307	Essential	412	Essential	2
3′-GCTGGTC-5′		Nonessential	2,825	Nonessential	68
		Total	3,237	Total	70
5′-CG**A**C(N7)TRGG-3′	614	Essential	82	Essential	1
3′-GCTG(N7)**A**YCC-5′		Nonessential	501	Nonessential	30
		Total	583	Total	31

*Adenines in boldface type are the methylated bases in the motifs.
†Essential and nonessential regions are from ref. 31.

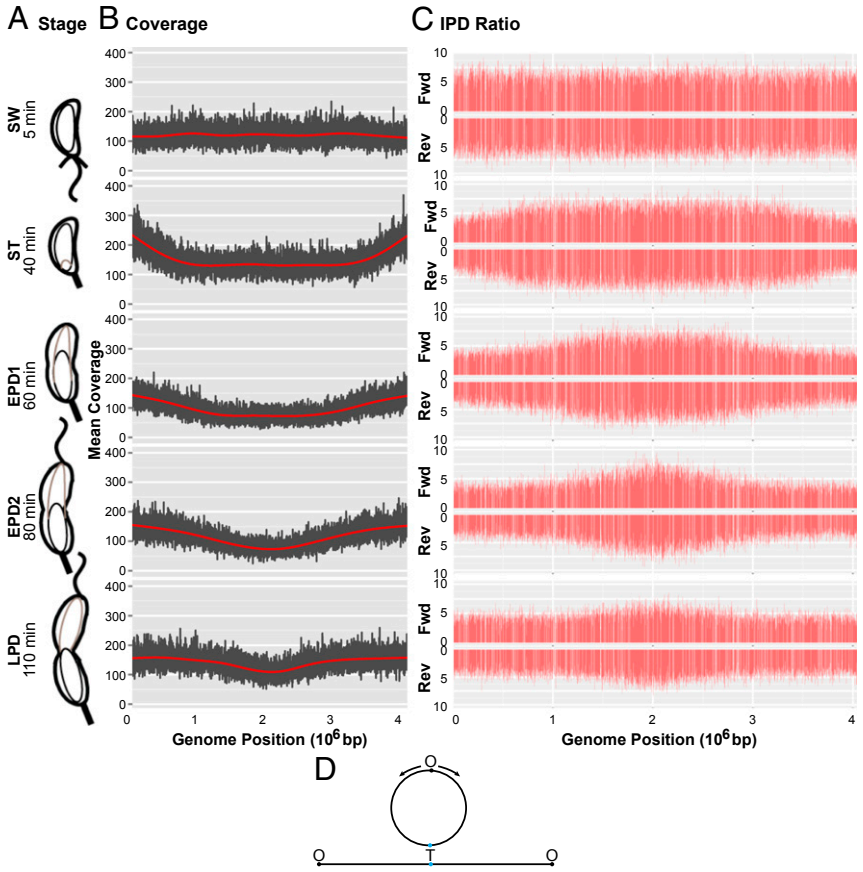


Fig. 3. Dynamic GANTC methylation profile of the genome as a function of the cell cycle. (*A*) Stages of the *Caulobacter* cell cycle. Progression of chromosome replication is indicated schematically within the cell. Chromosomal replication is initiated at the *Caulobacter* origin of replication (*cori*) at the swarmer to stalked cell transition. Thereafter, the two replication forks progress bidirectionally on the two arms of the chromosome from *cori* to the terminus. In *B* and *C*, this progression of the two forks moves from the right and left edges of the graphs toward the terminus in the center. The CcrM DNA methyltransferase is present only during a short cell-cycle interval near completion of chromosome replication in the late predivisional cell (LPD). (*B*) SMRT sequencing coverage versus genome position at five stages in the cell cycle. The bold red lines approximate the average coverage. The increased coverage of chromosomal loci as chromosome replication progresses extends bidirectionally from the origin. (*C*) Each vertical red line is the IPD ratio for a GANTC position. The transition from full methylation (larger IPD ratios) to hemimethylation (smaller IPD ratios) follows the bidirectional progression of the replication fork from the origin to the terminus. Forward-strand IPD ratios are are indicated by "Fwd," characterizing the "plus" strand. Reverse-strand IPD ratios are indicated by "Rev" and characterize the "minus" strand. (*D*) Diagram of the circular *Caulobacter* chromosome with the origin of DNA replication indicated as "O" and the terminus indicated as "T." The data shown in *B* and *C* are presented in a linearized chromosome cut at the origin.

cell-cycle stages (Fig. 3*C*). In the LPD stage, *ccrM* is expressed, and CcrM then begins to remethylate the hemimethylated GANTC sites so that the IPD ratios start to return to higher values. Thus, we observed a wave of increasing hemimethylation at the GANTC sites, reflecting the advancing replication fork in the cell-cycle samples taken during chromosome replication. The same pattern of genome transition from full to hemimethylation at GANTC sites as a function of DNA replication was observed upon analysis of the sequencing data using the percentage methylated metric (*Materials and Methods*; *SI Appendix*, Fig. S2).

It has been verified previously that there is a transition from full to hemimethylation at seven genomic locations (12, 23, 25). Here, we asked whether the decreased average IPD ratios observed after replication could in principle have arisen from averaging over two double-stranded DNA populations, one fully methylated and the other fully unmethylated, rather than from a single population of double-stranded hemimethylated DNA. The single-molecule nature of SMRT sequencing, coupled with its ability to sequence both strands of a DNA molecule in the same sequencing read, allows for demonstrating that the reduction in kinetic signal after passage of the replication fork is due to the generation of two double-stranded DNA molecules, each composed of one parental methylated strand and one daughter unmethylated strand. Correlation analysis of the kinetic signals from single molecules at the two target adenine positions *within* a GANTC′ site demonstrates that, after replication, one highly methylated strand (high IPD ratio) and one unmethylated strand (low IPD ratio) are present on the DNA molecules (Fig. 4), confirming that the decreased IPD ratios result from hemimethylated DNA. The additional CGACCAG and CGAC(N7)TRGG adenine methyltransferase recognition motifs that we identified did not change from full to hemimethylated over the course of the cell cycle (Fig. 5; *SI Appendix*, Figs. S3 and S4). This suggests that the CGACCAG and CGAC(N7)TRGG motifs are remethylated immediately after these regions of the chromosome have been replicated.

A Subset of GANTC Sites Remains Unmethylated Throughout the Cell Cycle. The statistical significance of IPD values for adenines is quantified by the kinetic score, which is $-10\log(P \text{ value})$ where the P value for each position is determined from a t test of the experimentally determined IPD values versus the control unmethylated IPD values (*Materials and Methods*). We observed that several GANTC sites had low kinetic scores at all five stages of the cell cycle, suggesting that they remain in the unmethylated

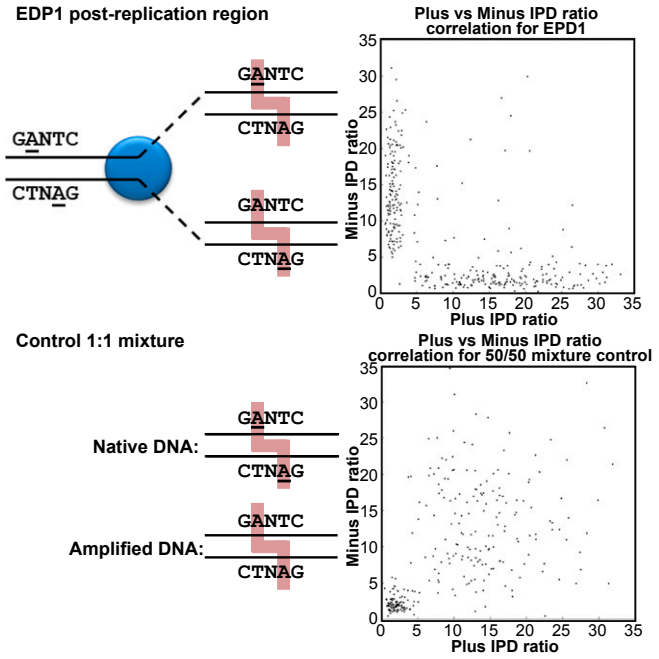


Fig. 4. Correlation analysis of the kinetic signals from single molecules at the two target adenine positions within a given GANTC site, demonstrating that, after replication, one highly methylated strand (high-IPD ratio) and one unmethylated strand (low-IPD ratio) are present on each DNA molecule and confirming that the decreased IPD ratios result from hemimethylated DNA. (*Upper*) The expected hemimethylation status after replication. The graph depicts IPD ratios for the plus and minus strands of GANTC sites from multiple subreads for single DNA molecules from a region 100 kb to the left and right of the origin of replication at the EPD1 cell-cycle stage, showing two populations of molecules, each with a high-IPD ratio on one strand and a corresponding low-IPD ratio on the opposite strand (i.e., hemimethylation). (*Lower*) A control showing that, when fully methylated, swarmer-cell-stage native DNA is artificially mixed with amplified (fully unmethylated) DNA and subjected to the same analysis. Two different populations are detected: one with high-IPD ratios on both strands (fully methylated) and one with low-IPD ratios on both strands (amplified DNA). Lower IPD ratios are due to hemimethylated DNA rather than a mixture of fully methylated and unmethylated DNA.

state. To further investigate the methylation state of these sites, we analyzed the SMRT sequencing data using two additional thresholds: one in which the cell-cycle average kinetic scores of the adenines of both strands of the GANTC motif are below 40 and one in which the cell-cycle average kinetic scores of both strands are below 70. A total of 27 GANTC motifs fall into either category and were classified as persistently unmethylated. These unmethylated motifs are located in a variety of genomic contexts (Fig. 6 *A*–*G*). Some are within ORFs, some are in intergenic regions between divergent genes, and some are both within ORFs and in the promoter regions of nearby genes (*SI Appendix*, Table S2). Of particular interest are three noncoding intergenic regions that were previously classified as "essential" because they were nondisruptable by transposon insertions (31) (Fig. 6*A*). These three regions contain unmethylated GANTC sites, and one of these regions contains three unmethylated GANTC sites (Fig. 6*A*). We deleted each of these three DNA regions that could not be disrupted by transposon mutagenesis and found that the deletion strains were viable. We interpret this result to mean that these sites are protected by some mechanism from both transposon insertion and CcrM-catalyzed methylation. We confirmed the unmethylated status of four of the 27 unmethylated sites by digesting the chromosome with HinfI, a restriction endonuclease that cleaves only at unmethylated GANTC sites by subcloning these unmethylated fragments into a vector backbone and sequencing the insert. The unmethylated sites found by this method in cultures grown in rich peptone yeast extract (PYE) media are listed in *SI Appendix*, Table S3. The additional putative unmethylated GANTC sites revealed by Hinf1 digestion may reflect the fact that unmethylated sites revealed by SMRT sequencing (*SI Appendix*, Table S2) were analyzed in cultures grown in minimal media.

Candidate Genes Whose Transcription May Be Modulated by DNA Methylation State. It has been reported previously that the methylation state of the GANTC motifs within the promoters of four *Caulobacter* genes—two genes mediating cell division, *mipZ* and *ftsZ*, and two master transcriptional regulators, *ctrA* and *dnaA*—modulates their transcriptional activation at specific times in the cell cycle (6, 8, 22). Progression of the replication fork through the promoter regions of these genes not only yields two hemimethylated promoters, but also affects each promoter's transcriptional activity. Because we know the time of activation of each cell-cycle–regulated transcription start site (TSS), the position of a GANTC site within the promoter region of the corresponding TSS, and the time window in the cell cycle when each GANTC site transitions from full to hemimethylation, we can identify genes that can be considered candidates for epigenetic regulation. Using criteria described in *Materials and Methods*, we identified 59 candidate genes (*SI Appendix*, Table S4). Among the candidates detected in this analysis were the expected *dnaA*, *ctrA*, and *ftsZ* genes. Several candidate genes are known to be involved in the regulation of cell-cycle progression.

Matching an Adenine Methyltransferase and a Cytosine Methyltransferase to Their Cognate Motifs. A search of the *Caulobacter* genome for genes homologous to known DNA methyltransferases in the REBASE database (33) identified six homologous genes (*SI Appendix*, Table S5) including the gene encoding CcrM. The motifs recognized by the other five methyltransferase genes had not been determined. One of these predicted methyltransferases is of Type I; the genes encoding the Type I DNA methyltransferases generally are in an operon with a gene encoding a specificity subunit used in conjunction with the methyltransferase as well as a gene encoding a cognate restriction enzyme. The other four are categorized as

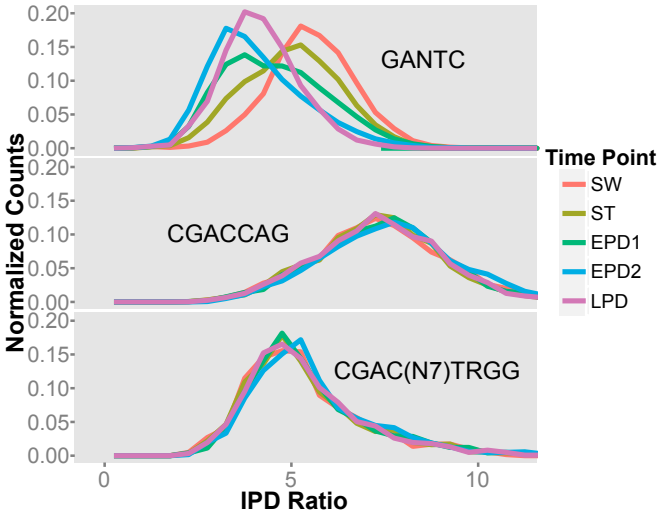


Fig. 5. Distributions of IPD ratios versus times in the cell cycle (colors) for three 6mA recognition motifs. Only the CcrM GANTC motif methylation state is dependent on the cell-cycle stage.

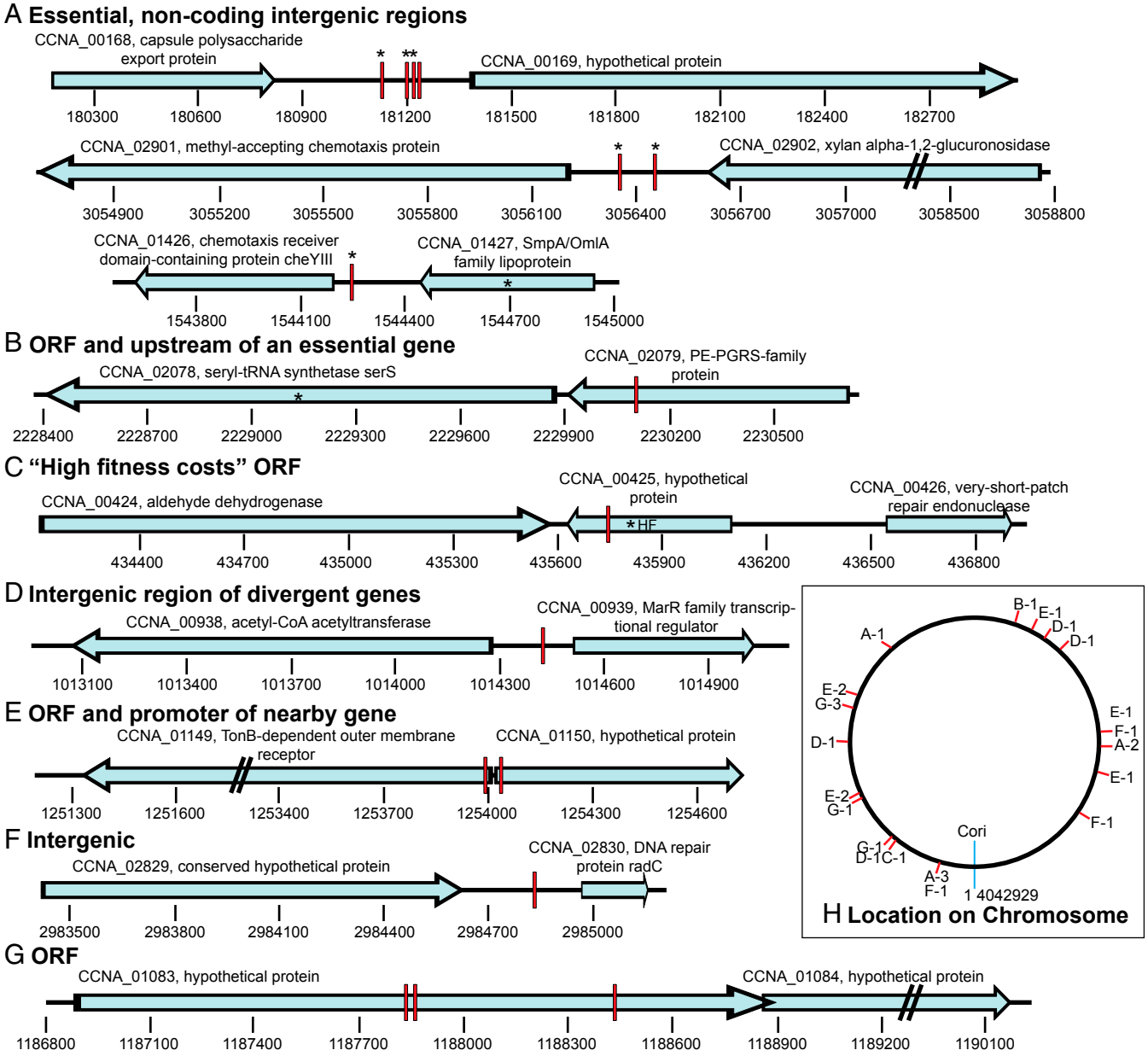


Fig. 6. Representative examples of unmethylated sites in alternative genomic location types. The classification of regions as "essential" or "high fitness cost" is as determined in ref. 31: (*A*) within three essential, noncoding intergenic regions; (*B*) within an ORF and upstream of an essential gene; (*C*) within a high-fitness-cost ORF; (*D*) in an intergenic region between divergent genes and in the promoters of both genes; (*E*) within an ORF and within the promoter of a nearby gene; (*F*) within a nonessential intergenic region; and (*G*) within an ORF. Arrows represent ORFs, and the red bars are unmethylated GANTC sites. Asterisks denote essential regions of the genome: an asterisk within an ORF denotes that the ORF is essential; an asterisk above a red bar represents an unmethylated GANTC in an essential region; and "*HF" denotes a high-fitness-cost gene. Diagonal lines represent part of a gene not shown. All diagrams are on the same genomic scale. (*H*) Red lines indicate the location of the 27 unmethylated GANTC sites on the circular chromosome. The type of genomic context is represented next to each red line in accordance with *A*–*G* with the number of unmethylated GANTC sites of that type at that location.

Type II methyltransferases. The Type II category includes orphan methyltransferases such as Dam and CcrM, as well as methyltransferases with a cognate restriction enzyme, but lacking a specificity factor. Two of the *Caulobacter* Type II methyltransferases (CCNA_01085 and CCNA_03741) are annotated in the NA1000 *Caulobacter crescentus* sequence as cytosine methyltransferases that lack a cognate restriction enzyme (34), and a third is an adenine methyltransferase with an associated restriction enzyme domain (CCNA_00869) (*SI Appendix*, Table S5). Of these putative DNA methyltransferase genes, only CcrM was found to be essential in rich media (31).

To identify the cognate methyltransferases for the newly identified motifs, we constructed plasmids that included potential target sequences and one of the predicted DNA methyltransferase genes, a procedure similar to a method described previously (29). Each motif was designed to have an overlapping methylation-sensitive restriction site (Fig. 7 *A* and *B*). These sites would be cleaved only if the motif was not methylated by the methyltransferase encoded in the plasmid construct. If, however, induction of the methyltransferase resulted in methylation of the motif, the restriction enzyme would not cleave at that site. Comparison of restriction digests of plasmids isolated from induced and uninduced

cultures then enabled us to pair each methyltransferase with its cognate motif (*Materials and Methods*).

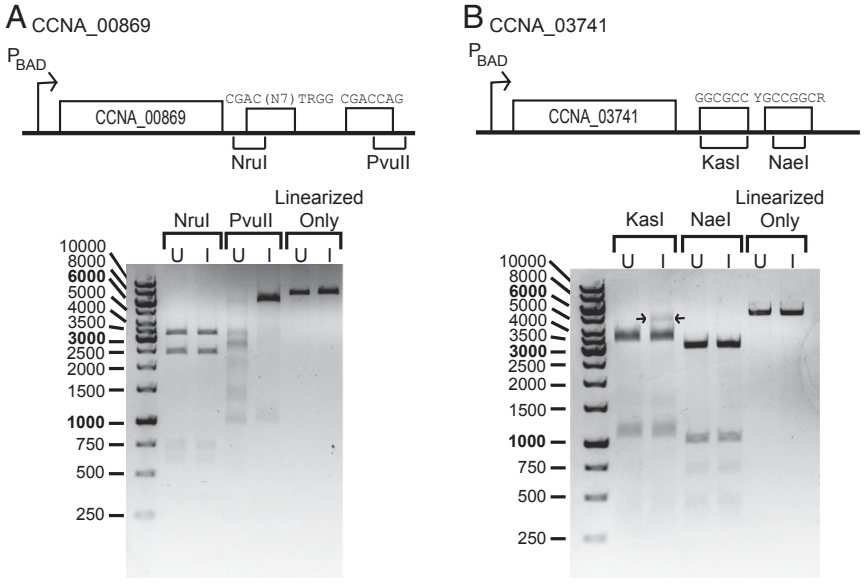
As shown in Fig. 7*A*, plasmid DNA isolated from cells containing CCNA_00869 was cut differently by the PvuII restriction endonuclease depending on whether the DNA was isolated from induced or uninduced cultures, but was not differentially cut by the NruI enzyme. [Gene references are to the *Caulobacter* NA1000 genome sequence (34).] Thus, the CGACCAG motif overlapping the designed PvuII site was methylated upon CCNA_00869 induction. The methylation state of the CGACCAG site was confirmed using SMRT sequencing, as indicated in Fig. 7*C*.

The GGCGCC motif was methylated by the methyltransferase encoded by CCNA_03741, as demonstrated by a difference in KasI cutting between DNA isolated from induced and uninduced cultures (Fig. 7*B*). The presence of the 5,000-bp band in the KasI digest of the plasmid DNA from the induced culture indicates that some of the plasmid isolated is methylated at the GGCGCC motif overlapping the KasI site, but the persistence of the 4,000- and 1,000-bp bands indicates that much of the plasmid isolated remained unmethylated at this site. This additional band was reproducible. Furthermore, SMRT sequencing of the plasmid DNA treated with Tet detected methylated cytosine in the GGCGCC motif, confirming that the protein encoded by CCNA_03741 methylated this motif (Fig. 7*C*).

Discussion

The promoter regions of the genes encoding DnaA and CtrA, two of the essential transcription factors in the *Caulobacter* cell-cycle control circuit (Fig. 1*A*), contain the GANTC motif. The *dnaA* gene is transcribed early in the cell cycle when the chromosome is in the fully methylated state (6), and the *ctrA* gene is preferentially transcribed later in the cell cycle when its promoter is in the hemimethylated state (8). This methylation-state dependence of *dnaA* and *ctrA* expression, along with the cell-cycle–dependent expression of the CcrM methyltransferase, contributes to synchronization of the *Caulobacter* cell-cycle control circuit with the progression of chromosome replication (6). Two additional genes, *ftsZ* and *mipZ*, encoding essential divisome genes with GANTC sites in their promoter regions, are preferentially activated when in the fully methylated state (22). Both of these genes are near the terminus of the chromosome. Their transcription levels decrease before the appearance of the CcrM DNA methyltransferase as DNA replication approaches completion and the two replicated strands of DNA become hemimethylated.

Here, we have used third-generation SMRT sequencing to show that there is a genome-wide progressive transition of GANTC methylation sites from the fully methylated state to the hemimethylated state coincident with passage of the replication fork. The GANTC sites then begin to be returned to the fully methylated state after a short interval of *ccrM* expression near



C

MTase	Differential Restriction	Motif	Confirmed by SMRT sequencing
CCNA_00869	PvuII	5′-CGACC**A**G-3′ 3′-GCTCGTC-5′	Yes
CCNA_03741	KasI	5′-GG**C**GCC-3′ 3′-CCG**C**GG-5′	Yes

Fig. 7. Genes encoding predicted DNA methyltransferases matched to their cognate motifs. (*A*) The methyltransferase encoded by CCNA_00869 methylates the second adenine in the CGACCAG motif. Diagram shows the plasmid construct with adenine motifs overlapping methylation-sensitive restriction sites downstream of the inducible CCNA_00869 methyltransferase gene. P_{BAD}, arabinose inducible promoter. (*Lower*) The 1.2% agarose gel shows the products of restriction digests of the plasmid DNA with the methyltransferase uninduced (U) or induced (I). (*B*) The methyltransferase encoded by CCNA_03741 methylates the first cytosine GGCGCC motif. Diagram shows the plasmid construct with cytosine motifs and overlapping methylation-sensitive restriction sites downstream of the inducible methyltransferase gene. (*Lower*) The 1.2% agarose gel shows restriction digests of the CCNA_03741 plasmid DNA as in *A*. In all lanes in both *A* and *B*, the plasmid DNA was linearized by EcoRV. (*C*) Summary of the methyltransferases that methylated each motif as shown by the restriction digest. We used SMRT sequencing of DNA isolated from induced cultures to detect the methylated bases on the plasmid. In both cases, the motifs detected as methylated by SMRT sequencing agreed with the restriction digest data, confirming that the CCNA_00869 and CCNA_03741 methyltransferases methylated the motifs shown.

the end of DNA replication. The single-molecule nature of SMRT sequencing allowed us to observe the global transition of each base pair from full to hemimethylation at GANTC sites. Thus, our results demonstrate that cell-cycle dependence of the hemi- or fully methylated state of GANTC site methylation is indeed a global phenomenon. The identification of the methylation state of each GANTC site over the course of the cell cycle, combined with identification of the global location of and activity of each transcription start site on the genome, enabled identification of 59 candidate genes whose expression may be affected by the methylation state of GANTC sites in their promoter regions. Notably, several of these genes play important roles in the regulation of cell-cycle progression.

Of the 4,542 GANTC sites in the *Caulobacter* chromosome, 4,515 sites had dynamically methylated adenines, and 27 GANTC sites were considered unmethylated throughout the cell cycle under our growth conditions. The unmethylated sites are in coding sequences, in intergenic regions, and in three noncoding intergenic DNA segments with no known regulatory function that are nondisruptable by transposon mutagenesis (31). When each of these three nondisruptable DNA segments was deleted, the strains remained viable, suggesting that either these DNA segments have a complex secondary structure that occludes both CcrM and transposon insertion or that a DNA-binding protein protects these segments from both methylation and transposon insertion. There are several examples in *E. coli* where a protein protects a site from Dam methylation, including the SeqA protection of methylation sites in the origin (35) and OxyR protection of three GATC sites in the 5′ UTR of *agn43* (36–39).

SMRT sequencing of the *Caulobacter* genome resulted in the identification of two 6mA motifs whose methylation states did not change over the course of the cell cycle. For one of these cases, we showed that the CCNA_00869 gene, predicted to function as a DNA methyltransferase with a cognate restriction domain, methylates the CGACCAG motif. The other 6mA motif is a Type I methyltransferase motif, characterized by a bipartite sequence separated by a nonspecific spacer of 7 bases. Because the methyltransferase encoded by CCNA_00656 is homologous to canonical Type I methyltransferases (33) and is in an operon with canonical Type I specificity and restriction subunits, we predict that this methyltransferase methylates the adenine in CGAC(N7)TRGG sequences. Because both of the adenine methyltransferases can be categorized as restriction-modification systems, it is not surprising that the methylation state of these sites does not vary with the cell cycle, as lack of methylation would make the DNA vulnerable to cleavage by the corresponding restriction endonuclease.

In addition to the adenine methylation in *C. crescentus*, we observed genomic cytosine methylation, previously reported in *Caulobacter bacteroides* (40). Two 5mC motifs were identified, and we demonstrated that methylation of the GGCGCC motif was performed by the methyltransferase encoded by CCNA_03741. Because only one other gene in the *Caulobacter* genome is homologous to cytosine methyltransferases and there is only one other cytosine motif, we predict that the methyltransferase encoded by CCNA_01085 methylates at YGCCGGCR motifs. Both cytosine motifs exhibited decreased levels of methylation, possibly due to inefficient Tet conversion used to detect 5mC modification. However, we found that Tet conversion was very efficient in an *E. coli* sample prepared in parallel with the *Caulobacter* sample, suggesting that a mechanism may protect these sites from methylation.

In conclusion, we have identified all 6mA and 5mC motifs at base-pair resolution in the *C. crescentus* genome. The 6mA methylome contains a motif, GANTC, which transitions sequentially from full to hemimethylation coincident with replication fork progression, and this transition in methylation state is known to modulate promoter activity. We also identified a subset of GANTC motifs that remain in the unmethylated state at all times in the cell cycle.

Materials and Methods

Bacterial Strains and Growth Conditions. *Caulobacter* cells were grown at 28 °C in PYE or M2-glucose minimal media (M2G) (41) with kanamycin (25 μg/mL solid media, 5 μg/mL liquid media) or no antibiotic as indicated. *E. coli* cells were cultured at 37 °C in Luria–Bertani broth media supplemented with kanamycin (50 μg/mL solid media, 30 μg/mL liquid media), chloramphenicol (30 μg/mL solid media, 20 μg/mL liquid media), or ampicillin (100 μg/mL solid media, 50 μg/mL liquid media) when appropriate. Strains and plasmids used are listed in *SI Appendix*, Tables S6 and S7, respectively.

Synchrony. Small-scale synchrony was performed as described previously (42). To have sufficient volume to remove large aliquots at each time point, swarmer cells were pooled from 16 small-scale synchronies. Cells were allowed to proceed synchronously through the cell cycle at 28 °C in M2G. Aliquots (4 mL) were removed for DNA isolation at the following time points: 5 min (swarmer), 40 min (stalked), 60 min (early predivisional cell 1), 80 min (early predivisional cell 2), and 110 min (late predivisional cell).

Genomic DNA Preparation. Genomic DNA was isolated using the Gentra Puregene Yeast/Bacteria Kit (Qiagen). Cells were removed from synchronous cultures growing in M2G. Genomic DNA was isolated from 4 mL of cells at each time point. The protocol provided with the kit was followed with a few modifications: 4 μL RNaseA was used instead of 1.5 μL, and after protein precipitation the samples were left on ice for 15 min and pelleted for 5 min at 4 °C. The DNA pellet was allowed to dry for 10 min and then resuspended in 25 μL 10 mM Tris, pH 8.0. Samples were incubated at 65 °C for 1 h and at room temperature overnight.

Cloning. Standard procedures were followed to insert each methyltransferase under the control of the arabinose-inducible promoter along with downstream methylation motifs and overlapping restriction enzymes. Plasmids were transformed into ER2925 that contained no Dam or Dcm methylation activity.

To verify the locations of unmethylated GANTC sites, 1 μg of *Caulobacter* NA1000 genomic DNA, extracted from cells cultivated in PYE, was digested with HinfI, filtrated on a Millipore membrane, and ligated to prehybridized HinfI-CasA1-XbaI/HinfI-CasA2 adaptors (1 μM). A secondary digestion was performed with PstI or SalI plus XhoI (Promega), and the products were cloned into pUC21 digested with PstI and XbaI or pGEMT digested with SpeI and SalI, respectively. Colonies containing a plasmid with an insert were selected by a blue/white screen, and the presence of an adaptor was confirmed by PCR with the CasA-specific and M13R primers. Sanger sequencing was performed with the M13F primer. The sequencing results were mapped to the *Caulobacter* genome using BLAST.

Three of the nondisruptable regions of the *Caulobacter* chromosome contain GANTC motifs that remain unmethylated throughout the cell cycle when grown in minimal media. These three regions of the chromosome are Gap_0007, Gap_0042, and Gap_0074 (31). We constructed *Caulobacter* NA1000 strains in which one of each of the Gap regions is deleted from the chromosome using double homologous recombination. Deletion candidates were PCR-screened and sequenced to identify and verify deletion strains.

Induction and Restriction Digests. Cells were induced with 0.2% arabinose for 2 h at 37 °C with shaking. As a control, cultures were also grown to the same density without arabinose. After induction, plasmid DNA was isolated with a Qiagen Miniprep kit and eluted from the column with water. All plasmids were linearized with EcoRV and cleaved with the appropriate restriction endonucleases for assaying methylation state: NruI, PvuII, KasI, or NaeI (New England Biolabs). Digests were performed for 2 h at 37 °C and run on 1.2% agarose gels.

Sequencing. SMRTbell template libraries were prepared as previously described (29, 43). Two differently sized SMRTbell template libraries were used. Genomic DNA samples were either sheared to an average size of ~800 bp via adaptive focused acoustics (Covaris) or to a target size of ~8–10 kb using Covaris g-TUBEs. Plasmid DNA was sheared to ~800 bp. Fragmented DNA was then end-repaired and ligated to hairpin adapters. Incompletely formed SMRTbell templates were digested with a combination of Exonuclease III (New England Biolabs) and exonuclease VII (Affymetrix). SMRT Sequencing was carried out on the Pacific Biosciences RS using C2 chemistry with standard protocols for either small- or large-insert SMRTbell template libraries.

Enhancement of the 5mC kinetic signatures was carried out as previously described (32). Plasmid DNA was fragmented to ~800 bp, and 5mC residues were converted to 5caC using the 5mC Tet1 Oxidation Kit (WiseGene). Tet1-treated fragmented DNA was purified with Agencourt AMPure XP Beads (Beckman Coulter). SMRTbell template libraries were then made as described above.

Analysis of SMRT Sequencing Results. Sequencing reads were processed and mapped to the respective reference sequences using the BLASR mapper (http://www.pacbiodevnet.com/SMRT-Analysis/Algorithms/BLASR) and the Pacific Biosciences' SMRT Analysis pipeline (http://www.pacbiodevnet.com/SMRT-Analysis/Software/SMRT-Pipe) using the standard mapping protocol. Upon mapping the sequence reads to the reference genome, several SNPs were detected relative to the *C. crescentus* reference genome (GenBank accession no. CP001340.1) (*SI Appendix*, Table S9).

Kinetic signals acquired during SMRT sequencing were measured as previously described (26) and processed as described (27) for all pulses aligned to each position in the reference sequence. To identify modified positions, we used Pacific Biosciences' SMRTPortal analysis platform, v. 1.3.1, which uses an in silico kinetic reference and a *t*-test–based kinetic score detection of modified base positions as described at http://www.pacb.com/pdf/TN_Detecting_DNA_Base_Modifications.pdf. A threshold value of 100 for the log-transformed *P* value from the *t* test [called kinetic score = −10 log(*P* value)] at each reference position was used for assigning the given position as methylated. This value was chosen based on the kinetic distribution observed in the data, showing a clear bimodal distribution arising from unmodified and methylated positions (*SI Appendix*, Fig. S1). The clustering of the sequence motifs was done using PacBio's Motif-Finder (https://github.com/PacificBiosciences/DevNet/wiki/MotifFinder). Additional data analysis was carried out in R. Figures were generated using Origin 8.1 (Origin Lab); logos plots were generated using MEME-Chip (44).

Percentage Methylated Metric. Another metric that quantifies the nucleotide methylation state is the percentage methylated (*SI Appendix*, Fig. S2). This metric is a maximum-likelihood estimate of the proportion of adenines modified at a certain position and time point, and it is calculated using mixture model analysis of the pooled kinetic data for a given adenine and time point.

Method for Determining Candidate Genes Whose Transcription May Be Modulated by DNA Methylation State. We have identified 109 TSSs that are both cell-cycle–regulated and have at least one GANTC site within 80 bp of the TSS. Transcriptional start sites were determined by identifying RNAs that carry a 5′ triphosphate group in exponential-growth-phase mixed populations using 5′ RACE and Illumina high-throughput sequencing. Cell-cycle profiles of transcriptional start sites were determined by performing 5′ RACE on synchronized cell populations that progress from swarmer cells to divided cells and by measuring relative RNA abundance using Illumina high-throughput sequencing. We then determined which of these TSSs are preceded by promoters that become hemimethylated in a window of the cell cycle that coincides with the window of time during which there is significant change in transcriptional levels at that TSS. The window of time when an adenine in a GANTC site upstream of a TSS changes from full to hemimethylation was taken as the interval in which the percentage methylated value of an adenine drops most dramatically from the previous time point. In the final step of the analysis, a TSS was considered a candidate if the time window of hemimethylation of at least one GANTC adenine upstream of a TSS matched the time interval during which that TSS's transcriptional level changed significantly during the cell cycle. Twofold or smaller increases in transcriptional activity of a TSS were considered insignificant. Those TSSs whose maximum fold change in transcription occurred after 100 min were not considered in the analysis. Unreplicated adenines near the terminus, as well as all poorly methylated adenines (score <30 at any time point), were also discarded. Any of the remaining TSSs whose transcriptional levels changed most dramatically between 0 and 20 min qualified as candidates if simply positioned within 100,000 bases of the origin.

Construction of Plasmid Test Systems to Match Methyltransferases to Their Cognate Motifs. We constructed two plasmid test systems: one for adenine methylation (Fig. 7*A*) and one for cytosine methylation (Fig. 7*B*). The test plasmid for adenine methylation contained a predicted adenine methyltransferase gene under the control of an arabinose-inducible promoter or an empty vector. Two adenine motifs, with overlapping methylation-sensitive restriction sites, were placed downstream of the methyltransferase gene (Fig. 7*A*). In this construct, NruI overlaps the Type I CGAC(N7)TRGG motif and PvuII overlaps the CGACCAG motif. Treatment of the plasmid with the methylation-sensitive NruI and PvuII enzymes would cut the plasmid DNA only if their respective recognition sites were unmethylated. The plasmids bearing the 5mC motifs were designed similarly, but with a predicted cytosine methyltransferase gene and the two cytosine motifs with overlapping methylation-sensitive restriction sites, KasI for GGCGCC and NaeI for YGCCGGCR (Fig. 7*B*). Both of these enzymes cleave only at unmethylated motifs. Each plasmid was inserted in an *E. coli* strain (ER2925) lacking both *dam* and *dcm* methyltransferase activity. Plasmid DNA was isolated from cultures in which the methyltransferase had been induced with arabinose as well as from uninduced cultures as a control. DNA was then linearized with the EcoRV restriction endonuclease that cut each plasmid only once, and restriction digests were performed with the enzymes shown in Fig. 7. If the downstream motif on the plasmid DNA was methylated in the induced culture, then the restriction site overlapping the cognate methylated motif would not be cleaved. We would thus have fewer bands and bands of higher molecular weight in the lane with the induced methyltransferase gene, compared with similarly digested DNA from an uninduced culture.

ACKNOWLEDGMENTS. This research was funded by Department of Energy Grant DE-FG0205ER64136 (to H.H.M.) and National Institutes of Health Grant RO1-GM051426 (to L.S.). M.D.M. was funded by National Science Foundation Predoctoral Fellowship DGE-114747, and B.Z. was funded by a Stanford Graduate Fellowship. This work was also supported in part by National Human Genome Research Institute Grant 1RC2HG005618-01 to Pacific Biosciences and by University of Lausanne and the Swiss National Science Foundation Fellowships 3100A0_122541 and 31003A_140758 (to J.C.).

1. Low DA, Casadesús J (2008) Clocks and switches: Bacterial gene regulation by DNA adenine methylation. *Curr Opin Microbiol* 11(2):106–112.
2. Casadesús J, Low D (2006) Epigenetic gene regulation in the bacterial world. *Microbiol Mol Biol Rev* 70(3):830–856.
3. Wion D, Casadesús J (2006) N6-methyl-adenine: An epigenetic signal for DNA-protein interactions. *Nat Rev Microbiol* 4(3):183–192.
4. Collier J (2009) Epigenetic regulation of the bacterial cell cycle. *Curr Opin Microbiol* 12 (6):722–729.
5. Srikhanta YN, Fox KL, Jennings MP (2010) The phasevarion: Phase variation of type III DNA methyltransferases controls coordinated switching in multiple genes. *Nat Rev Microbiol* 8(3):196–206.
6. Collier J, McAdams HH, Shapiro L (2007) A DNA methylation ratchet governs progression through a bacterial cell cycle. *Proc Natl Acad Sci USA* 104(43):17111–17116.
7. Collier J, Shapiro L (2009) Feedback control of DnaA-mediated replication initiation by replisome-associated HdaA protein in *Caulobacter*. *J Bacteriol* 191(18):5706–5716.
8. Reisenauer A, Shapiro L (2002) DNA methylation affects the cell cycle transcription of the CtrA global regulator in *Caulobacter*. *EMBO J* 21(18):4969–4977.
9. Kahng LS, Shapiro L (2001) The CcrM DNA methyltransferase of *Agrobacterium tumefaciens* is essential, and its activity is cell cycle regulated. *J Bacteriol* 183(10): 3065–3075.
10. Robertson GT, et al. (2000) The *Brucella abortus* CcrM DNA methyltransferase is essential for viability, and its overexpression attenuates intracellular replication in murine macrophages. *J Bacteriol* 182(12):3482–3489.
11. Ichida H, Matsuyama T, Abe T, Koba T (2007) DNA adenine methylation changes dramatically during establishment of symbiosis. *FEBS J* 274(4):951–962.
12. Zweiger G, Marczynski G, Shapiro L (1994) A *Caulobacter* DNA methyltransferase that functions only in the predivisional cell. *J Mol Biol* 235(2):472–485.
13. Wright R, Stephens C, Zweiger G, Shapiro L, Alley MR (1996) *Caulobacter* Lon protease has a critical role in cell-cycle control of DNA methylation. *Genes Dev* 10(12): 1532–1542.
14. Zhou P, et al. (1997) Gene transcription and chromosome replication in *Escherichia coli*. *J Bacteriol* 179(1):163–169.
15. McAdams HH, Shapiro L (2009) System-level design of bacterial cell cycle control. *FEBS Lett* 583(24):3984–3991.
16. McAdams HH, Shapiro L (2011) The architecture and conservation pattern of whole-cell control circuitry. *J Mol Biol* 409(1):28–35.
17. Hottes AK, Shapiro L, McAdams HH (2005) DnaA coordinates replication initiation and cell cycle transcription in *Caulobacter crescentus*. *Mol Microbiol* 58(5): 1340–1353.
18. Holtzendorff J, et al. (2004) Oscillating global regulators control the genetic circuit driving a bacterial cell cycle. *Science* 304(5673):983–987.
19. Laub MT, Chen SL, Shapiro L, McAdams HH (2002) Genes directly controlled by CtrA, a master regulator of the *Caulobacter* cell cycle. *Proc Natl Acad Sci USA* 99(7): 4632–4637.
20. Tan MH, Kozdon JB, Shen X, Shapiro L, McAdams HH (2010) An essential transcription factor, SciP, enhances robustness of *Caulobacter* cell cycle regulation. *Proc Natl Acad Sci USA* 107(44):18985–18990.
21. Nierman WC, et al. (2001) Complete genome sequence of *Caulobacter crescentus*. *Proc Natl Acad Sci USA* 98(7):4136–4141.

22. Gonzalez D, Collier J (2013) DNA methylation by CcrM activates the transcription of two genes required for the division of *Caulobacter crescentus*. *Mol Microbiol* 88(1): 203–218.
23. Marczynski GT (1999) Chromosome methylation and measurement of faithful, once and only once per cell cycle chromosome replication in *Caulobacter crescentus*. *J Bacteriol* 181(7):1984–1993.
24. Dingwall A, Shapiro L (1989) Rate, origin, and bidirectionality of *Caulobacter* chromosome replication as determined by pulsed-field gel electrophoresis. *Proc Natl Acad Sci USA* 86(1):119–123.
25. Stephens C, Reisenauer A, Wright R, Shapiro L (1996) A cell cycle-regulated bacterial DNA methyltransferase is essential for viability. *Proc Natl Acad Sci USA* 93(3): 1210–1214.
26. Flusberg BA, et al. (2010) Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods* 7(6):461–465.
27. Clark TA, et al. (2012) Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. *Nucleic Acids Res* 40(4):e29.
28. Fang G, et al. (2012) Genome-wide mapping of methylated adenine residues in pathogenic *Escherichia coli* using single-molecule real-time sequencing. *Nat Biotechnol* 30(12):1232–1239.
29. Murray IA, et al. (2012) The methylomes of six bacteria. *Nucleic Acids Res* 40(22): 11450–11462.
30. Lluch-Senar M, et al. (2013) Comprehensive methylome characterization of *Mycoplasma genitalium* and *Mycoplasma pneumoniae* at single-base resolution. *PLoS Genet* 9(1):e1003191.
31. Christen B, et al. (2011) The essential genome of a bacterium. *Mol Syst Biol* 7:528.
32. Clark TA, et al. (2013) Enhanced 5-methylcytosine detection in single-molecule, real-time sequencing via Tet1 oxidation. *BMC Biol* 11(1):4.
33. Roberts RJ, Vincze T, Posfai J, Macelis D (2010) REBASE—a database for DNA restriction and modification: Enzymes, genes and genomes. *Nucleic Acids Res* 38(Database issue):D234–D236.
34. Marks ME, et al. (2010) The genetic basis of laboratory adaptation in *Caulobacter crescentus*. *J Bacteriol* 192(14):3678–3688.
35. Waldminghaus T, Skarstad K (2009) The *Escherichia coli* SeqA protein. *Plasmid* 61(3): 141–150.
36. Henderson IR, Owen P (1999) The major phase-variable outer membrane protein of *Escherichia coli* structurally resembles the immunoglobulin A1 protease class of exported protein and is regulated by a novel mechanism involving Dam and *oxyR*. *J Bacteriol* 181(7):2132–2141.
37. Haagmans W, van der Woude M (2000) Phase variation of Ag43 in *Escherichia coli*: Dam-dependent methylation abrogates OxyR binding and OxyR-mediated repression of transcription. *Mol Microbiol* 35(4):877–887.
38. Waldron DE, Owen P, Dorman CJ (2002) Competitive interaction of the OxyR DNA-binding protein and the Dam methylase at the antigen 43 gene regulatory region in *Escherichia coli*. *Mol Microbiol* 44(2):509–520.
39. Wallecha A, Munster V, Correnti J, Chan T, van der Woude M (2002) Dam- and OxyR-dependent phase variation of agn43: Essential elements and evidence for a new role of DNA methylation. *J Bacteriol* 184(12):3338–3347.
40. Degnen ST, Morris NR (1973) Deoxyribonucleic acid methylation and development in *Caulobacter* bacteroides. *J Bacteriol* 116(1):48–53.
41. Ely B, Johnson RC (1977) Generalized transduction in *Caulobactor crescentus*. *Genetics* 87(3):391–399.
42. Goley ED, et al. (2011) Assembly of the *Caulobacter* cell division machine. *Mol Microbiol* 80(6):1680–1698.
43. Travers KJ, Chin CS, Rank DR, Eid JS, Turner SW (2010) A flexible and efficient template format for circular consensus sequencing and SNP detection. *Nucleic Acids Res* 38(15):e159.
44. Machanick P, Bailey TL (2011) MEME-ChIP: Motif analysis of large DNA datasets. *Bioinformatics* 27(12):1696–1697.
45. Collier J, Murray SR, Shapiro L (2006) DnaA couples DNA replication and the expression of two cell cycle master regulators. *EMBO J* 25(2):346–356.