Published online 14 October 2020 *Nucleic Acids Research, 2020, Vol. 48, No. 20* ***11589–11601***
doi: 10.1093/nar/gkaa844

Cell cycle regulated DNA methyltransferase: fluorescent tracking of a DNA strand-separation mechanism and identification of the responsible protein motif

Olivia Konttinen[1,†], Jason Carmody[2,†], Sarath Pathuri[2,†], Kyle Anderson[2], Xiaofeng Zhou[3,4] and Norbert Reich[1,2,*]

[1]Biomolecular Science and Engineering, University of California Santa Barbara, Santa Barbara, CA 93103, USA, [2]Chemistry and Biochemistry, University of California Santa Barbara, Santa Barbara, CA 93103, USA, [3]Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA 94305, USA and [4]Chan Zuckerberg Biohub, San Francisco, CA 94158, USA

Received July 21, 2020; Revised September 10, 2020; Editorial Decision September 17, 2020; Accepted September 23, 2020

ABSTRACT

DNA adenine methylation by *Caulobacter crescentus C*ell *C*ycle *R*egulated *M*ethyltransferase (CcrM) is an important epigenetic regulator of gene expression. The recent CcrM-DNA cocrystal structure shows the CcrM dimer disrupts four of the five base pairs of the (5′-GANTC-3′) recognition site. We developed a fluorescence-based assay by which Pyrrolo-dC tracks the strand separation event. Placement of Pyrrolo-dC within the DNA recognition site results in a fluorescence increase when CcrM binds. Non-cognate sequences display little to no fluorescence changes, showing that strand separation is a specificity determinant. Conserved residues in the C-terminal segment interact with the phospho-sugar backbone of the non-target strand. Replacement of these residues with alanine results in decreased methylation activity and changes in strand separation. The DNA recognition mechanism appears to occur with the Type II M.HinfI DNA methyltransferase and an ortholog of CcrM, BabI, but not with DNA methyltransferases that lack the conserved C-terminal segment. The C-terminal segment is found broadly in N4/N6-adenine DNA methyltransferases, some of which are human pathogens, across three Proteobacteria classes, three other phyla and in *Thermoplasma acidophilum*, an Archaea. This Pyrrolo-dC strand separation assay should be useful for the study of other enzymes which likely rely on a strand separation mechanism.

INTRODUCTION

Bacterial DNA methylation is involved in diverse functions including restriction/modification (R/M), the control of gene regulation, mismatch repair and replication timing (1). DNA methyltransferases (DNA MTases) can be organized by the reactions they catalyze (N6-adenine and N4-cytosine exocyclic amine methylation, C5-cytosine methylation), by their organization of conserved motifs (α, β, γ, ε, ζ) and whether they form part of a R/M system (e.g. M.EcoRI) or are 'orphan' enzymes that do not appear to have a partner endonuclease (e.g. cell cycle regulated methyltransferase, CcrM, DNA adenine methyltransferase, Dam) (2). DNA MTases methylate a specific recognition sequence, using the cofactor S-adenosyl methionine (AdoMet) to deliver a methyl group. The R/M enzymes comprise the majority of known MTases, whereas the CcrM enzymes are found in α-proteobacteria and Dam enzymes are found in α-proteobacteria (3,4). In addition to a set of highly conserved motifs, DNA MTases typically have a larger domain with conserved motifs, which binds AdoMet as well as stabilizes the base that undergoes methylation in an extrahelical position (base flipping) (5). The smaller domain frequently contains the target recognition domain (TRD), although residues distributed throughout the MTases contribute to DNA recognition (5). In some cases (e.g. M.HhaI), there is extensive conformational communication between these two domains in terms of DNA recognition, stabilization of the extrahelical base and the correct assembly of the active site (6–9).

The CcrM enzyme found in *Caulobacter crescentus* (CcrM) has homologs throughout the α-class of proteobacteria; these enzymes are important for several organisms,

*To whom correspondence should be addressed. Tel: +1 805 893 8368; Fax: +1 805 893 4120; Email: reich@chem.ucsb.edu
†The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.

including the human pathogen *Brucella abortus* (10–12). Based on the arrangement of conserved motifs, CcrM is a β-class adenine N6-MTase. CcrM (358 amino acids) and some β-class adenine N6-MTases, which recognize 5′GANTC3′ sites, have an additional C-terminal 80-residue extension of unknown function (13,14). CcrM displays unusual activities, which formed the basis of our hypothesis that it may rely on a novel DNA recognition mechanism (13,14). The enzyme is more discriminating ($k_{methylation}/K_D{}^{DNA}$) than other DNA MTases by many orders of magnitude with double-stranded (ds) DNA, and methylates single-stranded (ss) DNA efficiently, but with much less discrimination than dsDNA (13,14). Based on protein engineering efforts of CcrM and related enzymes (14), the highly conserved C-terminal segment of the protein is involved in sequence discrimination.

DNA MTases generally adhere to the recognition mechanisms now established for proteins that recognize unique DNA sequences, with some notable variations. Most importantly, and now demonstrated for most DNA MTases as well as other classes of enzymes, they stabilize their target base into an extrahelical position (base flipping) to gain stereochemical access for the delivery of the methyl group (5–9). Base flipping contributes to specificity since only when bound to the cognate sequence methyltransferases undergo conformational changes that facilitate base flipping (8,9). Numerous cocrystal structures and related functional studies have confirmed that DNA recognition by DNA MTases involves direct and indirect protein–DNA interactions with both strands of DNA (5). Interestingly, while the vast majority of DNA MTases act exclusively or largely on double-stranded DNA, a few reports describe enzymes whose biological role is to act on single-stranded DNA (15), or minimally, can methylate both single-stranded DNA and unpaired DNA *in vitro*, including human enzymes (16).

The recently described CcrM-DNA cocrystal structure provides insights into a new recognition mechanism, relying on the strand separation of four of the five base pairs within the enzyme's recognition site (17). Importantly, this strand displacement goes well beyond the base flipping mechanism observed with other DNA MTases, and unlike CRISPR-Cas9, recognition of the target strand relies solely on interactions with amino acids rather than nucleic acid hybridization. Each monomer has a core (residues 1–254), a flexible linker (residues 254–274) and a highly conserved C-terminal 83-residue segment (residues 275–358). Monomer A (cyan, Figure 1) makes nearly all of the contacts to the 'target' strand which is positioned to be methylated. Interactions to four of the five recognized bases are only possible because of the complete disruption of base pairing to the opposite strand. The 83-residue C-terminal segment of monomer A is connected through a disordered linker and faces away from the DNA. In contrast, monomer B interacts nearly exclusively with the non-target strand through interactions with phosphates. In spite of the detailed structural information, the underlying mechanism of the strand separation process and its relationship to specificity remain largely obscure. Here, we apply a fluorescence-based assay to interrogate the strand separation step to better understand these issues.

MATERIALS AND METHODS

DNA

Unmodified and modified DNA substrates were obtained from Integrated DNA Technologies and the Yale Keck Oligo Synthesis Facilities. Complementary oligos were annealed at 95°C for 5 min in annealing buffer (10 mM Tris HCl, 50 mM NaCl, 1 mM EDTA, pH 8) and subsequently cooled to room temperature. Once cooled, the annealing was analyzed by non-denaturing PAGE imaged on a GE Typhoon.

Equilibrium Pyrrolo-dC fluorescence

Equilibrium fluorescence was monitored at room temperature on a Horiba Scientific Fluoromax-4 spectrofluorometer. All measurements were conducted in reaction buffer (1 M HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)), 10 mM EDTA, 200 mM NaCl, pH 8). Background signal was determined by adding DNA (1 μM) and sinefungin (60 μM) to reaction buffer. Enzyme was added to a final concentration of 2.5 μM. The maximum excitation wavelength was determined to be 350 nm for Pyrrolo-dC (2 nm slit size). Emission data were collected over the interval 400–550 nm using an 8 nm slit size. Each measurement and background was averaged, and statistical outliers were eliminated. Statistical outliers were defined based on the maximum signal for each scan that was less than the first quartile (Q1) or greater than the third quartile (Q3) (18). Five scans were taken, and discarded reads were not replaced. About 3–5 scans were averaged for each sample. Samples were kept on ice and allowed to equilibrate to room temperature for 5–10 min prior to running scans.

Site-directed mutagenesis and protein purification

The wild-type CcrM plasmid containing kanamycin resistance was cloned using plasmid pXMCS2 as described previously (13). Mutant plasmids were constructed using the Agilent Quickchange Lightning Site-Directed Kit. The primers used in the PCR reactions are provided in Supplementary Table S3. Mutant plasmids were transformed into XL10 Ultracompetent *E. coli* cells (Agilent) and the plasmid isolated using an Agilent Mini Prep Kit. Plasmids were sequenced by the Berkeley DNA Sequencing Facility; confirmed plasmids were transformed into the NEB Nico21 (DE3) expression cells. Overnight cultures were grown at 37°C in LB broth and 30 μg/ml kanamycin, and 1 L cultures were initiated the next day in LB broth with 30 μg/ml kanamycin and shaken on a New Brunswick G10 Gyrotory shaker at 225 rpm at 37°C until an OD (600 nm) of 0.8 was reached. Cultures were placed on ice for 10 min and induced with 2 mM IPTG (Isopropyl β-D-1-thiogalactopyranoside, GoldBio) and shaken for 3 h at 225 rpm at room temperature. Cells were pelleted by centrifugation using a JA-10 rotor and a J2-21 centrifuge (Beckman) at 5000 rpm at 4°C for 20 min and stored at −80°C. Cell pellet was resuspended in lysis buffer containing 50 mM HEPES, 400 mM NaCl, 10% glycerol, and 50 mM Imidazole at pH 8.0 to a volume of ∼45 ml and sonicated with a

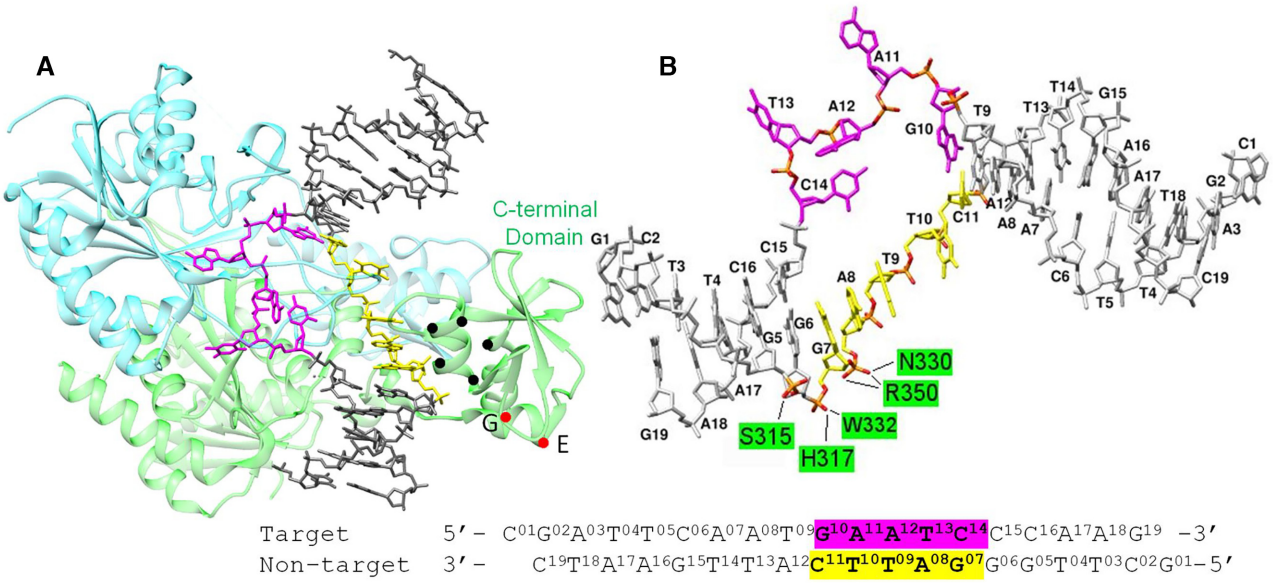

Figure 1. Cocrystal structure of CcrM-DNA. (**A**) Ccrm and double-stranded DNA (PDB: 6PBD). Monomer A is shown in cyan. Monomer B is shown in green. The DNA target strand and non-target strand are shown in magenta and yellow, respectively. Residues from the C-terminal segment of monomer B that interact with the non-target DNA strand are shown as black dots (S315, H317, N330, W332, and R350). Residues that do not interact with the non-target DNA are shown as red dots (G305 and E280). (**B**) The CcrM-DNA cocrystal structure reveals that four of the five base pairs within the cognate site are disrupted with the target strand bases positioned away from the complementary bases in the non-target strand. Amino acid residues S315, H317, W332, R350 and N330 make hydrogen bonds to the phosphate back bone of the non-target strand. The annotated 19mer double-stranded DNA used in the cocrystal structure shows the colored bases that represent the GANTC recognition site. Structural images were made with UCSF Chimera.

Branson digital sonifier in a water/ice slurry. The cell debris was separated from the lysate by centrifugation in a Beckman centrifuge at 11 000 rpm using a JA-20 rotor for 1 h. The clarified cell lysate was then passed through a 0.22 μM syringe driven filter unit and loaded onto a GE 5 ml HisTrap column using an AKTA Start FPLC system at a flow rate of 5 ml/min. The column was washed with several column volumes of the lysis buffer at a flow rate of 5 ml/min and fractions were eluted over a gradient from 50 to 250 mM imidazole over nine column volumes and 30, 1.5 ml fractions. Protein purity was assessed by SDS-PAGE, and purified proteins were dialyzed against dialysis buffer (similar to lysis buffer without imidazole) in Amicon Ultra 0.5 ml centrifugal filters (10 kDa) over four buffer exchanges. The protein was then stored in storage buffer (100 mM HEPES, 300 mM NaCl, 50% glycerol, 1 mM DTT (dithiothreitol), and 1 mM EDTA at pH 8.0) at −80°C. Subsequent densitometry analysis of wild-type CcrM and mutants by 12% SDS-PAGE imaged on a GE Typhoon revealed > 93% purity.

$k_{methylation}$ from radiochemical assays

DNA substrates containing a 5′ 6-fluorescein tag were purchased from Integrated DNA Technologies using standard desalting purification and N6-methyladenine containing substrates were ordered from the Yale Keck Oligo Synthesis Facility. Double-stranded substrates were created by annealing in 1× NEB Buffer 3.1 at 95°C and cooling to room temperature; substrates were then analyzed by native PAGE on a GE Typhoon Imager and showed >95% annealing success. The methylation reactions were performed as described previously (13). Single turnover reactions for substrates containing cognate recognition sites included 150 nM protein, 100 nM DNA, and 15 μM AdoMet, using hemimethylated double-stranded substrates. Non-cognate substrates used in single turnover experiments included 1.5 μM protein, 1 μM DNA and 15 μM AdoMet in all instances (^{3}H-CH_3 1 mCi [82.7 mCi/mmol]). Reactions were initiated with enzyme although the order of addition made no difference (data not shown). Samples (5 μl) were spotted in triplicate onto GE Amersham Hybond-XL nylon membrane blotting papers followed by placement in 400 ml of wash buffer (50 mM KH_2PO_4) to minimize background signal formation. The samples were shaken at room temperature for 5 min followed by two additional washes with the same buffer for 5 min. This was followed by a 5-min treatment with 400 ml of 80% EtOH, another wash for 5 min with 400 ml of 100% EtOH, and a final drying step for 5 min in 400 ml of ether in a fume hood. Samples were then placed into scintillation vials containing 3 ml of BioSafe II fluid. Radiochemical data were generated with a Beckman Coulter LS-6500 scintillation counter. Data for the single turnover reactions were fit to a one-phase decay model in GraphPad Prism 6.0. Substrate DNA sequences can be found in Supplementary Table S4.

K_d measurements

Dissociation constants were obtained from electrophoretic mobility shift assays using the fluorescein tagged DNA (13,14). Binding reactions consisted of 100 mM HEPES, 20 mM NaCl, 2 mM DTT and 1 mM EDTA, pH 8, and were performed using 10 nM DNA, and 60 μM sinefun-

gin (Sigma Aldrich). Reactions were incubated on ice for 30 min, diluted with an equal volume of 50% glycerol and loaded onto a 12% (75:1) Native PAGE gel and run for 60 min at 286 V in 0.5× TBE running buffer. The gels were then imaged on a GE Typhoon imager, and densitometry analysis using the Typhoon software was performed at the level of band disappearance (corresponding to the free DNA). The data were then fitted to a one-site specific binding model from which dissociation constants were obtained (13,14). Substrate DNA sequences can be found in Supplementary Table S4.

Strain construction and verification (*in vivo* experiments)

The deletion strains were constructed by electroporating plasmid pXMCS2-CcrMS315A and pXMCS2-CcrME280A into NA1000 (WT *C. crescentus* strain (12)). To construct pXMCS2-CcrM, the ccrM ORF was amplified and inserted into NdeI-KpnI digested pXMCS2 via Gibson assembly (19). The resultant plasmid was used to generate pXMCS2-CcrMS315A and pXMCS2-CcrME280A using Q5 mutagenesis (NEB). The plasmid was integrated into the *ccrM* locus on the chromosome by homologous recombination so that the endogenous CcrM is controlled by a xylose promoter and the exogenous CcrM mutant is controlled by the native promoter. The plasmid integration was verified by PCR using primer pair CcrMPro-Fwd (5′- GACT CAAAAGCGCCTGAAAGGC-3′) and pXMCS2-rev (5′-TTACCGCCTTTGAGTGAGCTG-3′), and the *CcrM* mutagenesis was confirmed by Sanger sequencing.

Microscopy

Caulobacter crescentus cells were grown in M2G media lacking xylose to deplete wild-type CcrM (19). Cells were collected at the exponential phase (OD600 < 0.3) and spotted on agarose pads (1.5%) containing M2G media prior to imaging. Phase-contrast images were obtained using a Leica DMi8 microscope with an HC PL APO 100 × /1.40 oil PH3 objective, Hamamatsu electron-multiplying charge-coupled device (EMCCD) C9100 camera, and Leica Application Suit X software. For computational image analyses, MicrobeJ (20) was used to determine cell outlines and lengths from phase images.

Bioinformatics

Searches in the UniProtKB reference proteomes plus SwissProt database were performed using BLAST, available at the Universal Protein Resource (UniProt). Proteins were identified using residues 272–358 (CcrM) and 275–358 (M.HinfI) as separate search seeds, with a search window of 500 hits. Multiple sequence alignments of the 500 resulting proteins were made using CLUSTAL O algorithm embedded in the JalView alignment editor. Alignments were visualized with ESPRIPT 3.0. Amino acid positions in both proteins are numbered relative to the CcrM sequence.

The phylogenetic tree was constructed from the results of the M.HinfI BLAST search. The tree includes 499 organisms, respresenting the organisms with a protein displaying a BLAST score greater than that of CcrM. The phylogenetic tree was constructed using NCBI CommonTree and the image was generated with iTOL. The logos were made with SeqLogo. The colors represent the chemical properties of each residue: polar residues (green), acidic (red), basic (blue), hydrophobic (black), and neutral (purple).

RESULTS

The CcrM-DNA cocrystal structure reveals that four of the five base pairs within the recognition site are disrupted upon CcrM binding, with the target strand positioned away from the complementary non-target strand (Figure 1) (17). Our interest was to provide a means to track the conformational changes within the DNA leading to this unusual complex. Our approach relies on the use of the cytosine analog, 6-methyl-3-(2-deoxy-β-D-ribofuranosyl)-3H-pyrrolo[2,3-d]pyrimidin-2-one (Pyrrolo-dC).

Pyrrolo-dC fluorescence is remarkably responsive to base stacking interactions which forms the basis of its use to study DNA repair, transcription factors, RNA polymerase, and nucleic acid conformations (21–26). When Pyrrolo-dC is positioned in the 'target' strand which undergoes methylation (Figure 2, $P1^{T}$) we observed a dramatic CcrM-dependent increase in fluorescence which is not observed when Pyrrolo-dC is positioned in the 'non-target' strand (Figure 2, $P0^{NT}$) which does not undergo methylation; note, CcrM is oriented on the dsDNA by using hemimethylated substrates, which is the form of DNA that CcrM interacts with biologically (3). We obtained similar results with the CcrM ortholog from *Brucella abortus,* as well as the β-class DNA MTase M.HinfI, both of which have the 80-residue C-terminal segment and methylate both single- and double-stranded DNA (Figure 3) (14). Thus, the Pyrrolo-dC assay shows similar protein-dependent fluorescence changes with enzymes that are functionally and structurally similar to CcrM (13). As a control, no fluorescence changes are observed with M.HhaII, which recognizes the same site (GANTC), lacks the C-terminal segment, and which shows no activity with single-stranded DNA (13). No CcrM-dependent changes in fluorescence are observed if Pyrrolo-dC is positioned outside the recognition site (Figure 2, $P2^{NT}$, $P3^{T}$, $P8^{T}$) which is consistent with the cocrystal structure (17). Placement of Pyrrolo-dC at the C position within the recognition site in the target and non-target strands (Figure 2, $P6^{NT}$, $P7^{T}$) also shows no to little change in fluorescence upon CcrM binding. $P6^{NT}$ replaces Cytosine 11 from the non-target strand with Pyrrolo-dC, and the G10:C11 base pair remains intact in the structure (Figure 1), consistent with the lack of significant fluorescence changes. In contrast, the base pairing of C14 to G7 is completely disrupted in the complex, suggesting that the Pyrrolo-dC fluorescence of $P7^{T}$ should increase upon CcrM binding. However, since we do not know the local environment around the Pyrrolo-dC in $P7^{T}$, it is conceivable that the fluorescence is effectively quenched by local interactions.

Our prior study of CcrM sequence specificity showed CcrM has no detectable activity when the G10:C11 base pair is switched to a non-cognate A10-T11 (14). Figure 2 ($P4^{T}$) shows that CcrM binding to this same non-cognate sequence in which the target strand contains the Pyrrolo-dC shows no evidence of strand separation, which is con-

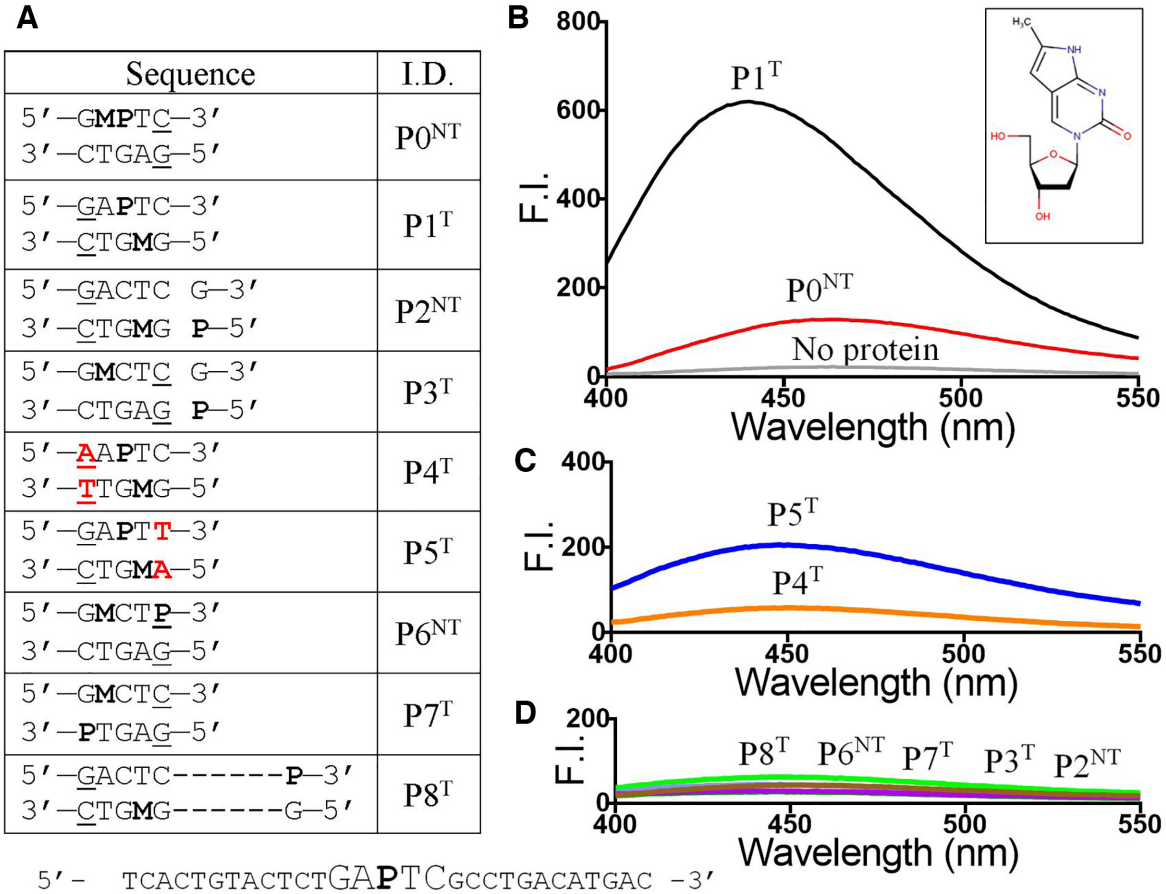

Figure 2. Pyrrolo-dC strand separation assay. P refers to Pyrrolo-dC, M refers to methylated adenine. The inset shows Pyrrolo-dC. (**A**) The parent sequence (bottom) is the 29mer DNA with P placed in the target strand and the GANTC recognition site is enlarged. Changes from the parent sequence are shown for all sequences. The underlined C:G base pair signifies that base pairing is maintained while the other four base pairs in the recognition site are disrupted (Figure 1A and B). The red A:T base pairs signifies a mutation in the recognition site. The superscript identifies the location of Pyrrolo-dC in the target or non-target strands for each sequence. (**B**) CcrM binding to DNA in which Pyrrolo-dC is positioned at the N position within the recognition site of the target ($P1^T$) is significantly greater that when placed in the N position of the non-target ($P0^{NT}$) strand. No protein represents the average of all DNA sequences without CcrM. (**C**) CcrM-dependent fluorescence is significantly less with the two non-cognate sequences ($P4^T$ and $P5^T$). (**D**) CcrM binding to DNA in which Pyrrolo-dC is positioned outside of the recognition site and at the C position in the recognition site shows no increase in fluorescence ($P8^T$, $P6^{NT}$, $P7^T$, $P3^T$, and $P2^{NT}$; maximum fluorescence intensity and error values provided in Supplementary Figure S6C). All fluorescence data contain DNA (1 μM), sinefungin (60 μM), and WT CcrM (2.5 μM, monomer concentration). All dsDNA are 29bp, hemimethylated, and contain one centrally located GANTC recognition site. Data collected at room temperature on a Horiba Scientific Fluoromax-4, with excitation at 350 nm. All traces average five scans. Background signal was subtracted from all traces. F.I. = Fluorescence Intensity. Variation in fluorescence intensity scans are provided in Supplementary Figure S6.

sistent with the prior activity data (14). CcrM affinity for this non-cognate site is only mildly altered from the cognate site binding (14), and these experiments are all done at high DNA and CcrM concentrations. The related non-cognate sequence in which the base pair at the other end of the recognition sequence is modified (C14 and G7, switched to T14 and A7) shows a minor increase in fluorescence (Figure 2 $P5^T$). These results show that the ability to induce strand separation and specificity are tightly correlated. We note that although positioning Pyrrolo-dC at the N position results in a small change in $k_{methylation}$ (compare sequence C, 1.38 min^{-1} and $P1^T$, 0.17 min^{-1}, Supplementary Table S2), this replacement appears to interfere with the discrimination against the two non-cognate sequences studied here (compare $P1^T$, $P4^T$ and $P5^T$).

C-terminal mutants and effects on strand separation and methylation

We initiated a mutational analysis of conserved residues in the C-terminal segment that make extensive contacts to the phosphates of the non-target DNA strand (Figure 1, S315, H317, N330, W332, R350). The C-terminal segment is folded as six antiparallel strands with three short α helices (Figure 1, C-terminal segment). This segment resembles a eukaryotic PWWP that can bind DNA non-specifically (27). Removal of the C-terminal segment in CcrM results in complete loss of ss and dsDNA activity and binding (10,11); interestingly, the same truncation of the β-class DNA MTase M.HinfI, which also methylates ss and ds-DNA, results in loss of dsDNA methylation activity with

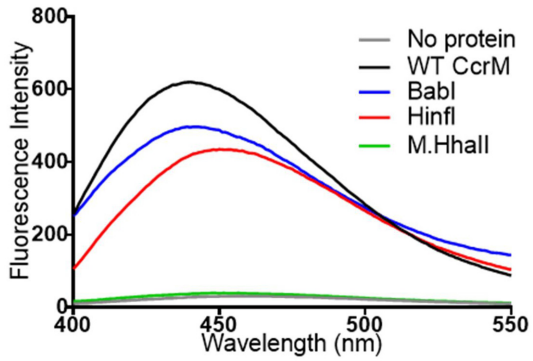

Figure 3. Strand separation occurs with CcrM ortholog BabI, the related M.HinfI, and does not occur with a control enzyme M.HhaII. All traces use sequence $P1^T$ (see Figure 2A) with Pyrrolo-dC at the N position in the target strand; WT CcrM, M.HinfI and HhaII 2.5 μM; BabI 6.13 μM. No protein represents the average of all DNA sequences without CcrM. Experimental conditions are the same as in Figure 2. Maximum fluorescence intensity and error values provided in Supplementary Figure S6D.

only minor impact on ssDNA methylation (14). Alanine substitution of a single conserved tryptophan (332) in this segment in CcrM results in complete loss (> 10^6-fold) of activities with ssDNA and dsDNA, without any detectable change in tertiary structure (13). W332 is part of a hydrophobic core that supports the integrity of the C-terminal segment (Figure 4) and the indole nitrogen contacts the phosphate to the 5′ side of guanosine (G7) within the non-target strand (5′pGACTC3′, Figure 4). The W332F and W332Y mutants both show nearly wild-type activity with ss DNA but are decreased 3- and 8-fold respectively with ds DNA (Supplementary Figure S5). These results provide strong support for the importance of the interaction between tryptophan 332 and the non-target phosphate.

Additional C-terminal segment mutants were designed, largely based on an analysis of known β-class methyltransferases, and *C. crescentus* homologs through a sequence alignment to see which residues are highly conserved (13). These include E280A, G305, S315, H317, N330, R350 and W332. Histidine 317 makes a 2.7Å hydrogen bond to the phosphate backbone of the non-target strand between bases G6 and G7 (Figure 5). Serine 315 contacts the phosphate between bases G5 and G6 (Figure 5). Asparagine 330 makes one hydrogen bond (3.2Å) to the phosphate between bases G7 and A8, and makes a hydrogen bond (3.2Å) to the peptide backbone of W332 (Figure 5). Arginine 350 makes two hydrogen bonds (2.8 and 3.0Å) to the phosphate backbone between G7 and G8 (Figure 5). G305 is highly conserved throughout CcrM homologs and its backbone amino is hydrogen bonded (3.4Å) to the backbone carbonyl oxygen of Arginine 302 (Figure 5). G305A is 8.6 Å removed from the DNA (Supplementary Figure S4); its role may be to maintain the loop configuration (Figure 5 inset). E280, while highly conserved, does not interact with either strand of DNA and the backbone carbonyl of E280 is 13.3Å removed from DNA. Figure 5 lacks structural information for E280 because the side chain was not resolved in the crystal structure.

Using the WT plasmid containing a C-terminal 6-Histidine tag (13,14), we mutated these residues to alanines via site-directed mutagenesis. Initial purification efforts of point mutants with an N-terminal His-tag resulted in severe proteolysis and contamination with proteolyzed fragments when using NiCo21 (DE3) expression cells from New England BioLabs. We therefore purified the WT and all mutants using a C-terminal His-tag, resulting in high concentrations and purity (Supplementary Figure S3). The purity of the proteins discussed here is significantly greater than previously published work which may contribute to the different results and conclusions (28,29).

Because our prior work showed CcrM has excellent activity with both ss and dsDNA (13,14), experiments were performed on ss and dsDNA (60 nucleotides, single stranded, 60 base pairs, double stranded) substrates that contain the cognate recognition site, 5′-GACTC-‘3, in the middle of the sequence (13,14). The mutant with the greatest change on ssDNA compared to WT (0.83 ± 0.067 min^{-1}) is S315A with a $k_{methylation}$ of 0.095 ± 0.005 min^{-1} (Supplementary Table S1 and Supplementary Figure S1A). The hemimethylated dsDNA was used to ensure the dimeric enzyme is positioned to methylate only one of the two strands. For the WT and mutants E280A and G305A, the experiments were performed with 150 nM protein, 100 nM DNA and 15 μM AdoMet. Experiments S315A, H317A, N330A and R350A used 300 nM protein, 100 nM DNA and 15 μM AdoMet. WT CcrM has a $k_{methylation}$ of 5.23 ± 0.65 min^{-1} while the mutant with the greatest kinetic perturbation, N330A, displayed $k_{methylation}$ of 0.011 min^{-1} indicating a maximal decrease of 476-fold on dsDNA. The mutants, S315A, H317A, and R350A also displayed significant decreases in $k_{methylation}$ on dsDNA (Supplementary Figure S1A and Supplementary Table S1). Our results suggest that residues which hydrogen bond to the non-target strand are critical for $k_{methylation}$, which is determined by methylation or a step preceding methylation (14).

We also determined if the C-terminal segment contributed to the sequence discrimination revealed with the WT CcrM. Our prior work with CcrM showed sequence discrimination ($k_{methylation}/K_D{}^{DNA}$, cognate versus non-cognate) of up to 10^7 fold on dsDNA (13) using the non-cognate recognition site, 5′AACTC’3, with ssDNA and dsDNA (hemi-methylated) substrates. WT CcrM has a $k_{methylation}$ of 5.5 × 10^{-4} min^{-1} on ssDNA while on dsDNA methylation was undetectable and is reported as <1 × 10^{-6} min^{-1}. These experiments were performed with 1.5 μM protein, 1 μM DNA, and 15 μM AdoMet for the WT and all mutants on both ss and ds DNA. For ssDNA, all of the mutants showed similar discrimination as the WT enzyme, with the exception of H317A that shows significantly greater discrimination against 5′AACTC’3 (Supplementary Figure S1A and Supplementary Table S1). In contrast, all of the mutants showed dramatic activity enhancements in comparison to the WT enzyme with the non-cognate dsDNA, resulting in significant losses of discrimination (Supplementary Figure S1B and Supplementary Table S1). While the WT had an undetectable $k_{methylation}$ of <1 × 10^{-6} min^{-1} for the non-cognate dsDNA substrate, the mutants showed rates ranging from at least ten to 150-fold greater than the WT enzyme (Supplementary Figure S1B and Supplementary Table S1). The stabilities of the mutant enzyme-DNA complexes with the cognate sequences

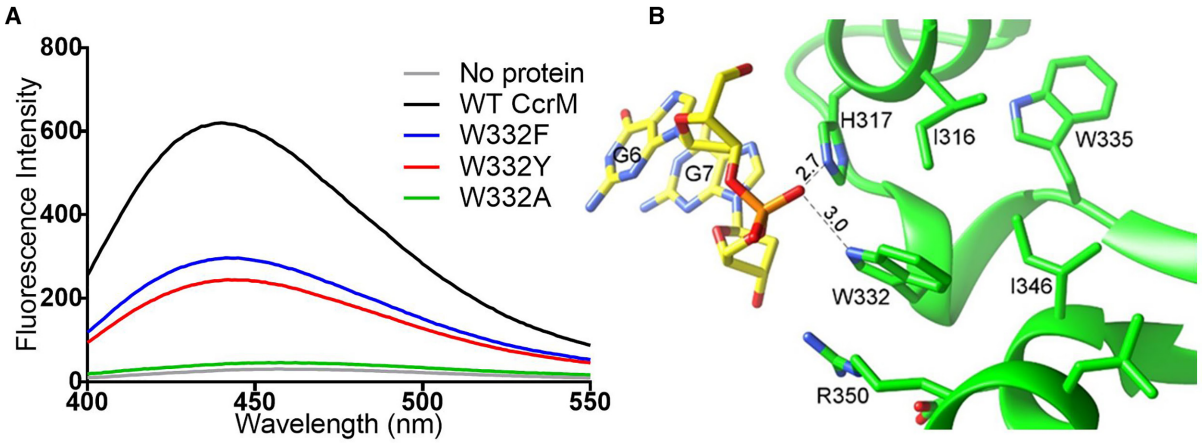

Figure 4. A single hydrogen bond between Tryptophan 332 and a non-target strand phosphate is important for strand separation. (**A**) Pyrrolo-dC fluorescence of WT CcrM and W332 mutants. W332A shows no increase in fluorescence. W332F and W332Y show an increase in fluorescence. No protein represents the average of all DNA sequences without CcrM. All traces contain DNA sequence $P1^T$ (see Figure 2A). Experimental conditions are the same as Figure 2. Maximum fluorescence intensity and error values provided in Supplementary Figure S6D. (**B**) W332 is hydrogen bonded to the phosphate backbone between DNA bases G6 and G7. W332 also contributes to the hydrophobic core of CcrM's C-terminal segment of Molecule B. PDB: 6PBD. Structural images were made with UCSF Chimera.

are compromised (Supplementary Figure S2). However, increasing the non-cognate DNA concentrations did not alter the methylation rate constants reported here (data not shown). Therefore, the $k_{methylation}$ values reported here are valid. Thus, the residues investigated here contribute to discrimination with dsDNA but make little to no contribution with ssDNA.

We performed electrophoretic mobility shift assays to determine dissociation constants (K_d) for the WT and mutants on ss and ds DNA on cognate substrates. On ssDNA, the WT displayed a K_d of 18.5 ± 5.7 nM while the mutants displayed relatively intact binding, with H317A showing the greatest change compared to WT with a K_d of 101 ± 32.1 nM (∼5-fold loss of stability) as seen in Supplementary Table S1 and in Supplementary Figure S2A. EMSA experiments on cognate dsDNA substrates revealed dramatic disruption in the ability of S315A, H317A, N330A, and R350A to bind dsDNA. H317A has a K_d of 1.42 ± 0.20 μM while R350A has the greatest reduction of binding with a K_d of 4.04 ± 0.41 μM (Supplementary Figure S2A and Supplementary Table S1). The mutants E280A and G305A display near-WT like *K*d's on dsDNA. Our results for the mutants S315A, H317A, N330A and R350A differ with prior work on these mutants, reporting improved binding affinity over the WT enzyme using the same DNA and similar buffer conditions (28,29). Importantly, the prior work reported an extremely weak DNA affinity for the WT CcrM enzyme (2–10 μM compared to 75 nM shown in Supplementary Table S1), which may have resulted from using impure protein or only partially active enzyme (13). The results in Supplementary Table S1 show that the sequence discrimination by CcrM at the ssDNA and dsDNA levels is in large part driven by changes in methylation or a limiting step prior to methylation.

Pyrrolo-dC interrogation of CcrM C-terminal mutants

We used this assay to determine if the conserved C-terminal residues play a role in the strand separation mechanism. S315A, H317A, N330A and R350A showed no Pyrrolo-dC fluorescence enhancements upon CcrM binding (Figure 5) suggesting that the strand separation step is impacted. Each of these residues forms hydrogen bonds to the phosphate backbone of the non-target DNA either within or outside of the recognition site (Figure 1). Thus, these interactions are strongly implicated in inducing the strand separation, stabilizing the conformation observed in the cocrystal structure (Figure 1), or both. E280A and G305A show an increase in signal upon binding of CcrM, but the signal change is less than WT CcrM (Figure 5). This suggests that E280A and G305A do not induce the same conformational changes in the DNA as WT CcrM. We made three mutations to Tryptophan 332. W332A showed no change in fluorescence upon CcrM binding, consistent with its complete lack of enzymatic activity, in spite of showing no conformational changes in the protein as determined by CD (14). W332Y and W332F showed a significant change in Pyrrolo-dC fluorescence upon CcrM binding that was less than WT CcrM (Figure 4), suggesting that hydrogen bonding to the non-target strand as well as interactions in the hydrophobic core of the C-terminal segment contribute to CcrM's ability to strand separate.

A CcrM mutant causes a filamentous phenotype

CcrM plays essential roles in controlling progression through the cell cycle (3,4). Disruption of this control results in an extensive bacterial filamentous growth phenotype (3–4,30). We imaged WT CcrM and two of the six mutated CcrM genes expressed from the native promoter on the *C. crescentus* chromosome (30). Figure 6 shows that the WT and one of the mutants (E280A) display normal growth, as determined by the average cell length (30). In contrast, S315A, which shows greater alterations in methylation activity (Supplementary Figure S1A), shows extensive filamentous growth, comparable to that observed with the CcrM deletion strain (30). These results suggest that the enzyme's ability to control filamentous growth can be al-

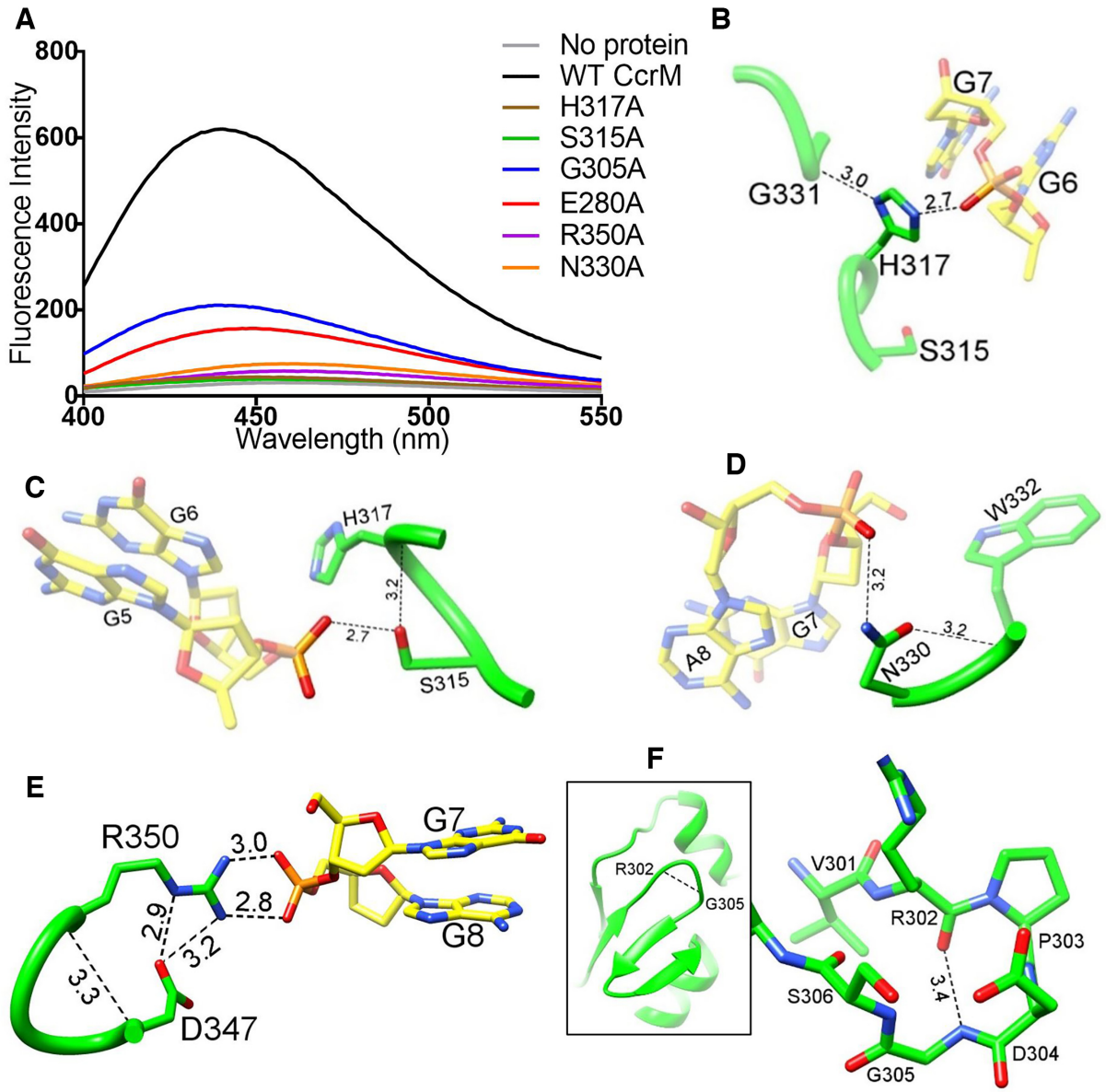

Figure 5. Pyrrolo-dC Fluorescence of CcrM mutants and structures of the WT residues G305, H317, S315, N330 and R350. (**A**) Conserved residue mutants dramatically alter the CcrM-dependent changes in Pyrrolo-dC DNA fluorescence. All experimental conditions as in Figure 2 using DNA sequence $P1^T$. Maximum fluorescence intensity and error values provided in Supplementary Figure S6D. (**B–F**) Structure of amino acid residues showing the hydrogen-bond interactions made by each side chain. Hydrogen bonds are shown as dashed lines and the distance is labeled in Angstroms; PDB: 6PBD. Structural images were made with UCSF Chimera.

tered by a single amino side change in the conserved segment, and moreover, this phenotype is correlated with the severity of changes to the methyltransferase activity of the mutated CcrM (Supplementary Table S1 and Supplementary Figure S1A).

Bioinformatics

The multiple sequence alignments using the 80 amino acid C-terminal segment of M.HinfI (Figure 7) reveals conservation among a variety of DNA methyltransferases in the N4/N6-methyltransferase family including the highlighted human pathogens *Mycoplasma girerdii* (31), *Bartonella bacilliformis* (32), *Bartonella tamaie* (33), *Capnocytophaga canimorsus* (34), *Helicobacter pylori* (35), *Brucella abortus* (36) and *Haemophilus influenzae* (37), animal pathogens *Mycoplasma nasistruthionis* (38), *Ureaplasma diversum* (39), *Campylobacter sputorum biovar sputorum* (40), *Brachyspira catarrhinii* (41), *Mycoplasma californicum* (42), *Moraxella lincolnii* (43), *Mycoplasma falconis* (44) and *Moraxella macacae* (45). These organisms highlight the widespread distribution of this protein segment, since it is present in three of the six classes in the Proteobacteria phylum, three other phyla (Spirochaetes, Bacteroidetes

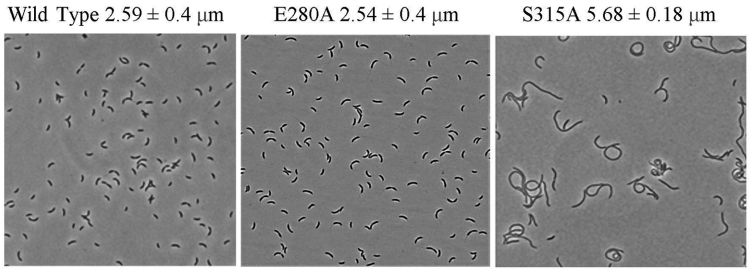

Figure 6. Mutations in the C-terminal segment of CcrM cause dramatic cellular changes in growth characteristics of *Caulobacter cresentus*. The WT and mutant CcrM genes were inserted into the correct genomic positions (60). Analysis of 300 bacteria used phase-contrast and fluorescence microscopy. The WT and E280A (which is 6- to 7-fold decreased in methylation activity) have normal growth phenotypes whereas S315A shows extensive elongation (filamentous growth), which was determined by average cell length (μm), indicating a severe disruption of the regulatory processes that control progression through the cell cycle (61).

and Tenericutes) as well as an archaea. These organisms are highlighted in the M.HinfI phylogenetic tree (Supplementary Figure S8).

We performed a full sequence alignment of the organisms highlighted in the M.HinfI search to investigate regions outside of the 80 amino acid segment that appear to be involved in strand separation. In addition to the 80 amino acid segment, there is conservation in loops that may be involved in strand separation. Loop-2B and Loop-45 are inserted within the separated DNA strands in the cocrystal structure (17). All of the highlighted proteins have fully conserved residues in Loop-45 (N120, P123, N124, F125, G127, R129 and N132). All proteins, except for *Mycoplasma girerdii,* have fully conserved residues in Loop 2B (L42, R44, W57 and D58).

We also determined that 477/500 of the proteins in the M.HinfI BLAST are β-class DNA methyltransferases based on the organisation of conserved motifs (2). Three proteins could not be assigned due to problems in identifying the catalytic domains.

The multiple sequence alignment from the 80 amino acid C-terminus of CcrM (Figure 7) reveals conservation among a variety of N4/N6-DNA methyltransferases from alphaproteobacteria. The organisms listed represent this variety and include human pathogens *Rhodobacter massiliensis* (46), *Pannonibacter pragmitetus* (47), Inquilinus limosus (48), *Haematospirillum jordaniae* (49) and *Methylocapsa palsarum* (50), a bacteria that contributes to biogas production *Rhodopseudomonas faecalis* (51), a plant-growth promoting bacteria *Azospirillum sp. RU38E* (52), two methanotrophs *Methylocapsa palsarum* (53) and *Methylocella silvestris* (54), and *Tepidicaulis Marinus*, which reduces nitrate (55). Also highlighted are a nitrogen fixing alphaproteobacterium *Hartmannibacter diazotrophicus* (56) and *Martelella endophytica*, which exhibits inhibitory activity against fungal plant pathogens (57).

The logo shows that the residues we mutated (E280, G305, S315, H317, N330, W332 and R350) are highly conserved among CcrM and M.HinfI (Figure 7). E280 and G305 have some variability with consensuses of 93.4% and 97.0%, respectively. This is unlike S315, H317, N330, W332, and R350, which have consensuses of 98–100%. This is correlated with the fluorescence data for each of these mutants suggesting that the strand separation process can accommodate changes at these positions. The less severely impacted strand separation ability of E280A is consistent with the *in vivo* results showing this mutant has wild-type growth characteristics (Figure 6).

DISCUSSION

The catalogue of characterized protein-nucleic acid recognition mechanisms is rich and diverse, informed by extensive cocrystal structures. Our prior biochemical work (13–14,17) and the recent CcrM-DNA cocrystal structure shows that CcrM relies on a new recognition mechanism in which the protein induces the unpairing of four out of five base pairs making up the recognition sequence (17). The potential contribution of this mechanism to the extreme sequence discrimination shown by CcrM further emphasizes its importance (14). Although the actual strand separation mechanisms may differ, enzymes such as CRISPR/Cas9 also carry out extensive strand separation without reliance on an exogenous energy source, although recognition is mediated through RNA/DNA hybridization. Strand separation of the DNA duplex by CRISPR/Cas9 to allow base pairing between the target DNA and crRNA guide sequence requires initial recognition of a short protospacer adjacent motif (e.g., 5′NGG3′). The strand separation step remains 'enigmatic, but must rely on thermally available energy' (58). A newly described human ß class adenine methyltransferase (MettL3-MettL14 complex) methylates single stranded DNA and unpaired regions of double-stranded DNA with reduced activity, and may rely on a strand separation mechanism (16).

The cocrystal structure of CcrM bound to double stranded DNA (Figure 1) leaves unanswered questions related to how such a massive perturbation of the duplex DNA conformation occurs. The ability to track strand separation could provide a basis for deconstructing how CcrM and other enzymes facilitate this. Pyrrolo-dC is remarkably responsive to its local environment (21) and thus provides a potential basis for tracking strand separation. The fluorescence enhancement resulting from CcrM binding to DNA in which the target strand contains a single Pyrrolo-dC positioned in the 'N' position (Figure 2, $P1^T$) is dramatic. While fluorescence changes can be caused by diverse effects, the DNA control presented here (Figure 2, no protein) provides strong evidence that this change results from the separation of the two strands that leads to the struc-

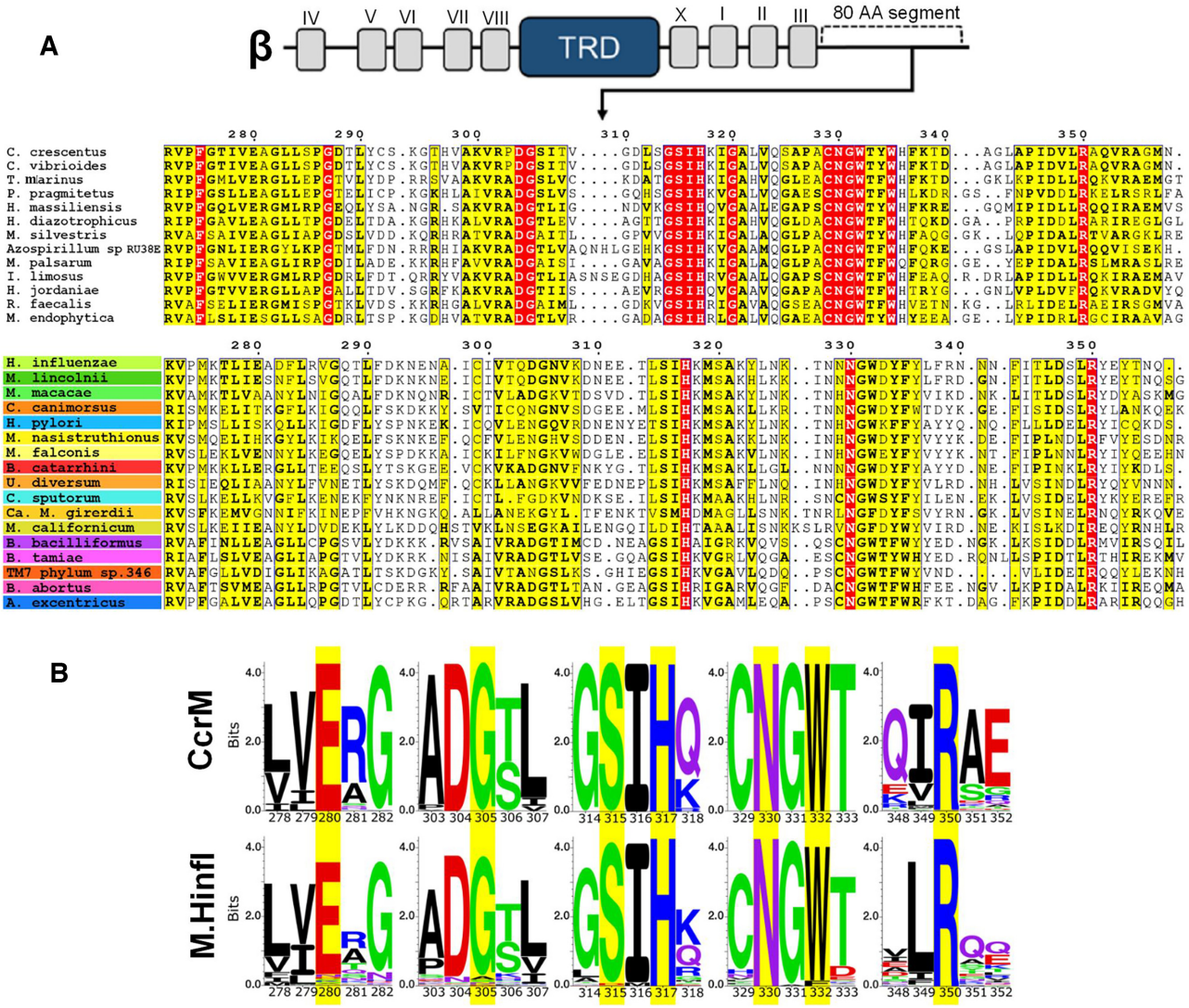

Figure 7. The 80 amino acid C-terminal segment is widespread amongst N4 and N6-DNA methyltransferases. (**A**) The motifs of β-class methyltransferases are depicted in relation to the C-terminal 80-residue segment (2). Proteins were collected through UniProt BLAST searches of the 80 amino acid segment (dashed line) from β–class CcrM (*C. crescentus,* top) and α-class M.HinfI (*H. influenza*, bottom), with a window of 500 sequences to generate the sequence alignments. The search seeds for CcrM and M.HinfI were residues 272–358 and 275–358, respectively. The displayed protein sequences represent organisms of scientific interest and are labeled by the organism containing each protein. The alignment suggests that conserved residues in the 80 amino acid segment are widespread among N4 and N6-DNA methyltransferases. Multiple sequence alignments were made using CLUSTAL O. Alignments were visualized with ESPRIPT 3.0 which depicts highly conserved residues in red and moderately conserved residues in yellow. Red arrows indicate residues that were mutated to alanine. Residue numbers for both alignments reflect the position when aligned to CcrM. The highlighted colors in the M.HinfI alignment correspond to the highlighted organisms in the phylogenetic tree (Supplementary Figure S8). The accession numbers for the displayed proteins in the CcrM alignment: *Caulobacter crescentus* (P0CAW2), *Caulobacter vibrioides* (B8GZ33), *T. mariunus* (A0A081BAH8), *P. pragmitetus* (A0A0U3N820), *H. massiliensis* (A0A086XXK7), *H. diazotrophicus* (A0A2C9D333), *M. silvestris* (B8EI64), *Azospirillum* sp. RU38E (A0A239AAC3), *M. palsarum* (A0A1I3Z5R1), *I. limosus* (A0A211ZSZ0), *H. jordaniae* (A0A143DFF5), *R. faecalis* (A0A318TWW6) and *M.endophytica* (A0A0D5LWP1). The accession numbers for the proteins in the M.HinfI alignment: *Haemophilus influenzae* (P20590), *Moraxella lincolnii* (A0A1T0CF71), *Moraxella macacae* (L2F796), *Capnocytophaga canimorsus* (F9YSB6), *Helicobacter pylori* (O25907), *Mycoplasma nasistruthionis* (A0A4Y6I7I2), *Mycoplasma falconis* (A0A501XAX8), *Brachyspira catarrhinii* (A0A4U7NEV9), *Ureaplasma diversum* (A0A084F1N9), *Campylobacter sputorum* biovar sputorum (A0A381DI05), *Candidatus Mycoplasma girerdii* (A0A097SSH4), *Mycoplasma californicum* (A0A059XRQ7), *Bartonella bacilliformis* (A1URX9), *Bartonella tamiae* (J0R4G4), *Brucella abortus* (B2S9Y5), TM7 phylum sp. oral taxon 346 (A0A563D6M2) and *Asticcacaulis excentricus* (E8RMI2). (**B**) Excerpts of the Logos of the 80 amino acid segment of CcrM (top) and M.HinfI (bottom) show that the conserved residues in CcrM are also conserved in M.HinfI. Residue numbers for both logos reflect the position when aligned to CcrM. Aligned sequences were cropped so that blocks of five residues are shown. Full sequence logos are provided in Supplementary Figure S7. The colors represent the chemical properties of each residue; polar residues (green), acidic (red), basic (blue), hydrophobic (black) and neutral (purple). Yellow highlighted residues were mutated in this study.

ture shown in Figure 1. Thus, compared to when the target strand is tracked, this enhancement is significantly reduced when the non-target strand contains the Pyrrolo-dC (Figure 2, $P0^{NT}$). Inspection of the CcrM-DNA structure (Figure 1) provides a plausible explanation for this strand-specific effect. While the base at the N position in the target strand (A12) is poorly base-stacked with adjacent bases, this base in the non-target strand (T9) is well stacked with the proximal bases. Thus, while the bases at this N position are no longer paired to the partner base in both cases, only the target strand base has the further stacking disruption that likely leads to the enhanced fluorescence increase (Figure 2, $P1^{T}$) (21).

Figure 2 also shows that positioning Pyrrolo-dC immediately outside the canonical site shows no increase in fluorescence upon CcrM binding (Figure 2; $P2^{NT}$, $P3^{T}$). Pyrrolo-dC (represented as P in Figure 2) was placed in the non-target strand adjacent to the canonical site ($P2^{NT}$) and in the target strand adjacent to the canonical site ($P3^{T}$); because the DNA is hemimethylated, CcrM should be oriented differently on $P2^{NT}$ and $P3^{T}$. Although the base pairs flanking the recognition site in the DNA sequences studied here (Figure 2) are different from those in the cocrystal structure (Figure 1), the lack of detectable changes in fluorescence upon binding by CcrM is consistent with little disruption of the base pairing at these two positions. We previously showed that the base seven bases outside the recognition site in ssDNA impacts activity which we tested with substrate $P8^{T}$ (14). However, $P8^{T}$ shows no evidence of strand separation consistent with the cocrystal structure. It remains possible that CcrM interacts differently with ssDNA and dsDNA, which will require additional structural studies.

The only base pair that is not completely disrupted in the cocrystal structure (G10:C11) was probed with $P6^{NT}$ (Figure 2). As expected, $P6^{NT}$ shows no change in fluorescence upon CcrM binding (Figure 2). However, replacement of Pyrrolo-dC within the G:C base pair (C14:G7) that is disrupted in the cocrystal structure (Figure 1 and Figure 2 $P7^{T}$) also did not show a significant change in fluorescence upon CcrM binding (Figure 2), which is likely the result of local quenching by residues proximal to Pyrrolo-dC at that site. Cytosine 14 (Figure 1) appears to be sandwiched between residues N124 and L42, which could be responsible for the signal quenching with $P7^{T}$.

Figure 2 provides compelling data that the strand separation step tracked by Pyrrolo-dC is a specificity determinant for CcrM. Sequence $P4^{T}$ is mutated from G:C to A:T at the base pair that is maintained within the recognition site (Figure 2) and shows a significantly reduced change in fluorescence (Figure 2, $P4^{T}$); note, Pyrrolo-dC is positioned at the center nucleotide position. Thus, the inability to stabilize a strand separated intermediate appears to contribute to the lack of activity with this non-cognate sequence. Similarly, sequence $P5^{T}$ has the G:C replaced by A:T at the base pair that is disrupted in the recognition site (Figure 2), also shows a much reduced fluorescence change upon CcrM binding (Figure 2). The slightly greater fluorescence change observed with $P5^{T}$ versus $P4^{T}$ suggests that interactions between CcrM and non-cognate sequences may resemble those with the cognate site.

M.HinfI, like CcrM, shows good activity with both single and double stranded DNA, which led to our prior suggestion that its DNA recognition mechanism is similar to CcrM (13). Further, M.HinfI, a β-class DNA adenine MTase, has a similar C-terminal sequence seen in CcrM. M.HinfI, like CcrM shows the same increase in Pyrrolo-dC fluorescence (Figure 3), suggesting that it relies on the same strand separation mechanism. Similarly, the CcrM ortholog BabI, also shows a similar increase when the Pyrrolo-dC is positioned in the target strand. Further validation of the assay comes from the observation that another enzyme, M.HhaII, which recognizes and methylates the same sequence as CcrM, shows no fluorescence enhancement (Figure 3). M.HhaII is unlikely to use the strand separation mechanism since it unable to methylate ssDNA and lacks the conserved C-terminal sequence seen in CcrM, its orthologs and M.HinfI.

The conserved C-terminal 83 residue segment that makes extensive contacts to the non-target strand through one of the two monomers (Figure 1, Monomer B in green) stands out as a potential contributor to the strand separation process. The dramatic decrease in methylation activity we observe when conserved residues S315, H317, R350, N330 and W332 (Figure 1 black dots, Figure 7) are replaced with alanine is certainly consistent with their importance. Residues G305 and E280 (Figure 1 red dots, Figure 7) do not show the same decrease in methylation and do not interact with DNA (Supplementary Figure S4). Moreover, inspection of the cocrystal structure reveals that this segment is distinct from the classical target recognition domain of CcrM (1–264). Similarly, this segment contains no residues known to be important for catalysis. This, along with our observation that disruption of single hydrogen bonding interactions between these residues and the non-target backbone phosphates (e.g. S315, H317, N330, R350 and W332) dramatically alters the strand separation step or stabilization of the strand separated intermediate (Figure 5) suggests these residues are important for the strand separation mechanism. The connection between the strand separation step and the dramatic sequence discrimination displayed by CcrM (14) is supported by the fact that although capable of binding non-cognate sequences (14), the WT enzyme shows little to no ability to induce strand separation of such sequences (Figure 2, $P4^{T}$ and $P5^{T}$). Many of the CcrM mutants which display enhanced methylation activity with non-cognate sequences (and thus, decreased discrimination) have lost single H-bonding contacts between the protein and phosphates within the non-target strand. Increased promiscuity resulting from alteration in recognition interactions is often observed (59), although in this case, the interactions are limited to the non-target strand.

We sought to determine if the functional changes resulting from the alterations in residues investigated here are important for the critical *in vivo* transcriptional regulation displayed by the WT CcrM (3). This regulation involves a complex array of proteins in combination with CcrM, that ultimately drives the phenotypic changes observed as the normal growth phenotype. We relied on the replacement of the WT CcrM we previously developed in *C. crescentus* (3). Figure 6 shows that WT and the E280A CcrM show the nor-

mal phenotype. E280A shows only minor alterations in its ability to methylate DNA (Supplementary Figure S1A and Supplementary Table S1) and shows some strand separation activity (Figure 5). In contrast, S315A, which shows a 75-fold loss in methylation activity and 2000-fold loss in $k_{methylation}/K_d$ (Supplementary Figure S1A and Supplementary Table S1) and no ability to stabilize the strand separated DNA, shows the same filamentous growth phenotype observed for the CcrM knockout (30). Thus, the ability to separate the target and non-target strands is essential for the biological CcrM-mediated regulation of the Caulobacter growth phenotype.

Based on the results presented here, the C-terminal segments of CcrM and M.HinfI contribute to the strand separation mechanism. Using the CcrM and M.HinfI segments, we searched for other organisms which have a protein with similar C-terminal segments and found N4/N6-DNA MTases with this segment in broadly distributed organisms. The residues we investigated in CcrM (Figure 7) show a high degree of conversation across this entire group which spans multiple Phyla (Supplementary Figure S8). The segment found in M.HinfI is observed in diverse bacteria, including one Archaea example (Supplementary Figure S8). This includes three of the six Classes of Proteobacteria (gamma-, epsilon-, and alpha-proteobacteria) as well as organisms in other Phyla (Spirochaetes, Bacteriodetes, Tenericutes) and numerous human and animal pathogens (Supplementary Figure S8). The segment found in CcrM is broadly distributed in alphaproteobacteria, including human pathogens *Rhodobacter massiliensis* (46), *Pannonibacter pragmitetus* (47), Inquilinus limosus (48) and *Haematospirillum jordaniae* (49).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

National Science Foundation [CHE 1413722 to N.R.].
Conflict of interest statement. None declared.

REFERENCES

1. Adhikari,S. and Curtis,P. (2016) DNA methyltransferases and epigenetic regulation in bacteria. *FEMS Microbiol. Rev.*, **40**, 575–591.
2. Malone,T., Blumenthal,R. and Cheng,X. (1995) Structure-guided analysis reveals nine sequence motifs conserved among DNA Amino-methyl-transferases, and suggests a catalytic mechanism for these enzymes. *J. Mol. Biol.*, **253**, 618–632.
3. Gonzalez,D., Kozdon,J.B., McAdams,H.H., Shapiro,L. and Collier,J. (2013) The functions of DNA methylation by CcrM in Caulobacter crescentus: a global approach. *Nucleic Acids Res.*, **42**, 3720–3735.
4. 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. Bacteriology*, **179**, 5869–5877.
5. Cheng,X. and Roberts,R.J (2001) AdoMet-dependent methylation, DNA methyltransferases and base flipping. *Nucleic Acids Res.*, **29**, 3784–3795.
6. Estabrook,R.A., Nguyen,T.T., Fera,N. and Reich,N.O (2009) Coupling sequence-specific recognition to DNA modification. *J. Biol. Chem*, **284**, 22690–22696.
7. Shieh,F.K., Youngblood,B. and Reich,N.O. (2006) The role of Arg165 towards base flipping, base stabilization and catalysis in M.HhaI. *J. Mol. Biol*, **362**, 516–527.
8. Matje,D., Krivacic,C., Dahlquist,F. and Reich,N.O. (2013) Distal structural elements coordinate a conserved base flipping network. *Biochemistry*, **52**, 1669–1676.
9. Matje,D., Zhou,H., Smith,D., Neely,R., Dryden,D., Jones,A., Dahlquist,F. and Reich,N.O. (2013) Enzyme-Promoted Base Flipping Controls DNA Methylation Fidelity. *Biochemistry*, **52**, 1677–1685.
10. 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*, **26**, 4658–4667.
11. Zhou,B., Schrader,J.M., Kalogeraki,V.S., Abeliuk,E., Dinh,C.B., Pham,J.Q., Cui,Z.Z., Dill,D.L., McAdams,H.H. and Shapiro,L. (2015) The global regulatory architecture of transcription during the Caulobacter cell cycle. *PLos Genet.*, **11**, e10043831.
12. Robertson,G.T., Reisenauer,A., Wright,R., Jensen,R.B., Jensen,A., Shapiro,L. and Roop,R.M. (2000) The Brucella abortus CcrM DNA methyltransferase is essential for viability, and its overexpression attenuates intracellular replication in murine macrophages. *J. Bacteriology*, **182**, 3842–3849.
13. Woodcock,C.B., Yakubov,A.B. and Reich,N.O. (2017) Caulobacter crescentus cell Cycle-Regulated DNA methyltransferase uses a novel mechanism for substrate recognition. *Biochemistry*, **56**, 3913–3922.
14. Reich,N.O., Dang,E., Kurnik,M., Pathuri,S. and Woodcock,C.B. (2018) The highly specific, cell cycle-regulated methyltransferase from Caulobacter crescentus relies on a novel DNA recognition mechanism. *J. Biol. Chem*, **293**, 19038–19046.
15. de la Campa,A.G., Kale,P., Springhorn,S.S. and Lacks,S.A. (1987) Proteins encoded by the DpnII restriction gene cassette: two methylases and an endonuclease. *J. Mol. Biol*, **196**, 457–469.
16. Woodcock,C.B., Yu,D., Hajian,T., Li,J., Huang,Y., Dai,N., Corrêa,I.R., Wu,T., Vedadi,M., Zhang,X. *et al.* (2019) Human MettL3–MettL14 complex is a sequence specific DNA adenine methyltransferase active on single-strand and unpaired DNA in vitro. *Cell Discovery*, **5**, 63.
17. Horton,J.R., Woodcock,C.B., Opot,S.B., Reich,N.O., Zhang,X. and Cheng,X. (2019) The cell cycle-regulated DNA adenine methyltransferase CcrM opens a bubble at its DNA recognition site. *Nat. Commun.*, **10**, 4600.
18. Ramm,B., Glock,P., Mücksch,J., Blumhardt,P., García-Soriano,D.A., Heymann,M. and Schwille,P. (2018) The MinDE system is a generic spatial cue for membrane protein distribution in vitro. *Nat. Commun.*, **9**, 3942.
19. Thanbichler,M., Iniesta,A.A. and Shapiro,L. (2007) A comprehensive set of plasmids for vanillate- and xylose-inducible gene expression in Caulobacter crescentus. *Nucleic Acids Res.*, **35**, e137.
20. Ducret,A., Quardokus,E.M. and Brun,Y.V. (2016) MicrobeJ, a tool for high throughput bacterial cell detection and quantitative analysis. *Nat. Microbiol*, **1**, 16077.
21. Berry,D.A., Jung,K.Y., Wise,D.S., Sercel,A.D., Pearson,W.H., Mackie,H., Randolph,J.B. and Somers,R.L. (2004) Pyrrolo-dC and pyrrolo-C: fluorescent analogs of cytidine and 2′-deoxycytidine for the study of oligonucleotides. *Tetrahedron Lett.*, **45**, 2457–2461.
22. Dash,C., Rausch,J.W. and Le Grice,S.F.J. (2004) Using pyrrolo-deoxycytosine to probe RNA/DNA hybrids containing the human immunodeficiency virus type208;1 3′ polypurine tract. *Nucleic Acids Res.*, **32**, 1539–1547.
23. Zhang,X. and Wadkins,R.M. (2008) DNA hairpins containing the cytidine analog Pyrrolo-dC: Structural, thermodynamic, and spectroscopic studies. *Biophys. J.*, **96**, 1884–1891.
24. Lee,C.Y., Park,K.S. and Park,H.G. (2017) Pyrrolo-dC modified duplex DNA as a novel probe for the sensitive assay of base excision repair enzyme activity. *Elsevier Biosensors Bioelectronics*, **98**, 210–214.
25. Sukackaite,R., Grazulis,S., Tamulaitis,G. and Siksnys,V. (2012) The recognition domain of the methyl-specific endonuclease McrBC flips out 5-methylcytosine. *Nucleic Acids Res.*, **40**, 7552–7562.
26. Zang,H., Fang,Q., Pegg,A.E. and Guengerich,P. (2005) Kinetic analysis of steps in the repair of damaged DNA by human O^6-Alkylguanine-DNA Alkyltransferase. *J. Biol. Chem.*, **280**, 30873–30881.

27. Qin,S. and Min,J. (2014) Structure and function of the nucleosome-binding PWWP domain. *Trends Biochem. Sci.*, **39**, 536–547.
28. Maier,J., Albu,R., Jurkowski,T. and Jeltsch,A. (2015) Investigation of the C-terminal domain of the bacterial, DNA-(adenine N6)-methyltransferase CcrM. *Biochimie*, **119**, 60–67.
29. Albu,R., Zacharias,M., Jurkowski,T. and Jeltsch,A. (2012) DNA interaction of the CcrM DNA methyltransferase: a mutational and modeling study. *ChemBioChem*, **13**, 1304–1311.
30. Stephens,C., Reisenauer,A., Wright,R. and Shapiro,L. (1995) A cell is dependent on DNA-facilitated proteolysis and substrate polar sequestrationcycle-regulated bacterial DNA methyltransferase isessential for viability. *Proc. Natl. Acad. Sci. USA*, **93**, 1210–1214.
31. Costello,E.K., Sun,C.L., Carlisle,E.M., Morowitz,M.J., Banfield,J.F. and Relman,D.A. (2017) *Candidatus Mycoplasma girerdii* replicates, diversifies, and cooccurs with Trichomonas vaginalis in the oral cavity of a premature infant. *Sci. Rep.*, **7**, 3764.
32. Gomes,C., Martínez-Puchol,S., Ruiz-Roldán,L., Pons,M.J., del Valle Mendoza,J. and Ruiz,J. (2016) Development and characterization of highly antibiotic resistant *Bartonella bacilliformis* mutants. *Sci. Rep.*, **6**, 33584.
33. Kosoy,M., Morway,C., Sheff,K.W., Bai,Y., Colborn,J., Chalcraft,L., Dowell,S.F., Peruski,L.F., Maloney,S.A., Baggett,H. *et al.* (2008) *Bartonella tamiae* sp. nov., a Newly Recognized Pathogen Isolated from Three Human Patients from Thailand. *J. Clin. Microbiol.*, **46**, 772–775.
34. Mader,N., Lührs,R., Langenbeck,M. and Herget-Rosenthal,H. (2020) *Capnocytophaga canimorsus* – a potent pathogen in immunocompetent humans – systematic review and retrospective observational study of case reports. *Infect. Dis.*, **52**, 65–74.
35. Beydoun,M.A., Beydoun,H.A., Weiss,J., Hossain,S., El-Hajj,Z.W. and Zonderman,A.B. (2020) *Helicobacter pylori*, periodontal pathogens, and their interactive association with incident all-cause and Alzheimer's disease dementia in a large national survey. *Mol. Psychiatry*, doi:10.1038/s41380-020-0736-2.
36. Atluri,V.L., Xavier,M.N., de Jong,M.F., den Hartigh,A.B. and Tsolis,R.M. (2011) Interactions of the human pathogenic *Brucella* Species with their hosts. *Annu. Rev. Microbiol.*, **65**, 523–541.
37. Price,E.P., Sarovich,D.S., Nosworthy,E., Beissbarth,J., Marsh,R.L., Pickering,J., Kirkham,L.S., Keil,A.D., Chang,A.B. and Smith-Vaughan,H.C. (2015) *Haemophilus influenzae*: using comparative genomics to accurately identify a highly recombinogenic human pathogen. *BMC Genomics*, **16**, 641.
38. Spergser,J., Botes,A., Nel,T., Ruppitsch,W., Lepuschitz,S., Langer,S., Ries,S., Dinhopl,N., Szostak,M., Loncaric,I. and Busse,H. (2019) *Mycoplasma nasistruthionis* sp. nov. and Mycoplasma struthionis sp.nov. isolated from ostriches with respiratory disease. *Elsevier System. Appl. Microbiol.*, **43**, 126047.
39. Andrade,Y., Santos-Junior,M.N., Rezende,I.S., Barbosa,M.S., Amorim,A.T., Silva,Í.B.S., Queiroz,E.C., Bastos,B.L., Campos,G.B. and Timenetsky,J. (2020) Multilocus sequence typing characterizes diversity of *Ureaplasma diversum* strains, and intra-species variability induces different immune response profiles. *BMC Vet. Res.*, **16**, 163.
40. Miller,W.G., Yee,E., Chapman,M.H. and Bono,J.L. (2017) Comparative genomics of all three *Campylobacter sputorum Biovars* and a novel Cattle-Associated *C. sputorum* clade. *Genome Biol. Evol.*, **9**, 1513–1518
41. Phillips,N.D., La,D. and Hampson,D.J. (2019) *Brachyspira catarrhinii* sp. nov., an anaerobic intestinal spirochaete isolated from vervet monkeys may have been misidentified as Brachyspira aalborgi in previous studies. *Elsevier Anaerobe*, **59**, 8–13.
42. Hewicker-Trautwein,M., Feldmann,M., Kehler,W., Schmidt,R., Thiede,S., Seeliger,F., Wohlsein,P., Ball,HJ., Buchenau,I., Spergser,J. *et al.* (2002) Outbreak of pneumonia and arthritis in beef calves associated with *Mycoplasma bovis* and *Mycoplasma californicum*. *Vet. Rec.*, **151**, 699–703.
43. Vandamme,P., Gillis,M., Vancanneyt,M., Hoste,B., Kersters,K. and Falsen,E.(1993) Moracxella ZincoZnii sp. nov., isolated from the human respiratory tract, and reevaluation of the taxonomic position of *Moraxella osloensis*. *Int. J. Syst. Bacteriol.*, **43**, 474–481.
44. Lierz,M., Hagen,N., Lueschow,D. and Hafez,H.M. (2008) Species-specific polymerase chain reactions for the detection of *Mycoplasma buteonis, Mycoplasma flconis, Mycoplasma gypis*, and *Mycoplasma corogypsi* in captive birds of prey. *Avian Dis.*, **52**, 94–99.
45. Whitehouse,C.A., Chase,K., Embers,M.E., Kulesh,D.A., Ladner,J.T., Palacios,G.F. and Minogue,T.D. (2015) Development of real-time PCR assays for the detection of *Moraxella macacae* associated with bloody nose syndrome in rhesus (*Macaca mulatta*) and cynomolgus (*Macaca fascicularis*) macaques. *J. Med. Primatol.*, **44**, 364–372.
46. Greub,G. and Raoult,D. (2003) Rhodobacter massiliensissp. nov., a new amoebae-resistant speciesisolated from the nose of a patient. *Elsevier Res. Microbiol.*, **154**, 631–635.
47. Wang,M., Zhang,X., Jiang,T., Hu,S., Yi,Z., Zhou,Y., Ming,D. and Chen,S. (2017) Liver abscess caused by Pannonibacter phragmitetus: case report and literature review. *Front. Med.*, **4**, 48.
48. McHugh,K.E., Rhoads,D.D., Wilson,D.A., Highland,K.B., Richter,S.S. and Procop,G.W. (2016) Inquilinus limosusin pulmonary disease: case report and review of the literature. *Elsevier Diagn. Microbiol. Infect. Dis.*, **86**, 446–449.
49. Hovan,G. and Hollinger,A. (2018) Clinical Isolation and Identification of Haematospirillum jordaniae. *Emerg. Infect. Dis.*, **24**, 1955–1956.
50. Benítez-Páez,A., Olivares,M., Szajewska,H., Pieścik-Lech,M., Polanco,I., Castillejo,G., Nuñez,M., Ribes-Koninckx,C., Korponay-Szabó,I.R., Koletzko,S. *et al.* (2020) Breast-milk microbiota linked to celiac disease development in children: a pilot study from the PreventCD cohort. *Front. Microbiol.*, **11**, 1335.
51. Liu,B., Ren,N., Ding,J., Xie,G. and Guo,W. (2009) The effect of Ni^{2+}, Fe^{2+} and Mg^{2+} concentration on photo-hydrogen production by *Rhodopseudomonas faecalis* RLD-53. *Elsevier International Journal of Hydrogen Energy*, **34**, 721–726.
52. Casán,F. and Diaz-Zorita,M. (2016) Azospirillum sp. in current agriculture: From the laboratory to the field. *Elsevier Soil Biol. Biochem.*, **103**, 117–130.
53. Miroshnikov,K.K., Didriksen,A., Naumoff,D.G., Huntemann,M., Clum,A., Pillay,M., Palaniappan,K., Varghese,N., Mikhailova,N., Mukherjee,S. *et al.* (2020) Draft genome sequence of Methylocapsa palsarum NE2T, an Obligate Methanotroph from Subarctic Soil. *Am. Soc. Microbiol. Prokaryot.*, **24**, 1–2.
54. Dedysh,S.N., Knief,C. and Dunfield,P.F. (2005) Methylocella species are facultatively methanotrophic. *J. Bacteriol.*, **187**, 4665–4670.
55. Takeuchi,M., Yamagishi,T., Kamagata,Y., Oshima,K., Hattori,M., Katayama,T., Hanada,S., Tamaki,H., Marumo,K., Maeda,H. *et al.* (2015) Tepidicaulis marinus gen. nov., sp. nov., a marine bacterium that reduces nitrate to nitrous oxide under strictly microaerobic conditions. *Int. J. System. Microbiol.*, **65**, 1749–1754.
56. Suarez,C., Ratering,S., Geissler-Plaum,R. and Schnel,S. (2014) Hartmannibacter diazotrophicus gen. nov., sp. nov., a phosphate-solubilizing and nitrogen-fixing alphaproteobacterium isolated from the rhizosphere of a natural salt-meadow plant. *Int. J. Syst. Evol. Microbiol.*, **64**, 3160–3167.
57. Bibi,F., Chung,E.J., Kha,A., Jeon,C.O. and Chung,Y.R. (2013) Martelella endophytica sp. nov., an antifungal bacterium associated with a halophyte. *Int. J. Syst. Evol. Microbiol.*, **63**, 2914–2919.
58. Sternberg,S., Redding,S., Jinek,M., Greene,E.C. and Doudna,J.A. (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, **507**, 62–67.
59. Rimseliene,R., Maneliene,Z., Lubys,A. and Janulaitis,A. (2003) Engineering of restriction endonucleases: Using methylation activity of the bifunctional endonuclease Eco571 to select the mutant with a novel sequence specificity. *J. Mol. Biol.*, **327**, 383–391.
60. 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.*, **98**, 4136–4141.
61. Collier,J., McAdams,H.H. and Shapiro,L. (2007) A DNA methylation ratchet governs progression through a bacterial cell cycle. *Proc. Natl Acad. Sci.*, **104**, 17111–17116.