UNIVERSITY OF CALIFORNIA

Santa Barbara

The kinetic mechanism of DNA strand separation by high-fidelity DNA methyltransferase, CcrM

A dissertation submitted in partial satisfaction of the requirements for the degree of

Doctor of Philosophy

in

Biochemistry and Molecular Biology

by

Olivia R. Konttinen

Committee in charge:

Professor Norbert Reich, Chair

Professor Brandon Greene

Professor Kevin Plaxco

June, 2024

The dissertation of Olivia R. Konttinen is approved

Professor Kevin Plaxco

Professor Brandon Greene

Professor Norbert Reich, Committee Chair

June, 2024

ACKNOWLEDGEMENTS

I joined Dr. Norbert Reich's lab because of his intense and inspirational approach to biochemical research and continuous learning. After 6 years, I have grown as a person and scientist because of Dr. Reich's mentorship. Thank you, Norbert, for the many years of spontaneous conversations, discussing exciting new results, always being available and engaged in my work, and guiding me through the exciting world of biochemistry and enzymology.

Several of my publications would not have been possible without the collaboration and mentorship from Kenneth A. Johnson and Tyler Dangerfield at The University of Texas at Austin. Ken and Tyler taught me numerous lessons in enzyme mechanisms and global data fitting.

I would like to thank my parents (Jari and Janet Konttinen) and siblings (Bridget, Lillie, and Patrick) for supporting and encouraging me through every chapter of my graduate studies and my 29 years of life that got me to this point. My advisors and mentors Kevin Plaxco, Martin Kurnik, Brandon Greene, John Lew, and Rick Dalquist amongst many other advisors. My undergraduate mentor, Christophe Monnier, who I worked with in Herb Waite's lab, introduced me to biophysical and biochemical research and inspired me to pursue a PhD. My lab-mates, Hanson Huang, Alexandra Lantz, Jennifer Vargas, Jason Carmody, Jonathan Sandoval, Ethan Ward, Jaqueline Roher, Claire Jin, Yifan Zhou, Yichen Yang, Kyle Anderson, Danea Palmer, Tyler Dangerfield and many others.

VITA OF OLIVIA R. KONTTINEN

June 2024

EDUCATION

Bachelor of Science in Biology, University of California, Santa Barbara, June 2017

Doctor of Philosophy in Biochemistry and Molecular Biology with an emphasis in Bioengineering and Biophysics, University of California, Santa Barbara, June 2024 (expected)

PROFESSIONAL DEVELOPMENT

2018-2024: Teaching assistant in Biochemistry, Drug design, Pharmacology, Introductory Biology lab, Enzyme kinetics, Biochemistry lab, University of California, Santa Barbara

2018-2024: Research mentor for undergraduate interns, The Reich Lab, University of California, Santa Barbara

2023: UC Leads summer research mentor for visiting undergraduate intern, The Reich Lab, University of California, Santa Barbara

2021-2023: Scientific advisor and module development lead, Scitrek outreach program

PUBLICATIONS

Konttinen, O., Carmody, J., Pathuri, S., Anderson, K., Zhou, X., and Reich, N. (2020) Cell cycle regulated DNA methyltransferase: fluorescent tracking of a DNA strand-separation mechanism and identification of the responsible protein motif. *Nucleic Acids Research*, 48, 20, 11589-11601.

Konttinen, O., Carmody, J., Kurnik, M., Johnson, K.A., and Reich, N. (2023) High fidelity DNA strand-separation is the major specificity determinant in the DNA methyltransferase CcrM's catalytic mechanism. *Nucleic Acids Research*, 51, 13, 6883-6898.

In preparation for submission to *Nucleic Acids Research*: Konttinen, O., Dangerfield, T., Vargas, J., Johnson, K.A., and Reich, N. (2024) Protein loops are major contributors to DNA strand separation and high-fidelity substrate recognition for DNA methyltransferase CcrM

In preparation for submission to *Nucleic Acids Research*: Konttinen, O., Dangerfield, T., Johnson, K.A., and Reich, N. (2024) Target adenine base-flipping follows DNA strand-separation and is the rate-determining step for methylation and catalysis by cell-cycle regulated DNA methyltransferase CcrM.

AWARDS

DeWolf Teaching Fellowship in Organic Chemistry 2022, The Department of Chemistry and Biochemistry, University of California Santa Barbara

CONERENCES

Enzyme Mechanism Conference - Tucson, Arizona January 2022. Poster presentation.

ABSTRACT

The kinetic mechanism of DNA strand separation by high-fidelity DNA methyltransferase, CcrM

by

Olivia R. Konttinen

DNA methyltransferases are responsible for transcriptional regulation, cell cycle progression, DNA repair, DNA protection, tumor suppression, and several other important biological processes. Aberrant bacterial DNA methylation can lead to cell death and loss of protection against viral infection; in humans this leads to cancer, autoimmune diseases, metabolic disorders, and neurological disorders. Thus, DNA methyltransferases are common drug targets for cancer therapeutics and novel antibiotics.

The conformational transitions in DNA and protein that govern recognition, substrate accessibility, and catalysis are fundamental to understanding the mechanisms that regulate biological processes. The bacterial N^6-adenine cell-cycle regulated DNA methyltransferase, CcrM, is the first DNA methyltransferase shown to rely on a unique DNA recognition mechanism in which the DNA strands are separated and most recognition interactions appear to involve the target strand. Strand-separation is emerging as a novel DNA recognition mechanism but the underlying mechanisms and quantitative contribution of strand-separation to fidelity remain obscure for any enzyme (CRISPR-Cas9 and RNA polymerase sigma factor).

This work uncovers the fundamental steps governing CcrM's DNA strand separation and high-fidelity DNA recognition mechanism. We relied on mutational analysis of highly conserved residues in the C-terminal domain, Loop-2B, Loop-45, and the active site to probe the function of structurally implicated protein moieties. We collected stopped-flow kinetic

fluorescence to monitor transitions in DNA and protein and relied on rigorous global data fitting to understand the states that regulate catalysis in CcrM.

We incorporated Pyrrolo-dC into cognate and noncognate DNA to monitor the kinetics of strand-separation and used tryptophan fluorescence to follow protein conformational changes. Both signals are biphasic and global fitting showed that the faster phase of DNA strand-separation was coincident with the protein conformational transition. Non-cognate sequences did not display strand-separation and methylation was reduced >300-fold, providing evidence that strand-separation is a major determinant of selectivity. Analysis of an R350A mutant (C-term domain) showed that the enzyme conformational step can occur without strand-separation, so the two events are uncoupled. A stabilizing role for the methyl-donor (SAM) is proposed; the cofactor interacts with a critical loop which is inserted between the DNA strands, thereby stabilizing the strand-separated conformation.

Loops 2B and 45 are inserted between the strand-separated DNA interface. During strand-separation, residues within Loops 2B, 45, and 6E contact the target DNA strand that undergoes methylation. R44 and R129 (Loop-2B and Loop-45, respectively) when mutated to Alanine, disrupt strand-separation and are catalytically inactive. The highly conserved Loop-45 residue F125, which is positioned between the separated DNA strands, is also essential for maintaining the strand-separated intermediate; replacement of F125 with Alanine, Leucine, and Tryptophan results in various perturbations of strand-separation that are correlated to the bulkiness of the substituted residue. Global fitting for each mutant shows that generation and stabilization of DNA strand-separation are perturbed, providing a functional role for these loops in generating and stabilizing the strand-separated intermediate, which is essential for discrimination and catalysis.

Employing a fluorescent adenine analog (6MAP) at the target position to monitor base flipping, we resolved that target adenine base flipping follows DNA strand separation and is followed by fast methylation and fast product DNA release. A W57F mutant (active site) displayed an unaltered rate of base flipping as monitored by 6MAP fluorescence but greatly reduced rate of methylation, showing that base-flipping and methylation can be uncoupled. In addition, single-stranded DNA bypasses the DNA strand separation step, while rates of base flipping measured by 6MAP fluorescence and DNA methylation are similar to dsDNA. Global data fitting for each model resolves that base flipping of the target adenine is the rate-limiting step in catalysis.

The results presented here are broadly applicable to the study of other N^6-adenine methyltransferases that contain the structural moieties implicated in strand-separation (Loop-2B, Loop-45, and the C-terminal domain), which are found widely dispersed across many bacterial phyla, including human and animal pathogens. Insights into CcrM's mechanism of DNA strand-separation are likely to clarify strand-separation mechanisms for other enzymes such as CRISPR-Cas9 and RNA polymerase sigma factor. Additionally, the elevated understanding of CcrM's strand-separation mechanism could be useful for the development of selective CcrM inhibitors as novel antibiotics.

TABLE OF CONTENTS

Chapter I: Cell-cycle regulated DNA methyltransferase, CcrM, is an important epigenetic regulator of gene expression and cell-cycle progression

I. CcrM biology

Bacterial DNA methylation is involved in diverse functions including restriction/modification (R/M), the control of gene regulation, mismatch repair, and replication timing[1.1]. The R/M enzymes (e.g. M.HinfI, M.EcoRI) comprise the majority of known bacterial MTases, whereas orphan CcrM enzymes are found in α-proteobacteria and orphan Dam enzymes are found in ɣ-proteobacteria[1.2, 1.3]. Orphan MTases do not appear to have a cognate endonuclease and control the complex network that regulates cell cycle regulatory genes[1.4]. *Caulobacter crescentus* is a gram-negative aquatic alphaproteobacterial and is an excellent model to study epigenetics because the *Caulobacter crescentus* chromosome replicates once per cell division[1.5].

N^6-adenine methylation by CcrM is an important epigenetic mark for regulation of gene expression and cell cycle progression in *Caulobacter crescentus*[1.5, 1.2, 1.6, 1.7]. At the beginning of the cell cycle, the chromosome is fully methylated at CcrM recognition sites (5′-GANTC-3′) in the G1 phase (Fig. 1.1)[1.4]. Upon replication, the two replication forks proceed, generating hemimethylated DNA in the S phase. CcrM is expressed only in the G2 phase for a short <20 minute window and rapidly remethylates 4515 GANTC sites on the daughter strand (Fig. 1.1)[1.4]. CcrM is then degraded by Lon-mediated proteolysis[1.8,1.9]. 27 GANTC sites remain unmethylated at all stages of the cell cycle[1.10, 1.11, 1.2]. Single-molecule real-time (SMRT) DNA

sequencing has elevated the analysis of methylomes which has elucidated the spatial and temporal methylation patterns during this cell cycle[1.10, 1.11, 1.2].

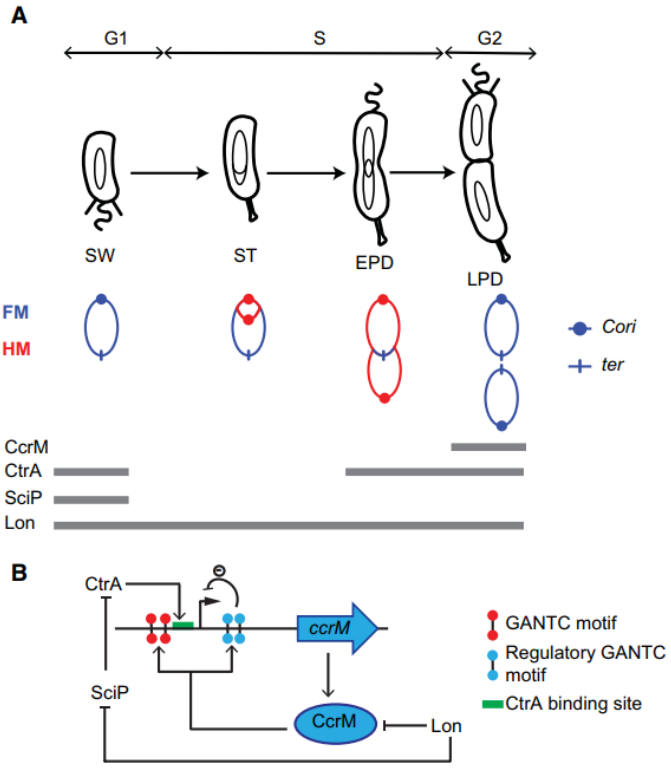


Figure 1.1. (Collier 2018) DNA methylation by CcrM during the *C. crescentus* cell cycle. **A.** Cell cycle progression of the swarmer cell (SW), stalked cell (ST), early pre-divisional stage (EPD) and late pre-divisional stage (LPD). Methylation state of most GANTC motifs throughout the cell cycle (FM = fully methylated and HM = methylated and hemimethylated). The temporal expression of different enzymes throughout the cell-cycle indicated by a grey bar. **B.** The ccrM promoter region carries four GANTC motifs (lollipops): two before and two after the transcriptional start site. The two blue motifs are supposedly important for the feedback regulation of ccrM transcription by CcrM[1.4].

CcrM-dependent methylation can direct the binding of transcription factors to specific methylated sequences and affect the expression of genes depending on the methylation state of their promoters[1.5]. CcrM, in concert with three transcription factors (DnaA, GcrA, and CtrA), orchestrates the cell cycle-regulated cell division in which a single genome gives rise to two distinct and heritable cell types (flagellated swarmer cell and a stalked cell)[1.5, 1.8].

CcrM was thought to be essential for viability in C. crescentus, Agrobacterium tumefaciens, B. abortus, and Rhizobium meliloti, but recent results under low-nutrient conditions indicate it to be dispensable in C. crescentus, presumably because of slow growing

conditions, allowing for a complete cell division. In contrast, growths in rich media cause the cells to deform and die[1.11, 1.2, 1.7, 1.12, 1.13]. In summary, CcrM is a key epigenetic player in gene expression and cell-cycle progression in *Caulobacter crescentus* and other bacterial organisms.

I. References

1.1. Adhikari,S. and Curtis,P. (2016) DNA methyltransferases and epigenetic regulation in bacteria. *FEMS Microbiol. Rev.,* 40, 575–591.

1.2. Gonzalez, D., Kozdon, J. B., McAdams, H. H., Shapiro, L., and Collier, J. (2014) The functions of DNA methylation by CcrM in Caulobacter crescentus: a global approach. *Nucleic Acids Res*. 42, 3720−3735.

1.3. Wright, R., Stephens, C., and Shapiro, L. (1997) The CcrM DNA methyltransferase is widespread in the alpha subdivision of proteobacteria, and its essential functions are conserved in Rhizobium meliloti and Caulobacter crescentus. *J. Bacteriol.* 179, 5869−5877.

1.4. Mouammine,A. and Collier,J. (2018) The impact of DNA methylation in Alphaproteobacteria. *Molecular Microbiology*, 110, 1, 1-10.

1.5. Mohapatra, S. S., Fioravanti, A., and Biondi, E. G. (2014) DNA methylation in *Caulobacter* and other Alphaproteobacteria during cell cycle progression. *Trends Microbiol*. 22, 528−535.

1.6. 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, e1004831.

1.7. Gonzalez, D., and Collier, J. (2013) DNA methylation by CcrM activates the transcription of two genes required for the division of Caulobacter crescentus. *Mol. Microbiol.* 88, 203−218.

1.8. Collier, J., McAdams, H. H., and Shapiro, L. (2007) A DNA methylation ratchet governs progression through a bacterial cell cycle. *Proc. Natl. Acad. Sci. U. S. A.* 104, 17111−17116.

1.9. Wright, R., Stephens, C., Zweiger, G., Shapiro, L., and Alley, M. R. (1996) Caulobacter Lon protease has a critical role in cell-cycle control of DNA methylation. *Genes Dev.* 10, 1532−1542.

1.10. Berdis, A. J., Lee, I., Coward, J. K., Stephens, C., Wright, R., Shapiro, L., and Benkovic, S. J. (1998) A cell cycle-regulated adenine DNA methyltransferase from *Caulobacter crescentus* processively methylates GANTC sites on hemimethylated DNA. *Proc. Natl. Acad. Sci. U. S. A*. 95, 2874−2879.

1.11. 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. Bacteriol.* 182, 3482−3489.

1.12. Stephens, C., Reisenauer, A., Wright, R., and Shapiro, L. (1996) A cell cycle-regulated bacterial DNA methyltransferase is essential for viability. *Proc. Natl. Acad. Sci. U. S. A.* 93, 1210−1214.

1.13. Kahng, L. S., and Shapiro, L. (2001) The CcrM DNA methyltransferase of Agrobacterium tumefaciens is essential, and its activity is cell cycle regulated. *J. Bacteriol.* 183, 3065−3075.

Chapter II: CcrM is orders of magnitude more discriminating than other DNA methyltransferases and relies on a novel DNA strand-separation mechanism

II. CcrM biochemical parameters

DNA methyltransferases can be organized by the reactions they catalyze (N^6-adenine and N^4-cytosine exocyclic amine methylation, C^5-cytosine methylation) and by their organization of conserved motifs (α, β, γ, ε, ζ)[2.1]. Based on the arrangement of conserved motifs, CcrM is a β-class adenine N^6-adenine Mtase[2.1].

Prior to the publication of the co-crystal structure, CcrM was thought to be a functional monomer and shown to rely on an unusual DNA recognition mechanism with unique kinetic parameters[2.2, 2.3, 2.4]. CcrM revealed a profound level of sequence discrimination between ssDNA, dsDNA, and ssRNA (Fig. 2.1)[2.2]. The enzyme is uniquely able to efficiently methylate

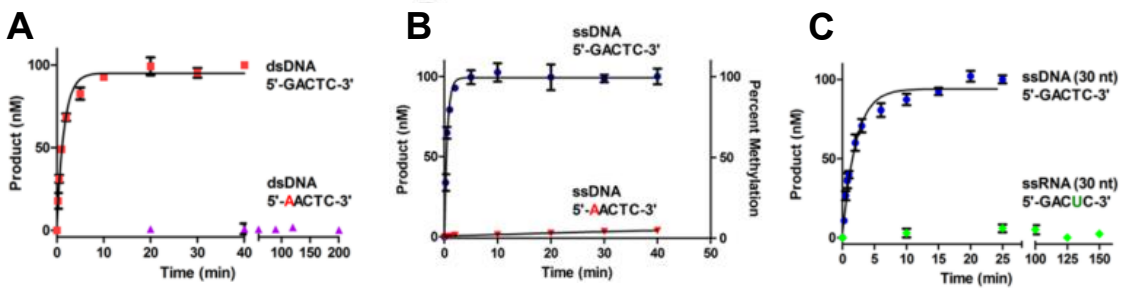


Figure 2.1 (Woodcock 2017[2.2]). CcrM can discriminate between ssDNA, dsDNA, and ssRNA. **A.** Single turnover on dsDNA (red squares) with 100 nM substrate and 150 nM CcrM and single turnover with the canonical sequence changed to AACTC (purple triangles) with 1 μM substrate and 1.5 μM CcrM. **B.** Single turnover on ssDNA (blue circles) with 100 nM substrate and 150 nM CcrM and single turnover with the canonical sequence changed to AACTC (red inverted triangles) with 1 μM substrate and 1.5 μM CcrM. **C.** Single turnover on 30-nucleotide ssDNA (blue circles) with 100 nM substrate and 150 nM CcrM and single turnover on ssRNA (green diamonds) with 5 μM substrate, 25 μM CcrM, and excess 100 μM AdoMet.

its cognate recognition site 5′-GANTC-3′ in single-stranded (ss) and hemimethylated double-stranded (ds) DNA, and can discriminate 10^6-fold against ssRNA (Fig. 2.1)[2.2, 2.3]. Recognition of both ssDNA and dsDNA is not novel for a DNA modifying enzyme, but it is seldom observed. CcrM's ability to methylate both ssDNA and dsDNA, presents a perplexing biological role for the enzyme, because the dominant cellular single-stranded nucleic acid is RNA.

Site-specific investigation of CcrM's discrimination against RNA revealed that the enzyme can accommodate riboses outside of the recognition site, but activity is negligible when all nucleic acids in the recognition site are riboses (Fig. 2.2a)[2.3]. Replacement of one or two recognition site deoxyribose nucleic acids with ribose results in dramatic decreases in activity (Fig. 2.2b)[2.3]. CcrM's discrimination against ssRNA is likely not due to the formation of secondary structures that occur in ssRNA folding due to its ability to recognize both ssDNA and dsDNA, implying that variation in secondary structure is accommodated by the enzyme. CcrM likely discriminates against ssRNA due to a chemical or spatial preference for deoxyribose over ribose in the nucleic acid backbone.

CcrM also stands out for its unprecedented level of discrimination against noncognate DNA; it discriminates 10^7-fold against single-base substitutions in noncognate sequences (5′-GANTC-3′ vs 5′-AANTC-3′), which is orders of magnitude greater than other Mtases (Fig. 2.1) [2.2]. CcrM methylates ssDNA efficiently ($k_{methylation}/K_D^{DNA}$), but with much less discrimination than dsDNA[2.2, 2.3]. The increased specificity with dsDNA compared to ssDNA suggests that a dsDNA substrate undergoes a step that is likely involved in substrate discrimination while the ssDNA substrate does not.

In support of this, mismatch DNA in which one noncognate base is substituted in only the nontarget strand enhances methylation, which further supports selective recognition of a single (target) strand[2.3]. In contrast, mismatches disrupt methylation by other enzymes such as Dam and HhaI[2.3]. The CcrM protein-DNA interface is unusually large and contacts 17-19 bases as determined by kinetic footprinting, which is dramatically large for a small 39kD enzyme (Fig. 2.3)[2.3]. Interestingly, CcrM has a 2-fold to 4-fold preference for a 5' cytosine in 17nt and 19nt ssDNA substrates, respectively (Fig. 2.3)[2.3]. The presence of a pyrimidine at the 5' end is likely to favor a conformational change in the DNA that is less favored with a purine.

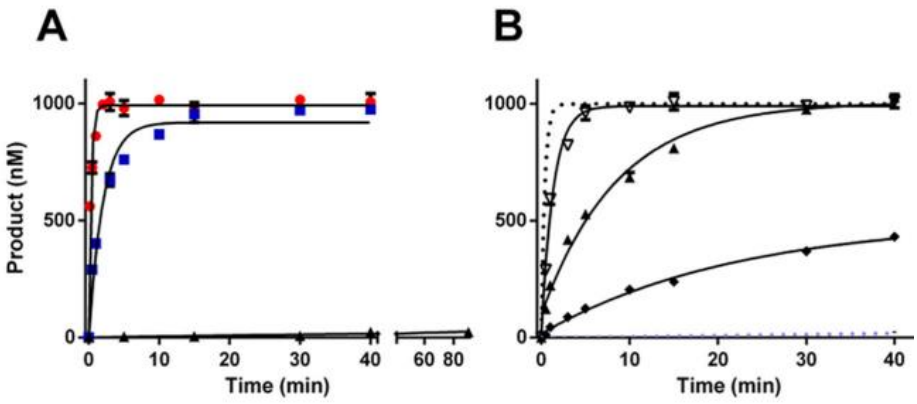


Figure 2.2 (Woodcock 2018[2.3]). CcrM is highly discriminating against RNA. **A.** CcrM single-turnover methylation assay with an oligonucleotide made up exclusively of deoxyriboses (red circles) or with an oligonucleotide with three-flanking riboses on either side of the GANTC (5'-rArGrGGACTCrGrCrC-3') (blue squares) or with an oligonucleotide that has riboses at all five internal positions (5'-AGGrGrArCrTrCGCC-3') (black triangels). **B.** The DNA control is given for perspective as a dotted black line (5'-AGGGACTCGCC-3'), and the all internal ribose is given as a light blue dotted line (5'-AGGrGrArCrTrCGCC-3'). The open upside-down triangles represent a single ribose substitution at the adenine (5'-AGGGrACTCGCC-3'), the right-side-up filled triangles represent a single ribose at the cytosine (5'-AGGGArCTCGCC-3'), and the black diamonds are both positions with a ribose (5'-AGGGrArCTCGCC-3').

Another unique kinetic result for CcrM was observed in multiple turnover methylation experiments which revealed that CcrM does not display a pre-steady state burst, indicating that

a step preceding methylation is slow or that methylation itself is slow[2.2]. In contrast, DNA methyltransferases typically display burst kinetics followed by a lag phase which corresponds to rapid enzyme-bound product formation followed by rate-limiting product release.

CcrM was shown to processively methylate both ssDNA and dsDNA (Fig. 2.4)[2.2], which is consistent with its ability to methylate 4515 GANTC sites within a <20 min timeframe[1.8]. CcrM is highly processive on ssDNA and dsDNA which contributes to CcrM's high level of efficiency. CcrM's processivity results reveal that the enzyme can bind nonspecific DNA, however methylation and catalysis are highly specific processes.

Bioinformatics and protein sequence analysis also revealed unique features of CcrM. CcrM has an unusual 83-amino acid C-terminal domain[2.3]. Mutational analysis of the CcrM C-term supported that the C-terminal domain was essential for binding and methylation[2.3]. A CcrM mutant that lacks the C-terminal domain (CcrM truncation) has no binding or methylation activity for ssDNA nor dsDNA (Table 2.1)[2.3]. W332 is a highly conserved C-terminal domain residue and mutant W332A has no binding nor methylation activity with ssDNA nor dsDNA (Table 2.1)[2.3]. Together, these results indicate that the C-terminal domain is essential and plays an important role in DNA recognition of ssDNA and dsDNA substrates.

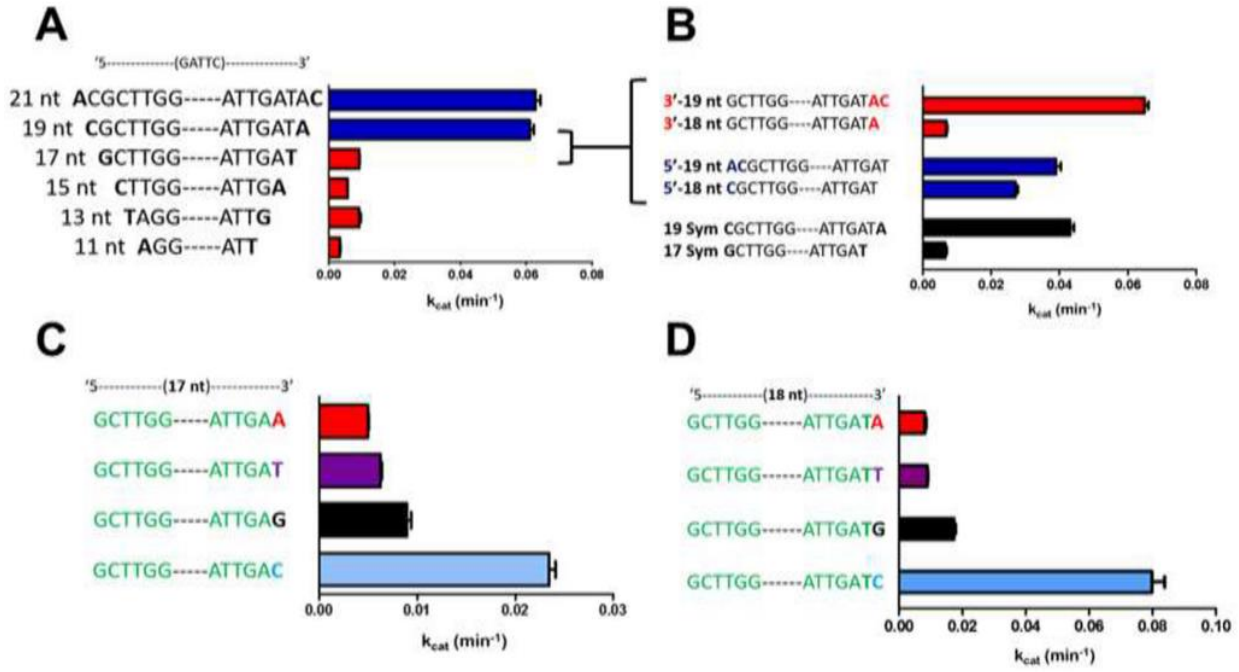


Figure 2.3 (Woodcock 2018[2,3]). CcrM C. crescentus has an unusually large DNA interface as determined by kinetic footprinting on ssDNA substrates. **A.** ssDNA substrates with varying length show that CcrM requires 19 or more nucleotides to perform efficient catalysis. **B.** The asymmetric 19-nt substrate is more active than the symmetric 19-nt substrate. Removal of the 3' C results in significant loss of activity. **C.** CcM has approximately 2-fold more activity with a substrate that contains a cytosine at the 3' end of 17-nt symmetric substrate. **D.** CcM has approximately 4-fold more activity with a substrate that contains a cytosine at the 3' end of 18-nt asymmetric substrate.

The CcrM co-crystal structure was published in 2019 and resolved that CcrM is a dimer on dsDNA (Fig. 2.5)[2,5]. The most profound observation from the crystal structure was the distortion of the DNA molecule in which four of the five bases of the recognition site are strand separated, losing Watson-Crick base-

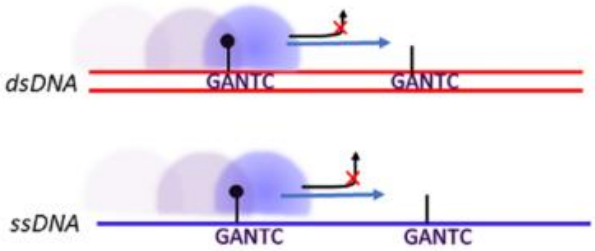


Figure 2.4 (Woodcock 2018[2,3]). CcrM processively methylates ssDNA and dsDNA substrates.

pairing and base-stacking interactions (Fig. 2.5). Structurally, CcrM consists of 358 amino acids and each monomer has a core (residues 1-254), a flexible linker (residues 254-274) and a C-terminal 83-residue segment (residues 275-358). Monomer A (Cyan, Fig. 2.5) makes nearly all the contacts to the "target" strand (magenta) which is positioned to be methylated. Interactions to four of the five recognized bases are only possible because of the complete disruption of basepairing to the nontarget strand (yellow) (Fig. 2.5). 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 (Fig. 2.5).

The newly observed DNA strand-separated structure supported the previous biochemical data and provided meaningful interpretations. For example, CcrM's ability to methylate ssDNA does not suggest that CcrM's biological role includes ssDNA methylation in *Caulobacter*. Rather, CcrM is capable of ssDNA recognition and methylation because after dsDNA strand-separation, it selects for recognition and modification of a single strand (the target strand). Additionally, the nontarget strand mismatch DNA substrates where a mismatch was positioned at the "N" position could have enhanced specificity due to lack of energy required to break the Watson-Crick basepair at this position.

Enzyme	Double-stranded DNA			Single-stranded DNA		
	$k_{methylation}$	K_d	$k_{methylation}/K_d$	$k_{methylation}$	K_d	$k_{methylation}/K_d$
	min^{-1}	nM	nM min^{-1}	min^{-1}	nM	nM min^{-1}
CcrM (*C. crescentus*)	2.61 ± 0.21	70.8 ± 7.2	0.037 ± 0.005	3.33 ± 0.30	39.2 ± 4.5	0.085 ± 0.012
CcrM *C. crescentus* truncation	No activity	No activity	No activity	No activity	No activity	No activity
CcrM *C. crescentus* W332A	No activity	No activity	No activity	No activity	No activity	No activity
CcrM A. tumefaciens	0.50 ± 0.05	89.0 ± 3.1	0.006 ± 0.001	0.73 ± 0.07	34.2 ± 5.0	0.021 ± 0.004
CcrM *B. abortus*	0.42 ± 0.03	41.5 ± 7.4	0.010 ± 0.002	0.46 ± 0.06	77.1 ± 14.6	0.006 ± 0.001
M.HinfI	1.27 ± 0.10			0.060 ± 0.003		
M.HinfI truncation	No activity	>1 μM	No activity	0.39 ± 0.03		
M.HhaII	2.30 ± 0.21	31.5 ± 3.0	0.073 ± 0.010	>300 times slower		
DpnA	0.020 ± 0.001			0.051 ± 0.004		

Table 2.1 (Woodcock 2018[2,3]). Summary of kinetic data and substrate specificity constants, single-turnover assays, and thermodynamic constants for CcrM, orthologs, mutants, and non-ortholog enzymes. Bold type indicates a lack of the C-terminal segment.

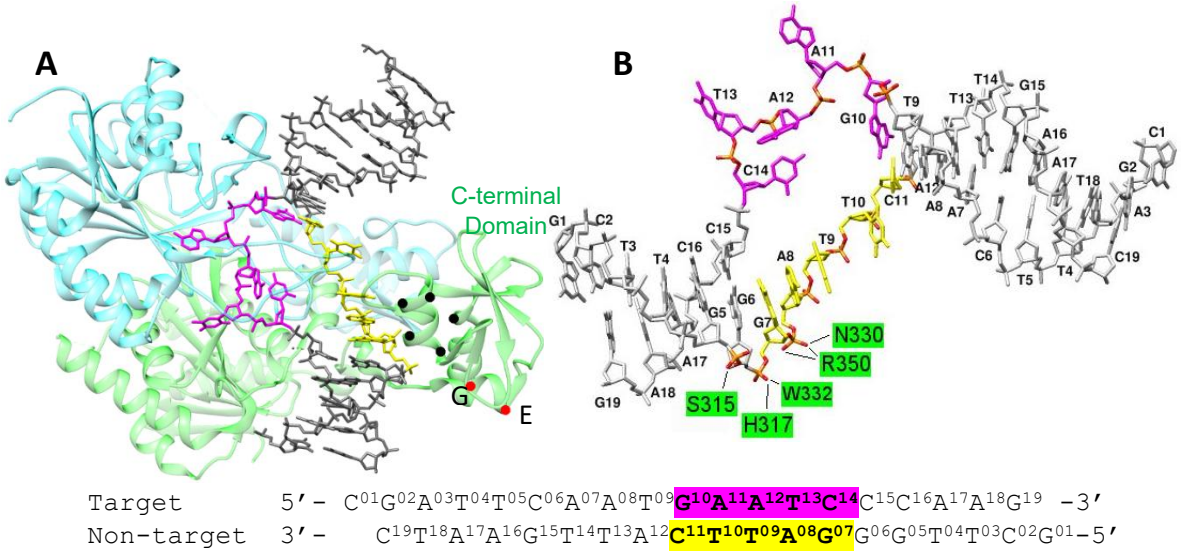


Figure 2.5. Cocrystal stucture of CcrM-DNA. A. Ccrm and double stranded DNA (PDB: 6PBD). Monomer A is shown in cyan. Monomer B is shown in green. The DNA target strand and non-target strand are shown in magenta and yellow, respectively. Residues from the C-terminal segment of monomer B that interact 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 basepairs 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.

While the cocrystal structure vindicated several of CcrM's unique kinetic characteristics, it also prompted several mechanistic questions regarding the many details of CcrM's DNA strand separation mechanism. The following chapters of this dissertation provide novel data that resolve the DNA recognition mechanism of strand-separation utilized by CcrM.

II. References

2.1. 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.

2.2. 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.

2.3. 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.

2.4. Shier,V.K., Hancey,C.J., and Benkovic,S.J. (2001) Identification of the active oligomeric state of an essential adenine DNA methyltransferase from *Caulobacter crescentus*. *The Journal of Biological Chemistry*, 276, 18, 14744-14751.

2.5. 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. *Nature Communications,* 10, 4600.

Chapter III: Fluorescent tracking of CcrM-dependent DNA strand-separation and identification that the C-terminal domain is essential for this process

III. Abstract

The recent CcrM-DNA cocrystal structure showed the CcrM dimer strand-separates four of the five basepairs 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 CcrM-dependent fluorescence increase. 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 reduction in strand separation monitored by PydC. 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 methyltransferase (HhaII) that lacks the conserved C-terminal segment. The results presented here demonstrate that strand separation can be tracked fluorescently and that the C-terminal domain is essential for this process.

III. Introduction

DNA MTases typically have a large domain with conserved motifs, which binds SAM as well as stabilizes the base that undergoes methylation in an extrahelical position (base flipping)[3.1]. A smaller domain frequently contains the target recognition domain (TRD),

although residues distributed throughout the MTases contribute to DNA recognition[3.1]. 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[3.2, 3.3, 3.4, 3.5.] CcrM follows this organization; the larger CcrM N-term contains the conserved catalytic DPPY motif and the active site, while the smaller CcrM C-term is structurally involved in DNA binding and stabilization of the strand-separated intermediate. Based on protein engineering efforts of CcrM and related enzymes the highly conserved C-terminal segment of the protein is involved in sequence discrimination[2.3].

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[3.1, 3.2, 3.3, 3.4, 3.5, 3.24]. Base flipping contributes to specificity since only when bound to the cognate sequence do methyltransferases undergo conformational changes that facilitate base flipping[3.4, 3.5]. 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[3.1]. Interestingly, while the vast majority of DNA MTases act exclusively or largely on dsDNA, a few reports describe enzymes whose biological role is to act on single stranded DNA[3.6], or minimally, can methylate both single stranded DNA and unpaired DNA *in vitro*, including human enzymes[3.7].

The CcrM-dsDNA 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[2,5]. 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**.** Here we apply a fluorescence-based assay to interrogate the strand separation step to better understand these issues.

III. 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 minutes 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 (1M HEPES, 10mM EDTA, 200mM 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 was collected over the interval 400 to 550 nm using an 8 nm slit size. Each measurement and background were 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). Five scans were taken, and discarded reads were not replaced. 3-5 scans were averaged for each sample. Samples were kept on ice and allowed to equilibrate to room temperature for 5-10 minutes 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[2.2]. Mutant plasmids were constructed using the Agilent Quickchange Lightning Site-Directed Kit. The primers used in the PCR reactions are provided in SI Table 3[3.0]. 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, 1L 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 OD600nm of 0.8 was reached. Cultures were placed on ice for 10 min and induced with 2 mM IPTG (GoldBio) and shaken for 3 hours at 225 rpm at room temperature. Cells were pelleted by centrifugation using a JA-10 rotor and a J2-21 centrifuge (Beckman) at 5,000 rpm at 4°C for 20 min and stored at -80°C. Cell pellet were 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 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 one hour. 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 mM 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, 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 1X 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. 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 (five 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 five minutes followed by two additional washes with the same buffer for five minutes. This was followed by a five-minute treatment with 400 mL 80 % EtOH, another wash for five minutes with 400 mL of 100 % EtOH, and a final drying step for five minutes in 400 mL of ether in a fume hood. Samples were then placed into scintillation vials containing three mL of BioSafe II fluid. Radiochemical data was 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 SI Table 4[3.0].

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 sinefungin (Sigma Aldrich). Reactions were incubated on ice for 30 minutes, diluted with an equal volume of 50 % glycerol and loaded onto a 12 % (75:1) Native PAGE gel and run for 60 minutes at 286 V in 0.5 X 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 was then fitted to a one-site specific binding model from which dissociation constants were obtained. Substrate DNA sequences can be found in SI Table 4[3.0].

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). To construct pXMCS2-CcrM, the ccrM ORF was amplified and inserted into NdeI-KpnI digested pXMCS2 via Gibson assembly[3.8]. 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'-GACTCAAAAGCGCCTGAAAGGC-3') and pXMCS2-rev (5'-TTACCGCCTTTGAGTGAGCTG-3'), and the *CcrM* mutagenesis was confirmed by Sanger sequencing.

Microscopy

C. crescentus cells were grown in M2G media lacking xylose to deplete wild type CcrM[3.8]. 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[3.9] was used to determine cell outlines and lengths from phase images.

III. Results

The CcrM-DNA cocrystal structure reveals that four of the five basepairs within the recognition site are disrupted upon CcrM binding, with the target strand positioned away from the complementary non-target strand (Fig. 2.5)[2.5]. 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)-3*H*-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[3.10, 3.11, 3.12, 3.13, 3.14, 3.15]. When Pyrrolo-dC is positioned in the "target" strand which undergoes methylation (Figure 3.1A, $P1^{T}$) we observed a dramatic CcrM-dependent increase in fluorescence (Figure 3.1B) which is not observed when Pyrrolo-dC is positioned in the "non-target" strand (Figure 3.1A, $P0^{NT}$) which does not undergo methylation (Figure 3.1B); note, CcrM is oriented on the dsDNA by using hemimethylated substrates, which is the form of DNA that CcrM interacts with biologically[1.2]. 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 3.2)[2.3]. Thus, the Pyrrolo-dC assay shows similar protein-dependent fluorescence changes with enzymes that are functionally and structurally similar to CcrM[2.2]. 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[2.2]. No CcrM-dependent changes in fluorescence are observed if Pyrrolo-dC is positioned outside the recognition site (Figure 3.1A, $P2^{NT}$, $P3^{T}$, $P8^{T}$, Figure 3.1D) which is consistent with the cocrystal structure[2.5]. Placement of Pyrrolo-dC at the C position within the recognition site in the target and non-target strands (Figure 3.1A, $P6^{NT}$, $P7^{T}$, Figure 3.1D) also show 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 basepair remains intact in the structure (Figure 2.5), consistent with the lack of significant fluorescence changes. In contrast, the basepairing 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.

A

Sequence	I.D.
5′–G**M**P**T**C–3′ 3′–CTGAG–5′	P0^{NT}
5′–GA**P**TC–3′ 3′–CTG**M**G–5′	P1^{T}
5′–GACTC G–3′ 3′–CTG**M**G **P**–5′	P2^{NT}
5′–G**M**CTC G–3′ 3′–CTGAG **P**–5′	P3^{T}
5′–**A**A**P**TC–3′ 3′–**T**TG**M**G–5′	P4^{T}
5′–GA**P**T**T**–3′ 3′–CTG**M****A**–5′	P5^{T}
5′–G**M**CT**P**–3′ 3′–CTGAG–5′	P6^{NT}
5′–G**M**CTC–3′ 3′–**P**TGAG–5′	P7^{T}
5′–GACTC------**P**–3′ 3′–CTG**M**G------G–5′	P8^{T}

5′- TCACTGTACTCTGA**P**TCGCCTGACATGAC -3′
3′- AGTGACATGAGACTG**M**GCGGACTGTACTG -5′

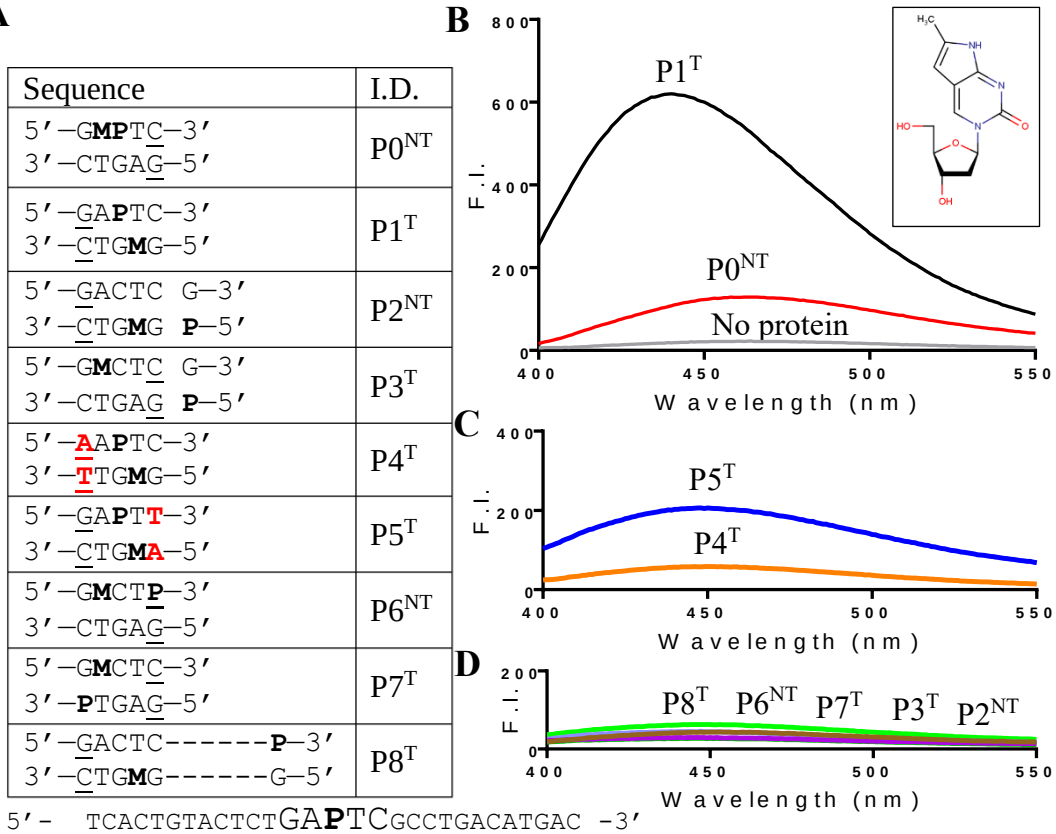


Figure 3.1. Pyrrolo-dC strand separation assay. P refers to Pyrrolo-dC, M refers to methylated adenine. The inset shows Pyrrolo-dC. **A.** The parent sequence (bottom) is the 29mer DNA with P placed in the target strand and the GANTC recognition site is enlarged. Changes from the parent sequence are shown for all sequences. The underlined C:G basepair signifies that base pairing is maintained while the other four basepairs in the recognition site are disrupted (Figure 1A and B). The red A:T basepairs 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 SI Fig. 6C$^{3.0}$. All fluorescence data contains DNA (1μM), sinefungin (60μM), and WT CcrM (2.5μM, monomer concentration). All dsDNA is 29bp, hemimethylated, and contains 1 centrally located GANTC recognition site. Data collected at room temperature on a Horiba Scientific Fluoromax-4, with excitation at 350nm. All traces

average five scans. Background signal was subtracted from all traces. F.I. = Fluorescence Intensity.

Our prior study of CcrM sequence specificity showed CcrM has no detectable activity when the G10:C11 basepair is switched to a non-cognate A10-T11[2.3]. Figure 3.1C ($P4^T$) 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 consistent with the prior activity data[2.3]. CcrM affinity for this non-cognate site is only mildly altered from the cognate site binding[2.3], and these experiments are all done at high DNA and CcrM concentrations. The related non-cognate sequence in which the basepair 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 3.1C $P5^T$). 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}, SI Table 2[3.0] 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 which make extensive contacts to the phosphates of the non-target DNA strand (Figure 2.5B, S315, H317, N330, W332, R350). The C-terminal segment is folded as six antiparallel strands with three short α helices (Figure 2.5A, C-terminal segment). This segment resembles a eukaryotic PWWP s which can bind DNA non-specifically[3.16]. Removal of the C-terminal segment in CcrM results in complete loss of ss and dsDNA activity and binding[2.2, 2.3];

interestingly, the same truncation of the β-class DNA MTase M.HinfI, which also methylates ss and dsDNA, results in loss of dsDNA methylation activity with only minor impact on ssDNA methylation[2.3]. 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[2.2]. W332 is part of a hydrophobic core that supports the integrity of the C-terminal segment (Figure 3.3B) and the indole nitrogen contacts the phosphate to the 5' side of guanosine (G7) within the non-target strand (5'pGACTC3', Figure 3.3B). 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 (SI Fig. 5[3.0]). These results provide strong support for the importance of the interaction between tryptophan 332 and the non-target phosphate.

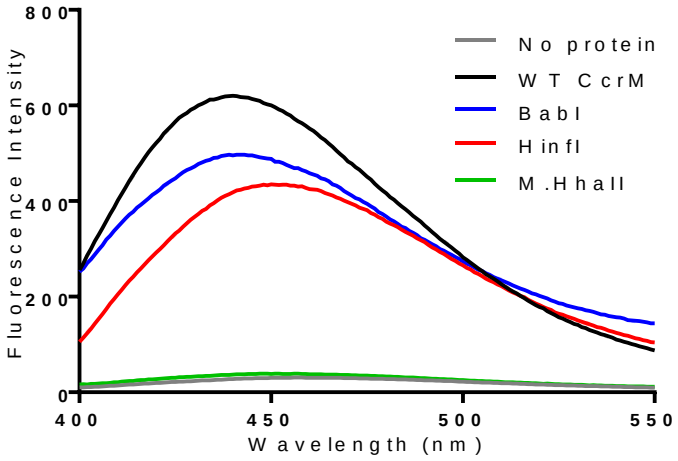


Figure 3.2. Strand separation occurs with CcrM ortholog BabI, the related M.HinfI, and does not occur with a control enzyme M.HhaII. All traces use sequence $P1^T$ (see Figure 2A) with Pyrrolo-dC at the N position in the target strand. WT CcrM, M.HinfI, and HhaII 2.5μM. BabI 6.13μM. No protein represents 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 SI Fig. 6C[3.0].

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[2.2]. 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 3.4B). Serine 315 contacts the phosphate between bases G5 and G6 (Figure 3.4C). 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 3.4D). Arginine 350 makes two hydrogen bonds (2.8 and 3.0Å) to the phosphate backbone between G7 and G8 (Figure 3.4E). 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 3.4F). G305A is 8.6Å removed from the DNA (SI Fig. 4[3.0]); its role may be to maintain the loop configuration (Figure 3.4F 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 3.4 lacks structural information for E280 because the side chain was not resolved in the crystal structure.

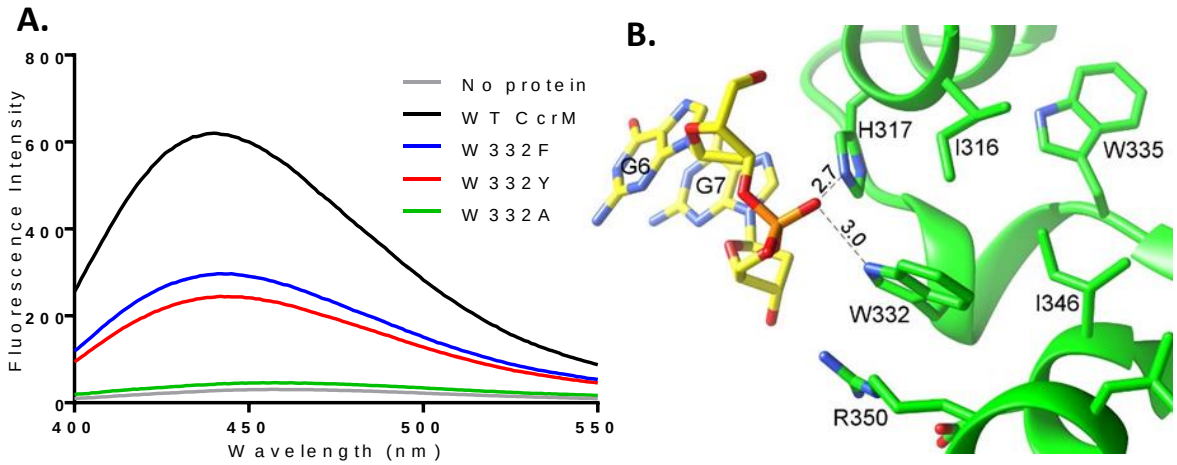


Figure 3.3. A single hydrogen bond between Tryptophan 332 and a non-target strand phosphate is important for strand separation. A. Pyrrolo-dC fluorescence of WT CcrM and W332 mutants. W332A shows no increase in fluorescence. W332F and W332Y show an

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 SI Fig. 6C[3.0]. 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.

Using the WT plasmid containing a C-terminal 6-Histidine tag[2.2, 2.3] 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 (SI Fig. 3[3.0]). The purity of the proteins discussed here is significantly greater than previously published work which may contribute to the different results and conclusions[3.17, 3.18].

Because our prior work showed CcrM has excellent activity with both ss and dsDNA[2.2, 2.3], experiments were performed on ss and dsDNA (60 nucleotides, single stranded, 60 basepairs, double stranded) substrates that contain the cognate recognition site, 5'-GACTC-'3, in the middle of the sequence[2.2, 2.3]. 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} (SI Table 1, SI Fig. 1A[3.0]). 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 (SI Fig. 1A, SI Table 1[3.0]). 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[2.3].

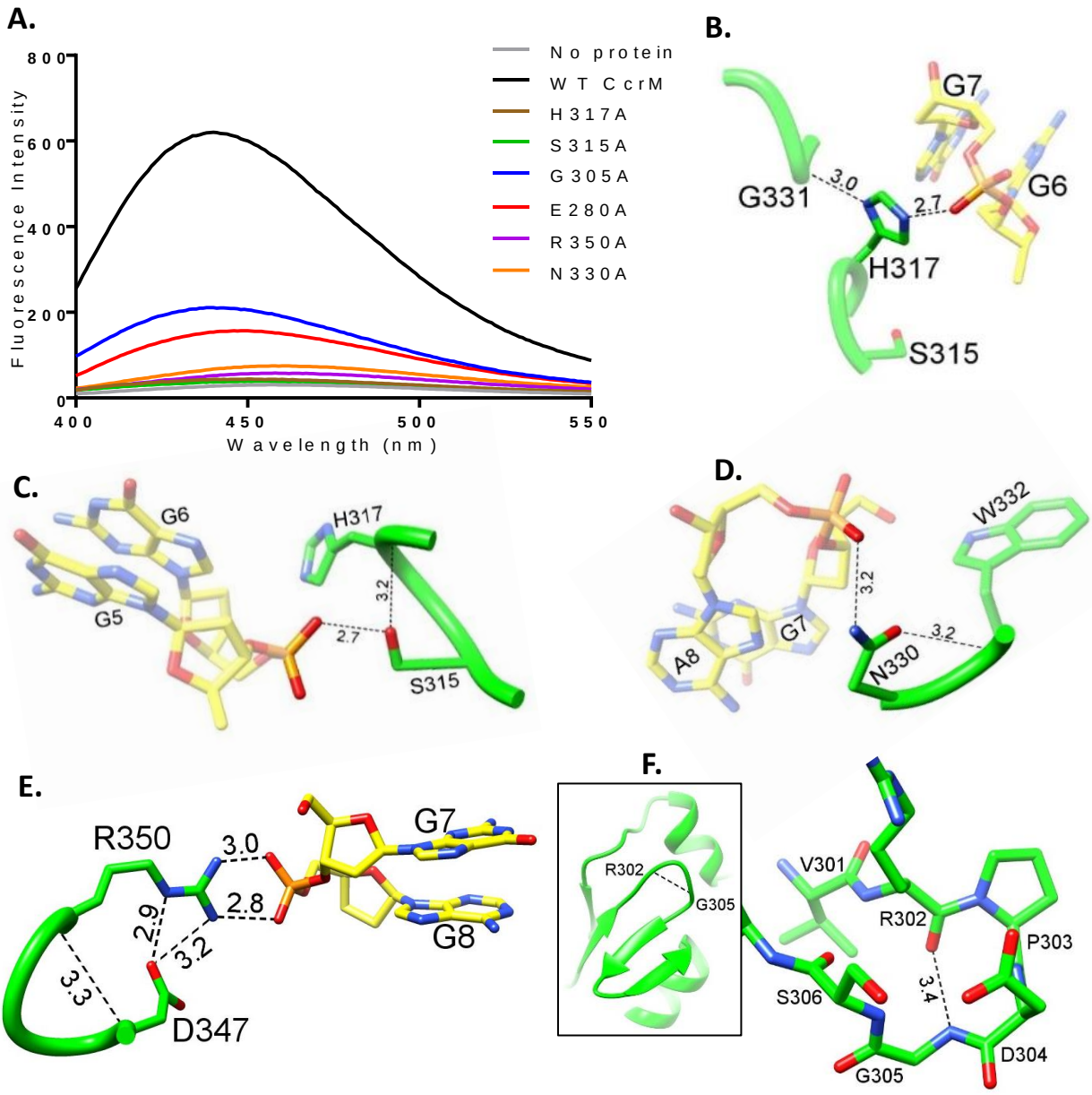


Figure 3.4. Pyrrolo-dC Fluorescence of CcrM mutants and structures of the WT residues G305, H317, S315, N330, and R350. A. Conserved residue mutants dramatically alter the CcrM-dependent changes in Pyrrolo-dC DNA fluorescence. All experimental conditions as in Figure 2 using DNA sequence $P1^{T}$. Maximum fluorescence intensity and error values provided in SI Fig. $6C^{3.0}$. B-F. Structure of amino acid residues showing the hydrogen-bond interactions made by each side chain. Hydrogen bonds are s hown as dashed lines and the distance is labeled in Angstroms. PDB: 6PBD. Structural images were made with UCSF Chimera.

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[2.2] 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 x 10^{-4} min^{-1} on ssDNA while on dsDNA methylation was undetectable and is reported as < 1 x 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 which shows significantly greater discrimination against 5'AACTC'3 (SI Fig. 1A, SI Table $1^{3.0}$). 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 (SI Fig. 1B, SI Table $1^{3.0}$). While the WT had an undetectable $k_{methylation}$ of less than 1 x 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 (SI Fig. 1B, SI Table $1^{3.0}$). The stabilities of the mutant enzyme-DNA complexes with the cognate sequences are compromised (SI Fig. $2^{3.0}$). 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 (~ five fold loss of stability) as seen in SI Table 1[3.0] and in SI Fig. 2A[3.0]. EMSA experiments on cognate dsDNA substrates revealed dramatic disruption in the ability of S315A, H317A, N330A, and R350A to bind dsDNA. H317A has a Kd of 1.42 ± 0.20 μM while R350A has the greatest reduction of binding with a K_D of 4.04 ± 0.41 μM (SI Fig. 2A, SI Table 1[3.0]). 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[3.17, 3.18]. Importantly, the prior work reported an extremely weak DNA affinity for the WT CcrM enzyme (2-10 μM compared to 75 nM shown in SI Table 1[3.0]), which may have resulted from using impure protein or only partially active enzyme[2.2]. The results in SI Table 1[3.0] 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 3.4A) suggesting that the strand

separation step is impacted. Each of these residues form hydrogen bonds to the phosphate backbone of the non-target DNA either within or outside of the recognition site (Figure 2.5). Thus, these interactions are strongly implicated in inducing the strand separation, stabilizing the conformation observed in the cocrystal structure (Figure 2.5B), or both. E280A and G305A show an increase in signal upon binding of CcrM, but the signal change is less than WT CcrM (Figure 3.4A). 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[2.3]. W332Y and W332F showed a significant change in Pyrrolo-dC fluorescence upon CcrM binding that was less than WT CcrM (Figure 3.4A), 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 C-term mutant causes a filamentous phenotype

CcrM plays essential roles in controlling progression through the cell cycle[1.2]. Disruption of this control results in an extensive bacterial filamentous growth phenotype[1.2, 3.19]. We imaged WT CcrM and two of the six mutated CcrM genes expressed from the native promoter on the *C. crescentus* chromosome[3.19]. Figure 3.5 shows that the WT and one of the mutants (E280A) display normal growth, as determined by the average cell length[3.19]. In contrast, S315A, which shows greater alterations in methylation activity (SI Fig. 1A[3.0]), shows extensive filamentous growth, comparable to that observed with the CcrM deletion strain[3.19]. These results suggest that the enzyme's ability to control filamentous growth can be altered 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 (SI Table 1, SI Fig. 1A[3.0]).

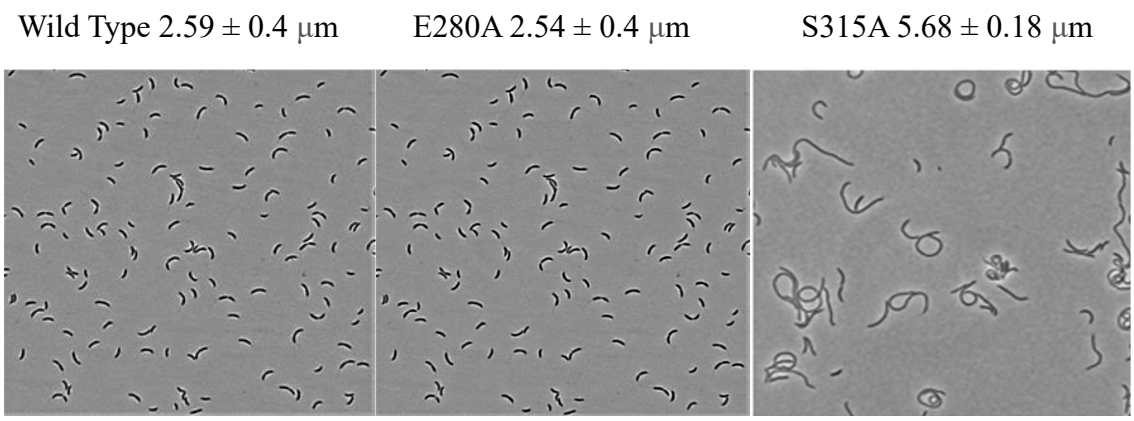


Figure 3.5. Mutations in the C-terminal segment of CcrM cause dramatic cellular changes in growth characteristics of *Caulobacter cresentus*. The WT and mutant CcrM genes were inserted into the correct genomic positions[3.22]. Analysis of 300 bacteria used phase-contrast and fluorescence microscopy. The WT and E280A (which is six to seven 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[3.23].

III. Discussion

The catalogue of characterized protein-nucleic acid recognition mechanisms is rich and diverse, informed by extensive cocrystal structures. Our prior biochemical work[2.2, 2.3, 2.5] 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 basepairs making up the recognition sequence[2.5]. The potential contribution of this mechanism to the extreme sequence discrimination shown by CcrM further emphasizes its importance[2.3]. 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 (eg., 5'NGG3'). The strand separation step remains "enigmatic, but must rely on thermally available energy"[3.20]. 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[3.7].

The cocrystal structure of CcrM bound to double stranded DNA (Figure 2.5) 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[3.10] 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 3.1B, $P1^{T}$) is dramatic. While fluorescence changes can be caused by diverse effects, the DNA control presented here (Figure 3.1, no protein) provides strong evidence that this change results from the separation of the two strands that leads to the structure shown in Figure 2.5A. Thus, compared to when the target strand is tracked, this enhancement is significantly reduced when the non-target strand contains the Pyrrolo-dC (Figure 1B, $P0^{NT}$). Inspection of the CcrM-DNA structure (Figure 2.5A,B) 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 3.1B, $P1^{T}$)[3.10].

Figure 3.1D also shows that positioning Pyrrolo-dC immediately outside the canonical site shows no increase in fluorescence upon CcrM binding (Figure 3.1D; $P2^{NT}$, $P3^{T}$). Pyrrolo-dC (represented as P in Figure 3.1A) 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 basepairs flanking the recognition site in the DNA sequences studied here (Figure 3.1A) are different from those in the cocrystal structure (Figure 2.5B), the lack of detectable changes in fluorescence upon binding by CcrM is consistent with little disruption of the basepairing 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 $P8^{T}$ [2.3]. 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 basepair that is not completely disrupted in the cocrystal structure (G10:C11) was probed with $P6^{NT}$ (Figure 3.1A). As expected, $P6^{NT}$ shows no change in fluorescence upon CcrM binding (Figure 3.1D). However, replacement of Pyrrolo-dC within the G:C basepair (C14:G7) that is disrupted in the cocrystal structure (Figure 2.5B, Figure 3.1A $P7^{T}$) also did not show a significant change in fluorescence upon CcrM binding (Figure 3.1D), which is

likely the result of local quenching by residues proximal to Pyrrolo-dC at that site. Cytosine 14 (Figure 2.5) appears to be sandwiched between residues N124 and L42, which could be responsible for the signal quenching with $P7^T$.

Figure 3.1C 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 basepair that is maintained within the recognition site (Figure 3.1A) and shows a significantly reduced change in fluorescence (Figure 3.1C, $P4^T$); 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 basepair that is disrupted in the recognition site (Figure 3.1A), also shows a much reduced fluorescence change upon CcrM binding (Figure 3.1C). 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[2.2]. Further, M.HinfI, a ß-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 3.2), 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 3.2). 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 which makes extensive contacts to the non-target strand through one of the two monomers (Figure 2.5A, 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 2.5A) are replaced with alanine is certainly consistent with their importance. Residues G305 and E280 (Figure 2.5A) do not show the same decrease in methylation and do not interact with DNA (SI Fig. $4^{3.0}$). 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 3.4A) suggests these residues are important for the strand separation mechanism. The connection between the strand separation step and the dramatic sequence discrimination displayed by $CcrM^{2.3}$ is supported by the fact that although capable of binding non-cognate sequences[2.3], the WT enzyme shows little to no ability to induce strand separation of such sequences (Figure 3.1C, $P4^T$ 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[3.21], 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[1.2]. 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*[1.2]. Figure 3.5 shows that WT and the E280A CcrM show the normal phenotype. E280A shows only minor alterations in its ability to methylate DNA (SI Fig. 1A, SI Table 1[3.0]) and shows some strand separation activity (Figure 3.4A). In contrast, S315A, which shows a 75-fold loss in methylation activity and 2000-fold loss in $k_{methylation}/K_d$ (SI Fig. 1A, SI Table 1[3.0]) and no ability to stabilize the strand separated DNA, shows the same filamentous growth phenotype observed for the CcrM knock out[3.19]. Thus, the ability to separate the target and non-target strands is essential for the biological CcrM-mediated regulation of the Caulobacter growth phenotype.

III. References

3.0. Konttinen, O., Carmody, J., Pathuri, S., Anderson, K., Zhou, X., and Reich, N. (2020) Cell cycle regulated DNA methyltransferase: fluorescent tracking of a DNA strand-separation mechanism and identification of the responsible protein motif. *Nucleic Acids Research*, 48, 20, 11589-11601.

3.1. Cheng,X., and Roberts,R.J. (2001) AdoMet-dependent methylation, DNA methyltransferases and base flipping. *Nucleic Acids Research*, 29, 3784-3795.

3.2. Estabrook R.A., Nguyen T.T., Fera N., Reich N.O. Coupling sequence-specific recognition to DNA modification. *J. Biol. Chem*. 2009; 284:22690-22696.

3.3. Shieh F.K., Youngblood B., Reich N.O. The role of Arg165 towards base flipping, base stabilization and catalysis in M.HhaI. *J. Mol. Biol*. 2006; 362:516-527.

3.4. Matje D., Krivacic C., Dahlquist F., Reich N.O. Distal structural elements coordinate a conserved base flipping network. *Biochemistry*. 2013; 52:1669-1676.

3.5. Matje D., Zhou H., Smith D., Neely R., Dryden D., Jones A., Dahlquist F., Reich N.O. *Biochemistry*. 2013; 52:1677-1685.

3.6. 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.

3.7. Woodcock 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

3.8. 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 Research*, Vol. 35, No. 20 e137.

3.9. 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.

3.10.
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 Letters* 45, 2457–2461.

3.11. 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 type-1 3′ polypurine tract. *Nucleic Acids Research* 32, 4, 1539–1547.

3.12. Zhang,X. and Wadkins,R.M. (2008) DNA hairpins containing the cytidine analog Pyrrolo-dC: Structural, thermodynamic, and spectroscopic studies. *Biophysical J.*, 96, 5, 1884-1891.

3.13. 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 and Bioelectronics*, 98, 210-214.

3.14. Sukackaite, R., Grazulis, S., Tamulaitis, G., Siksnys, V., (2012) The recognition domain of the methyl-specific endonuclease McrBC flips out 5-methylcytosine. *Nucleic Acids Research*, 40, 15, 7552-7562.

3.15. Zang, H., Fang, Q., Pegg, A.E., Guengerich, P., (2005) Kinetic Analysis of Steps in the Repair of Damaged DNA by Human O^6-Alkylguanine-DNA Alkyltransferase. *Journal of Biological Chemistry*, 280, 35, 30873-30881.

3.16. Qin,S. and Min,J. (2014) Structure and function of the nucleosome-binding PWWP domain. *Trends in Biochemical Sciences*, 39, 11, 536-547.

3.17. 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.

3.18. 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.

3.19. Zhou,X. and Shapiro,L. (2018) Cell cycle-controlled clearance of the CcrM DNA methyltransferase by Lon is dependent on DNA-facilitated proteolysis and substrate polar sequestration. *bioRxiv*.

3.20. Sternberg,S., Redding,S., Jinek,M. et al. (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* 507, 62–67.

3.21. 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. *Journal of Molecular Biology,* 327, 2, 383-391.

3.22. Nieman et al. (2001) Complete genome sequence of Caulobacter crescentus, *PNAS,* 98, 7, 4136-4141.

3.23. Collier J., McAdams H.H., Shapiro L. A DNA methylation ratchet governs progression through a bacterial cell cycle. *PNAS.* 2007; 104:17111-17116.

3.24. Woodcock,C.B., Horton,J.R., Zang,X., Blumenthal,R.M. and Cheng X. (2020) Beta class amino methyltransferase from bacteria to humans: evolution and structural consequences. *Nucleic Acids Research,* 48, 18, 10034-10044.

Chapter IV: High fidelity DNA strand-separation is the major specificity determinant in DNA methyltransferase CcrM's catalytic mechanism

IV. Abstract

Chapter III in combination with the CcrM/dsDNA cocrystal structure shows that CcrM relies on a strand-separation DNA recognition mechanism and that this process can be monitored via fluorescence. However, the underlying mechanisms and quantitative contribution of strand-separation to fidelity remain obscure. To further explore this recognition mechanism, we collected stopped-flow fluorescent kinetic data to monitor changes in DNA and protein and built an enzyme kinetic model. Again, we incorporated Pyrrolo-dC into cognate and noncognate DNA to monitor the kinetics of strand-separation and used tryptophan fluorescence to follow protein conformational changes. Both signals are biphasic and global fitting showed that the faster phase of DNA strand-separation was coincident with the protein conformational transition. Non-cognate sequences did not display strand-separation and methylation was reduced >300-fold, providing evidence that strand-separation is a major determinant of selectivity. Analysis of an R350A mutant showed that the enzyme conformational step can occur without strand-separation, so the two events are uncoupled. A stabilizing role for the methyl-donor (SAM) is proposed; the cofactor interacts with a critical loop which is inserted between the DNA strands, thereby stabilizing the strand-separated conformation.

IV. Introduction

Of the known enzyme/DNA cocrystal structures that display strand-separated complexes, CRISPR-Casp/gRNA structures are the most characterized[4.2, 4.3]. Despite numerous cocrystal structures for enzymes that perform DNA strand separation, the events leading up to the strand-separated structures remain enigmatic[4.2, 4.3]. The transient intermediates that govern specificity are challenging to stabilize and observe via structural methods. Therefore, utilizing fluorescent probes in both DNA and enzyme, and global fitting can support the existence of intermediates as well as assign kinetic parameters to the transitions that lead up to catalysis.

In Chapters II and III, we show that the DNA strand separation complex for CcrM/dsDNA can be monitored fluorescently, and that the C-terminal domain is essential for strand-separation. Substitution of conserved residues in the C-terminus of CcrM negatively impacts both DNA recognition and strand separation[3.0]. We also showed that CcrM is extraordinarily discriminating, compared to other DNA MTases, having at least 10^6-fold loss in specificity for non-cognate recognition sequences (e.g., 5'AANTC3' compared to 5'GANTC'3)[2.3] and strand-separation is not observed with noncognate sequences[3.0].

Expanding on our previous equilibrium measurements, here we use stopped-flow fluorescence to monitor the kinetics of a protein conformational transition using intrinsic tryptophan fluorescence and the DNA strand separation event using PydC fluorescence. We measured these transitions for WT CcrM with cognate and noncognate DNA and for CcrM variant (R350A) with cognate DNA. Our results provide a basis for investigating the ordered events that lead up to the strand-separated CcrM-DNA complex and shed light on which events contribute to the enzyme's DNA sequence-specificity. For example, conformational transitions

such as base-flipping can be rate-limiting and contribute significantly to sequence discrimination[2.1], so the first questions to address is whether DNA strand separation by CcrM is rate-limiting and whether it is readily reversible.

In spite of numerous structural studies demonstrating changes in enzyme structure after substrate binding, the fidelity contribution of substrate-induced conformational changes has been controversial. For example, Fersht argued that a two-step binding contributes the same free energy change as a one-step binding mechanism and therefore the two-step induced-fit pathway cannot add anything to selectivity beyond a simple binding step[4.4]. Moreover, Warshel argued that a conformational change cannot influence specificity unless it is rate-limiting[4.5]. Despite these theoretical arguments, comprehensive kinetic analysis of the conformational changes preceding nucleotide incorporation by moderate to high fidelity DNA polymerases finally showed that the nucleotide-induced change in structure is the primary determinant of fidelity even though it is not rate-limiting[4.6, 4.7, 4.8, 4.9]. Accordingly, in this study we provide direct measurements of the rate constants governing strand-separation and subsequent methylation in order to define the role of DNA strand separation in fidelity of CcrM. We show that enzyme-DNA interactions leading to strand separation are the primary determinant of fidelity even though the rates of strand separation are much faster than DNA methylation.

IV. Materials and Methods

Equilibrium Tryptophan Fluorescence

Equilibrium tryptophan (Trp) fluorescence measurements were performed at room temperature on a Horiba Fluoromax-4 spectrofluorometer using 296 nm excitation wavelength

and 1 nm slit size. Emission data was collected at 311 nm to 430 nm using 5 nm slit size. Samples were kept on ice and allowed to equilibrate to room temperature for 5-10 minutes before measurements were taken. All readings were performed in 100 mM HEPES, 1 mM EDTA, 20 mM NaCl, 2 mM DTT, pH 8. Measurements were then taken in quintuplet for the buffer, additions of protein (500 nM, monomer concentration), addition of sinefungin (SFG) (60 μM), and addition of cognate or non-cognate DNA (5 μM). Each measurement and background were averaged, and statistical outliers were eliminated. Statistical outliers were defined based on the maximum signal for each scan that was outside of the lower or upper bounds. The lower bound = Q1-1.5*IQR and the upper bound = Q3+1.5*IQR, where IQR = inner quartile range[4.10]. Five scans were taken, and discarded reads were not replaced. The cognate sequence is 5'-TCACTGTACTCTGACTCGCCTGACATGAC-3' and the non-cognate sequence is 5'-TCACTGTACTCTAACTCGCCTGACAGAC-3'. Underlined bases indicate the recognition sequence of CcrM.

Kinetic measurements by Trp Fluorescence

Kinetic constants were measured on an Applied Photophysics SX.18MV stopped-flow spectrometer (Leatherhead, UK) temperature controlled to 22 ± 1 ˚C, using 296 nm excitation and a 320 nm emission cutoff filter. Final concentrations after 1:1 mixing were 500 nM enzyme, 60.0 μM SAM (S-adenosyl-L-methionine), and DNA varied 2.5, 5.0, 7.5, and 10μM, and the buffer composition was the same as in the equilibrium fluorescence experiments. Kinetic traces were collected in triplicate and averaged. The first 3ms were truncated out of each trace to account for the deadtime of the stopped-flow instrument (SI Fig. 1[4.1]). The

deadtime determination of the stopped-flow instrument was based on Peterman's protocol (1979)[4.26]. Data were fit to double-exponential functions in KinTek Global Kinetic Explorer. The function used for fitting was $y = a_0 + a_1(1 - e^{-b1t}) + a_2(1 - e^{-b2t})$, where y = fluorescence intensity (arbitrary units), t = time (seconds), a_1 = the amplitude of the first phase, b_1 = the rate of the first phase, a_2 = the amplitude of the second phase, b_2 = the rate of the second phase, and a_0 = the initial fluorescence amplitude (arbitrary units).

Kinetic measurements by PydC Fluorescence

PydC fluorescence kinetics were measured in the same buffer as in the equilibrium Trp fluorescence experiments at 22 ± 1 ˚C using an Applied Photophysics SX.18MV stopped-flow spectrometer (Leatherhead, UK). The excitation wavelength was 350 nm, and emission was collected with a 385-nm cut-off filter. The concentrations after 1:1 mixing were 2.5, 5.0, 7.5, and 10µM CcrM, 60.0 µM SAM and 1 µM dsDNA. Data collection and analysis was identical to the procedure used for the Trp fluorescence experiments.

Kinetic measurements by PydC Fluorescence over 2000 seconds

Kinetic measurements over 2000 seconds were performed at room temperature on a Horiba Fluoromax-4 spectrofluorometer in kinetics mode. The monochromator excitation wavelength was 350nm with a 2nm bandpass and the emission monochromator was parked at 450nm with a 2nm bandpass. The reaction was initiated with protein and hand-mixing was carefully completed in a quartz cuvette before data acquisition began. The mixing-time (10 seconds) was considered the deadtime for this experiment. The reaction consisted of P1-DNA [1.0 µM], SAM [60.0 µM], and enzyme [5.0 µM].

Fluorescence Anisotropy to estimate K_D

Fluorescein (56-FAM) DNA oligos were obtained from Integrated DNA Technologies. DNA was annealed by melting at 95 degrees C for 5 minutes followed by passive cooling to room temperature for 3 hours. Native PAGE gels determined >98% purity of annealed double-stranded DNA. Binding reactions consisted of 10nM FAM-DNA, 0-1000nM protein, 30µM SAH, in 50µl volumes in the following buffer: 100mM HEPES, 1mM EDTA, 20mM NaCl, pH 8. Reactions were loaded into a Corning 96-well flat-bottom black plate and incubated at room temperature for 20 minutes. Fluorescence anisotropy was monitored using a Tecan Spark microplate reader at room temperature. Excitation and emission polarizers were 485nm and 535nm, respectively. Anisotropy (a.u.) was plotted vs. protein concentration. Anisotropy data was background subtracted (background = FAM-DNA in buffer alone) and fit to a specific binding with Hill slope equation ($Anisotropy = Anisotropy_{max}*[CcrM]^h/(Kd^h + [CcrM]^h)$) using Graphpad Prism 7.00.

Radiochemical methylation assay

Single turnover methylation reactions consisted of 250nM enzyme, 100nM DNA, and 15µM AdoMet, using hemimethylated double-stranded substrates. Tritiated AdoMet in all instances had a specific activity of (3H-CH3 1 mCi [82.7 mCi/mmol]). Reactions were initiated by adding enzyme. 5µl of reaction time points were spotted in triplicate onto GE Amersham Hybond-XL nylon membranes followed by three 5min washes in wash buffer (50 mM KH_2PO_4). The washes were followed by a 5 min dehydration step with 80% Ethanol, another

wash for 5 min in 100% Ethanol, and a final drying step for 10 min under a heat lamp. Samples were then placed into scintillation vials containing 3 ml of BioSafe II Scintillation cocktail. Radiochemical data were generated with a Hidex 300SL scintillation counter. Data for the single turnover reaction were background-subtracted and fit to a one-phase decay model in GraphPad Prism 10.0.2.

Global fitting

Global data fitting was performed in KinTek Global Kinetic Explorer Version v11.0.1. The reaction scheme used as the unifying model to describe the experimental data was E + S = FS = GS^{I} = GS^{II} = GS_p^{II} = E + S_p. Data from four experiments were input and conditions were consistent with how the data was collected. Time-dependent data were corrected for the measured dead time of the instrument (2.6 ms) by excluding the first two data points that were collected within the first 3ms. Each experiment had a unique observable signal which relates the experimental data to the model. For example, Experiment 1 is PydC kinetics over 2 seconds, where PydC fluorescence depends on the following observable signal: a1*(S + FS + e*GS^{I} + f*(GS^{II} + GS_p^{II}) + h*S_p). This observable signal describes that GS^{I} partially contributes to the change in PydC signal, while GS^{II} and GS_p^{II} also contribute, and that GS^{II} and GS_p^{II} contribute equally to the change in fluorescence. Experiment 4 is the Trp kinetic trace which depends on the following observable signal: a4*(E + k*FS + m*GS^{I} + n*(GS^{II} + GS_p^{II})). Exp 3 is PydC data over a long time-course to observe the annealing of the methylated DNA strands after methylation. Experiment 3 had an observable signal identical to experiment 1 because they are both monitoring PydC signal, but over different time-bases. Experiment 3

had an observable signal of a3*(S + FS + e*GS^{I} + f*(GS^{II} + GS_p^{II}) + h*S_p). Experiment 2 is the radiochemical methylation assay, and we therefore have an observable signal of GS_p^{II} + S_p + bkg2.

Data from all four experiments were fit globally based on numerical integration of the rate constants (computer simulation). Initial values were estimated based on the fitting of each data by equation. Fluorescence scaling factors were applied to the data in experiments 1 and 4 to normalize variability in the starting amplitudes for each trace within a concentration series. Scaling factors were close to unity and therefore did not influence the concentration dependence of these traces. Some values were locked during the simulation, while others were allowed to float, as described. Locked values were chosen based on the parameter's lower limit beyond which there is no affect on the fitted curves.

Three individual models were built (WT/cognate DNA, R350A/cognate DNA, WT/Noncognate DNA) with identical model architecture for direct comparison. An alternative presentation for the R350A and Noncognate models would have been to truncate the models only showing phases that were described by the data. However, we opted for the 5-step model for a direct comparison to the WT/cognate model. Confidence contour analysis represent the 1D Fitspace calculated for each rate constant. The 95% Chi^2 limits were calculated in KinTek Explorer.

IV. Results

Equilibrium DNA binding

Fluorescence-based anisotropy using a 29mer hemimethylated DNA with one centrally located 5'GANTC3' site and a 5'-FAM tag was used to estimate the apparent K_d (K_d^{app}). The K_d^{app} for WT CcrM with cognate DNA (C1-FAM) was determined to be 150 ± 5 nM (SI Fig. 2A[4.1]. The K_d^{app} for WT CcrM with noncognate DNA (NC-FAM) was 135 ± 8 nM (SI Fig. 2B[4.1]. The binding affinity for NC-FAM DNA is comparable to C1-FAM DNA suggesting that substrate binding is nonspecific and does not contribute to CcrM's highly discriminating mechanism. DNA sequences are listed in Table 4.1.

Equilibrium Trp fluorescence

Equilibrium tryptophan fluorescence was measured upon the binding of cognate (C1) and noncognate (NC) DNA (Table 4.1). There are 18 Trp residues per dimer of CcrM (Fig. 4.1A). As a result of CcrM binding to DNA, the protein fluorescence is significantly decreased (Fig. 4.2). Consistent with our prior finding that CcrM methylates ssDNA as efficiently as dsDNA, we observe similar reductions in protein fluorescence with both ssDNA and dsDNA (Fig. 4.2C[2.2]).

DNA name and sequence	
	C1 5′-TCACTGTACTCTGACTCGCCTGACATGAC-3′ 3′-AGTGACATGAGACTGMGCGGACTGTACTG-5′
	P1 5′-TCACTGTACTCTGAPTCGCCTGACATGAC-3′ 3′-AGTGACATGAGACTGMGCGGACTGTACTG-5′
	NC 5′-TCACTGTACTCTGACTGGCCTGACATGAC-3′ 3′-AGTGACATGAGACTGMCCGGACTGTACTG-5′
	PNC 5′-TCACTGTACTCTGAPTGGCCTGACATGAC-3′ 3′-AGTGACATGAGACTGMCCGGACTGTACTG-5′
	C1-FAM 5′-(56-FAM)TCACTGTACTCTGACTCGCCTGACATGAC-3′ 3′-AGTGACATGAGACTGMGCGGACTGTACTG-5′
	NC-FAM 5′-(56-FAM)TCACTGTACTCTGACTGGCCTGACATGAC-3′ 3′-AGTGACATGAGACTGMCCGGACTGTACTG-5′

Table 4.1. Names and sequences of 29mer DNA substrates. C1 is cognate DNA, P1 is cognate DNA with PydC inserted at the N-position of the recognition site, NC is noncognate DNA, PNC is noncognate DNA with PydC at the N-position of recognition site. C1-FAM is cognate DNA with a FAM-tag on the 5' end of the target strand. NC-FAM is noncognate DNA with a FAM-tag on the 5' end of the target strand. Underlined based identify the recognition site. P = Pyrrolo-dC, M = N6-methyl adenine, red bases = noncognate substitutions.

Interestingly, we observe similar changes in protein fluorescence with noncognate DNA, suggesting nonspecific protein conformational changes occur even though noncognate DNA shows no evidence of strand separation (Fig. 4.2A-B[3.0]). Therefore, the protein conformational change monitored by Trp fluorescence is not significantly influenced by the event of strand separation. Rather, as we show subsequently, Trp fluorescence monitors a protein conformational change that precedes DNA strand separation. We know from the crystal structure and previous PydC data[3.0] that cognate dsDNA undergoes strand separation, while non-cognate dsDNA is not strand separated. Therefore, the equilibrium Trp fluorescence occurs regardless of DNA strand separation. Binding of ssDNA induced smaller

conformational changes than dsDNA, independent of whether the strands were cognate or non-cognate (Fig. 4.2C-D).

W332 and W335 are the only two Trp residues in the C-terminal segment (Fig. 4.1A). The DNA-dependent change in protein fluorescence observed with the W332Y/W335Y double variant is significantly attenuated compared to the WT (Fig. 4.2F), suggesting that these residues and the C-terminal segment contribute to the fluorescence changes observed upon DNA binding. WT and W332Y binding to ssDNA give the same result (Fig. 4.2E, red trace). However, only when bound to dsDNA is the fluorescence of W332Y different than wild type (Fig. 4.2E, green traces). This suggests that protein conformational changes associated with dsDNA are dependent on interactions from the C-terminal domain, while changes associated with ssDNA are not dependent on such interactions.

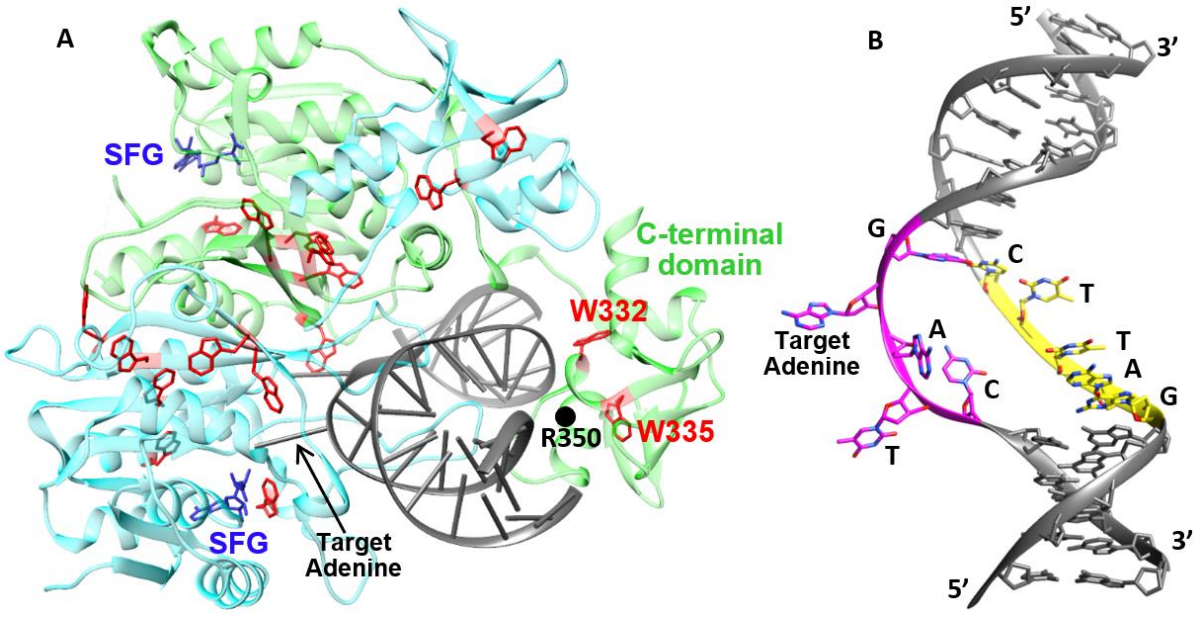


Figure 4.1: The CcrM cocrystal structure depicts GS^{II} in our kinetic model. A. The structure shows the CcrM dimer with 18 Trp residues (red), strand-separated DNA (grey), and two Sinefungin molecules (SFG; blue). R350, W332 and W335 are located in the C-terminal 83-residue segment of monomer B (green), and the N-terminus of monomer A (cyan) is positioned to catalyze methyl transfer on the target adenine. **B.** DNA strand separation at the

recognition site. The target strand recognition site is colored magenta and the nontarget strand recognition site is colored yellow. Four of the five basepairs of the recognition site are disrupted. The 5' G of the target strand maintains base pairing to the 3' C of the nontarget strand. Base-stacking interactions between non-target strand bases are maintained, while base-stacking of target strand bases is lost. Molecular graphics were made with UCSF Chimera.

Kinetic model for WT and Cognate DNA

As detailed below, our data support a kinetic model to describe the conformational changes in the protein and DNA that lead up to methylation by CcrM (Fig. 4.3). The model was derived to account for data from four experiments: Trp fluorescence, PydC fluorescence over 2s, PydC fluorescence over 2000s, and a radiochemical single-turnover methylation assay. Global data fitting of experiments performed with WT CcrM and cognate DNA defined the kinetically significant steps leading to DNA methylation as shown in Figure 4.4. Enzyme conformational states detected by changes in Trp fluorescence are designated as *E*, *F*, and *G*. DNA strand-separated states detected by PydC fluorescence are designated as *S* (duplex), S^{I} (partially strand-separated intermediate), and S^{II} (fully strand-separated). *Sp* designates the methylated DNA product. In step 2, the protein isomerization and the first DNA strand-separation step appear to be coincident.

Kinetics for WT and Cognate DNA

Trp fluorescence is biphasic, and the fast phase (k_1) is concentration-dependent and represents the association of CcrM and DNA. The second slower phase of Trp fluorescence (k_2) monitors an isomerization event in the protein after DNA binding (Fig. 4.4A). The first fast phase (k_1) is evident at the lowest DNA concentration, but lost in the dead time of the stopped flow at the

higher DNA concentrations. The fluorescence amplitude of the first Trp phase decreases with increasing enzyme, suggesting that at higher enzyme concentrations, more FS complex has formed during the deadtime of the experiment (SI Fig. 3A-B[4.1]). The deadtime of the stopped-flow is approximately 2.6 ms, therefore two data points within the first 3ms were excluded from each trace (SI Fig. 1[4.1]).

Global fitting reveals that when the Trp data is extrapolated back to zero time, there is a concentration-dependent loss of signal amplitude. Global data fitting accounting for the observed signal and loss of amplitude in the fast phase provides a minimal estimate of the DNA binding rate constant of approximately 100 $\mu M^{-1}s^{-1}$. During global data fitting k_1 is locked because it is not well-defined. Accordingly, fitting to derive k_{-1} affords an estimate of the K_d for formation of the initial DNA-bound state. Importantly, the data describe two changes in tryptophan fluorescence: one occurring simultaneously with DNA binding and a second change in fluorescence occurring after DNA binding.

Incorporation of PydC in place of N in the target strand (5'GANTC3') provides a strong signal for strand separation as supported by extensive prior controls[3.0]. As is often the case in analog studies, this alters the enzyme activity ($k_{methylation}$) (SI Fig. 9[4.1]). The apparent $k_{methylation}$ with C1-DNA is 1.4 min^{-1} which is decreased with P1-DNA to 0.05min^{-1}. Furthermore, the prior controls and those included here provide strong support for the relevance of using PydC in our studies.

PydC fluorescence monitoring DNA strand-separation is also biphasic with a fast change in fluorescence followed by a slower signal with a lower amplitude (Fig. 4.4B). The first phase of PydC fluorescence and the second phase of Trp fluorescence appear to be

coincident, therefore we have defined this coincident step as k_2. In the WT/CognateDNA model, GS^I represents an intermediate in which both the protein has undergone an isomerization event and the DNA has been partially strand-separated. The second phase of PydC kinetics is the second phase of DNA strand separation which we have defined as k_3. The apparent rates of these two phases are independent of concentration over the range tested, indicating that DNA strand-separation is a first-order process (SI Fig. 3C-D[4.1]).

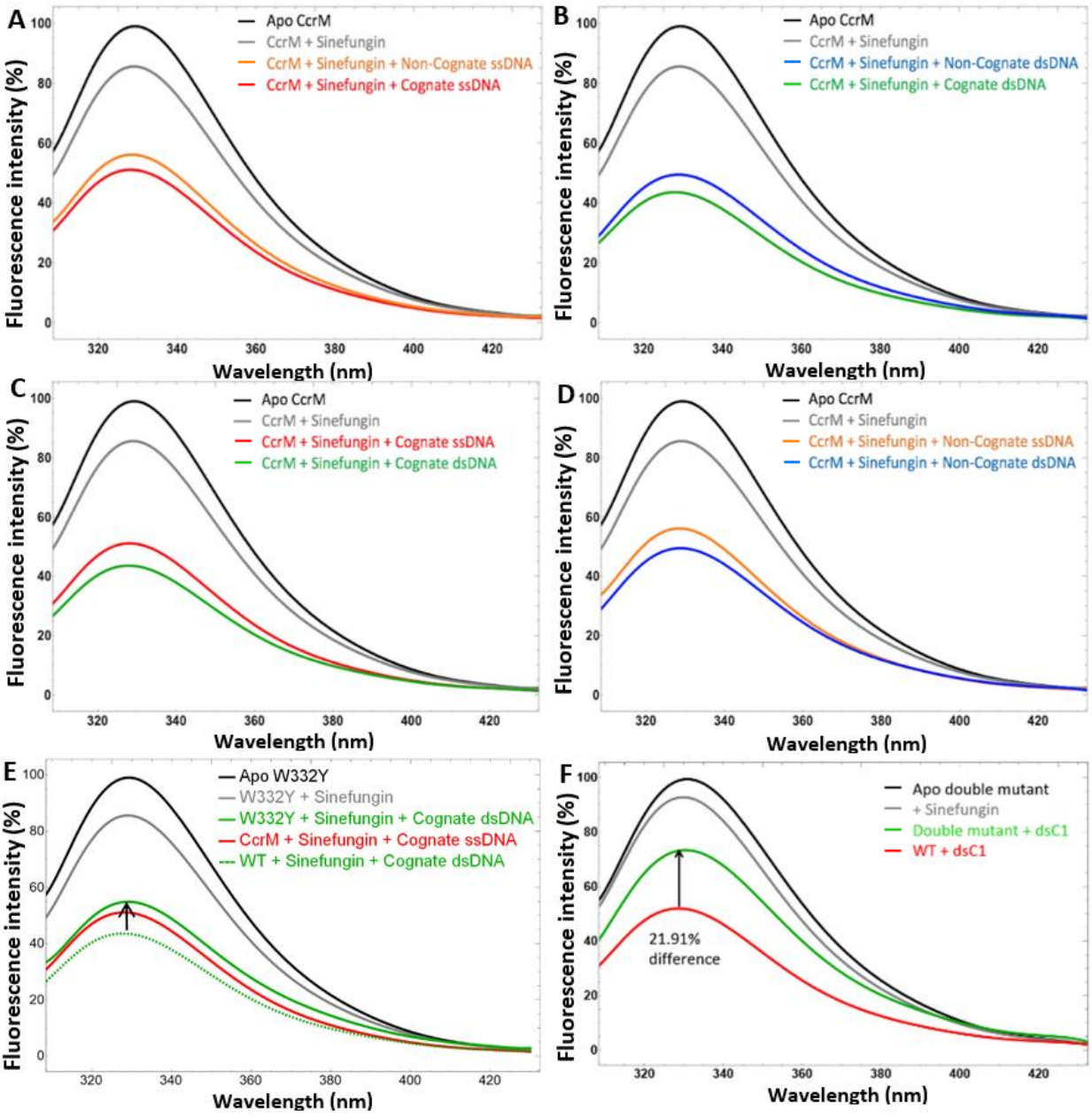


Figure 4.2. Equilibrium Trp fluorescence shows minor differences in the presence of ssDNA vs dsDNA and cognate vs noncognate DNA. Equilibrium fluorescence was measured for the binding of cognate or non-cognate DNA. **A.** Difference in cognate vs non-cognate ssDNA. **B.** Difference in cognate vs non-cognate dsDNA. **C.** Difference in cognate ssDNA or dsDNA. **D.** Difference in non-cognate ssDNA or dsDNA. Binding of ssDNA induced smaller fluorescence changes than dsDNA independent of whether the strands were cognate or non-cognate. **E.** WT and W332Y binding to ssDNA give the same result (red). The arrow shows that the fluorescence of W332Y is only different than that of WT when bound to dsDNA. **F.**

The double variant W332Y/W335Y displays lesser relative fluorescence change than WT for binding to dsDNA, suggesting that the WT effect is due to contributions from both the N-terminus and C-terminus.

On a longer timescale (Fig. 4.4D) the PydC signal decreases as the product DNA is released from the enzyme and re-anneals to form duplex, although we cannot define the order of these two steps. The data can be fit with fast product release by assigning $k_5 = 1\ s^{-1}$ so it is greater than k_4. Accordingly, the decrease in the PydC signal on the long timescale defines the rate of DNA methylation (Fig. 4.4E) consistent with the direct measurement of DNA methylation in a single-turnover using radiochemical methods (Fig. 4.4C).

Globally fitting these data show that the DNA binding is fast and reaches equilibrium as measured by the fast tryptophan fluorescence change. The second change in tryptophan fluorescence indicating a change in enzyme structure is correlated with the first DNA strand-separation step to form an unknown intermediate state, preceding to a second change in PydC signal that is then followed by DNA methylation (Fig. 4.4E).

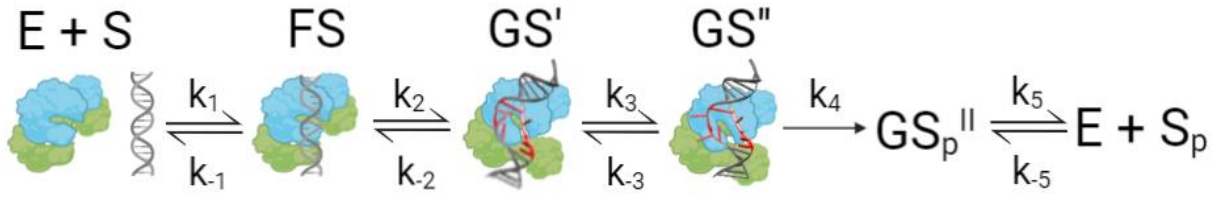


Figure 4.3. Conformational changes in the protein and DNA that lead up to methylation by WT CcrM. Enzyme conformational states detected by changes in Trp fluorescence are designated as *E*, *F*, and *G*. DNA strand-separated states detected by PydC fluorescence are designated as *S* (duplex), S^{I} (partially strand-separated intermediate), and S^{II} (fully strand-separated). *Sp* designates the methylated DNA product. In step 2, the protein isomerization and the first DNA strand-separation step appear to be coincident.

Kinetic model for WT and Noncognate DNA

Figure 4.5 shows the global fitting of data describing the reaction of WT CcrM and noncognate DNA (5'-GANTG-3'), which has one basepair changed from the cognate sequence. For direct comparison with cognate DNA, we fit the data globally using the same model as was derived for cognate DNA, although the steps and rate constants were not well resolved.

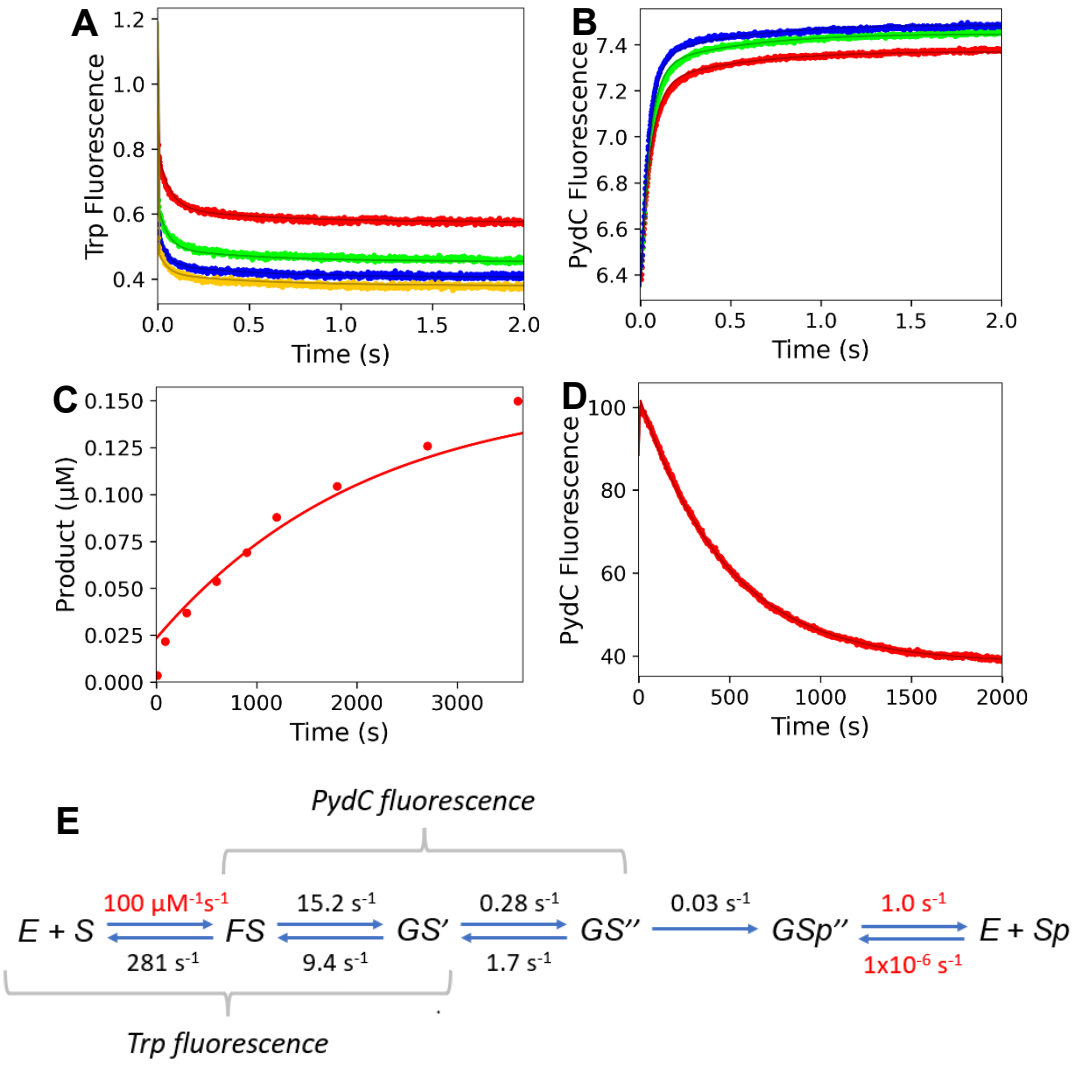


Figure 4.4. Global fitting of the data defining the kinetic steps for WT CcrM and Cognate DNA. A. Trp kinetics consisted of WT CcrM [500 nM], C1-DNA [2.5 (red), 5.0 (green), 7.5 (blue), and 10.0 (yellow) μM], and μM]. **B.** DNA strand-separation kinetics monitored by PydC fluorescence consisted of P1-DNA [1.0 μM], WT CcrM [5.0 (red), 7.5 (green) and 10.0 (blue) μM], and SAM [60 μM]. **C.** WT and cognate DNA single turnover methylation monitored via radiochemical methods. P1-DNA [100 nM], WT [250 nM], and SAM [15 μM]. **D.** PydC kinetics over 2000 seconds monitored product DNA annealing and/or release, and consisted of WT [5 μM], DNA [1 μM], and SAM [60 μM]. All experiments in each model were globally fit in KinTek Explorer. Solid lines represent the simulated traces from the global fit. **E.** Rate constants derived from globally fitting the data for WT and Cognate DNA. Red values indicate rates that were locked during the simulation.

Trp kinetics with WT and noncognate DNA are similar to that observed with cognate DNA (Fig. 4.5A) but with a decreased amplitude in both phases (SI Fig. 4A-B[4.1]). PydC fluorescence with noncognate DNA shows no change indicating that noncognate DNA does not undergo DNA strand separation (Fig. 4.5B, SI Fig. 4C-D[4.1]). The most obvious results shown in Figure 5 are that DNA strand-separation does not occur (Fig. 4.5B), the rate of DNA methylation is reduced at least 300-fold (Fig. 4.5C), and subsequent product annealing and release is not observed (Fig. 4.5D). Thus, although the second tryptophan fluorescence change appears to be coincident with the first DNA strand-separation step for cognate DNA, these data demonstrate that the two events are not coupled.

The noncognate DNA has cytosine14 replaced with guanine on the target strand and G7 is replaced with cytosine on the nontarget strand (Fig. 4.5E). Cognate Cytosine-14 makes 3 hydrogen bonds from its base to the peptide backbone of K126, F125, and N124 and to the pentose ring of A12. Moreover, C14 appears to be stacked between N124 and L42 (Fig. 4.5E).

The WT/Noncognate DNA model reveals that the substrate discriminating step is DNA strand separation. The equilibrium constant K_1 is slightly greater with noncognate DNA

implying that CcrM binds noncognate DNA tighter than cognate DNA. This is consistent with the apparent Kd estimated from anisotropy data (SI Fig. 2[4.1]) and prior Kd estimates via EMSA methods[2.3]). The equilibrium constant K_2 in the noncognate model favors the forward progression in the pathway slightly more than K_2 in the cognate model, but for noncognate DNA, this step is defined solely by the Trp fluorescence signal. Although global data fitting derives an estimate for k_3, this is based on a small signal that is comparable to the noise in the data (Fig. 4.5B) and is therefore not deemed to be reliable.

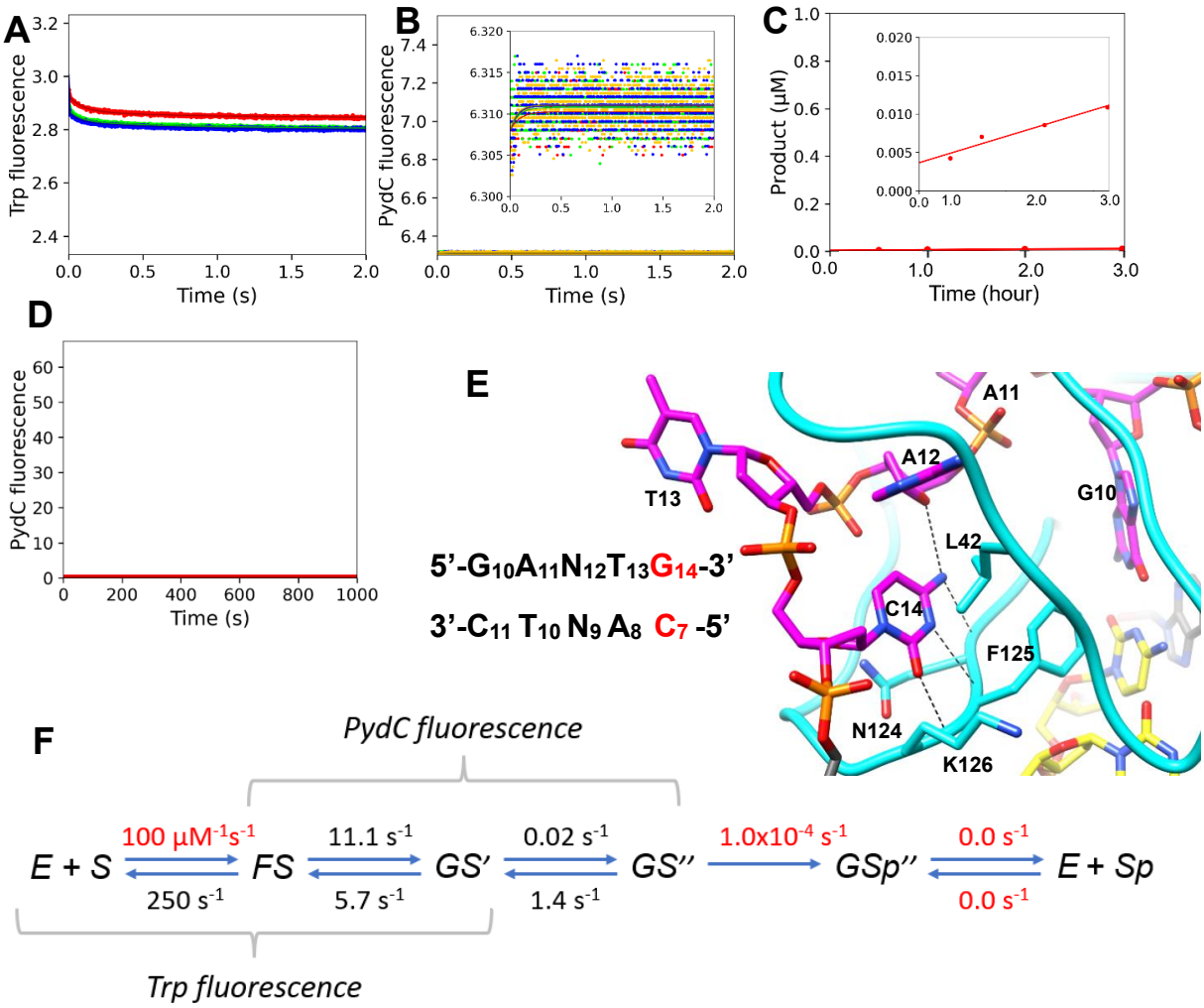


Figure 4.5. WT and Noncognate DNA Trp kinetics are only slightly perturbed while PydC kinetics are not observed, suggesting that DNA strand-separation is tightly coupled to substrate discrimination. A. Trp kinetics consisted of WT [500 nM], NC-DNA [2.5 (red), 7.5 (green), and 10.0 (blue) μM], and SAM [60 μM]. **B.** DNA strand-separation monitored by PydC fluorescence consisted of PNC-DNA [1.0 μM], WT [2.5 (red), 5.0 (green), 7.5 (blue), and 10.0 (yellow) μM], and SAM [60 μM]. **B inset.** Increased resolution in the y-axis of B. **C.** WT and noncognate DNA single turnover methylation assay. An important distinction in the conditions for the noncognate methylation assay; enzyme and DNA concentrations were increased to bring the radiochemical signal above baseline levels and the reaction went longer. NC-DNA [1.0 μM], WT [2.5 μM], SAM [15 μM]. C inset. Increased resolution in the y-axis of C. **D.** PydC kinetics over 1000 seconds consisted of WT [5 μM], NC-DNA [1 μM], and SAM [60 μM]. **E.** The noncognate DNA has one base pair substitution in the recognition site. Cytosine14 is replaced with guanine on the target strand and G7 is replaced with cytosine on the nontarget strand. The red bases represent the noncognate substitution. Cytosine14 makes 3 hydrogen bonds from its base to the peptide backbone of K126, F125, and N124 and to the

pentose ring of A12. C14 appears to be stacked between N124 and L42. **F.** Rate constants derived from globally fitting the data for WT and Noncognate DNA. Red values indicate rates that were locked during the simulation.

The lack of an observable PydC signal indicates that GS^{I} does not progress to GS^{II} with noncognate DNA, which is consistent with the low level of methylation activity (Fig. 4.5F), which is at least 300-fold slower than with cognate DNA.

Kinetic model for R350A and Cognate DNA

R350A displays Trp kinetics similar to WT, while not displaying a change in PydC signal (Fig. 4.6A-B, SI Fig. 5B-$D^{4.1}$), providing further evidence that the protein isomerization and DNA strand separation are uncoupled. The protein fluorescence change still reflects fast DNA binding and a possible change in enzyme conformation occurring in two steps. R350A displays biphasic Trp kinetics with apparent rates slightly less than WT, and the fluorescence amplitudes in both phases are slightly reduced (Fig. 4.6A, SI Fig. 5A-$B^{4.1}$). R350A displays severely attenuated PydC signal (Fig. 4.6B).

Surprisingly, the methylation by R350A still proceeds as a rate comparable to WT, but the amplitude is reduced to approximately 15% of the DNA forming product (Fig. 4.6C) in this single turnover experiment. This mutant reveals unexpected behavior in that less than the expected amount of product is generated (only 15%) and DNA strand-separation is negligible. The mutant may allow a fraction of the adenine in the target strand to reach the active site without DNA strand-separation. The parameter k_{-4} was proposed in the R350A/CognateDNA model to account for the incomplete methylation reaction by allowing the chemistry step to

come to equilibrium with slow product release. Thus, the single-turnover methylation assay does not go to completion (Fig. 4.6C), and subsequent product annealing and release is not observed (Fig. 4.6D). Alternatively, R350A could have an off-pathway intermediate in which we did not incorporate into the model. However, we opted for the 5-step model to keep the models consistent. Further studies are needed to define the origin of this phenomenon.

R350 from the C-terminal domain makes two hydrogen bonds to the phosphate backbone of the non-target strand (Fig. 4.6E). Substitution of R350 with alanine disables these hydrogen bonds, resulting in destabilization of both FS and GS^{I} preventing formation of GS^{II} (Fig. 4.6F). The functional consequences for R350A primarily are due to the increase in the rate constant k_{-1} and the decrease in the equilibrium constant K2. The equilibrium constant K2 favors the reverse of DNA binding for R350A, while WT favors going forward in the pathway. This explains the lack of PydC signal observed for R350A. Global fitting resolved a rate of DNA strand separation to be 3.5-fold slower for R350A (Fig. 4.6F). Therefore, if GS^{I} is formed, R350A is still able to DNA strand-separate. However, the lack of PydC signal is due to the minimal population of GS^{I}. R350A can methylate DNA, however product turnover is limited by the unfavorable formation of the strand-separated state (GS^{II}).

Calculations of k_{cat}, K_M, and k_{cat}/K_M

Values for k_{cat}, K_M, and k_{cat}/K_M were calculated for each model based on equations 1, 2, and 3, respectively. These values are listed in Table 4.2. The specificity constant is defined by k_{cat}/K_M. These values show that the decreased specificity constant for R350A is mostly due to an increase in K_M, while the larger decrease for noncognate is mostly due to a decreased k_{cat}.

Because the binding steps are much faster than DNA methylation, the Km reflects the equilibrium binding of DNA including isomerization steps and DNA strand separation, which are unfavorable thermodynamically so only one third of the enzyme-DNA complexes reach the fully strand-separated state required for catalysis.

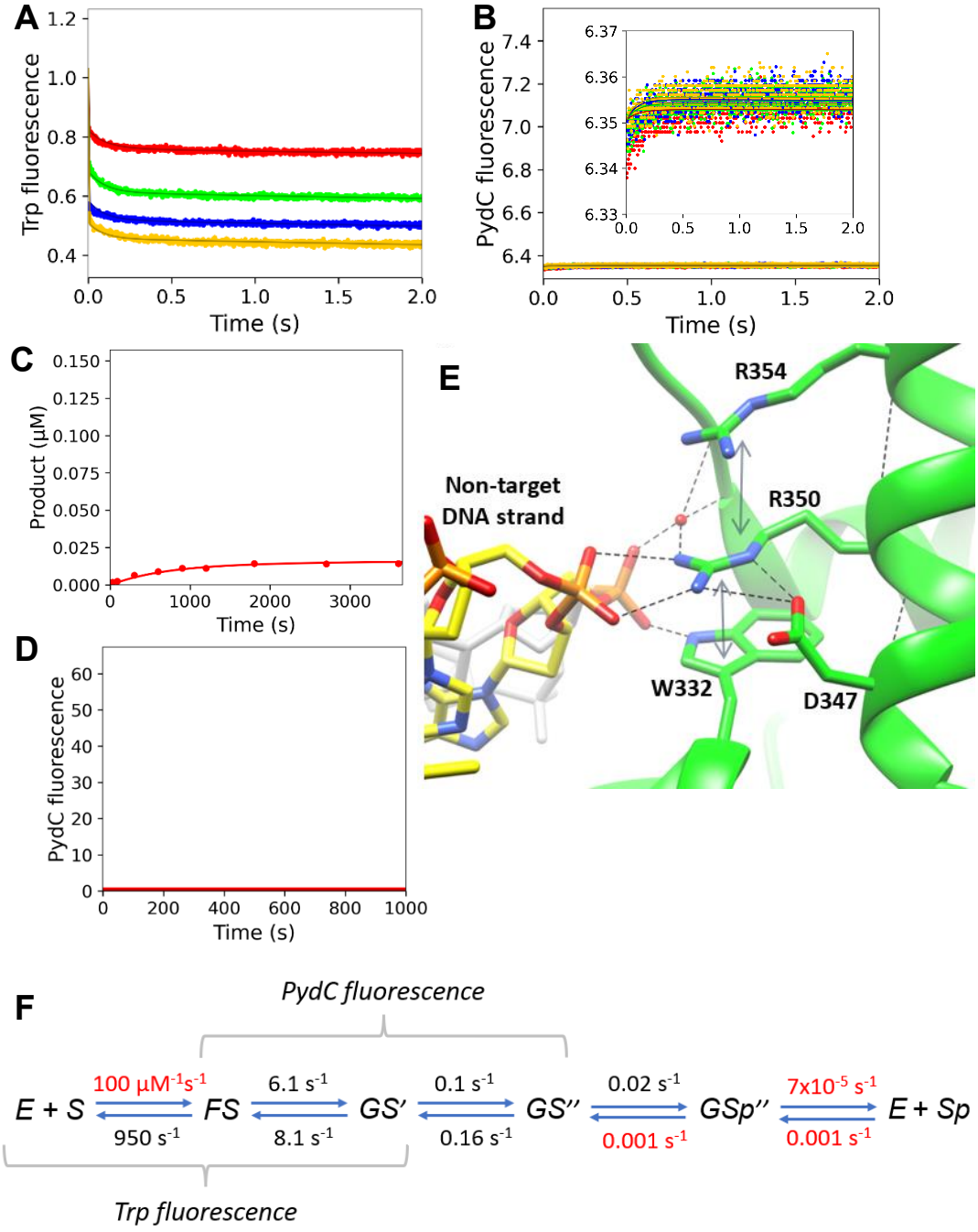


Figure 4.6. R350A has severely attenuated PydC fluorescence while Trp kinetics are only slightly perturbed, suggesting that the protein isomerization step and DNA strand separation are mechanistically uncoupled. A. Trp kinetics consisted of R350A [500 nM], C1-DNA [2.5 (red), 5.0 (green), 7.5 (blue), and 10.0 (yellow) μM], and SAM [60 μM]. **B.** DNA strand-separation monitored by PydC fluorescence consisted of P1-DNA [1.0 μM], R350A [5.0 (red), 7.5 (green) and 10.0 (blue) μM], and SAM [60 μM]. **C.** R350A and cognate DNA single

turnover methylation monitored via radiochemical methods. P1-DNA [100 nM], R350A [250 nM], and SAM [15 μM].**D.** PydC kinetics over 1000 seconds consisted of R350A [5 μM], P1-DNA [1 μM], and SAM [60 μM]. **E.** R350 from the C-terminal domain makes two hydrogen bonds to the phosphate backbone of the non-target strand (yellow DNA). Hydrogen bonds are black dashed lines. Stacking interactions are depicted with grey arrows. Molecular graphics were made with UCSF Chimera. **F.** Rate constants derived from globally fitting the data for WT and Noncognate DNA. Red values indicate rates that were locked during the simulation.

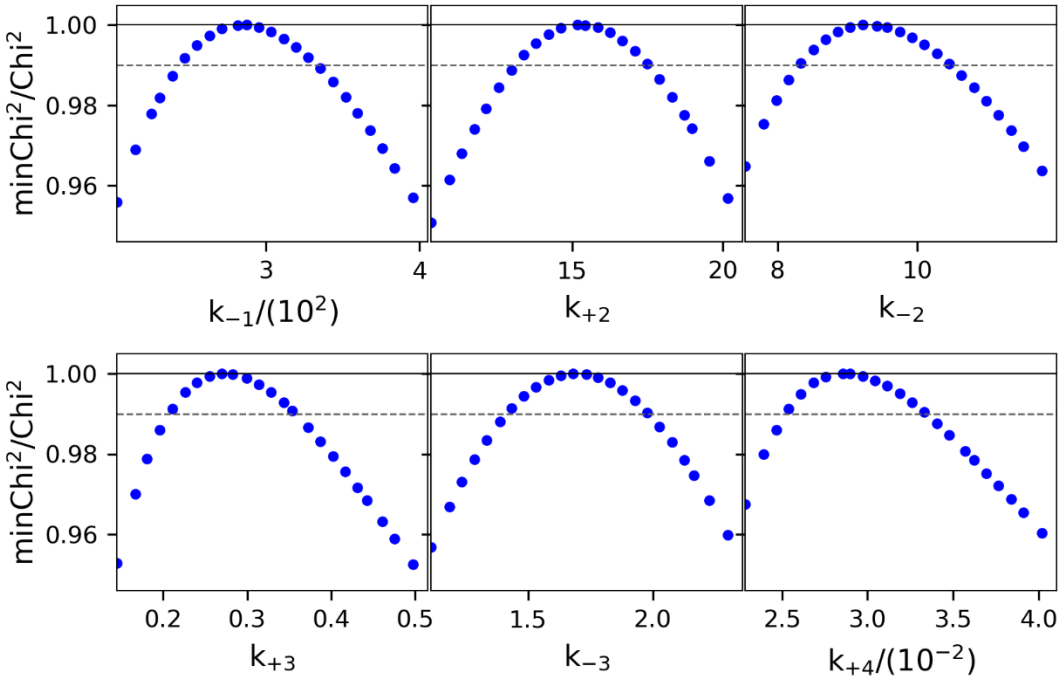
Confidence contour analysis

Confidence contour analysis show that the kinetic parameters are well-defined for the five step model with WT enzyme with cognate DNA (Fig. 4.7). The data in Figure 7 represent the 1D FitSpace calculated for each rate constant. The dashed line establishes the 95% confidence interval at the 0.99 normalized Chi^2 threshold. The 95% Chi^2 limits were calculated in KinTek Explorer and are shown in Figure 7. The confidence contour 1D FitSpace for the WT/Noncognate DNA and R350A/CognateDNA models are shown in SI Fig. 6, and SI Fig. 7, respectively[4.1]).

Kinetics of PydC in the absence of cofactor

In order to understand the role that cofactor plays in DNA strand separation we compared the changes in fluorescence of the PydC signal in the presence and absence of the cofactor. In the absence of cofactor, PydC signal is 5-fold smaller than in the presence of SAH (the product of the reaction with SAM) (Fig. 4.8A). When fit to a double-exponential function, the initial increasing phase has an apparent rate of $11.4s^{-1}$ with a fluorescence amplitude of 0.14, and the second decreasing phase has an apparent rate of $1.07s^{-1}$ with a fluorescence amplitude of 0.02 (Fig. 4.8A). In the presence of SAH, (Fig. 4.8B) the PydC signal follows a double-exponential function with a fast initial phase at an apparent rate of $13.3s^{-1}$ with an

amplitude of 2.4 and the second slower phase has an apparent rate of $2.3s^{-1}$ with an amplitude of 0.5 (Fig. 4.8B). While we do not understand the mechanistic basis for the small increase then decrease in fluorescence of PydC in the absence of cofactor, it is clear that a much larger signal for DNA strand separation occurs with the cofactor analog.



Rate	95% confidence interval
k_{-1}	247-327
k_2	13.4-17.5
k_{-2}	8.3-10.5
k_3	0.21-0.35
k_{-3}	1.43-1.98
k_4	0.025-0.033

Figure 4.7. Confidence contour analysis show that the data are well-defined for the WT and cognate DNA model. The data represent the 1D Fitspace calculated for each rate constant. The dashed line establishes the 95% confidence interval at the 0.99 Chi^2 threshold. The 95% Chi^2 limits were calculated in KinTek Explorer and are shown in the table.

The similar rate of the first increasing phase with and without SAH could suggest that some strand separation can occur independent of cofactor, but the approximately 5-fold

increase in the amplitude with SAH implies a much greater fraction of the DNA reaches the strand-separated state (GS^{II}). Because the steps leading up to GS^{II} reach equilibrium before DNA methylation, interactions between the cofactor, enzyme and DNA will affect the fraction of bound DNA in the GS^{II} state and the observed rate for the PydC signal will be the sum of rate constant for the forward and reverse reactions[4.11]. Our data suggest that cofactor (SAH) stabilizes the strand-separated conformation (GS^{II}) leading to a much larger amplitude in the PydC signal.

	Global fitting model		
Parameter	**WT/Cognate DNA**	**R350A/Cognate DNA**	**WT/Noncognate DNA**
k_{cat} (s^{-1})	0.0027	*0.0039	$1.0x10^{-6}$
K_M (μM^{-1})	0.99	4.39	0.85
k_{cat}/K_M (μM^{-1} s^{-1})	0.0027	0.00088	$1.2x10^{-6}$

$$k_{cat} = \frac{k_2k_3k_4}{k_2(k_3+k_{-3}+k_4)+k_{-2}(k_{-3}+k_{-4})+k_3k_4} \quad \textbf{Eq.1}$$

$$K_M = \frac{k_{-1}(k_{-2}k_{-3}+k_{-2}k_4+k_3k_4)+k_2k_3k_4}{k_1[k_2(k_3+k_{-3}+k_4)k_{-2}(k_{-3}+k_4)+k_3k_4]} \quad \textbf{Eq.2}$$

$$k_{cat}/K_M = \frac{k_1k_2k_3k_4}{k_{-1}(k_{-2}k_{-3}+k_{-2}k_4+k_3k_4)+k_2k_3k_4} \quad \textbf{Eq.3}$$

Table 4.2. k_{cat}, K_M, and k_{cat}/K_M were calculated for each model. k_{cat}, K_M, and k_{cat}/K_M were calculated from equations 1, 2, and 3, respectively. k_{cat}/K_M is the specificity constant. An asterisk indicates a value that is not well defined. k_{cat} for R350A does not include the parameter k_{-4} which is greater than zero, therefore k_{cat} for R350A is not well-defined.

The cocrystal structure relies on sinefungin (a nonreactive structural analog of SAM) to stabilize the GS^{II} complex and shows that the cofactor stabilizes GS^{II} with interactions involving loop residues. For example, W57 at one terminus of Loop2B makes a stacking interaction to sinefungin, and the catalytic DPPY motif at the other terminus of the loop makes

hydrogen bonds to sinefungin (Fig. 4.8C)[2.1]. P45 at the proximal end of Loop2B is intercalated between two thymine bases of the nontarget strand and the peptide backbone of Loop2B makes an intra-loop hydrogen bond to the sidechain of K126 of Loop45 (Fig. 4.8C). Therefore, when cofactor is bound, Loop2B is positioned between the separated strands of DNA and makes interactions that stabilize the strand-separated conformation (GS^{II}).

It is important to note which cofactor analog was used in the experiments throughout this work and the structures of each cofactor are shown in Figure 9. SFG was used in equilibrium Trp fluorescence (Fig. 4.2) as well as the crystal structure[2.5]. SAM was used in all experiments included in global fitting (Figs. 4-6) and SAH was used to compare PydC kinetics in the absence of cofactor (Fig. 4.8).

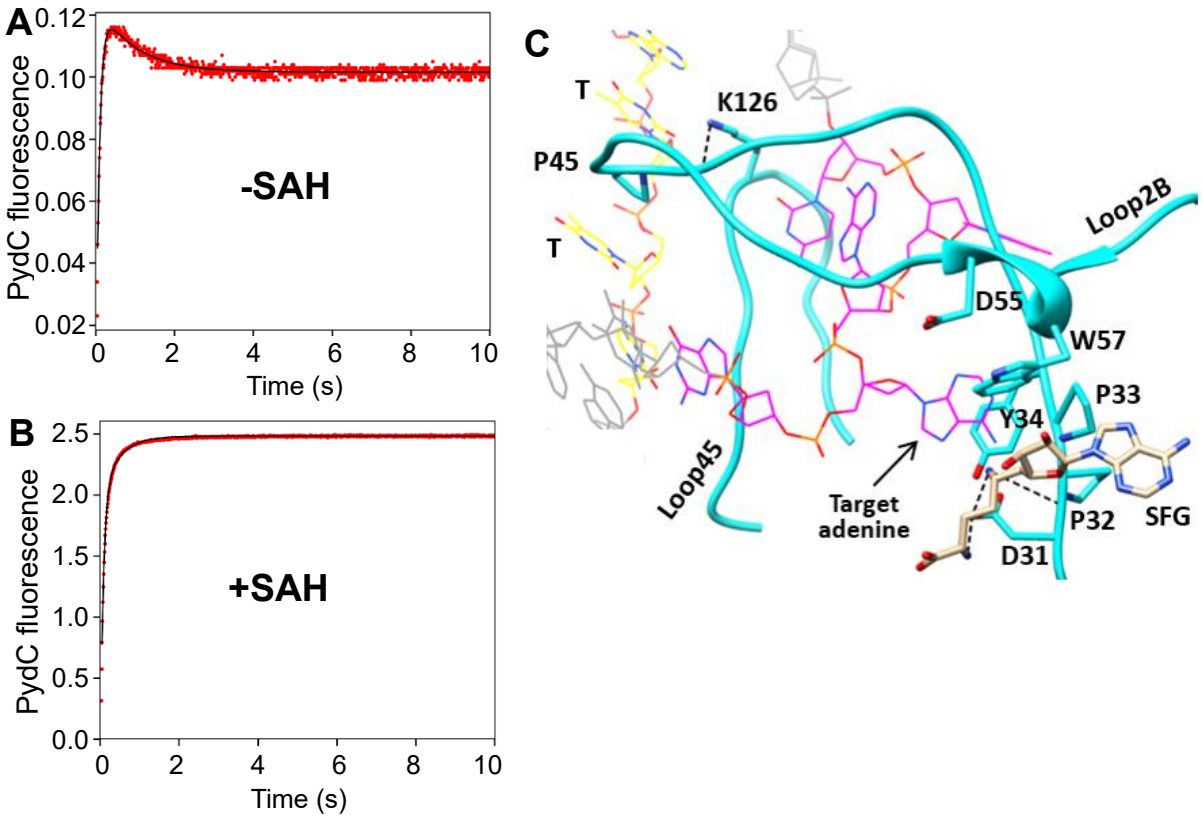


Figure 4.8. PydC kinetics in the absence of cofactor suggest that the cofactor stabilizes the strand-separated DNA conformation (GS^{II}). A. We observe biphasic PydC kinetics for WT CcrM in the absence of SAH. When fit to a double-exponential function, the initial increase in fluorescence has an apparent rate of $11.4s^{-1}$ with an amplitude of 0.14, and the second decreasing phase has an apparent rate of $1.07s^{-1}$ with an amplitude of 0.02. **B.** In the presence of SAH, we do not observe the reversal of the signal. When fit to a double-exponential function, the fast initial increase in fluorescence has an apparent rate of $13.3s^{-1}$ with an amplitude of 2.4 and the second slower phase has an apparent rate of $2.3s^{-1}$ with a fluorescence amplitude of 0.5. The conditions consisted of WT CcrM [5 µM], P1-DNA [1µM], and SAH [0 or 60µM]. **C.** The crystal structure suggests that SAH stabilizes the strand-separated conformation (GS^{II}) as seen by interactions between the analog sinefungin (SFG) and the DPPY (D31-Y34) motif in Loop2B. Magenta and yellow DNA represent the recognition site of the target and non-target strands, respectively, while grey DNA is outside of the recognition site. Sinefungin (tan) makes hydrogen bonds (black dashed lines) to D31 and the peptide backbone of P32. W57 makes a stacking interaction to the dihydroxyoxolan of sinefungin. P45 is intercalated between two thymine bases of the non-target DNA strand (yellow). Molecular graphics were made with UCSF Chimera.

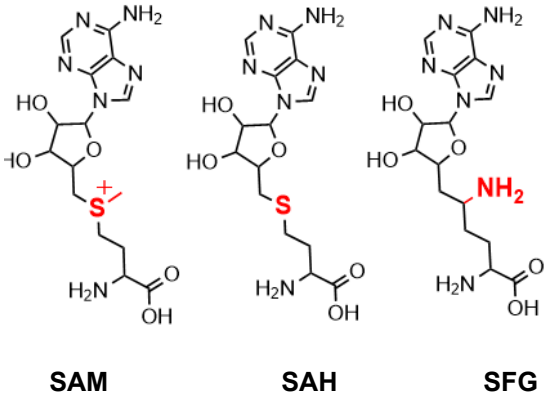


Figure 4.9. Structures of SAM, SAH, and SFG. The structure of SAM and analogs SAH and SFG are shown with the structural differences colored red. Structural images were made with ChemDraw.

IV. Discussion

How proteins bind to specific DNA sequences is well understood and sequence recognition by DNA endonucleases and methyltransferases has been extensively studied[4.12, 4.13]. For example, DNA methyltransferases gain stereochemical access to the target adenine or cytosine within a recognition sequence by "base-flipping" of the target base[2.1, 4.14].

Conformational transitions such as base-flipping can be rate-limiting and contribute significantly to sequence discrimination[2.1]. CcrM stands out in this context for several reasons. Its sequence discrimination is orders of magnitude more stringent than previously reported DNA methyltransferases, perhaps reflecting selective pressures associated with its essential role in gene regulation in *C. crescentus*[1.7, 2.3]. Additionally, rather than causing a single base to undergo a base-flipping transition, CcrM induces the extrahelical positioning of four of the five bases in its recognition sequence. Data presented here support our working hypothesis that this unprecedented repositioning of multiple bases into distinct recognition pockets on the protein is a major determinant of specificity. The assays and concepts being presented here for CcrM have potential relevance to other enzymes using a strand displacement recognition mechanism (e.g., CRISPR-Cas9)[4.2, 4.3] or enzymes that recognize and modify a single strand within DNA lesions or mismatches (e.g., Human MettL3-L14)[4.15].

Relationship between DNA strand separation and methylation

Unlike many DNA methyltransferases, product release is not rate-limiting for CcrM[2.3, 2.2], and our data and model (Fig. 4.3 and Fig. 4.4) support this conclusion. To define the rate-limiting and specificity-determining steps we measured the rate constants of the steps leading up to the first kinetically irreversible step (methylation), including DNA strand separation. Our proposed kinetic model (Fig. 4.3) describes a precatalytic protein conformational change followed by reversible DNA strand separation. The strand-separated conformation is the precatalytic intermediate where the target adenine is positioned for catalysis (GS^{II}) as observed in the cocrystal structure depicting GS^{II} where four of the five bases of the target sequence are

disrupted (Fig. 4.1). The rates of DNA strand separation (k_2 and k_3) are an order of magnitude faster than methylation and they are readily reversible with both k_{-2} and k_{-3} faster than k_4. DNA methylation (k_4) is the rate-determining step and the actual rate is attenuated by the fact that only 10% of the enzyme-bound DNA is in the fully strand-separated state poised for catalysis (fraction = $K_2K_3/(1 + K_2 + K_2K_3)$). The thermodynamically unfavorable strand separation may contribute to the enzyme's high fidelity in that any weakening of the interactions of the DNA with the enzyme would greatly reduce the amount of strand-separated DNA. In contrast, if the reaction went to 99% completion, a small change in free energy of strand separation due to noncognate DNA may still allow 90% strand separation. This phenomenon is similar to the weak binding of catalytic Mg^{2+} to DNA polymerases which occurs only after the induced-fit recognition of a canonical base-pair[4.9]. Tighter binding of the catalytic Mg^{2+} would provide additional binding energy to stabilize and incorporate a mismatch.

Relationship between protein and DNA conformational changes

We sought to monitor the steps leading up to GS^{II}. We monitored equilibrium Trp fluorescence changes upon binding of different DNA substrates (Fig. 4.2) which corresponds to the transition from E to GS^{I}. Because our previous work showed, surprisingly, that CcrM methylates ssDNA and dsDNA with similar efficiency[2.3, 2.2], we showed that both cognate and non-cognate ssDNA induce the same changes in Trp fluorescence of CcrM (Fig. 4.2A). The same result was observed with cognate and non-cognate dsDNA (Fig. 4.2B). Together, these results suggest that protein conformational changes induced by binding DNA are not inherently dependent on the form nor sequence of the substrate. Figure 2E shows that WT and W332Y,

which directly contact the non-target DNA strand (Fig. 4.4C), have similar Trp fluorescence changes upon binding ssDNA, whereas this is not observed upon binding dsDNA by the mutant. This suggests the fluorescence changes deriving from W332 make a significant contribution to the changes observed upon binding dsDNA, and implicates the C-terminal domain's importance in these changes.

We utilized stopped-flow fluorescence to quantify the rates of the changes in protein conformation and DNA strand separation. Our model suggests that a protein conformational change occurs after the association of enzyme and DNA, followed by DNA strand-separation, generating the intermediate poised for methylation (GS^{II}). The apparent rates of Trp and PydC kinetics are effectively identical for WT CcrM with cognate DNA (C1) (Fig. 4.4A-B). Our interpretation of the coincident rates is that the protein conformational change precedes and limits the rate of the first DNA strand-separation step to form the intermediate state (GS^{I}). Thus, the first change in PydC fluorescence is actually reflecting the rate of the protein conformational change.

Relationship between DNA strand separation and substrate recognition

We suggest that DNA discrimination is dependent on strand-separation, while protein conformational changes do not contribute significantly to DNA sequence discrimination. Similar Trp kinetics are observed with both cognate and noncognate DNA, suggesting that CcrM undergoes similar protein conformational changes with both substrates (Fig. 4.5A compared to Fig. 4.4A). Also, PydC kinetics are not observed with noncognate DNA (Fig. 4.5B). The inability to strand-separate noncognate DNA while the protein conformational

change is unperturbed, provides additional evidence that these processes are independent. The interactions between features of CcrM and either the target or non-target DNA strands observed in the cocrystal structure, suggest potential roles in inducing or stabilizing the strand separated intermediate (GS^{II}, Fig. 4.2B). For example, three loops (2B, 45, and 6E) interact with specific bases of the target strand 5'GANTC'3 site[2,5]. Loop2B and Loop45 are inserted between the separated DNA, suggesting they contribute to strand separation or maintain the strand separated form (GS^{II}). Loop6E may contribute to the stabilization of GS^{II} and recognition of specific nucleic acid bases since it is not positioned between the strand-separated DNA and it makes base-specific interactions with the recognition site.

Noncognate DNA is able to bind to the enzyme, but the DNA strand-separated state is destabilized and does not accumulate, leading to reduced methylation. This observation supports the conclusion from studies on cognate DNA that discrimination against noncognate DNA is facilitated by equilibria that disfavor DNA strand-separation. Thus, a small change in equilibrium constant for stabilizing the strand-separated state at the active site translates into a large factor of discrimination against DNA methylation of noncognate DNA. Note that the rate constants governing steps 1 and 2 with noncognate DNA were derived solely from the tryptophan fluorescence data while steps 3 and 4 are not well defined because their amplitudes are so small.

Global fitting reveals that DNA strand separation is the substrate discriminating step. We explored a model that included an additional and much faster rate of DNA strand separation (200-300s^{-1}); however, the fast phase was not described kinetically by the data. This taught us an important lesson: when deciding how many phases to include in global fitting, only steps

that are described by the data may be included in defining a minimal model. The additional step may occur but it is not defined by the data. If there is a fast phase of DNA strand separation, it is not resolved with PydC.

CRISPR/Cas9 also relies on a DNA strand displacement mechanism for target DNA recognition and access[4.16]. Unlike CcrM, CRISPR/Cas9 is not highly discriminating against off-target DNA, and once bound to an off-target site Cas9 performs DNA cleavage unless the rate of hydrolysis is exceedingly slow[4.16, 4.17]. CcrM's discrimination, however, is tightly coupled to the strand-separation event, and methylation will not occur on non-cognate sites due to its highly selective DNA strand separation and recognition mechanism. Similarly to CcrM, however, CRISPR-Cas9 recognition and cleavage depends on the conformation of the guide RNA-DNA duplex[4.16]. CRISPR-Cas9 favors cleavage of a kinked cognate duplex, and favors release of a linear mismatched duplex[4.17]. The design of CRISPR/Cas9 variants with higher selectivity is limited by the stability of the RNA/DNA duplex. This contrast further illuminates the role of DNA strand-separation in sequence fidelity.

CcrM's mechanism of discrimination can also be compared to that of T7 DNA polymerase, in which a conformational change in the protein that is much faster than chemistry selects the correct nucleotide via an induced-fit mechanism[4.6]. Similarly, CcrM's discriminating strand-separation step is faster than chemistry. However, in the case of DNA polymerase, dissociation of the bound nucleotide is modulated so that a correct base pair is captured and committed to incorporation, while a mismatch is rapidly released rather than incorporated. In the case of CcrM the strand-separated state with noncognate DNA is so disfavored that it cannot be observed. Presumably, this is due primarily to a fast dissociation

rate that we cannot measure because we can't form the bound strand-separated state with noncognate DNA. The arguments are similar to the controversies over the role of induced-fit in specificity where substrate-induced changes in enzyme structure were thought to be unimportant for specificity unless they were rate-limiting[4.5]. Here we show that the rate of DNA strand-separation is much faster than DNA methylation, but still constitutes a major determinant of specificity.

This model tells a story where DNA strand-separation is 100-fold faster than DNA methylation which does not occur until the DNA is fully strand-separated. However, DNA strand-separation is not thermodynamically favorable; rather, the unwinding comes to equilibrium with a net equilibrium constant defined by $K_2K_3 = 0.27$. The fraction of DNA in the fully strand-separated state is $= K_2K_3/(1 + K_2 + K_2K_3) = 0.09$. Thus, DNA strand-separation does not drive the reaction forward toward the reactive state. Rather, a small fraction of DNA is unwound and aligned for catalysis. There may be additional steps leading to alignment of the adenine, SAM and catalytic residues that may limit the net rate of methylation. Because DNA binding to the enzyme does not drive DNA strand-separation, discrimination against noncognate DNA may occur by allowing DNA release from any bound state prior to catalysis.

Our results with the R350A CcrM mutant support our proposed separation of protein conformational changes and DNA strand separation (Fig. 4.6). R350 is highly conserved in the C-terminal domain and makes two hydrogen bonds to the phosphate backbone of the nontarget DNA strand[3.0] (Fig. 4.6E). Our prior mutational analysis shows that highly conserved residues in the C-terminus that make hydrogen bonds to the phosphate backbone of the nontarget DNA strand are essential for strand-separation[3.0]. Our results with R350A show that the processes of

protein conformational changes and DNA strand-separation can be uncoupled. R350A is able to undergo protein conformational changes (k_2) similar to WT (Fig. 4.4A), while DNA strand separation (k_3) is massively reduced in rate and amplitude (Fig. 4.6B).

Global fitting reveals the functional consequences for R350A are primarily due to the increase in k_{-1} and the decrease in K_2. The rate of DNA strand separation is 2.8-fold slower for R350A compared to WT (Fig. 4.6F). However, due to the diminished PydC amplitude, the rates for this process may not be well defined. R350A generates a small amplitude of PydC signal and can form product, so we present a few possible explanations.

One possible explanation of R350A's diminished PydC signal is that the population of GS^I is too low to detect PydC signal due to the destabilization of FS and GS^I. The equilibrium constant K2 favors the reverse for R350A, while WT favors going forward in the pathway. Therefore, if GS^I is formed, R350A is still able to DNA strand-separate, however, the lack of PydC signal is due to the fast reverse reaction, resulting in less methylation.

Another possible explanation for R350A's low PydC signal and slower product formation is an off-pathway non-methylatable intermediate. k_{-4} is a parameter that was included in the R350A/Cognate-DNA model due to the incomplete methylation reaction. The single-turnover methylation assay does not go to completion (Fig. 4.6C). Therefore, k_{-4} was included in the model to account for the incomplete amplitude in this reaction. To reconcile this parameter, R350A could have an off-pathway intermediate that is not capable of methylation. This off-pathway intermediate would not have strand-separated DNA and may not revert to a state which can undergo methylation during the time course of the reaction, thereby explaining the diminished PydC signal and incomplete substrate turnover.

A third possible explanation for R350A's low PydC signal and slower product formation is that this mutant undergoes an entirely different mechanism of gaining stereochemical access to the target adenine. Most N^6-adenine DNA methyltransferases (unlike CcrM) solely flip the target adenine outside of the DNA helix, in a well-known process called base-flipping[4.18, 4.19, 4.20, 4.21, 4.22]. R350A might be unable to separate four of the five bases of the recognition site, but may be able to base-flip the target adenine. Thus, PydC at the N-position would maintain Watson-Crick base-pairing, while the target adenine is extra-helical, explaining the diminished PydC signal, and slow rate of methylation. Interestingly, R350A is less discriminating against noncognate substrates[3.0], which is further evidence that strand separation is responsible for substrate discrimination.

A new role of cofactor is proposed

S-adenosylhomocysteine (SAH) is formed as a product of the reaction and is an inhibitor of CcrM (23). SAH or the cofactor analog, sinefungin, help stabilize the strand-separated conformation (GS^{II}) through interactions with Loop2B (Fig. 4.8C). The crystal structure uses sinefungin and our study used SAH. We show that SAH and SFG have similar effects on equilibrium Trp fluorescence changes and therefore can both be used to monitor possible changes in enzyme structure as revealed by changes in Trp fluorescence (SI Fig. 8[4.1]). In contrast stopped-flow fluorescence in the presence of SAH shows an increase in PydC signal (Fig. 4.8B). In the absence of SAH, we observe a small increase, followed by a smaller decrease of PydC fluorescence (Fig. 4.8A). The decrease of the signal without SAH suggests that GS^{II} may undergo a partial reversal, or a transition to another, off pathway intermediate.

Trp kinetics shows similar biphasic kinetics in the presence and absence of SAH when fit to double exponential functions (SI Fig. 10[4.1]). This preliminary result suggests that the transition from E to GS^{I} is not massively dependent on cofactor. Further analysis of the kinetics in the absence of cofactor may resolve the contributions of the cofactor to formation, stabilization and alignment of catalytic residues in the presence of cofactor.

The interactions seen with the cofactor in the crystal structure support our explanation of how SAH stabilizes GS^{II}. Sinefungin interacts with Loop2B (Fig. 4.8C) which consists of residues D31-E61 and the DPPY motif (D31-Y34). The DPPY motif is a conserved SAM binding motif in N^6 DNA methyltransferases that is commonly located in loop or disordered secondary structures and forms part of the active site[2.1]. Uniquely, CcrM's Loop2B, which contains the DPPY motif, is inserted within the strand-separated DNA (Fig. 4.8C). We suggest that the interactions that occur at both termini of Loop2B to cofactor (SAH or sinefungin) stabilize the position of Loop2B within the strand-separated DNA, therefore stabilizing GS^{II}. Other interactions from Loop2B, such as the intercalation of P45 between two thymine bases of the non-target DNA and the intraloop hydrogen bond made by Loop2B's peptide backbone at L43 to the sidechain of K126 in Loop45 also contribute to stabilization of GS^{II} (Fig. 4.8C). These stabilizing interactions are favored only when cofactor is bound to Loop2B and are likely to provide key interactions responsible for the high specificity of CcrM.

Other cocrystal structures of MTase-dsDNA-cofactor complexes reveal similar and variable roles of the loop containing the DPPY or NPPY motif. T4Dam's DPPY is in an 8-residue loop that does not interact with DNA[4.12]. EcoP151I's DPPY motif is in a 19-residue loop that does not interact with DNA[4.23]. CamA's NPPY motif is in a 10-residue loop that

makes base-specific interactions to target and non-target strand DNA bases[4.24]. M.Taq1's NPPY motif is in a 17-residue loop that makes base-specific interactions to three non-target strand bases[4.25]. Thus, the cofactor-associated positioning of a loop containing the (D/N)PPY motif may not have a conserved functional role, but in the context of CcrM, the cofactor plays a unique and indirect role in stabilizing strand-separated DNA via interactions with Loop2B's DPPY motif.

Summary

The results presented in this work contribute to our understanding of the complex mechanism that governs DNA discrimination by means of DNA strand-separation. CcrM relies on conformational intermediates in both protein and DNA, and the changes in DNA significantly contribute to substrate discrimination. We identified the conserved residues that are essential for DNA strand-separation and are found in diverse bacterial phyla and in human and animal pathogens[3.0]. The mechanism of strand-separation and its relationship to DNA discrimination could be insightful for the similar utilization of strand-separation by enzymes that rely on recognizing dsDNA but only modify a single strand.

IV. References

4.1. Konttinen, O., Carmody, J., Kurnik, M., Johnson, K.A., and Reich, N. (2023) High fidelity DNA strand-separation is the major specificity determinant in the DNA methyltransferase CcrM's catalytic mechanism. *Nucleic Acids Research*, 51, 13, 6883-6898.

4.2. Sternberg,S.H., 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.

4.3. Jiang,F. and Doudna,J.A. (2015) The structural biology of CRISPR-Cas systems. *Current Opinion in Structural Biology*, 30, 100-111.

4.4. Fersht, A.R. (1999) *Enzyme Structure and Mechanism*. 3rd ed. Freeman, New York.

4.5. Prasad,B.R., Kamerlin,S.C.L., Florian,J. and Warshel,A. (2012) Prechemistry barriers and checkpoints do not contribute to fidelity and catalysis as long as they are not rate limiting. *Theor. Chem. Acc*, 131, 1288.

4.6. Dangerfield,T.L. and Johnson,K.A. (2021) Conformatinal dynamics during high-fidelity DNA replicaiton and translocation defined using a DNA polymerase with a fluorescent amino acid. *J. Biol. Chem.*, 296, 100143.

4.7. Kellinger,M.W. and Johnson,K.A. (2010) Nucleotide-dependent conformational change governs specificity and analog discrimination by HIV reverse transcriptase. *PNAS*, 107, 7734-7739.

4.8. Kirmizialtin,S., Nguyen,V., Johnson,K.A. and Elber,R. (2012) How conformational dynamics of DNA polymerase select correct substrates: experiments and simulations. *Structure*, 20, 618-
627.

4.9. Gong,S., Kirmizialtin,S., Chang,A., Mayfield,J.E., Zhang,Y.J. and Johnson,K.A. (2021) Kinetic and thermodynamic analysis defines roles for two metal ions in DNA polymerase specificity and catalysis. *J Biol Chem*, 296, 100184.

4.10. Ramm,B., Glock,P., Mucksch,J., Blumhardt,P., Garcia-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.

4.11. Johnson, K.A. (2019) *Kinetic Analysis for the New Enzymology: Using computer simulation to learn kinetics and solve mechanisms*. KinTek Corporation, Austin, USA.

4.12. Halford,S.E., and Marko,J.F. (2004) How do site-specific DNA-binding proteins find their target?, *Nucleic Acids Research*, 32, 10, 3040-3052.

4.13. Rohs,R., Jin,X., West,S.M., Joshi,R., Honig,B., and Mann,R.S. (2010) Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem.*, 79, 233-69.

4.14. Horton,J.R., Liebert,K., Hattman,S., Jeltsch,A. and Cheng,X. (2005) Transition from nonspecific to specific DNA interactions along the substrate-recognition pathway of Dam methyltransferase. *Cell*, 121, 3, 349-361.

4.15. Cheng,X. et. al. (2021) Human MettL3-MettL14 RNA adenine methyltransferase complex is active on double-stranded DNA containing lesions. *Nucleic Acids Research*, 49, 20, 11629-11642.

4.16. Liu,M., Gong,S., Yu,H., Jung,K., Johnson,K.A. and Taylor, D.W. (2020) Engineered CRISPR/Cas9 enzymes improve discrimination by slowing DNA cleavage to allow release of off-target DNA. *Nature Communications*, 11, 3576.

4.17. Bravo,J.P.K., Liu,M., Hibshman,G.N., Dangerfield,T.L., Jung,K., McCool,R.S., Johnson. and Taylor,D.W. (2022) Structural basis for mismatch surveillance by CRISPR-Cas9, *Nature*, 603, 343-347.

4.18. Roberts,R.J. and Cheng,X. (1998) Base flipping. *Annual Review Biochemistry*, 67, 181-198.

4.19. Allan,B.W., Beechem,J.M., Lindstrom,W.M. and Reich, N.O., (1998) Direct real time observation of base flipping by the *EcoRI* DNA methyltransferase. *Journal of Biological Chemistry*, 273, 4, 2368-2373.

4.20. Liebert,K., Hermann,A., Schlickenrieder,M. and Jeltsch,A., (2004) Stopped-flow and mutational analysis of base flipping by the *Escherichia coli* Dam DNA (adenine-N6)-methyltransferase. *Journal of Molecular Biology*, 341,2, 443-454.

4.21. Cheng,X. and Roberts,R.J. (2001) AdoMet-dependent methylation, DNA methyltransferases and base flipping. *Nucleic Acids Research*, 29, 3784-3795.

5.22. Hong,S. and Cheng,X. (2016) DNA base Flipping: A general mechanism for writing, reading, and erasing DNA modifications. *Advances in Experimental medicine and biology*, 945, 321-341.

4.23. Gupta,Y.K., Chan,S., Xu,S. and Assarwal,A.K. (2015) Structural basis of asymmetric DNA methylation and ATP-triggered long-range diffusion by EcoP15I. *Nature Communications*, 6, 7363.

4.24. Zhou,J., Horton,J.R., Blumenthal,R.M., Zhang,X. and Cheng,X. (2021) *Clostridioides difficile* specific DNA adenine methyltransferase CamA squeezes and flips adenine out of DNA helix. *Nature Communications*, 12, 3436.

4.25. Lenz,T. et. al. (2007) 2-Aminopurine flipped into the active site of the adenine-specific DNA methyltransferase M.TaqI: crystal structures and time-resolved fluorescence. *J. Am. Chem. Soc*., 129, 6240-6248.

4.26. Peterman,B.F. (1979) Measurement of the dead time of a fluorescence stopped-flow instrument. *Analytical Biochemistry*, 93, 442-444.

Chapter V: Highly conserved catalytic loops are responsible for recognition, and generating and stabilizing the DNA strand separated state

V: Abstract

Two highly conserved loops in CcrM, Loops 2B and 45, are inserted between the strand-separated DNA interface; we hypothesize that these loops generate and stabilize the strand-separated conformation. During strand-separation, residues within Loops 2B, 45, and 6E contact the DNA strand that undergoes methylation (target). Highly conserved loop residues R44 and R129 when mutated to Alanine, disrupt strand-separation and are catalytically inactive. The highly conserved Loop-45 residue F125, which is positioned between the separated DNA strands, is also essential for maintaining the strand-separated intermediate; replacement of F125 results in various perturbations of strand-separation that are correlated to the bulkiness of the substituted residue. Global fitting for each mutant shows that generation and stabilization of DNA strand-separation are perturbed, providing a functional role for these loops in generating the strand-separated intermediate, which is essential for discrimination and catalysis. Insights into CcrM's mechanism of DNA strand-separation are likely to be helpful in understanding strand-separation mechanisms for other enzymes such as CRISPR-Cas9 and RNA polymerase sigma factor.

V: Introduction

Protein recognition of DNA sequences is fundamental to all known organisms and the underlying structural mechanisms are now well-understood[5.1]. For example, zinc finger proteins contain a conserved structural motif that coordinates zinc ions that stabilize a DNA

binding fold[5.2]. Leucine zipper proteins contain two dimerized alpha helices which contain basic amino acids at one terminus that recognize the major and minor DNA grooves[5.1]. Helix-turn-helix proteins contain a structural motif comprised of two alpha helices connected by a short flexible peptide in which the recognition helix is inserted within the DNA major groove where protein residues participate in base-specific DNA recognition[5.3, 5.4, 5.5]. These and other motifs are found in diverse proteins that recognize a vast variety of DNA sequences[5.6].

DNA methyltransferases recognize specific sequences and modify adenines or cytosines[5.7,5.8]. The majority of these bacterial and eukaryotic enzymes have been shown to rely on a "base flipping" mechanism in which the target base is flipped out of the DNA double helix; in the case of the well-studied M.HhaI C^5-cytosine MTase, a two-loop DNA binding motif is used for both sequence recognition and base-flipping[5.9, 5.10, 5.11, 5.12]. Movement of one of these loops is coupled with base-flipping[5.9, 5.10, 5.11, 5.12]. The homo-tetrameric human DNMT3A is a C^5-cytosine DNA methyltransferase and relies on a target recognition domain (TRD) which is made up of a loop that makes major groove contacts to the recognition sequence CpG[5.13, 5.14]. The TRD loop is inserted into the DNA major groove where it makes specific contacts[5.15]. The TRD loop's flexibility has also been proposed via molecular dynamic simulations to be important for DNMT3A recognition and catalysis[5.16].

The bacterial N^6-adenine DNA methyltransferase CcrM dimer (methylates adenine in 5'GANTC3') is the first DNA methyltransferase shown to rely on a unique DNA recognition mechanism in which the DNA strands are separated and most recognition interactions are limited to only one of the two strands (target)[2.5]. Four of the five base pairs within the recognition site are no longer paired, where Loop-2B approaches DNA from the major groove

and makes base-specific interactions to the target strand (Fig. 5.1[2.5]). Loop-45 approaches DNA from the minor groove and is positioned between the DNA strands. Together, Loop-2B and Loop-45 are inserted between the DNA strands, making contacts with target-strand bases, and appear to be stabilized by interactions with other protein moieties (Fig. 5.1)[2.5].

The target strand recognition sequence ($G_{10}A_{11}A_{12}T_{13}C_{14}$) is structurally perturbed in the strand-separated state (Fig. 5.2a). A_{11}, T_{13}, and C_{14} are flipped outside of the DNA helix (Fig. 5.2a). G_{10} maintains Watson-Crick base-pairing to C_{11} of the nontarget strand and G_{10} is recognized by R44 from Loop-2B which is a highly conserved residue necessary for catalysis. G_{10} also makes a hydrogen bond to the peptide backbone of Loop-45 (Fig. 5.2b). A_{11} (the target adenine) is positioned for methyl-transfer and the conserved catalytic DPPY-motif (D31, P32, P33, and Y34) within Loop-2B makes hydrogen bonds and stacking interactions to A_{11}. K193 from Loop-6E makes a hydrogen bond to A_{11}, likely contributing to base-specific recognition and stabilization. R179 also hydrogen bonds to the phosphate backbone of A_{11} (Fig. 5.2c). A_{12} is not recognized by a protein residue which is not surprising due to CcrM's ability to accommodate any base at this N-position (5'GANTC3') (Fig. 5.2d). T_{13} is recognized by R129 from Loop-45 via hydrogen bonds and W109 via stacking (Fig. 5.2e). C_{14} makes three hydrogen bonds to the peptide backbone of Loop-45. Lastly, C_{14} is recognized by N124 from Loop-45 via stacking interactions (Fig. 5.2f).

We previously showed that CcrM's highly conserved and unusual 83 amino acid C-terminus is essential for DNA binding and strand-separation[3.0] and that DNA strand-separation is tightly coupled with substrate discrimination[4.1]. Global fitting resolved an enzyme kinetic model in which CcrM binds DNA, the protein undergoes a conformational change, followed

by DNA strand-separation and methylation[4.1]. However, the mechanistic details that lead up to DNA strand separation remain obscure. In this work, we rely on mutational analysis of two highly conserved loops (Loop-2B and Loop-45), stopped-flow fluorescent kinetics, and global fitting to understand how Loop-2B and Loop-45 govern strand-separation, which is essential for substrate discrimination and catalysis.

V: Materials and methods

All materials and methods in Chapter V are reported in previous chapters.

V: Results

Loop-2B and Loop-45 are positioned within the strand-separated DNA and make base-specific contacts to the target strand recognition sequence

The CcrM co-crystal structure reveals that Loop-2B, Loop-45, and Loop-6E are implicated in DNA strand-separation and base-recognition (Fig. 5.1[2.5]). Loop-2B and Loop-45 are positioned within the strand-separated DNA interface and make base-specific contacts with the target-strand bases, while Loop-6E is positioned outside of the strand-separated interface and interacts with target-strand bases (Fig. 5.1 and Fig. 5.2). We wanted to understand how these closely packed and busy loops contribute to generating or stabilizing DNA strand-separation, or both.

Again, we previously showed that Trp fluorescence monitored enzyme/DNA association and a subsequent protein conformational change (k_1, k_{-1}, k_2, and k_{-2}), PydC fluorescence monitored biphasic DNA strand-separation (k_2, k_{-2}, k_3, and k_{-3}) where the second

phase of Trp kinetics is kinetically coupled to the first phase of PydC kinetics, and a radiochemical methylation assay monitored product formation (k_4)[4.1]. Expanding on our previously published CcrM kinetic model[4.1], we collected fluorescence kinetic data for loop mutants to understand how loop-2B and loop-45 regulate DNA strand-separation. Residues selected for mutation are highly conserved and are structurally implicated in generating and/or stabilizing DNA strand-separation.

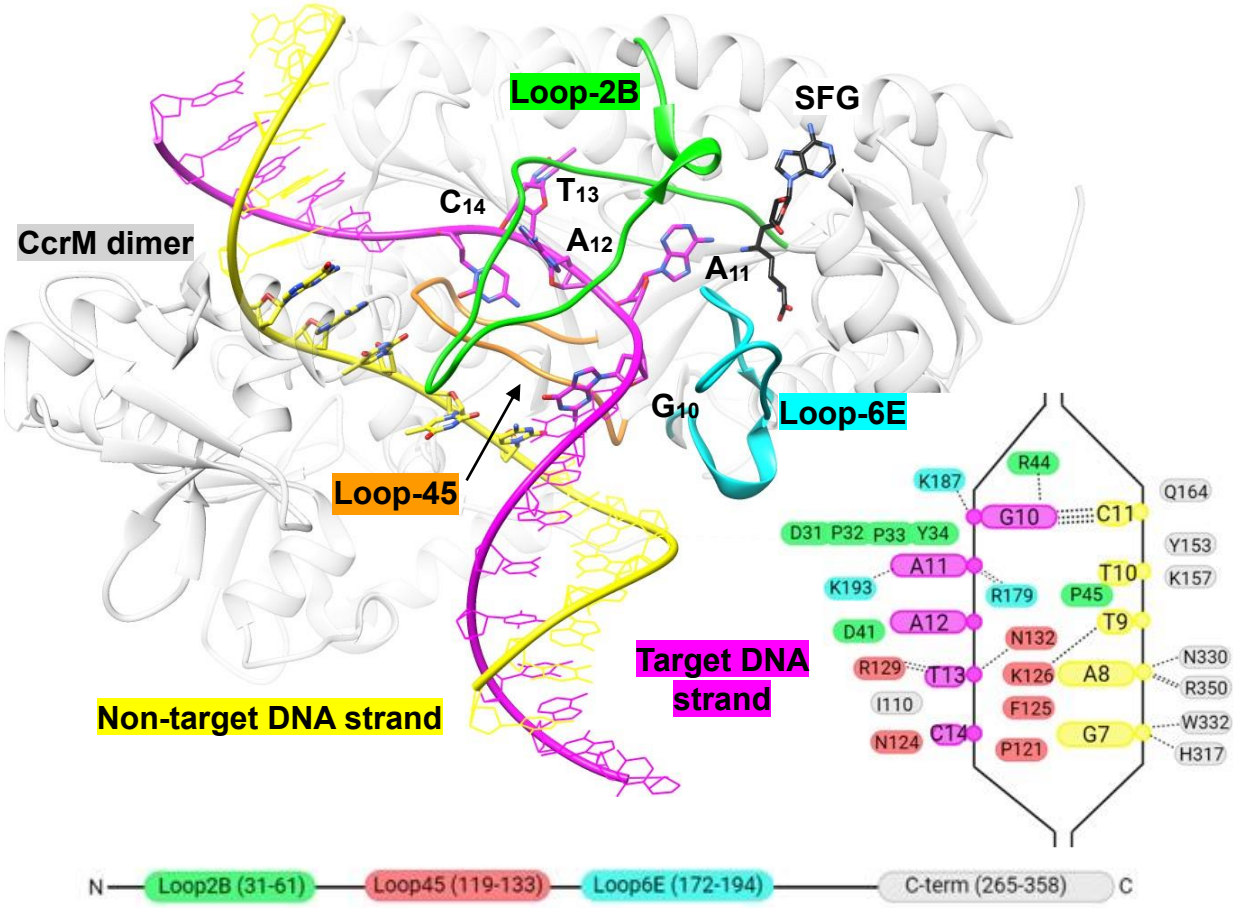


Figure 5.1. Cocrystal structure of CcrM complexed with double stranded DNA and sinefungin. Loop-2B and Loop-45 are positioned within the separated interface of DNA. Loop-6E is positioned near the target adenine and methyl-donor cofactor sinefungin (SFG). The GANTC recognition site is annotated 5'-$G_{10}A_{11}A_{12}T_{13}C_{14}$-3'. The structural image was made with UCSF Chimera. **Inset:** the strand-separated DNA phosphate backbone is depicted by the black outline. The DNA phosphates are depicted as circles and DNA bases as ovals.

Amino acids that interact with nucleic acids are depicted as ovals. Hydrogen bonds are dashed black lines. **Bottom:** The 2D positioning of loops within the CcrM N-terminal domain relative to the C-terminus. The structural image was made with UCSF Chimera and the cartoon was made with Biorender.

Bulky loop residue F125 primarily contributes to generation and stabilization of GS'

F125 in Loop-45 was selected for mutation due to its positioning within the DNA strand-separated bubble. F125 does not make any base-specific interactions and is therefore not directly involved in base-specific recognition (Fig. 5.3). We predicted that the bulkiness of F125 could contribute to either generation of DNA strand-separation, stabilization of DNA strand-separation, or both.

F125L binds DNA with a slightly weaker affinity to WT; F125L K_d^{app} = 185.3 ± 19.1, WT K_d^{app} = 149.9 ± 5.4nM (Fig. 5.10). F125L has a lessened k_2 and increased k_{-2} monitored by Trp fluorescence (F125L $k_2 = 10.8s^{-1}$ and $k_{-2} = 17.0s^{-1}$, WT $k_2 = 15.2s^{-1}$ and $k_{-2} = 9.4s^{-1}$) which results in a two-fold destabilization of the GS' intermediate (Fig. 5.4a, Table 5.1). F125L has a similar strand separation rate to WT monitored by PydC fluorescence (F125L $k_3 = 0.43s^{-1}$ and $k_{-3} = 1.1s^{-1}$, WT $k_3 = 0.28s^{-1}$ and $k_{-3} = 1.7s^{-1}$) (Fig. 5.4b, Table 5.1). F125L and WT have similar methylation rate constants; F125L $k_4 = 0.038s^{-1}$, WT $k_4 = 0.03s^{-1}$ (Fig. 5.4c, Table 5.1). The F125L kinetic model suggests that destabilization of GS' is the perturbation that limits this mutant's catalytic ability (Fig. 5.4d). Confidence contour 1D FitSpace analysis for F125L shows that the kinetic parameters are well-defined by the data (Fig. 5.11).

We sought to probe the role of F125 further with a smaller amino acid replacement (F125A). F125A binds DNA with a slightly weaker K_d^{app} compared to WT; F125A K_d^{app} = 190.8 ± 30.4, WT K_d^{app} = 149.9 ± 5.4nM (Fig. 5.10). The lessened Trp kinetics for F125A

indicates that the GS' intermediate is further destabilized when compared to F125L; F125A k_2 = 8.9s^{-1} and k_{-2} = 17.1s^{-1} (Fig. 5.5a, Table 5.1). F125A does not show a PydC fluorescence signal indicating destabilized GS''; F125A k_3 = 0.15s^{-1} and k_{-3} = 2.4s^{-1}, WT k_3 = 0.28s^{-1} and k_{-3} = 1.7s^{-1} (Fig. 5.5b, Table 5.1), which explains the lack of methylation (Fig. 5.5c). The F125A kinetic model suggests that the increased destabilization of GS' is so severe that the intermediate cannot progress to GS'' (Fig. 5.5d). Confidence contour 1D FitSpace analysis for F125A shows that the kinetic parameters are well-defined by the data (Fig. 5.11).

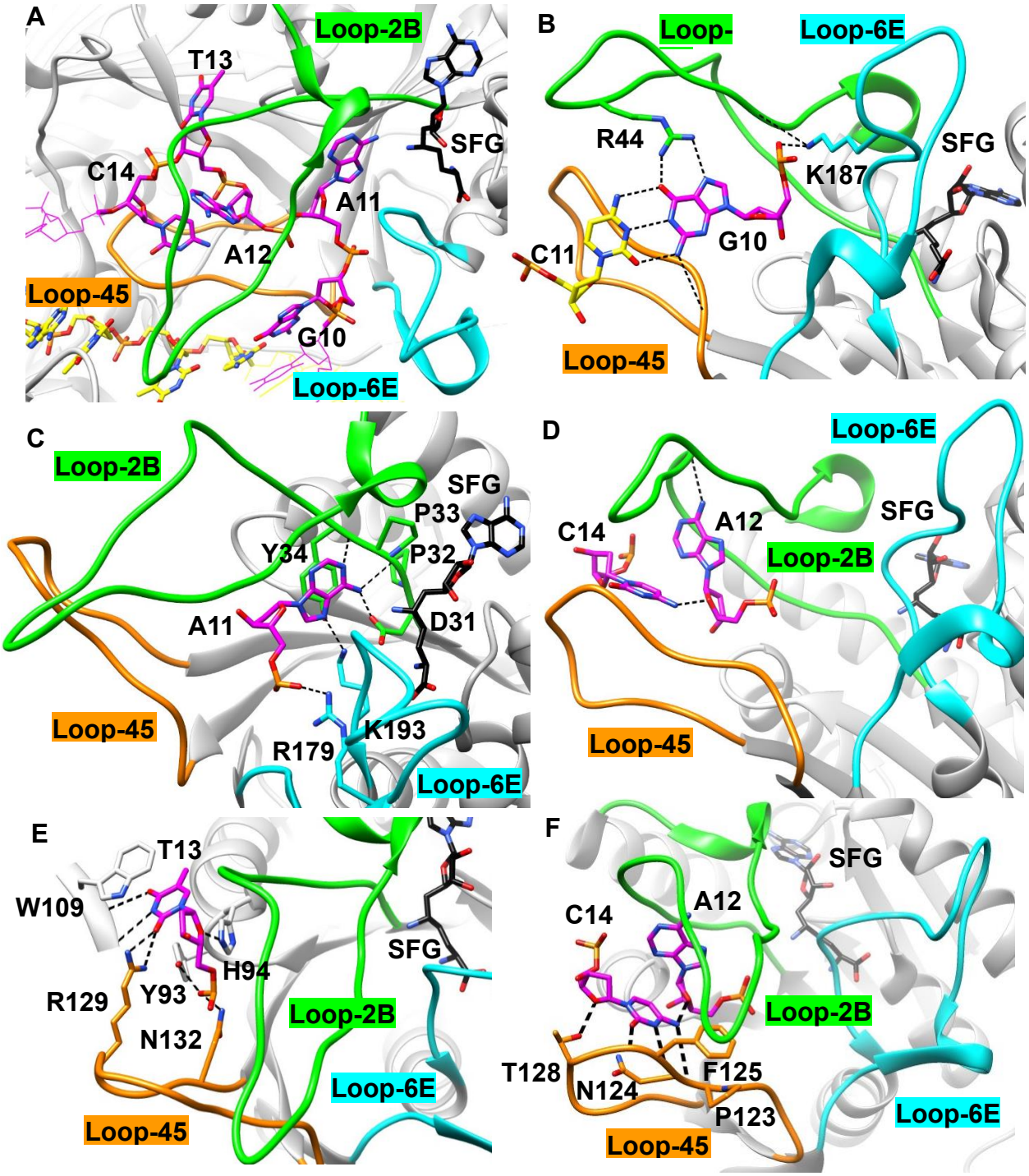


Figure 5.2. Base-specific recognition by Loop-405, Loop-2B, and Loop-6E. A. The target strand recognition sequence ($G_{10}A_{11}A_{12}T_{13}C_{14}$) is strand separated with Loop-45 and Loop-2B positioned between the DNA strands. Loop-6E is near the target adenine and methyl-donor cofactor sinefungin (SFG). **B.** G_{10} maintains Watson-Crick base-pairing to C_{11} of the nontarget strand. G_{10} is recognized by R44 from Loop-2B which is a highly conserved residue necessary

for catalysis. K167 makes a hydrogen bond to the phosphate backbone of G_{10}. G_{10} also makes a hydrogen bond to the peptide backbone of Loop-45 **C.** A_{11} (the target adenine) is positioned for methyl-transfer. The DPPY-motif (D31, P32, P33, and Y34) within Loop-2B makes hydrogen bonds and stacking interactions to A_{11}. K193 from Loop-6E makes a hydrogen bond to A_{11}, likely contributing to base-specific recognition and stabilization. R179 makes a hydrogen bond to the phosphate backbone of A_{11}. **D.** A_{12} is not recognized by a protein residue which is not surprising due to the ability of CcrM to accommodate any base at this N-position (5'GANTC3'). **E.** T_{13} is recognized by R129 from Loop-45 via hydrogen bonds and W109 via stacking. **F.** C_{14} makes 3 hydrogen bonds to the peptide backbone of Loop-45. C14 also makes a stacking interaction to N124.

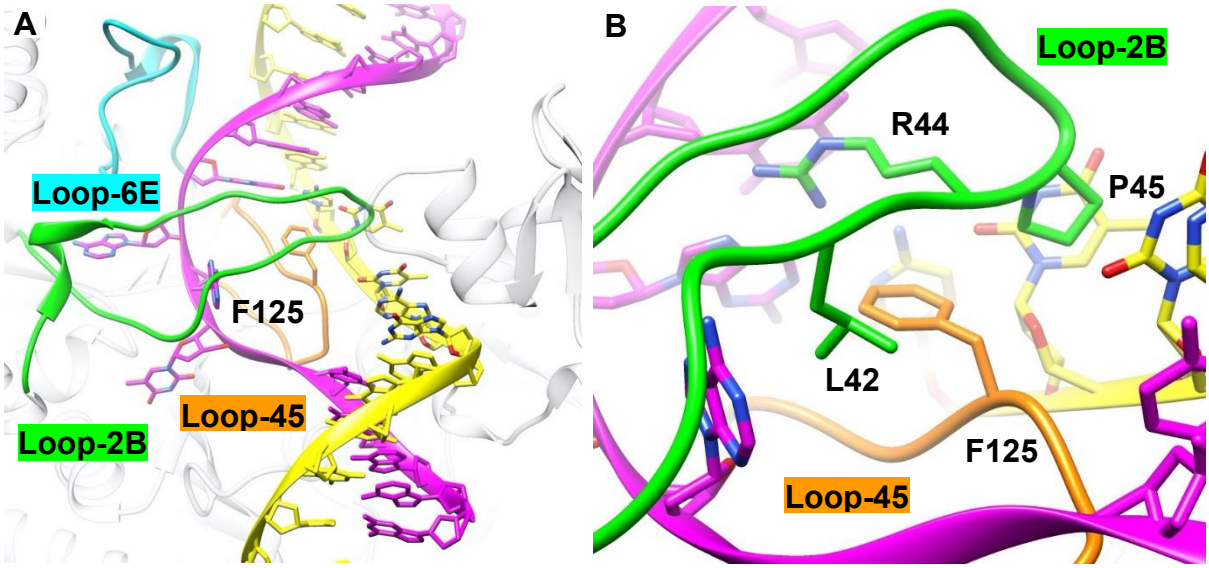


Figure 5.3. F125 from Loop-45 is positioned between the strand separated DNA and does not make base-specific contacts to nucleic acids of the recognition site. A. Loop-45 (orange) approaches the strand-separated DNA from the minor groove and Loop-2B (green) approaches the strand-separated DNA from the major groove. F125 protrudes out of Loop-45 and is positioned between the DNA strands. **B.** A closer view of F125 protruding between the strand-separated interface.

We introduced greater bulkiness at position 125 with the F125W mutant. This mutant binds DNA with smiliar affinity to WT; F125W K_d^{app} = 136.3 ± 15.7, WT K_d^{app} = 149.9 ± 5.4nM (Fig. 5.10). F125W undergoes the initial protein conformational change monitored by Trp fluorescence faster than WT; F125W $k_2 = 20.0s^{-1}$ and $k_{-2} = 8.9s^{-1}$, WT $k_2 = 15.2s^{-1}$ and k_{-2}

= 9.4s^{-1} (Fig. 5.6a, Table 5.1). F125W is able to strand-separate monitored by PydC fluorescence to a lesser extent than WT; F125W k_3 = 0.16s^{-1} and k_{-3} = 0.6s^{-1}, WT k_3 = 0.28s^{-1} and k_{-3} = 1.7s^{-1} (Fig. 5.6b, Table 5.1). F125W is able to methylate the fastest compared to F125L and F125A with a k_4 = 0.04s^{-1} (Fig. 5.6c, Table 5.1). The F125W kinetic model reveals that the formation of the GS'' species is destabilized and slows down the rate of catalysis for this mutant (Fig. 5.6d). Confidence contour 1D FitSpace analysis for F125W shows that the kinetic parameters are well-defined by the data (Fig. 5.11).

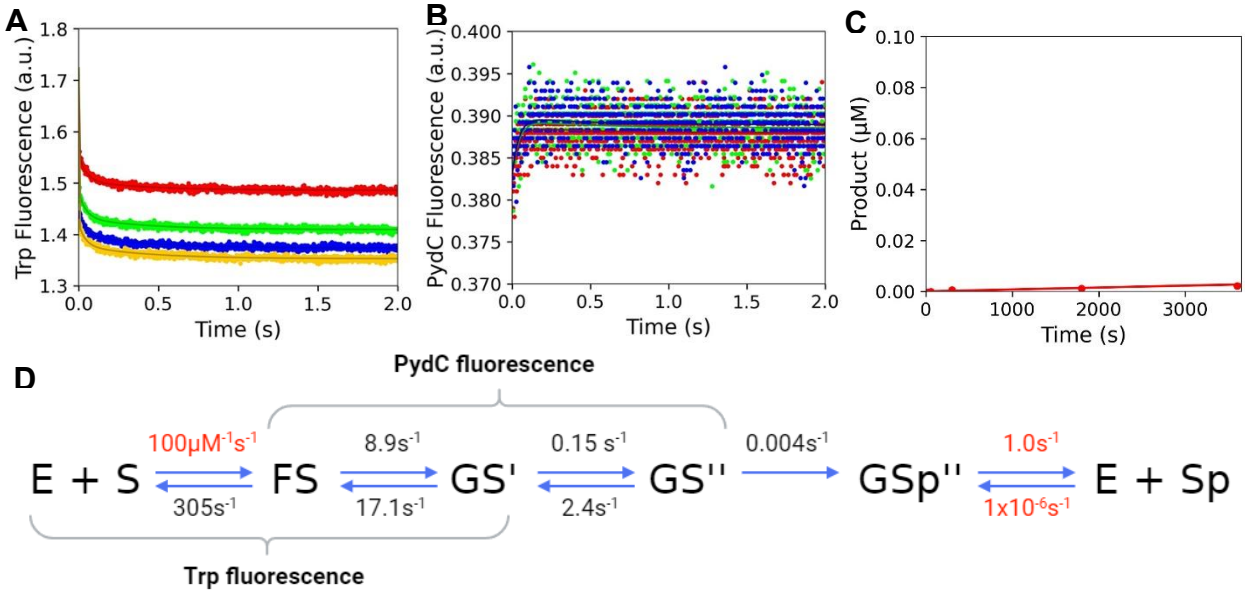


Figure 5.4. F125A does not show PydC fluorescent signal due to destabilization of GS' and GS''. A. F125A Trp kinetics are biphasic. **B.** F125A does not show PydC fluorescence. **C.** F125A shows little to no product turnover**. D.** F125A kinetic model shows the rate constants derived from global fitting and reveals that GS' and GS'' are massively destabilized which precludes detection of the strand-separated intermediate via PydC fluorescence and F125A can not go forward in the pathway to achieve product-turnover.

Loop residues that interact with specific recognition site bases are responsible for generation and stabilization of GS' and GS''

R129 is a highly conserved residue from Loop-45 that makes two hydrogen bonds to T_{13} within the target strand recognition site (Fig.7a). R129A is able to bind DNA with a similar affinity to WT; R129A K_d^{app} = 135.3 ± 7.1, WT K_d^{app} = 149.9 ± 5.4nM (Fig. 5.10). R129A is perturbed approximately two-fold compared to WT in the protein conformational change monitored by Trp fluorescence; R129A $k_2 = 6.0s^{-1}$ and $k_{-2} = 8.9s^{-1}$, WT $k_2 = 15.2s^{-1}$ and $k_{-2} = 9.4s^{-1}$ (Fig. 7b, Table 5.1). The strand-separation event monitored by PydC fluorescence is reduced for R129A relative to WT (R129A $k_3 = 0.05s^{-1}$, WT $k_3 = 0.28s^{-1}$) while k_{-3} is similar to WT (R129A $k_{-3} = 1.5s^{-1}$, WT $k_{-3} = 1.7s^{-1}$) (Fig. 7c, Table 5.1). PydC fluorescence is not observed for R129Adue to the inability to form and stabilize the strand-separated intermediate (GS'') (Fig. 7c), explaining the lack of product formation (Fig. 7d). The enzyme kinetic model for R129A reveals that the GS' intermediate is destabilized, therefore limiting the mutant's ability to go forward in the pathway (Fig. 7e). Confidence contour 1D FitSpace analysis for R129A shows that the kinetic parameters are well-defined by the data (Fig. 5.11).

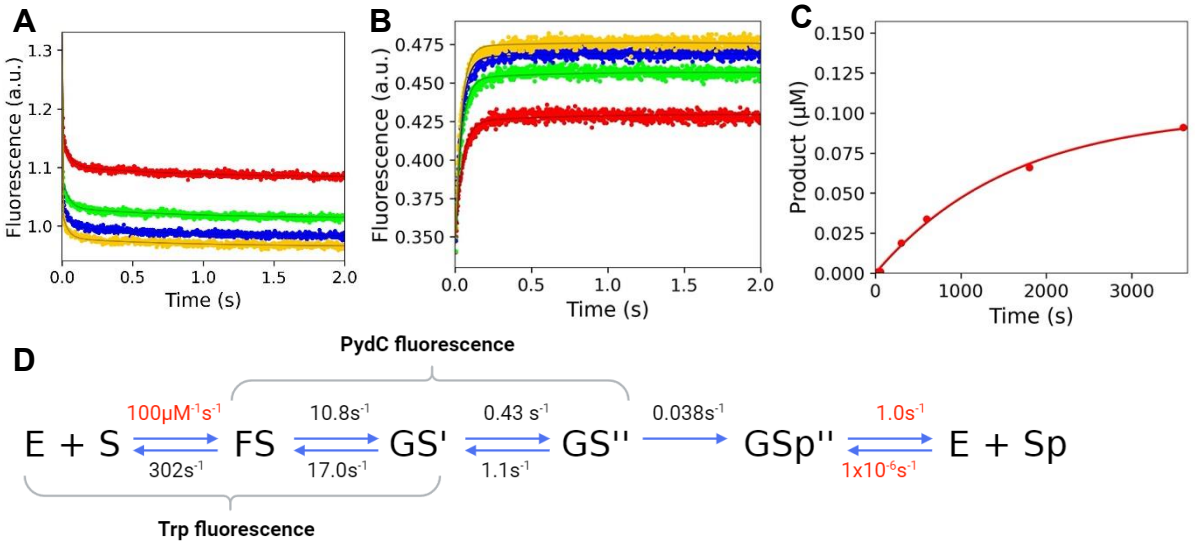


Figure 5.5. F125L is perturbed in which GS' is destabilized which decreases kcat. A. Trp fluorescence by F125L is biphasic. **B.** PydC fluorescence for F125L is biphasic. **C.** F125L can methylate DNA but at a slower rate than WT. **D.** The enzyme kinetic model for F125L shows the rate constants derived from global fitting, where K2 is now favoring the reverse direction therefore limiting F125L from going forward in the pathway.

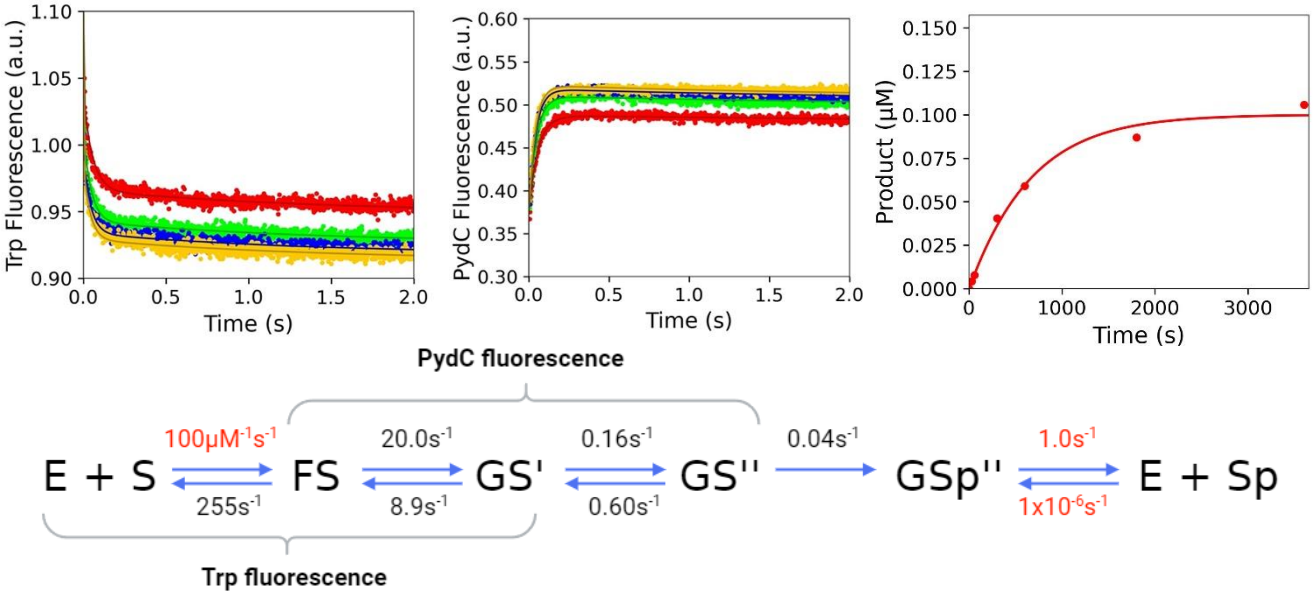


Figure 5.6. F125W has faster protein conformational change but slower DNA strand-separation. A. F125W Trp kinetics are biphasic. **B.** F125W PydC kinetics are monophasic. **C.** Product formation for F125W is faster than WT. **D.** The F125W kinetic model shows that K3 is perturbed, thus making this mutant have slower chemistry.

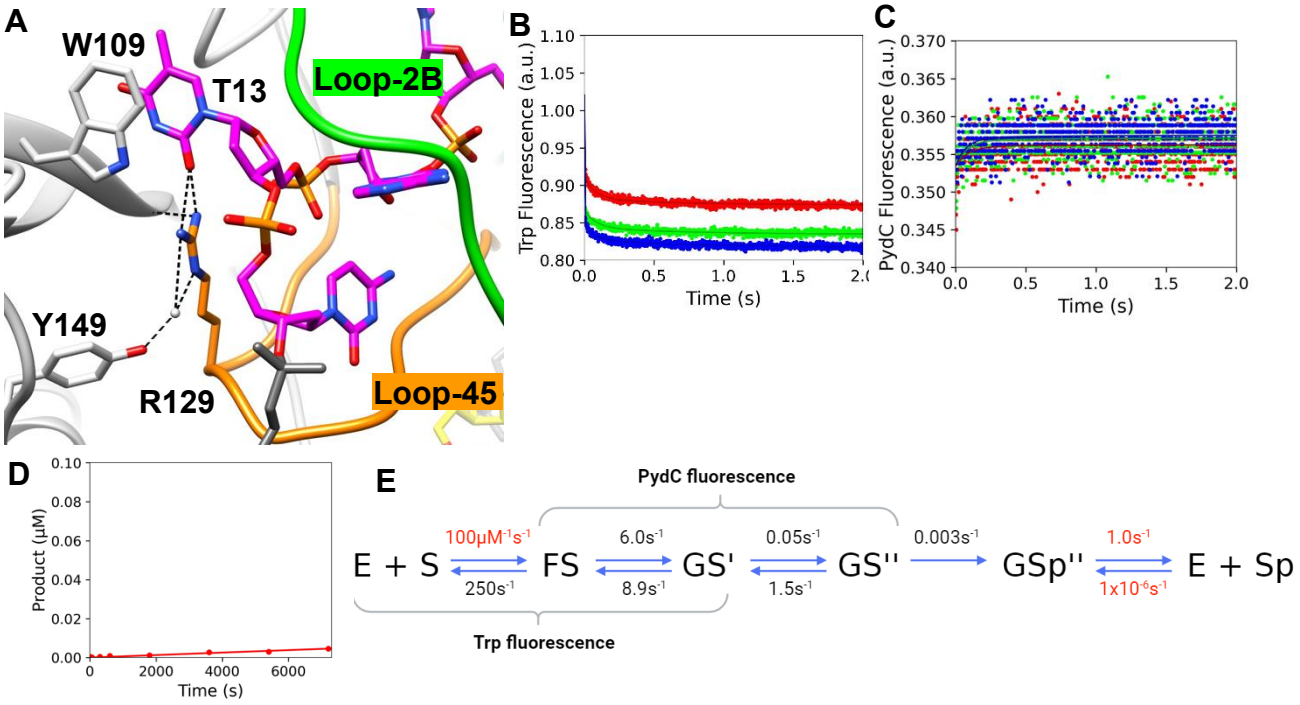


Figure 5.7. R129A is unable to form or stabilize the strand-separated intermediate resulting in lack of methylation. A. R129 from Loop-45 makes two hydrogen bonds to T_{13} of the recognition sequence. **B.**Trp kinetics **C.** PydC kinetics **D.** Methylation **E.** R129A enzyme kinetic model with rate constants derived from global fitting.

N124 is a highly conserved residue from Loop-45 that does not make any hydrogen bonds to nucleic acids but recognizes C_{14} of the recognition site via stacking interactions (Fig.8a). N124A is able to bind DNA with a similar affinity as WT; N124A K_d^{app} = 137.5 ± 17.8nM, WT K_d^{app} = 149.9 ± 5.4nM (SI Fig.1). N124A has perturbed protein conformational forward and reverse kinetics monitored by Trp fluorescence; N124A $k_2 = 6.4s^{-1}$ and $k_{-2} = 8.7s^{-1}$, WT $k_2 = 15.2s^{-1}$ and $k_{-2} = 9.4s^{-1}$ (Fig. 8b, Table 5.1). The strand-separation step monitored by PydC fluorescnce is also perturbed for N124A; N124A $k_3 = 0.10s^{-1}$ and $k_{-3} = 2.1s^{-1}$, WT $k_3 = 0.28s^{-1}$ and $k_{-3} = 1.7s^{-1}$ (Fig. 8c, Table 5.1). Product formation is minimal for N124A with $k_4 = 0.015s^{-1}$ (Fig. 8d, Table 5.1). Thus, the N124A mutant intermediate GS' is destabilized and

disfavors going forward in the pathway as seen in the mutant kinetic model (Fig. 8e). Confidence contour 1D FitSpace analysis for N124A shows that the kinetic parameters are well-defined by the data (Fig. 5.11).

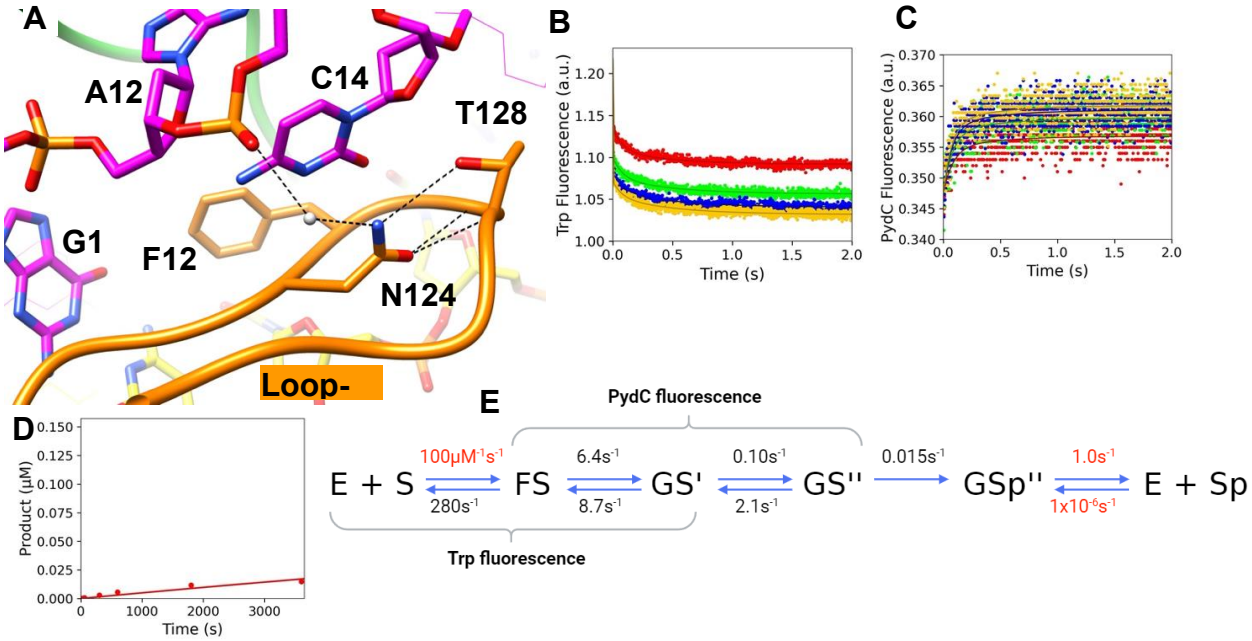


Figure 5.8. N124A destabilizes strand separation which limits catalysis. N124 from Loop-45 makes a stacking interaction to C_{14} and is perturbed in its ability to strand separate. **A.** Trp kinetics. **B.** PydC kinetics. **C.** Methylation. **D.** Structural image of N124. N124 makes a stacking interaction to C_{14} of the target strand recognition site, and makes a water-mediated hydrogen bond to the phosphate backbone of T_{13}. N124 also makes interloop hydrogen bonds to T128 of Loop-45 and the peptide backbone of Loop-45. **E.** Confidence contour fitspace. **F.** N124A kinetic model.

R44 is a highly conserved residue from Loop-2B that makes hydrogen bonds to G_{10} of the target strand which is the only base of the recognition site that maintains Watson-Crick hydrogen bonds during strand-separation (Fig. 9a). R44A binds DNA with a similar affinity as WT; R44A K_d^{app} = 144.5 ± 9.5 nM, WT K_d^{app} = 149.9 ± 5.4nM (Fig. 5.10). R44A has similar protein conformational forward kinetics monitored by Trp fluorescence; R44A k_2 = 15.4s^{-1} ,

WT $k_2 = 15.2s^{-1}$ (Fig. 9b). R44A has increased k_{-2} indicating destabilization of GS'; R44A $k_{-2} = 25.4s^{-1}$, WT $k_{-2} = 9.4s^{-1}$ (Fig. 9b). R44A can strand separate but with much slower forward rate constants and the major perturbation is in the destabilization of the GS'' complex; R44A $k_3 = 0.04s^{-1}$ and $k_{-3} = 1.2s^{-1}$, WT $k_3 = 0.28s^{-1}$ and $k_{-3} = 1.7s^{-1}$ (Fig. 9c). R44A is unable to methylate DNA (Fig. 9d). The enzyme kinetic model for R44A shows that the destabilization of GS' and GS'' are the steps that massively disfavor catalysis (Fig. 9e). Confidence contour 1D FitSpace analysis for R44A shows that the kinetic parameters are well-defined by the data (Fig. 5.11).

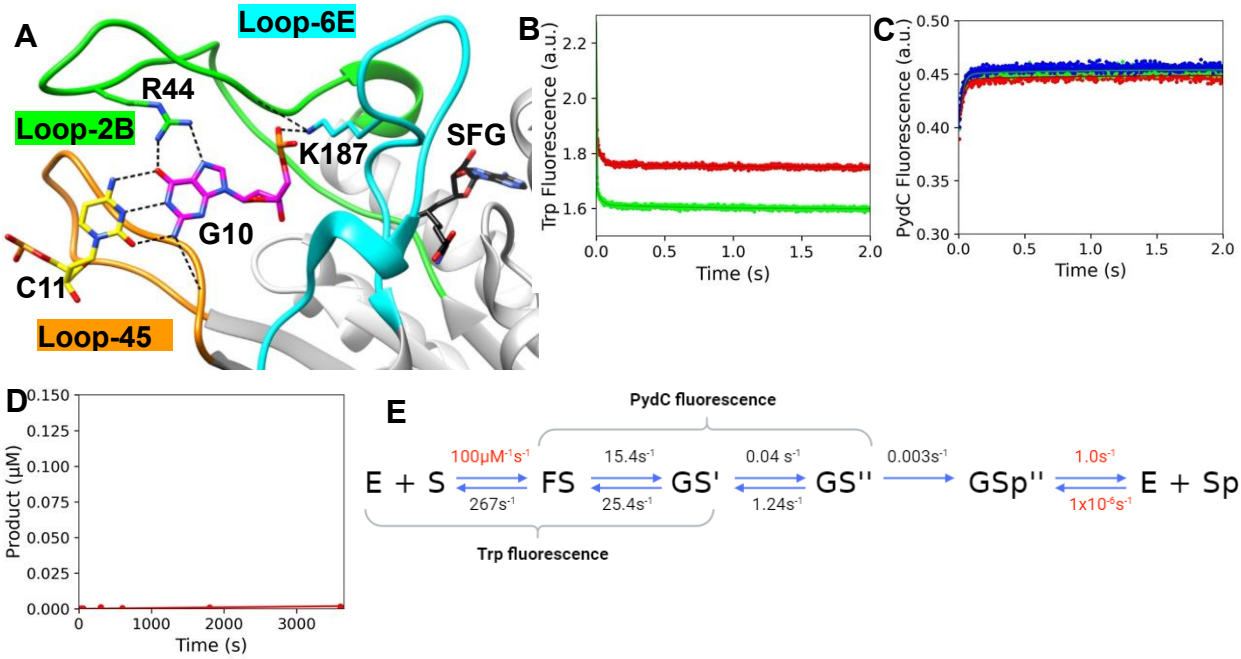


Figure 5.9. R44A destabilizes the strand-separated intermediate which disables methylation. A. R44 makes hydrogen bonds to G_{10} of the target strand which is the only base of the recognition site that maintains Watson-Crick hydrogen bonds during strand-separation. R44A disables these hydrogen bonds resulting in lack of recognition of G_{10}. **B.** R44A is able to bind dsDNA and undergo the protein isomerization event with similar rates to WT (k_1, k_1, k_2). However, k_{-2} increases compared to WT therefore destabilizing the GS' complex. **C.** R44A is able to strand separate but with much slower forward rate constants and the major difference

is in the destabilization of the GS'' complex. Note the amplitude difference of R44A PydC fluorescence compared to WT. **D.** R44A is unable to methylate dsDNA. **E.** The enzyme kinetic model for R44A shows that the destabilization of GS' is the step that massively disfavors catalysis for this mutant.

Enzyme	k_1 ($\mu M^{-1}s^{-1}$)	k_{-1} (s^{-1})	k_2 (s^{-1})	k_{-2} (s^{-1})	k_3 (s^{-1})	k_{-3} (s^{-1})	k_4 (s^{-1})	k_{-4} (s^{-1})	k_5 (s^{-1})	k_{-5} (s^{-1})
WT	100	281 (247-327)	15.2 (13.4-17.5)	9.4 (8.3-10.5)	0.28 (0.21-0.35)	1.7 (1.4-2.0)	0.03 (0.025-0.033)	0	1.0	$1x10^{-6}$
F125A	100	305 (296-314)	8.9 (7.3-10.1)	17.1 (14.2-19.0)	0.15 (0.12-0.17)	2.4 (2.0-2.8)	0.0035 (0.0033-0.0038)	0	1.0	$1x10^{-6}$
F125L	100	302 (296-308)	10.8 (10.2-11.4)	17.0 (16.2-17.9)	0.43 (0.32-0.59)	1.1 (0.71-1.55)	0.038 (0.033-0.044)	0	1.0	$1x10^{-6}$
F125W	100	255 (244-266)	20.0 (19.2-21.0)	8.9 (8.5-9.2)	0.16 (0.13-0.20)	0.6 (0.28-0.94)	0.04 (0.024-0.051)	0	1.0	$1x10^{-6}$
R129A	100	250 (238-266)	6.0 (4.5-8.3)	8.9 (7.1-11.6)	0.05 (0.04-0.08)	1.5 (0.96-2.35)	0.003 (0.0023-0.0033)	0	1.0	$1x10^{-6}$
N124A	100	280 (275-287)	6.4 (5.8-7.1)	8.7 (7.8-9.6)	0.10 (0.09-0.12)	2.1 (1.81-2.38)	0.015 (0.014-0.016)	0	1.0	$1x10^{-6}$
R44A	100	267 (265-270)	15.4 (14.2-17.4)	25.4 (23.2-29.0)	0.04 (0.03-0.07)	1.2 (0.63-2.17)	0.003 (0.0026-0.0038)	0	1.0	$1x10^{-6}$

Table 5.1 Summary of rate constants derived from global fitting for loop mutants. Rate constants derived from global fitting. Rates in red were locked during the continuous simulation while rates in black were allowed to float freely without constraints.

Enzyme	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($\mu M^{-1}s^{-1}$)
WT	0.003 ± 0.0009	0.99 ± 0.18	0.003 ± 0.001
R129A	$4x10^{-5}$ ± $2.6x10^{-5}$	1.47 ± 0.25	$3x10^{-5}$ ± $1.8x10^{-5}$
N124A	$3x10^{-4}$ ± $5.9x10^{-5}$	1.58 ± 0.10	$2x10^{-4}$ ± $3.9x10^{-5}$
F125A	$8x10^{-5}$ ± $2.5x10^{-5}$	1.96 ± 0.16	$4x10^{-5}$ ± $1.3x10^{-5}$
F125L	0.005 ± 0.003	1.61 ± 0.14	0.003 ± 0.002
F125W	0.006 ± 0.003	0.67 ± 0.07	0.009 ± 0.005
R44A	$4x10^{-5}$ ± $3x10^{-5}$	1.64 ± 0.10	$2x10^{-5}$ ± $1.9x10^{-5}$

Table 5.2. kcat, Km, kcat/Km. Steady-state parameters were calculated from the rate constants in Table 5.1. The variants R129A, F125A, and R44A have approximately 100-fold

decrease in kcat compared to WT. N124A has 10-fold decrease in kcat compared to WT. The variants have similar Km compared to WT. The catalytic efficiency (kcat/Km) for F125L may be similar or worse than WT, while catalytic efficiency for F125W is equal to or worse than WT. The other variants have 10-100 fold decreased catalytic efficiency relative to WT.

Enzyme	K1	K2	K3	K4
WT	0.35	1.6	0.17	0.029
R129A	0.4	0.67	0.03	0.003
N124A	0.36	0.74	0.05	0.015
F125A	0.33	0.52	0.06	0.004
F125L	0.33	0.64	0.39	0.038
F125W	0.39	2.25	0.27	0.04
R44A	0.37	0.61	0.03	0.003

Table 5.3. Equilibrium constants were derived from the rate constants in Table 5.1.

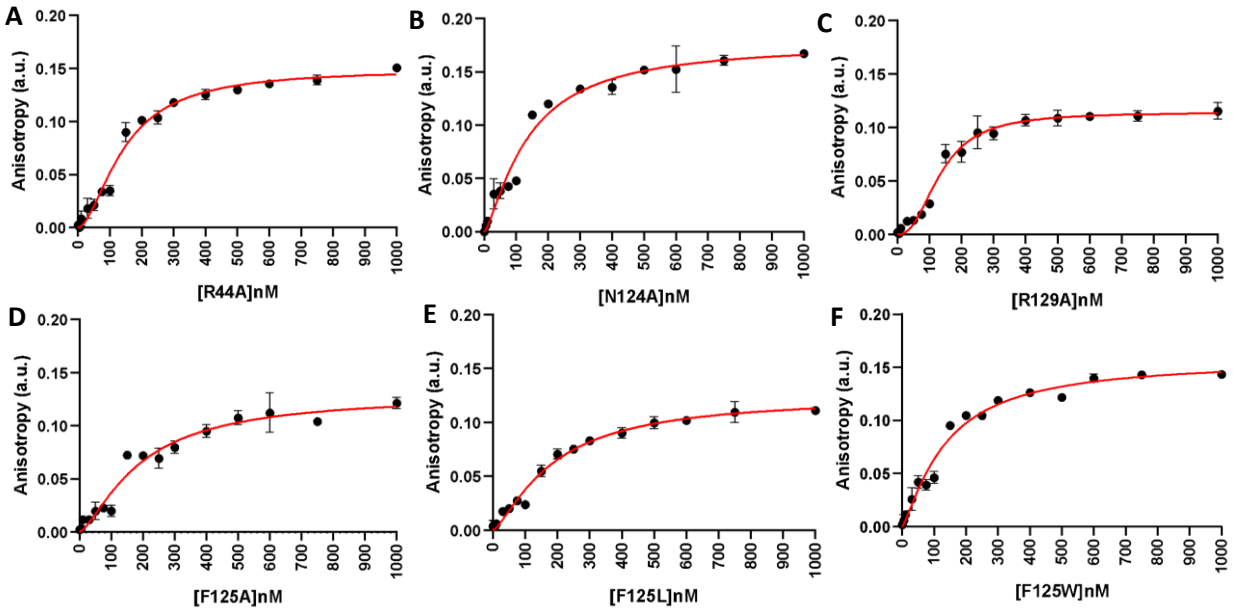


Figure 5.10. Fluorescence anisotropy to determine the K_d^{app} of loop mutants. Anisotropy conditions were 10nM FAM-tagged DNA, 15µM SAH, and CcrM monomer concentration (1, 10, 30, 50, 75, 100, 150, 200, 250, 300, 400, 500, 600, 750, 1000nM). **A.** R44A K_d^{app} = 144.5 ± 9.5, h=1.8. **B.** N124A K_d^{app} = 137.5 ± 17.8, h = 1.4. **C.** R129A K_d^{app} = 135.3 ±7.1, h = 2.4. **D.** F125A K_d^{app} = 190.8 ±30.4, h = 1.4. **E.** F125L K_d^{app} = 185.3 ±19.1, h = 1.4. **F.**

F125W K_d^{app} = 136.3 ±15.7, h = 1.3. Data were fit in Graphpad Prism 10.0.2 to a specific binding with Hill slope model where Anisotropy = $Anisotropy_{max}$*$[CcrM]^h$ /(K_d^h + $[CcrM]^h$.

DNA name and sequence	C1	5′-TCACTGTACTCTGACTCGCCTGACATGAC-3′ 3′-AGTGACATGAGACTG**M**GCGGACTGTACTG-5′
	P1	5′-TCACTGTACTCTGA**P**TCGCCTGACATGAC-3′ 3′-AGTGACATGAGACTG**M**GCGGACTGTACTG-5′
	NC3	5′-TCACTGTACTCTCACTCGCCTGACATGAC-3′ 3′-AGTGACATGAGAGTG**M**GCGGACTGTACTG-5′
	NC4	5′-TCACTGTACTCTGACTAGCCTGACATGAC-3′ 3′-AGTGACATGAGACTG**M**TCGGACTGTACTG-5′
	NC5	5′-TCACTGTACTCTGACTGGCCTGACATGAC-3′ 3′-AGTGACATGAGACTG**M**CCGGACTGTACTG-5′
	NC6	5′-TCACTGTACTCTGACTTGCCTGACATGAC-3′ 3′-AGTGACATGAGACTG**M**ACGGACTGTACTG-5′
	NC7	5′-TCACTGTACTCTAACTCGCCTGACATGAC-3′ 3′-AGTGACATGAGATTG**M**GCGGACTGTACTG-5′
	NC8	5′-TCACTGTACTCTTACTCGCCTGACATGAC-3′ 3′-AGTGACATGAGAATG**M**GCGGACTGTACTG-5′

Table 5.4. Names and sequences of 29mer DNA substrates. C1 is cognate DNA, P1 is cognate DNA with PydC inserted at the N-position of the recognition site, NC3, NC4, NC5, NC6, NC7, and NC8 are different noncognate DNA sequences. P = Pyrrolo-dC, M = N6-methyl adenine, red bases = noncognate substitutions.

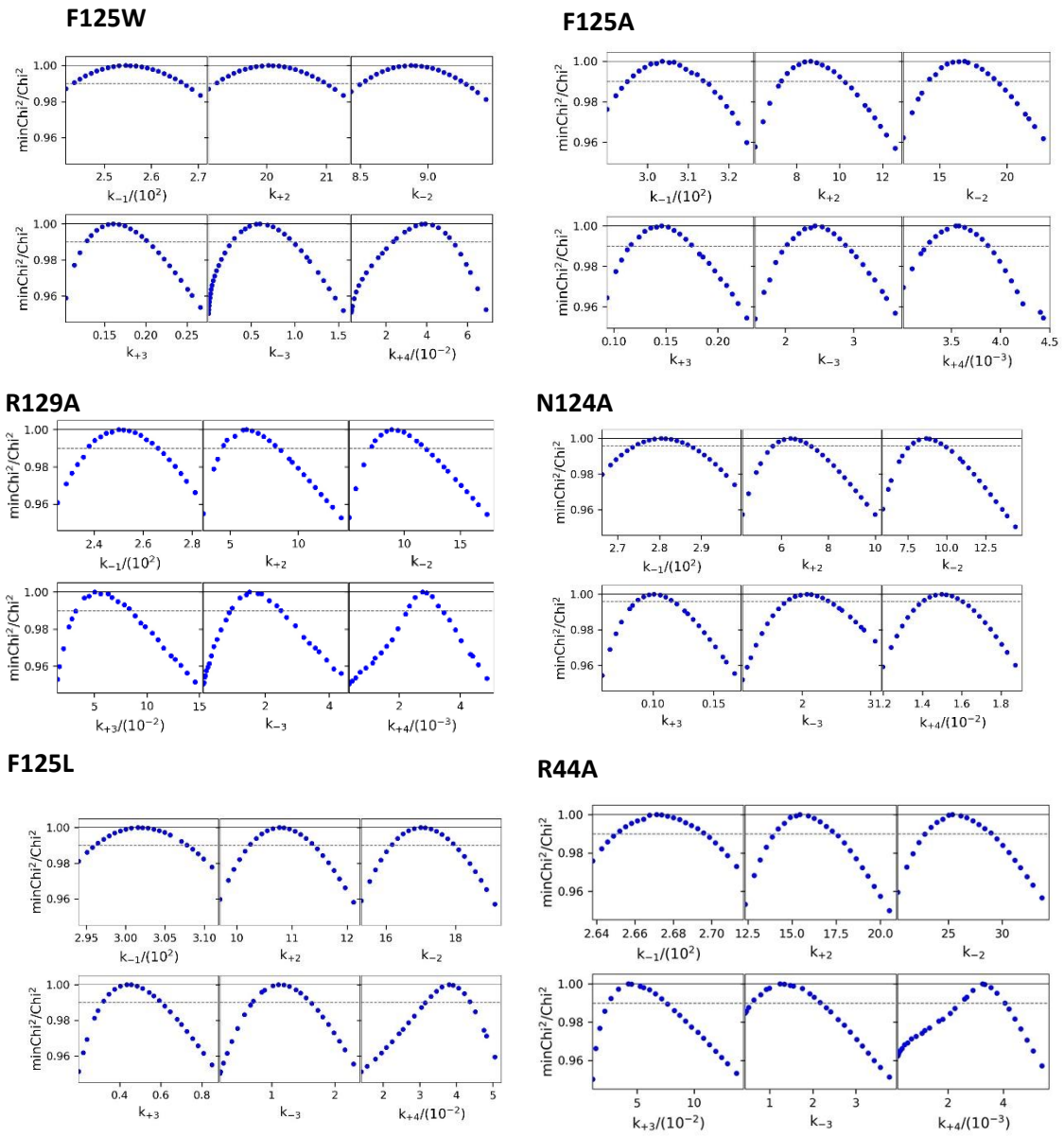


Figure 5.11. Confidence contour Fitspace analysis from global fitting for mutants F125W, F125L, F125A, R124A, and R129A. The data represent the 1D Fitspace calculated for each rate constant. The dashed line establishes the 95% confidence interval at the 0.99 Chi^2 threshold.

V: Discussion

The DNA strand-separation mechanism used by CcrM is essential for its extreme sequence discrimination[4.1,3.0], and we sought to better understand this process. This has relevance both for our understanding of a new DNA recognition mechanism, and how the kinetic partitioning of reaction intermediates can contribute to enzyme specificity. A mechanistic understanding of how enzymes induce conformational changes in their substrates is particularly challenging since the obligate intermediates are transient and require sophisticated probes that allow kinetic assignments. Moreover, there are now numerous examples showing that such transitions can be key determinants of enzyme specificity, and rigorous global fitting has been essential to resolving these pathways[5.17,5.18,5.19,5.20]. Understanding such mechanisms can lead to the design of highly specific cell biology tools and drugs[5.21].

We utilized stopped-flow fluorescence to quantify the rates of the changes in protein conformation and DNA strand separation, as well as radiochemical methods to monitor methyl-transfer. Global fitting of all the data resolved the rate constants that were assigned to transitions between known intermediates. Specifically, we sought to understand the roles that Loop-2B and Loop-45 played in regulating DNA strand separation. Mutational analysis of R44A, N124A, R129A, F125A, F125L, and F125W allowed us to interrogate the role of these loops in strand separation.

Loop-2B and Loop-45 are positioned within the strand-separated DNA, make interactions with target-strand bases, and appear to be stabilized by interactions with other protein residues (Fig. 5.2)[2.5]. The structural positioning of these loops supports that they are important motifs for regulating DNA strand separation, and we showed previously that strand-

separation was tightly coupled with substrate discrimination[4.1]. Our criteria for mutational analysis were high conservation and structural implication in strand-separation. R44, N124, and R129 make specific interactions to target strand bases in the strand-separated state (Fig. 5.7, 5.8, and 5.9). We previously interrogated CcrM's DNA specificity by making noncognate nucleic acid substitutions in the recognition site[4.1]. Here our focus is on protein residues and motifs that could contribute to recognition (R44A, N124A, and R129A) while F125 was selected because it does not make base-specific interactions and therefore could play a role in generating or stabilizing strand separation, or both.

Mutational analysis of highly conserved F125 revealed interesting functional responsibilities of this bulky residue. F125L destabilizes the GS' intermediate (Fig. 5.4) and our kinetic model suggests that this is responsible for the decreased rate of catalysis (Fig. 5.4). Noteworthy, the equilibrium constant K2 favors the reverse direction in the pathway, compared to WT which favors going forward (F125L K2 = 0.64, WT K2 = 1.6, Table 5.3). Thus, the global fitting suggests this bulky residue contributes to both generation and stabilization of DNA strand-separation. Consistent with this dual role, the F125A mutant reduced hydrophobicity and size even further and the resultant equilibrium constant K2 (0.52) further favors the reverse direction (Table 5.3). WT K3 favors the reverse direction, but F125A K3 further favors going in the reverse direction (Table 5.3). Analysis of F125A supports our working hypothesis that the bulky loop residue F125 is essential for both generation and stabilization of strand separation.

We also interrogated the bulkiness of position 125 in Loop-45 by adding bulk to this position with the F125W mutant. The F125W kinetic model reveals that the formation of the

GS'' species is destabilized and slows down the rate of catalysis for this mutant (Fig. 5.6). F125W K2 (2.25) is greater than WT (1.6), indicating that the added bulk at this position allows for easier formation and stabilization of GS'. However, the F125W GS' intermediate is less able to go forward and k_3 is decreased appropriately 3-fold (Table 5.3). Analysis of F125W likely suggests that the bulkiness contributes to generating partial strand-separation, and that the size of this bulk is essential for separating the strands to the correct position so that other residues can make base-specific interactions that are essential for going forward to methylation. In summary, global fitting resolves that K2 is the most perturbed by mutations of F125, which suggests that bulky residues in Loop-45 are responsible for generating and stabilizing DNA strand separation. Furthermore, a bulky loop residue (F125) that does not appear to make any base-specific interactions can contribute to recognition by stabilizing the GS' intermediate that precedes the recognition state, which is predominantly GS'' in the CcrM kinetic model.

On the other hand, we also wanted to understand how conserved loop residues that do make base-specific interactions contribute to strand-separation and substrate recognition. We approached this by designing loop mutants that make base-specific contacts to target strand bases. R129 in Loop-45 makes two hydrogen bonds to T_{13} within the target strand recognition site (Fig. 5.7a). PydC fluorescence changes, which track DNA strand separation, are not observed for R129A due to the inability to form and stabilize the strand-separated intermediate (GS''). The result of the destabilized GS'' is minimal product formation for R129A (Fig. 5.7). R129A is unable to recognize T_{13}, thus perturbing the protein's ability to form and stabilize the strand-separated state (GS''). The enzyme kinetic model for R129A reveals that the GS' and

GS'' intermediates are destabilized, therefore limiting the mutant's ability to go forward in the pathway (Fig. 5.7). Interestingly, the recognition interaction appears to influence the formation and stabilization of the GS' and GS'' states, indicating that recognition interactions contribute to the formation of the strand separated state in addition to stabilization.

N124 is another highly conserved residue from Loop-45 that does not make any hydrogen bonds to specific nucleic acids but contributes to recognition of C_{14} via stacking interactions (Fig. 5.8a). N124A results in the destabilization of the intermediate GS' and disfavors going forward in the pathway as seen in the mutant kinetic model (Fig. 5.8). The destabilization of GS' and GS'' by N124A supports our hypothesis that recognition interactions are essential to generate and stabilize strand-separation which is essential for catalysis.

Interestingly, R44A shows a small PydC signal indicating that this mutant is capable of strand separation (Fig. 5.9C). R44 stands out amongst the residues selected for mutation because it recognizes G_{10} of the 5'GANTC'3 recognition site. G_{10} is the only base within the recognition site where Watson-Crick base-pairing is maintained in the strand-separated GS'' state. R44A has similar strand separation forward kinetics as WT (R44A $k_2 = 15.4s^{-1}$, WT $k_2 = 15.2s^{-1}$) but the GS' state is destabilized as seen in the reverse kinetics (R44A $k_{-2} = 25.4s^{-1}$, WT $k_{-2} = 9.4s^{-1}$) (Table 5.1). Unlike the other loop residue's recognition interactions, R44's recognition of G_{10} does not contribute to the generation of GS', but this recognition interaction is required for stabilization of GS' and generation of GS''. Thus, DNA strand separation can begin prior to R44's recognition of G_{10}. In agreement with the other loop mutants that disable recognition, R44A's K2 is 2-3 fold reduced from WT (R44A K2 = 0.61, WT K2 = 1.6) and 5-6 fold reduced in K3 compared to WT (R44A K3 = 0.03, WT K3 = 0.17) (Table 5.3). Analysis

of R44A supports our model in which recognition interactions are essential for generation and stabilization of the strand-separated state which is essential for catalysis.

Together, the results with N124A, R129A, and R44A which disable base-specific recognition between loop residues to target strand bases, reveal that these interactions contribute to formation and stabilization of the strand-separted intermediate. GS' and GS'' are the most destabilized for N124A, R129A, and R44A suggesting that specificity occurs at these transitions (K2 and K3). This further supports our previous understanding of this system where we showed that DNA strand separation is tightly coupled with substrate discrimination[4.1]. In our previous work, we approached this relationship from the perspecitive of the DNA by building a model with WT CcrM and noncognate DNA[4.1]. Here, we approached this from the perspective of the protein where we built models with mutant CcrM and cognate DNA. In both cases, the substrate recognition at a single base is removed and the resulting step in the enzyme kinetic model are the strand-separation steps in the forward and reverse directions, constituting K2 and K3. The perturbations of DNA strand-separation are also correlated with methylation efficency.

In summary, the motifs and mechanisms that govern DNA strand-separation and substrate fidelity are relevant to other proteins such as CRISPR-Cas9 and RNA polymerase sigma factor[5.21,5.22]. The mechanisms described for CcrM, CRISPR, and sigma factor, in which an enzyme strand-separates double-stranded DNA, represent a new DNA recognition mechanism. In contrast to CRISPR and sigma factor, CcrM's DNA strand-separation is tightly coupled to the extreme sequence selectivity, providing a compelling opportunity to study this newly described recognition strategy[5.21]. The results presented here for CcrM can be useful for

understanding substrate discrimination in other systems such as CRISPR/Cas9 where gene editing tools require overcoming barriers of selectivity between enzymes and DNA substrates.

V: References

5.1. Rohs,R., Jin,X., West,S.M., Joshi,R., Honig,B., and Mann,R.S. (2010) Origins of specificity in protein-DNA recognition. *The Annual Review of Biochemistry*, 79, 233-269.

5.2. Klug,A. and Schwabe,J.W.R. (1995) Zing fingers. *The FASEB Journal*, 9, 8, 597-604.

5.3. Brennan,R.G. and Matthews,B.W. (1989) The Helix-Turn-Helix DNA Binding Motif. *The Journal of Biological Chemistry*, 264,4, 1903-1906.

5.4. Ohlendorf,D.H., Anderson,W.F., Fisher,R.G., Takeda,Y., and Mathews,B,W. (1982) The molecular basis of DNA-protein recognition inferred from the structure of cro repressor. *Nature*, 298, 718-723.

5.5. Suzuki,M., Yagi,N., and Gerstein,M. (1995) DNA recognition and superstructure formation by helix-turn-helix proteins. *Protein Engineering, Design, and Selection*, 8, 4, 329-338.

5.6. Wright,P.E., Lee,B.M., Laity,J.H. (2001) Zinc finger proteins: new insights into structural and functional diversity. *Current Opinion in Structural Biology*, 11, 39-46.

5.7. Nasrullah, Hussain,A., Ahmed,S,. Rasool,S., and Shah,A.J. (2022) DNA methylation across the tree of life, from micro to macro-organisms. *Bioengineered*, 13, 1, 1666-1685.

5.8. Mattei.A.L., Bailly,N., and Meissner,A. (2022) DNA methylation: a historical perspective. *Trends in Genetics*, 38, 7, 676-707.

5.9. Cheng,X. (1995) DNA modification by methyltransferases.*Structural Biology*, 5,4-10.

5.10. Estabrook,R.A. and Reich,N. (2006) Observing an Induced-fit mechanism during sequence-specific DNA methylation. *Journal of Biological Chemistry*, 281, 48, 37205-37214.

5.11. Youngblood.B., Buller,F., Reich,N.O. (2006) Determinants of sequence-specific DNA methylation: Target recognition and catalysis are coupled in M.HhaI. *Biochemistry*, 45, 51, 15563-15572.

5.12. Matje,D.M., Coughlin,D.F., Connolly,B.A., Dahlquist,F.W., and Reich,N.O. (2011) Determinants of precatalytic conformational transitions in the DNA cytosine methyltransferase M.HhaI. *Biochemistry*, 50, 9, 1465-1473.

5.13. Zhang,Z., Lu,R., Wang,P., Yu,Y., Chen,D., Gao,L., Liu,S., Ji,D., Rothbart,S.B., Wang,Y., Wang,G., and Song,J. (2018) Structural basis for DNMT3A-mediated de novo DNA methylation. *Nature*, 554, 7692, 387-391.

5.14. Anteneh,H., Fang,J., and Song,J. (2020) Structural basis for impairment of DNA methylation by the DNMT3A R882H mutation. *Nature Communications*, 11, 2294, 1-12.

5.15. Zhang,Z., Wang,P., Yu,Y., Chen.D., Gao,L., Liu,S., Ji.D., Rothbart,S.B., Wang,Y., Wang,G.G., and Song,J. (2018) Structural basis for DNMT3A-mediated *de novo* DNA methylation. *Nature*, 554, 387-391.

5.16. Zhao,H., Yu,J., Weng,G., Yu,J., Wang,E., Gao,J., Liu,H., Hou,T., Wang,Z., and Kang,Y. (2022) Structural view on the role of the TRD loop in regulating DNMT3A activity: a molecular dynamics study. *Royal Society of Chemistry*, 24,15791-15801.

5.17. Dangerfield,T.L. and Johnson,K.A. (2021) Conformatinal dynamics during high-fidelity DNA replicaiton and translocation defined using a DNA polymerase with a fluorescent amino acid. *J. Biol. Chem.,* 296, 100143.

5.18. Kellinger,M.W. and Johnson,K.A. (2010) Nucleotide-dependent conformational change governs specificity and analog discrimination by HIV reverse transcriptase. *PNAS,* 107, 7734-7739.

5.19. Kirmizialtin,S., Nguyen,V., Johnson,K.A. and Elber,R. (2012) How conformational dynamics of DNA polymerase select correct substrates: experiments and simulations. *Structure,* 20, 618-627.

5.20. Gong,S., Kirmizialtin,S., Chang,A., Mayfield,J.E., Zhang,Y.J. and Johnson,K.A. (2021) Kinetic and thermodynamic analysis defines roles for two metal ions in DNA polymerase specificity and catalysis. *J Biol Chem,* 296, 100184.

5.21. Liu,M., Gong,S., Yu,H., Jung,K., Johnson,K.A. and Taylor,D.W. (2020) Engineered CRISPR/Cas9 enzymes improve discrimination by slowing DNA cleavage to allow release of off-target DNA. *Nat. Commun.,* 11, 3576.

5.22. Bravo,J.P.K., Liu,M., Hibshman,G.N., Dangerfield,T.L., Jung,K., McCool,R.S., Johnson,K.A. and Taylor,D.W. (2022) Structural basis for mismatch surveillance by CRISPR-Cas9. *Nature*, 603, 343–347.

Chaper VI: Target adenine base flipping follows DNA strand separation and is the rate-determining step for methylation and catalysis CcrM

VI: Abstract

The Cell-cycle regulated Methyltransferase (CcrM) from *Caulobacter crescentus* stands out for its remarkable specificity and distinct mechanism employing an atypical DNA recognition method involving strand-separation. This study uncovers the fundamental steps governing CcrM's catalysis, in which DNA strand separation leads to base-flipping (docking at the active site) which is followed by fast methylation and DNA product release. Employing a fluorescent adenine analog (6MAP) at the target position to monitor base flipping, we have expanded our previous model of CcrM catalysis and used global data fitting to define intrinsic rate constants for each step leading to DNA unwinding and base flipping (docking), followed by fast methylation and product DNA release. DNA unwinding and base docking are the primary specificity-determining steps. A W57F mutant displayed an unaltered rate of base flipping as monitored by 6MAP fluorescence but greatly reduced rate of methylation, showing that base-flipping and methylation can be uncoupled. In addition, single-stranded DNA bypasses the DNA unwinding step, while rates of base flipping measured by 6MAP fluorescence and DNA methylation are similar to dsDNA. This study highlights the pivotal role of target-adenine positioning as the rate-limiting step in catalysis.

VI: Introduction

DNA methyltransferases typically rely on a base-flipping mechanism for catalysis. Base-flipping is a well-characterized mechanism by which the target base is flipped out of the

DNA helix via rotation of phosphate bonds and Watson-Crick base-melting[6.1,6.2,6.3,3.1,4.20,6.4,6.8]. Upon base-flipping, enzymes obtain stereochemical access to the target base for DNA modification. Co-crystal structures have resolved structures of base-flipping for several DNA methyltransferases[6.8,6.9,6.10,6.11,6.12] and fluorescent nucleic acid analogs have monitored the kinetics of target base-flipping to elucidate a variety of mechanisms[6.34.20,6.4,6.8]. 2-Aminopurine (2AP), which is a fluorescent analog of adenosine, has been extensively used as a probe for the well-characterized mechanism of base-flipping for DNA methyltransferases[6.3,4.20,6.4,6.8].

HhaI C^5-cytosine bacterial DNA methyltransferase was the first methyltransferase to reveal a base-flipping mechanism and has been extensively studied via various co-crystal structures, mechanistic studies, and mutational analyses[3.3,6.8,6.9,6.16,5.11,6.17,6.18,5.12]. HhaI recognizes 5'GCGC'3 and flips the internal 5'-cytosine into an extrahelical position, which is then stabilized by a catalytic loop closure, followed by methylation. HhaI base flipping kinetics are $2s^{-1}$ and base flipping is faster than methyl-transfer[5.12]. HhaI methyl-transfer ($k_{meth} = 0.26s^{-1}$) is faster than turnover ($k_{cat} = 0.04s^{-1}$) and product-release ($k_{off} = 0.04s^{-1}$) is rate-determining for catalysis ($k_{cat} = 0.04s^{-1}$)[6.16,6.17,5.12]. HhaI is highly discriminating (9,000-80,000-fold) against noncognate DNA substrates[5.11] and base-recognition begins prior to and accelerates target base flipping[5.12,6.2]. HhaI displays a pre-steady-state burst of product formation indicating rapid enzyme-bound product formation and a rate-limiting product release step, but this burst is diminished with noncognate DNA[5.11]. Taken together, these results with HhaI reveal a mechanism where base-flipping is coupled with recognition, faster than methyl-transfer, and product release is rate-determining for turnover[6.16,6.17,6.18,5.12].

EcoRI is an example of an N^6-adenine bacterial DNA methyltransferase that recognizes 5′-GAATTC-3′. EcoRI base-flipping kinetics monitored with 2AP are biphasic with 80% of the trace assigned to coupled DNA binding and base-flipping with a base-flipping rate of $195s^{-1}$. The additional 20% base-flipping signal with a rate of $0.55s^{-1}$ occurs after binding and corresponds to an intramolecular isomerization reaction. The initial binding and base-flipping steps are faster than methyl-transfer ($k_{methylation} = 41s^{-1}$)[6.5], while product release is slow and rate-determining for turnover ($k_{product\ release} = 0.18s^{-1}$)[6.4]. Another study showed that base flipping by EcoRI followed a DNA bending event and preceded an intercalation event[6.6]. Global fitting resolved that base flipping (k_{flip}) = $550s^{-1}$ and intercalation ($k_{intercalation}$) = $97s^{-1}$, where all pre-catalytic forward rate constants are faster than chemistry[6.6]. Additionally, EcoRI displays 3,500-23,000-fold decreased specificity for noncognate DNA and specificity is accounted for by increased reverse rate constants for intercalation and base flipping [6.6], suggesting that base-flipping is coupled to specificity.

Another example of base flipping is for the *Escherichia coli* Dam (Ecodam) N^6-adenine DNA methyltransferase, which recognizes 5'GATC'3 with high specificity. Studies with this enzyme show that base-flipping is a biphasic process where the first phase is very fast ($240s^{-1}$) and binding of the flipped-out target base into the active site pocket is slow ($0.1\text{-}2s^{-1}$). Base-rearrangement is slow and rate-determining for methylation ($k_{methylation} = 0.44s^{-1}$) and DNA recognition of GATC site starts before flipping. Base-flipping is coupled with recognition; forward flipping kinetics are reduced and reverse kinetics of flipping increase with noncanonical DNA sequences containing 2AP[4.20].

CcrM stands out in the context of base flipping because the target adenine is flipped out in addition to other bases of the recognition site (Fig. 6.1) [2.5]. Therefore, DNA strand-separation likely plays a unique role in the transition of the target adenine from the B-form duplex to the active site. The co-crystal structure reveals the DNA strand-separated intermediate (Fig. 6.1)[2.5] and we previously reported on the kinetic mechanism of DNA strand separation and methylation[4.1]. We monitored DNA strand separation with Pyrrolo-dC (PydC) placed at the N-position (5'GANTC3') because CcrM does not discriminate at the N-position and this probe could therefore monitor the enzyme-dependent transition from duplex to strand-separated DNA without attenuating methylation. However, this N-position monitored PydC fluorescence failed to provide a signal associated with DNA recognition and formation of intermediates associated with base-flipping. We relied on rigorous global data fitting to resolve that DNA strand separation is involved in substrate recognition, is faster than and precedes methylation, suggesting that DNA strand-separation is a facile and efficient process for CcrM[4.1].

Our prior kinetic study failed to resolve a pre-catalytic step that might limit the rate-of methylation. Typically, DNA methyltransferases show a pre-steady-state burst of product formation followed by rate-limiting product release[6.16,5.11,6.6]. CcrM, uniquely, does not show a burst, indicating that a step preceding product release is slow and rate-limiting in catalysis. Based on CcrM's unique mechanism of DNA recognition, high level of discrimination, and lack of a pre-steady state burst, we sought to interrogate base-flipping in the context of DNA strand-separation to understand if target base-flipping into the active site was rate-determining for CcrM catalysis.

2AP (2-aminopurine) has spectral overlap with protein fluorescence, lending itself to DNA methyltransferase that contain only one or two Trp residues that are unresponsive to DNA binding. CcrM has 18 Trp residues per dimer, therefore 2AP fluorescence was dominated and convoluted by Trp fluorescence (data not shown). Thus, we opted to use 6MAP (4-amino-6-methyl-8-(2′-deoxy-â-D-ribofuranosyl)-7(8H)-pteridone), a well-characterized but less well-known fluorescent analog of adenosine that is spectrally isolated from Trp fluorescence[6.14,6.15.] 6MAP has been shown to monitor base-flipping by DNA repair enzyme DNA photolyase[6.13] and is a robust probe for monitoring enzyme-dependent conformational changes in DNA substrates. We incorporated 6MAP into the target-adenine position (5'GANTC3'). We observed a larger signal when 6MAP was excited via FRET from Trp residues, therefore the 6MAP kinetics used in this study consist of 6MAP FRET excitation data. Energy transfer is highly efficient for 6MAP compared to 2AP[6.13]. The novel mechanism of DNA strand-separation and base-flipping by CcrM is useful for the study of other enzymes that rely on base-flipping and/or DNA strand-separation with high selectivity to perform efficient catalysis.

VI: Materials and methods

DNA

6MAP DNA oligos were obtained from Fidelity Oligos and the all others were obtained from the Yale Keck Oligo Synthesis Facilities. The oligos were annealed at 95 °C for 5 minutes in 10 mM Tris HCl, 50 mM NaCl, 1 mM EDTA, pH 8 and subsequently cooled passively to room temperature. Annealing was analyzed by 10% native PAGE. Gels were imaged on a Bio-

Rad Gel Doc EZ Imager. Densitometry was performed with FIJI ImageJ which determined a >98% annealing success.

Kinetic measurements using 6MAP Fluorescence

The adenosine analog 6MAP has an excitation maximum at 320nm and emission maximum at 430[6.13,6.14,6.15]. 6MAP has a high fluorescence quantum yield as a monomer ($\Phi_f \sim$ 0.4) and a lower quantum yield when incorporated into single- or double-stranded duplexes due to base-stacking quenching interactions ($\Phi_f >$ 0.01 to 0.11)[6.14,6.15]. Initial experiments were performed by direct excitation of the 6MAP (Figs. 6.3 and 6.5). To get a larger signal by only exciting 6MAP bound to the enzyme, we measured FRET from tryptophan to 6MAP using an excitation wavelength of 295nm and emission at 445nm (45 nm bandpass; Semrock). Subsequent analysis showed that the large FRET signal was attributable to Trp residues close to the 6MAP in the base-flipped state. Kinetic measurements in Figs. 6.6D, 6.7C and 6.7D were performed on a KinTek SF-300x stopped flow instrument (Austin, TX, USA) at 22˚C. Final concentrations after 1:1 mixing were 0.5, 1, 2, 4, and 6 μM CcrM, 60.0 μM SAM, and 200 nM 6-MAP dsDNA. At least three traces were collected and averaged. The dead time of the instrument was 1.365 ms. Data were fit to a single-exponential function in KinTek Global Kinetic Explorer. The function used for fitting was $y = a_0 + a_1(1 - e^{-b1t})$, where y = fluorescence intensity (arbitrary units), t = time (seconds), a_1 = the amplitude, b_1 = the rate, and a_0 = the initial fluorescence amplitude (arbitrary units).

Radiochemical methylation assay

Single turnover methylation reactions consisted of 250nM enzyme, 100nM DNA, and 15μM AdoMet, using hemimethylated double-stranded substrates. Tritiated AdoMet in all instances had a specific activity of (3H-CH3 1 mCi [82.7 mCi/mmol]). Reactions were initiated by adding enzyme. 5μl of reaction time points were spotted in triplicate onto GE Amersham Hybond-XL nylon membranes followed by three 5min washes in wash buffer (50 mM KH_2PO_4). The washes were followed by a 5 min dehydration step with 80% Ethanol, another wash for 5 min in 100% Ethanol, and a final drying step for 10 min under a heat lamp. Samples were then placed into scintillation vials containing 3 ml of BioSafe II Scintillation cocktail. Radiochemical data were generated with a Hidex 300SL scintillation counter. Data for the single turnover reaction were background-subtracted and fit to a one-phase decay model in GraphPad Prism 10.0.2.

Global fitting

Global data fitting was performed in KinTek Global Kinetic Explorer Version v11.0.1 (Austin, TX, USA). The reaction scheme used as the unifying model to describe the experimental data is shown in Figure 7: $E + S = FS = GS^{I} = GS^{II} = GS^{III} + GSp^{III} = E + Sp^{III}$. Data from five experiments were input and initial concentrations were entered according to how the data was collected. Time-dependent data for the Trp and pyrollo-dC signal were corrected for the measured dead time of the instrument (2.6 ms). Each experiment had a unique observable signal which relates the experimental data to the model. For example, Experiment 1 shows PydC kinetics over 2 s, where PydC fluorescence depends on the following observable signal: $a1*(S + FS + e*GS^{I} + f*(GS^{II} + GS^{III} + GSp^{III}) + h*Sp)$. This observable signal

describes that the GS^{I} species contributes to the change in PydC signal according to the scaling factor *e*, while GS^{II}, GS^{III} and GSp^{III} contribute by a scaling facor *f*, and that GS^{II}, GS^{III} and GSp^{III} contribute equally to the change in fluorescence. Experiment 4 shows the Trp fluorescence kinetic trace which depends on the following observable signal: a4*(E + k*FS + m*GS^{I} + n*(GS^{II} + GS^{III} + GSp^{III})). Exp 3 is PydC data over a long time-course to observe the annealing of the methylated DNA strands after methylation. Experiment 3 had an observable signal identical to experiment 1 because they are both monitoring PydC signal, but over different timescales. Experiment 3 had an observable signal of a3*(S + FS + e*GS^{I} + f*(GS^{II} + GS^{III} + GSp^{III} + h*Sp). Experiment 2 is the radiochemical methylation assay, and we therefore have an observable signal of GSp^{III} + Sp + bkg2. Experiment 5 is 6MAP kinetics over 200 s, in which we built a linked parallel pathway that describes DNA binding and base flipping without product formation because, even in the presence of SAM, control experiments showed that the 6MAP substrate (S2) does not get methylated. Therefore, 6MAP kinetics depend on the following observable signal: a5*(S2 + b1*FS2 + b1*$GS2^{I}$ + b1*$GS2^{II}$ + b4*$GS2^{III}$), where E + S2 = FS2 = $GS2^{I}$ = $GS2^{II}$ = $GS2^{III}$ was linked to the model above to describe the 6MAP experiment. Data from all five experiments were fit globally based on numerical integration of the rate constants (computer simulation). Initial values were estimated based on the fitting of each data set using simplified equations. Fluorescence scaling factors were applied to the data in experiments 1, 4 and 5 to normalize variability in the starting amplitudes for each trace within a concentration series. Scaling factors were close to unity for experiments 1 and 4 and therefore did not influence the concentration dependence of these traces. Scaling factors were more significant for the 6MAP fluorescence amplitudes because

the FRET excitation of 6MAP increased as protein concentration increased. Some values were locked during the simulation, while others were allowed to float, as described when fitting the data. Locked values were chosen based on the parameter's lower limit beyond which there is no effect on the fitted curves. Individual models were built for WT and W57F with ds6MAP-DNA with identical model architecture for direct comparison to the WT model. Confidence contour analysis was performed with the 1D Fitspace function for each rate constant, which outlines the space over which parameters can vary to provide a 95% confidence interval for each parameter.

VI: Results

6MAP is an analog of adenine with a second 6-membered to provide a fluorescence signal (Fig. 6.2a). The N^6 position of adenine is unmodified in 6MAP, so it can form a Watson-Crick base-pair with thymine (Fig. 6.2b). The target adenine (A_{11}) carbon-8 position in the CcrM/DNA co-crystal structure does not clash with other atoms. However, an H_20 molecule may be displaced to accommodate the additional atoms of 6MAP. (Fig. 6.2c).

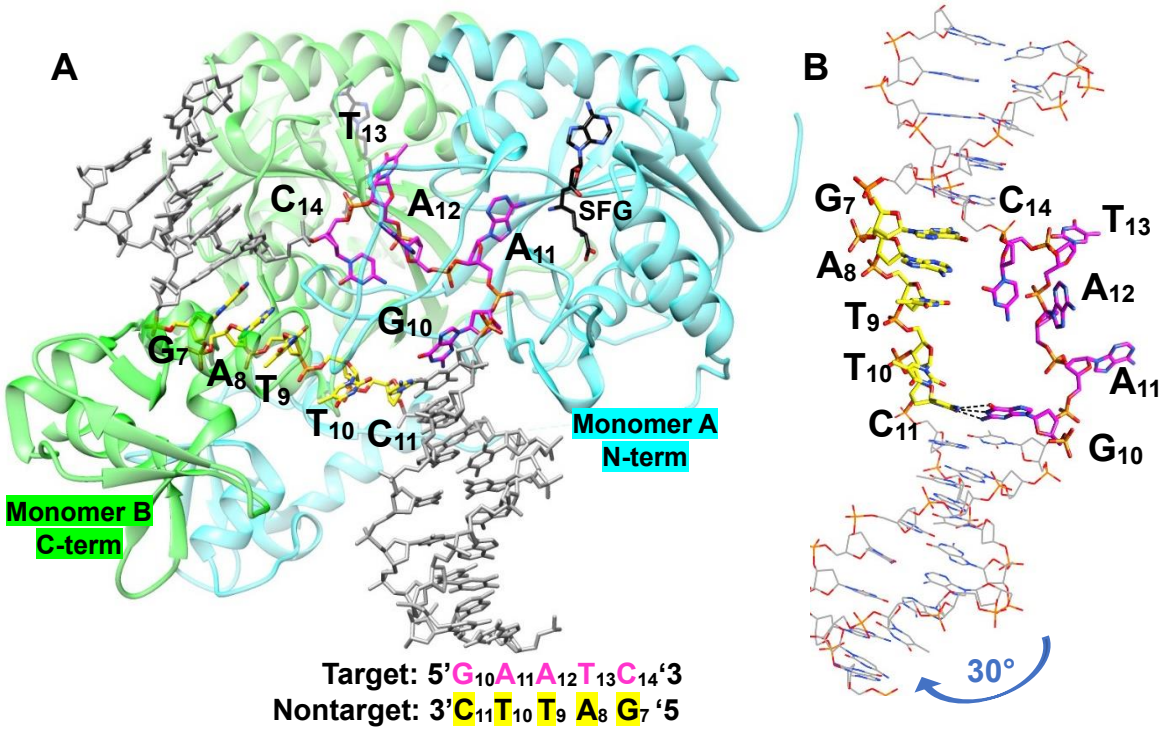


Figure 6.1. Base-flipping by CcrM is uniquely observed in the DNA strand-separated state. A. The CcrM/dsDNA/SFG co-crystral stucutre reveals the DNA strand-separation state. The target adenine (A_{11}) is positioned next to SFG in the catalytic pocket. The target strand recognition site (magenta bases) looses Watson-Crick base pairing and base-stacking interactions, while the nontarget strand (yellow bases) maintains base-stacking interactions. **B.** The CcrM-bound DNA molecule in the cocrystal structure is perturbed from B-form DNA with a 30° bend. G_{10} and C_{11} maintain Watson-Crick base-pairing. The strucutral images were made with UCSF Chimera, PDB: 6PBD.

6MAP fluorescence is low in a DNA duplex alone due to quenching interactions of base-stacking and base-pairing, and increases upon the addition of WT CcrM (Fig. 6.3). Upon binding DNA, CcrM separates 4 of the 5 base pairs within the 5'GANTC3' recognition so that 6MAP loses base-pairing and base-stacking interactions, resulting in an increase in fluorescence (Fig. 6.3). Upon the addition of demethylated cofactor product (SAH), the total fluorescence of 6MAP is slightly decreased (Fig. 6.3).

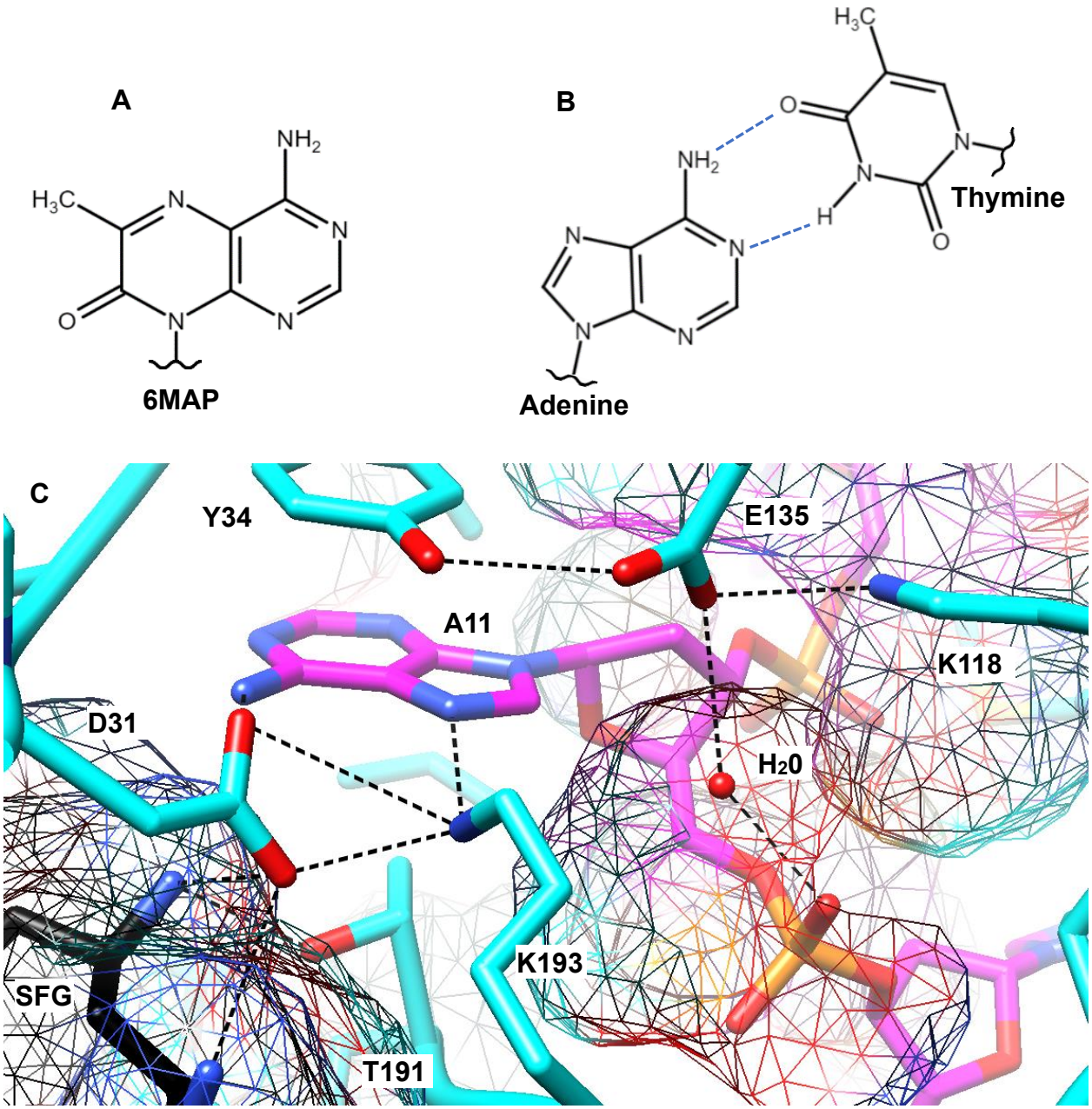


Figure 6.2. 6MAP is a fluorescent adenine analog positioned at the target adenine (A_{11}) position as a probe for base flipping. A. 6MAP relative to adenine has an additional 2 carbons, 1 oxygen, and 2 hydrogens with an introduced 6-membered ring. These additions are located at the C-8 position of adenine. **B.** Adenine makes two hydrogen bonds to thymine via N1 and N6, which are conserved in 6MAP. Thus, 6MAP can form a Watson-Crick base-pair with thymine. **C.** The target adenine's (A_{11}) carbon-8 position in the CcrM/DNA co-crystal

structure does not clash with other atoms. The structural image was made with UCSF Chimera and grids represent molecular surfaces.PDB: 6PBD.

Data in Fig. 6.4 show that 6MAP-modified DNA (ds6MAP) cannot be methylated by CcrM (Fig. 6.4). A single-turnover methylation assay shows little to no product formation over 60 minutes (Fig. 6.4). Unlabeled-DNA (dsC1) is fully turned over into product within 3 minutes with $k_{methylation} = 0.02 \pm 0.004\ s^{-1}$ (Fig. 6.4 inset). This suggests that methylation or a step leading up to methylation is perturbed by the modification of the 6MAP analog.

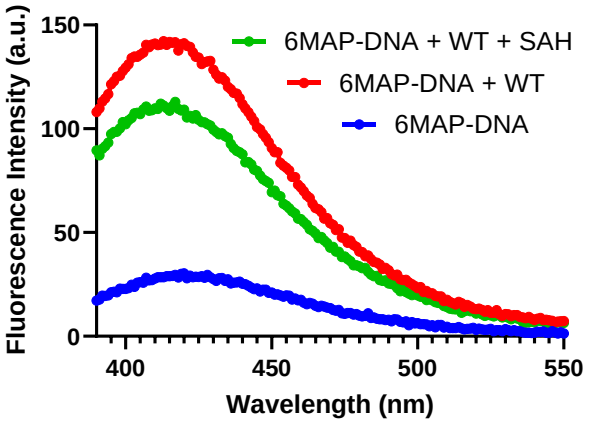


Figure 6.3. 6MAP equilibrium fluorescence increases in the presence of CcrM and decreases with SAH. DNA emission max = 420nm, DNA + WT CcrM emission max = 413nm, DNA + WT CcrM + SAH emission max = 417nm. The fluorophore was excited directly and 330nm; the excitation and emission bandpasses were 2nm. Conditions consisted of 1 µM 6MAP-DNA , 5 µM WT CcrM, and 60 µM.

When 6MAP is excited directly at 330nm, 6MAP fluorescent kinetics are biphasic and consist of an initial increase followed by a slow decrease and both rates are protein concentration dependent but the observed amplitude is small (Fig. 6.5). The initial increase monitored over 2s has a rate of 20-45s^{-1} with a marginal amplitude (>0.2) (Fig. 6.5a-c). The second phase is observed over 200 s with a rate of 0.002-0.012s^{-1} with an amplitude of >0.1 (Fig. 6.5d-f).

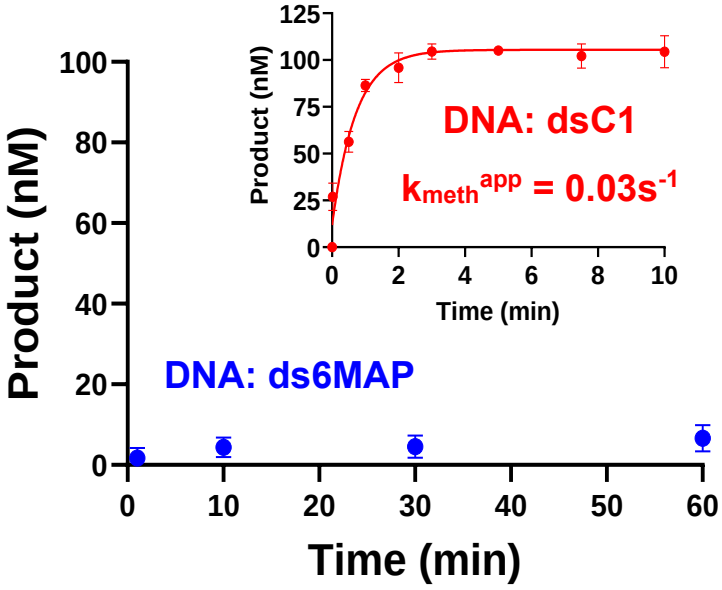


Figure 6.4. ds6MAP-containing-DNA is not methylatable. Radiochemical single-turnover assay shows that 6MAP (ds6MAP: 5'--G(6MAP)CTC--3') is not a methylatable substrate for CcrM. **Inset:** Unlabeled-DNA (dsC1: 5'--GACTC--3') is fully turned over into product within 3 minutes with $k_{methylation}$ = 1.4 ± 0.22 min^{-1}. Conditions consisted of 100nM DNA, 250nM WT CcrM, and 15µM 1:10 (hot:cold) SAM. Timepoints were spotted in triplicate and error bars represent standard deviation.

A larger 6MAP signal was observed when 6MAP was excited via FRET from tryptophans in CcrM (Fig. 6.6). DNA containing 6MAP in place of the targetted adenine gives a large FRET signal over 200 seconds with a slow single-exponential increase (Fig. 6.6d) which has a hyperbolic concentration-dependence on amplitude but not on rate (Fig. 6.6e-f). This is the kinetic signature of a readily reversible reaction, and can be fit better by simulation. The rate of 6MAP kinetics via FRET excitation is approximately 0.022-0.025 s^{-1} (Fig. 6.6E) which is the faster than the observed rate of methylation (0.0005s^{-1}) (Fig. 6.7D). However, global data fitting shows that these observed reactions are the result of 4-5 steps in sequence with reversible steps and comparable rate constants (see model in Fig. 6.7). Therefore, fitting

data to a single exponential is not easily interpretable, and in this case leads to incorrect conclusions.

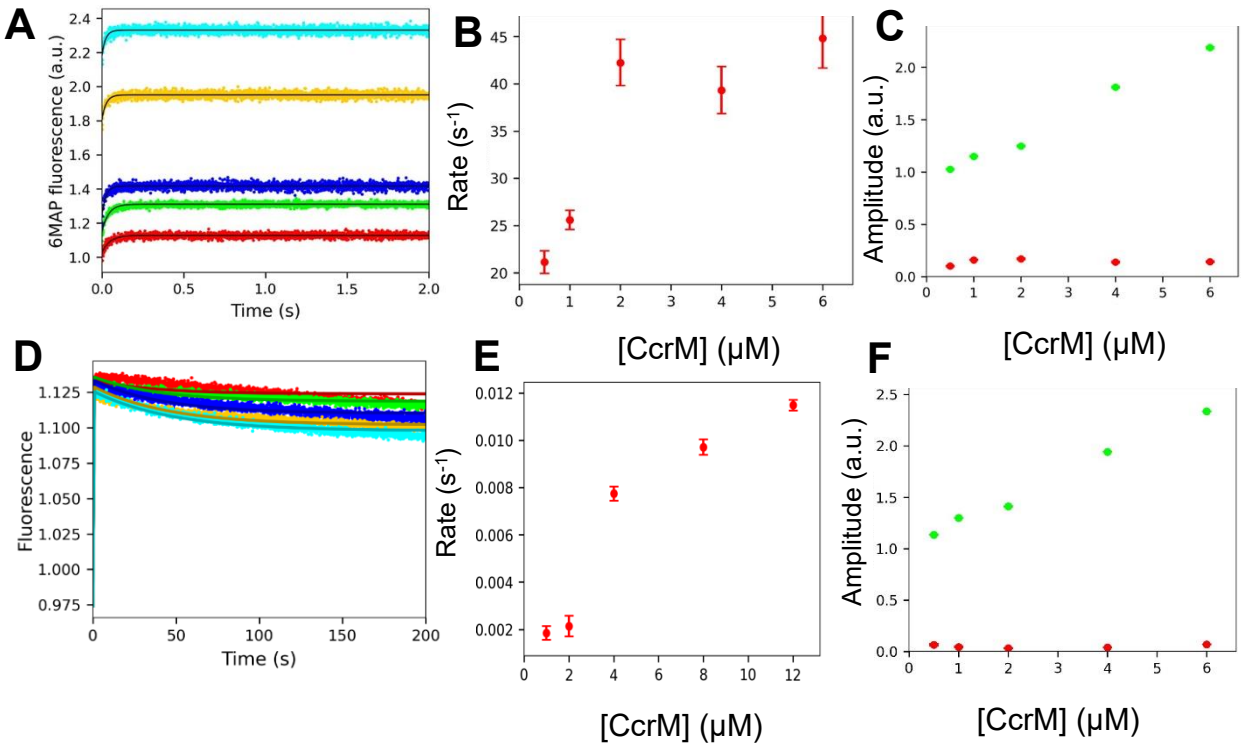


Figure 6.5. ds6MAP kinetics are biphasic and consist of an initial increase followed by a slow decrease and both rates are protein concentration dependent. Excitation wavelength = 330nm and emission wavelength = 445nm. Final concentrations after 1:1 mixing were ds6MAP-DNA = 200nM, WT CcrM = 0.5, 1.0, 2.0, 4.0, and 6.0μM, SAM = 60μM. . **A-C:** 6MAP kinetics over 2 seconds shows a single-exponential increase that has a concentration dependence on rate and starting amplitude amplitude. **D-F:** 6MAP kinetics over 200 seconds shows a fast initial increase followed by a slow single-exponential decrease that has a concentration dependence on rate and starting amplitde. The red data points in the amplitude graphs (C and F) are the total amplitude differences and the green data points are the starting amplitudes.

We relied on 6MAP excitation via FRET for global data fitting due to the much larger signal. Global fitting with 6MAP data shows that target-adenine base-flipping is the rate-determining step for methylation but the steps leading up to base-flipping are unfavorable

thermodynamically (Fig. 6.7). Trp fluorescence kinetics are biphasic and monitor DNA binding and a protein isomerization event (Fig. 6.7a). PydC kinetics are biphasic and monitor DNA strand separation; the rate of the protein conformational change (trp fluorescence) is correlated with the first phase of DNA strand-separation (Fig. 6.7b) measured by PydC fluorescence. The signal from ds6MAP excitation via FRET kinetics over 200 s are monophasic (Fig. 6.7c). The large signal giving a slow increase in fluorescence is monitoring target-adenine base-flipping in the active site due to the close approach of two Trp residues (Fig. 6.8a). Single-turnover methylation assay with PydC-labeled DNA gives the observed rate for methylation but is a function of multiple steps (Fig. 6.7d). Susequent PydC kinetics over to 2000s to monitor DNA product dissociation from CcrM and re-annealing of the duplex (k_6) (Fig. 6.7e).

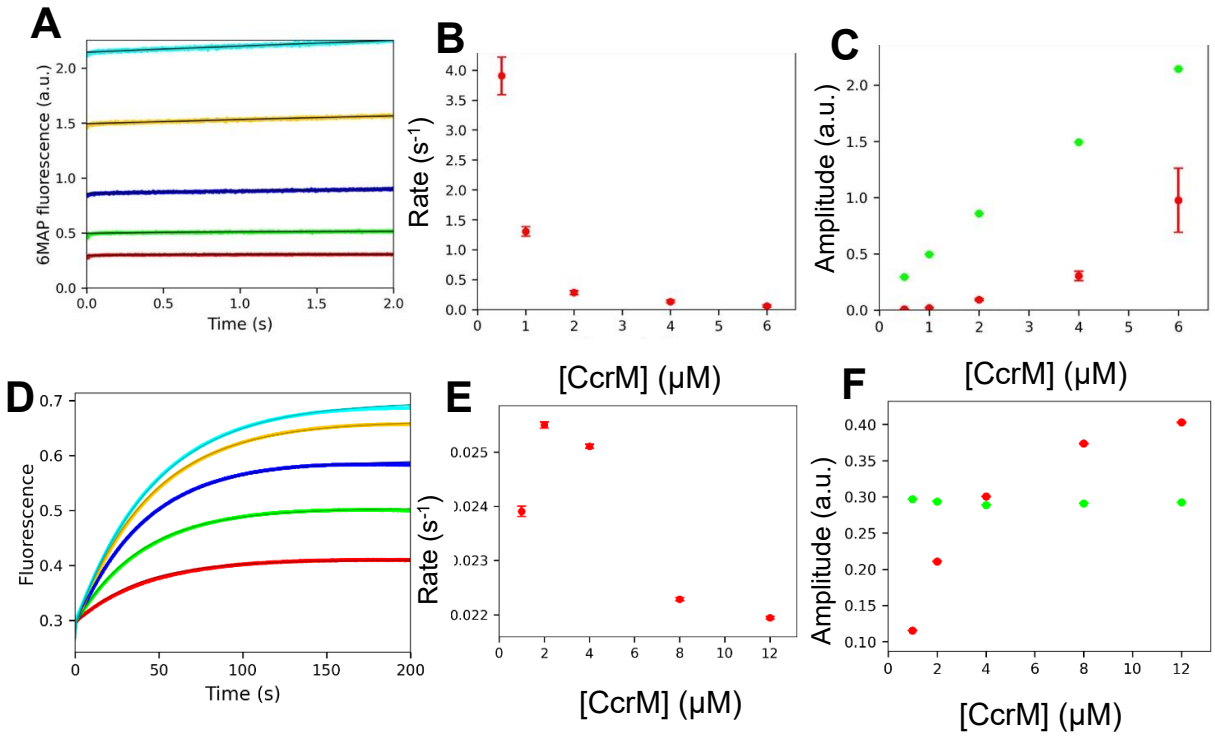


Figure 6.6. A large 6MAP signal is observed when 6MAP is excited via FRET from tryptophan. The excitation wavelength was 295nm and emission was monitored at 445nm. Final concentrations after 1:1 mixing were 200 nM ds6MAP-DNA, 0.5, 1, 2, 4, and 6 μM WT

CcrM, and 60 μM SAM which was preincubated in both syringes, and temperature controlled to 22˚C.**A-C:** 6MAP kinetics over 2 seconds show near-linear kinetics that do not fit well to an exponential function. When fit to an exponential function, there is a rate and amplitude dependence on CcrM concentration. **D-F:** 6MAP kinetics via FRET excitation over 200 seconds shows a slow single-exponential increase which has an amplitude-dependence on concentration. The rate of 6MAP kinetics via FRET excitation is protien concentration-independent and approximately 0.022-0.025 s^{-1} which is the same rate of $k_{methylation}$.

Global data of the data in Fig. 6.7 provides a single unifying model accounting for all of the data and the interplay between successive steps. For example the FRET 6MAP signal is attributed to a small increase in fluorescence upon DNA binding but then a large change in fluorescence with base flipping, but the observed kinetics are a function of all steps, including reverse reactions which increase the observed rate of approach to equilibrium but reduce the amplitude. Moreover, steps 2, 3, and 4 are all unfavorable thermodynamically so only a small fraction of the enzyme-DNA complex has the DNA unwound and the base flipped into the active site. Although the observed rates (from a single exponential fit) for methylation and product release are slow, according to the modeling, methylation of product release must be faster following base flipping to account for all of the data.

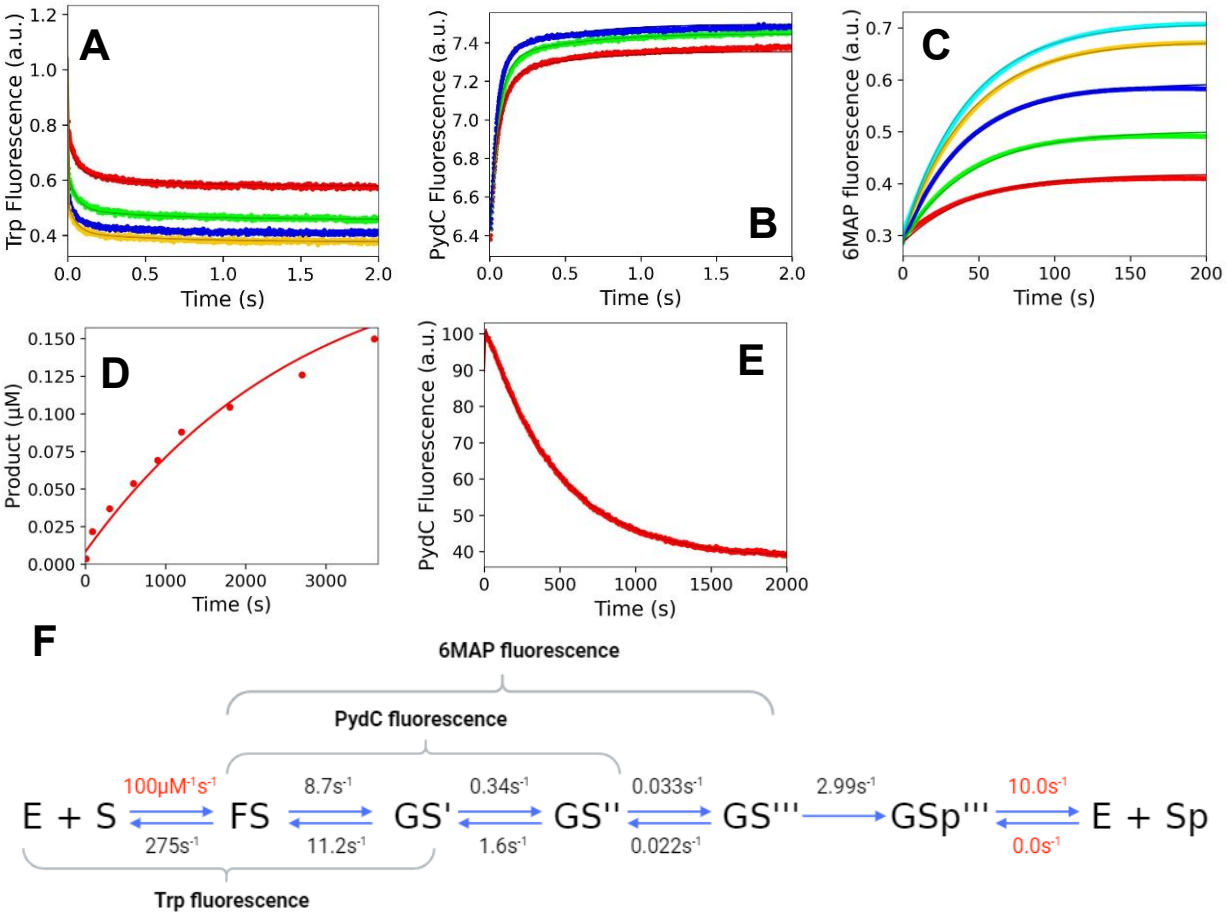


Figure 6.7. Global fitting with ds6MAP FRET data shows that target-adenine base-rearrangement is the rate-determining step for methylation. A. Trp fluorescence is biphasic and monitors binding and a protein isomerization step. B. PydC is biphasic and monitors DNA strand-separation. C. 6MAP fluorescence via FRET excitation from Trp residues in monophasic and the slow increase is rate-determining for CcrM catalysis. D. Radiochemical single-turnover assay monitors methylation. E. PydC fluorescence over 2000s monitors DNA product release and/or product re-annealing. F. WT/dsDNA/SAM enzyme kinetic model reveals that base-flipping (k_4) is rate-determining for methylation.

Confidence contour analysis shows that the rate contants derived from the simulation are well constrained by the data (Fig. 6.11). Combining all of these signals into a unifying kinetic model shows that target adenine base-flipping (k_4) occurs after DNA strand separation (k_3) and limits the rate of methylation (Fig. 6.7f). Although the global fitting requires the rate

of methylation to be limited by the rate of base flipping, the intrinsic rate constant for methylation is not known with certainty. Rather, the global data fitting only set a lower limit of $k_5 \sim 3\ s^{-1}$ to account for the data. However, it should be noted that the net rate of methylation is much less than k_5 due to the unfavorable equilibria for steps leading up to DNA methylation so that only a small fraction of the E-DNA complex is in the GS''' state poised for catalsys.

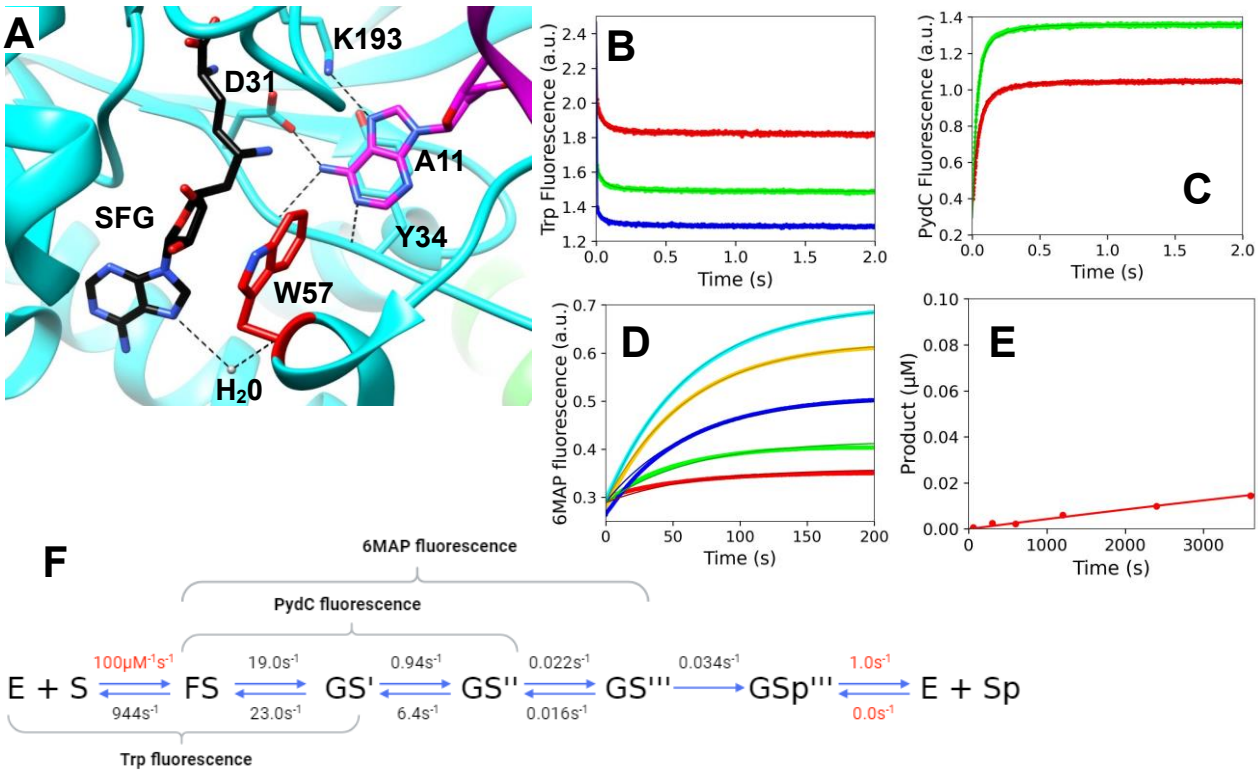


Figure 6.8. W57F is capable of strand separation and base flipping, but methylation is disfavored due to destabilized cofactor binding. A. W57 makes a water-mediated hydrogen bond via its peptide backbond to SFG's indole nitrogen. A_{11} hydrogen bonds to D31 and makes a stacking interaction to Y34 from the $D_{31}P_{32}P_{33}Y_{34}$ motif. Hydrogen bonds are dipicted as black dashed lines. The structural image was made with UCSF Chimera. **B.** W57F Trp kinetics are biphasic and monitor binding and a protein conformational change. **C.** W57F PydC kinetics are biphasic and monitor DNA strand-separation. **D.** W57F 6MAP kinetics monitor base-flipping. **E.** W57F has attenuated methylation. **F.** Global fitting of W57F data reveals that methylation is the perturbed step in catalysis.

We next used mutational analysis to further probe the kinetic mechanism of target-adenine base-flipping and repositioning. W57 is a highly conserved residue that is positioned near the target adenine and cofactor in the active site. W57 makes a water-mediated hydrogen bond via its peptide backbond to the indole nitrogen of Sinefungin (SFG) (Fig. 6.8a). A_{11} hydrogen bonds to D31 and makes a stacking interaction to Y34 from the $D_{31}P_{32}P_{33}Y_{34}$ motif (Fig. 6.8a). The W57F mutant fails to undergo the protein isomerization step (k_{-1} = 944s^{-1}, k_2=19.0s^{-1}, k_{-2} = 23.0s^{-1}) and strand-separate DNA (k_3 = 0.94s^{-1}, k_{-3} = 6.4s^{-1}), and methylates very poorly (k_5= 0.034s^{-1}). Target adenine base rearrangement is slightly slower than WT (W57F k_4 = 0.022s^{-1}, k_{-4} = 0.016s^{-1}) (Fig.6.8f). Target adenine base-rearrangement for W57F is also rate-determining and sets an upper limit on the derived $k_{methylation}$ (k_5). Confidence contour analysis shows that the rate contants derived from the simulation are well constrained by the data (Fig. 6.11).

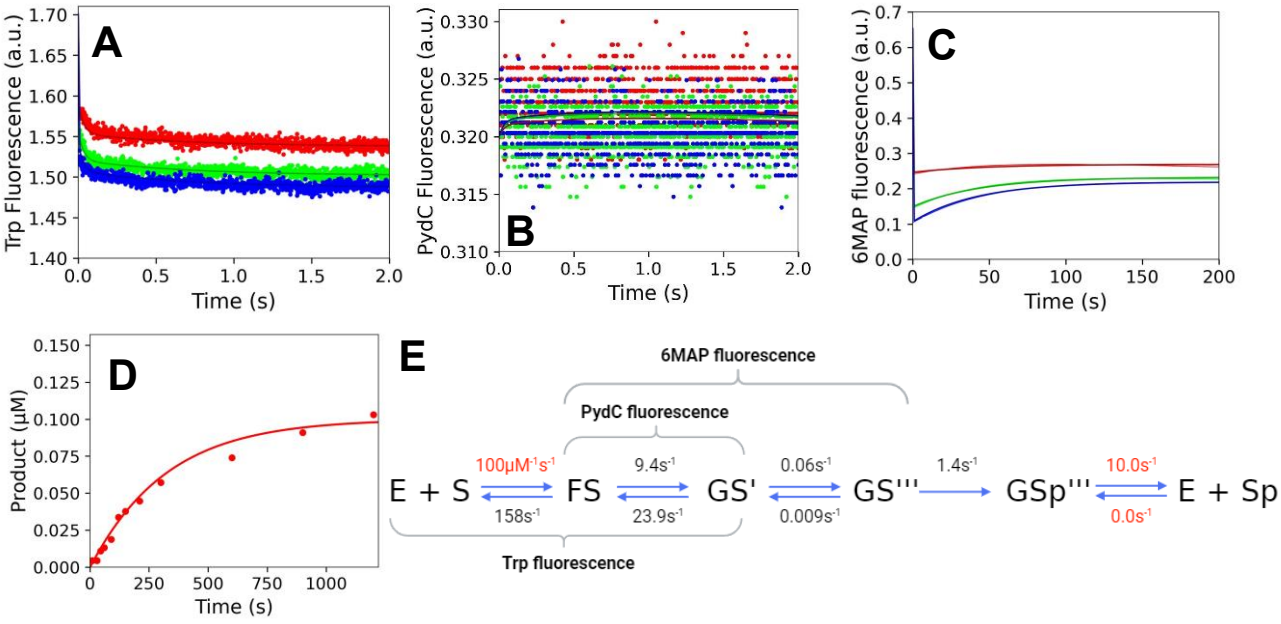


Figure 6.9. Single-stranded 6MAP kinetics have the same rate as single-stranded $k_{methylation}$, providing further evidence that target-adenine base-rearrangement is rate-determining for catalysis. A. Trp fluorescence is biphasic and monitors binding and a protein isomerization step. **B.** PydC fluorescence is not observed indicating no strand separation **C.** 6MAP fluorescence via FRET excitation from Trp residues. **D.** Radiochemical single-turnover assay monitors methylation. **E.** WT/ssDNA/SAM enzyme kinetic model reveals that base-flipping (k_4) is rate-determining for methylation.

We monitored the CcrM-dependent kinetics of 6MAP via FRET excitation in single-stranded DNA (ss6MAP) to understand how base-flipping occurs independently of DNA strand-separation (Fig. 9). Single-stranded 6MAP kinetics have the same rate as single-stranded $k_{methylation}$, providing further evidence that target-adenine base-flipping is rate-determining for catalysis (Fig. 11). Uniquely, CcrM is able to methylate single-stranded DNA. k_{cat} for ssDNA (unlabeled) is approx. $0.003s^{-1}$. Since the rate of ss6MAP kinetics are the same as ssC1 methylation, this strongly suggests that target-adenine base-flipping is rate-determining for methylation.

Confidence contour analysis shows that the data are well-defined by the models (Fig. 6.11). These contours are shown for each rate constant, which outlines the space over which parameters can vary. The dashed line establishes the 95% confidence interval at a Chi^2 threshold that is set based on the number of data points and parameters used in the fitting. Free energy profiles were calculated in Kintek Explorere and each free energy profile shows that the state with the highest free energy is the base-flipped intermediate (Fig. 6.10).

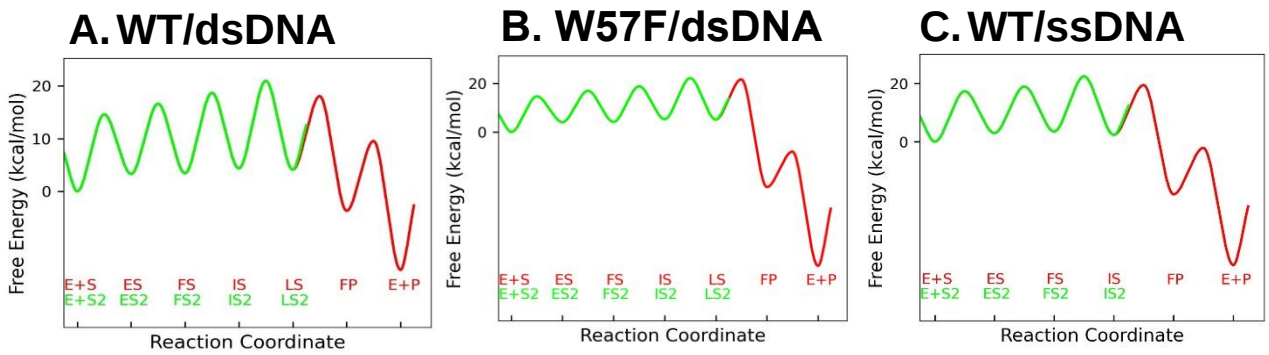


Figure 6.10. Free energy profiles show that the pre-catalytic base-flipped state (IS) has the highest free energy for each model. A. WT/dsDNA model shows that target adenine base-flipping is the rate-determining step for catalysis. **B.** W57F/dsDNA model shows that target adenine base-flipping is the rate-determining step for catalysis. **C.** WT/ssDNA model shows that target adenine base-flipping is the rate-determining step for catalysis. Free energy profiles were calculated in KinTek Explorer with the following parameters: temperature = 295K, transmission coefficient = 0.01, ligand concentrations = 0.01 (10nM).

Model	k_1 ($\mu M^{-1}s^{-1}$)	k_{-1} (s^{-1})	k_2 (s^{-1})	k_{-2} (s^{-1})	k_3 (s^{-1})	k_{-3} (s^{-1})	k_4 (s^{-1})	k_{-4} (s^{-1})	k_5 (s^{-1})
WT/ds DNA	100	275 (272-278)	8.7 (8.48-9.05)	11.2 (10.8-11.7)	0.34 (0.27-0.40)	1.6 (1.34-1.83)	0.033 (0.031-0.036)	0.022 (0.0215-0.022)	>1.2 (1.22-29.9)
W57F/ dsDNA	100	944 (931-953)	19.0 (17.7-20.3)	23.0 (20.5-25.2)	0.94 (0.46-1.47)	6.4 (4.1-8.1)	0.022 (0.018-0.029)	0.0156 (0.0154-0.0158)	0.034 (0.027-0.054)
WT/ss DNA	100	158 (154-163)	9.4 (6.4-11.8)	23.9 (15.3-31.3)	NA	NA	0.065 (0.060-0.071)	0.0088 (0.0085-0.0094)	1.39 (1.20-1.92)

Table 6.1. Rate constants derived from global fitting. k_1 and k_{-1} monitor the binding step (Trp fluorescence). k_2 and k_{-2} monitor a protein conformational change (Trp fluorescence) and the first phase of DNA strand separation (PydC fluorescence). k_3 and k_{-3} monitor the second phase of DNA strand separation (PydC fluorescence). k_4 and k_{-4} monitor target adenine base flipping (6MAP fluorescence). k_5 monitors methylation (single turnover radiochemical assay). k_6 monitors product release. Rate constants were derived from global fitting. Rates in red were locked during the continuous simulation while rates in black were allowed to float freely without constraints. k_6 was locked at $10s^{-1}$ in all models.

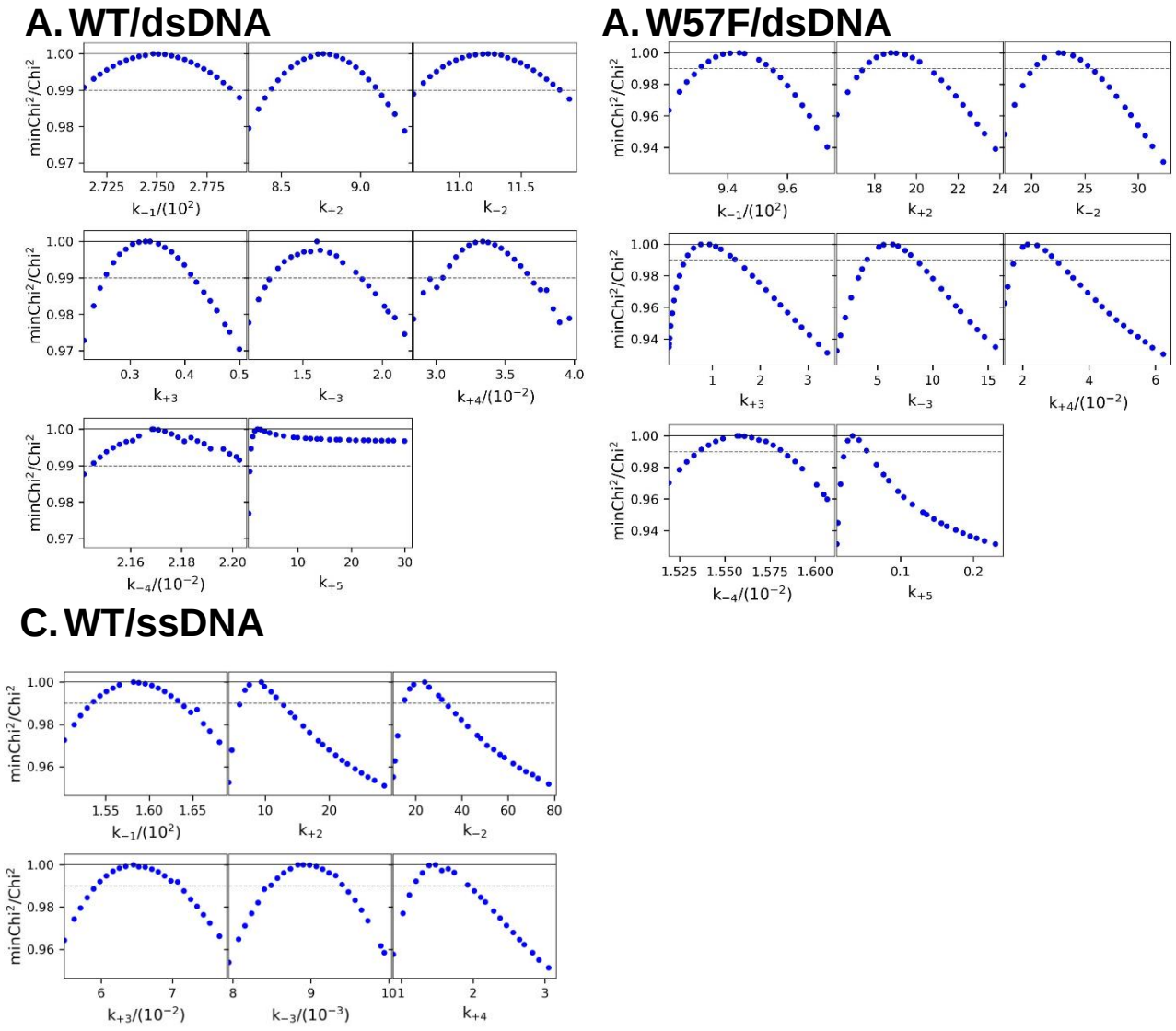


Figure 6.11 Confidence contour analysis show that the data are well defined by the models. The data represent the 1D Fitspace calculated for each rate constant, which outlines the space over which parameters can vary. The dashed line establishes the 95% confidence interval at the 0.99 Chi^2 threshold. The 95% Chi^2 limits were calculated in KinTek Explorer.

VI: Discussion

Understanding the detailed steps of enzyme mechanisms is essential to understanding biological processes and disease. The CcrM enzyme mechanism is an excellent example of

how modulation of free energy barriers are essential for efficient on-target catalysis. We previously reported that CcrM employs a unique mechanism relative to other DNA methyltransferases in that CcrM strand-separates its DNA recognition sequence to achieve an extraordinary level of sequence-recognition (Fig. 6.1)[4.1].

Here we deconstruct CcrM's mechanism of base-flipping which stands out from other enzymes because base-flipping for CcrM transitions from single-stranded DNA into the flipped-state, while other characterized DNA methyltransferases transition from duplex DNA to the flipped-state[6.1,6.2,6.3,3.1,4.20,6.4,6.8].

Previously, base-flipping was questioned for being a passive or active process. A passive mechanism would depend on intrinsic DNA breathing where the enzyme would passively capture the extra-helical target base into the active site for modification[6.1,6.2,4.20]. An active process would involve enzyme-promoted base-flipping of the target, followed by target base repositioning and catalysis. A few examples have argued in favor of an active mechanism for base-flipping: uracil DNA glycosylase[6.1], EcoDam DNA methyltransferase[4.20], and HhaI DNA methyltransferase[6.2]. Base flipping from duplex DNA could be facilitated by DNA bending.

For CcrM, the question of active or passive base-flipping may not have the same meaning because the target adenine (A_{11}) enters the active site from a segment of ssDNA so base flipping may not be an appropriate descriptor. The unfavorable equilibrium constant for the entry of 6MAP into the enzyme active site high FRET state indicates that the process is not driven by strong favorable interactions. One is left with the counterintuitive conclusion that it is precisely the weak binding of the target adenine at the active site that contributes to

specificity. A tighter binding could hold a mismatched DNA at the active site long enough even for a slow methylation reaction to proceed to completion leading to errors in methylation.

6MAP is not a methylatable substrate for CcrM, although 6MAP conserves the position of N^6 and the analog modifications occur at the C^8 position (Fig. 6.2 and Fig. 6.4). Two possible structural explanations are that 6MAP is structurally perturbed in the active site and is not in a conformation that allows chemistry to occur or that the active site cannot accommodate the additional space of the 6MAP analog. Another possible explanation is that the addition of an electron-withdrawing carbonyl in 6MAP could make 6MAP incapable of deprotonation by a basic amino acid sidechain in the active site (D31). Thus, 6MAP would be a much weaker nucleophile and incapable of performing the S_N2 reaction for the methyl-thiol on the SAM substrate which is essential for methyl-transfer chemistry.

Here we build on our previously published CcrM kinetic model[4.1]. We monitored CcrM-dependent target base-flipping with 6MAP positioned at the target position (5'-G(6MAP)CTC-'3) via FRET excitation from Trp residues (Fig. 6.6).

W57F was interrogated because it can DNA strand-separate but has very poor methylation (Fig. 8). W57F can base-flip, but likely has destabilized cofactor binding due to W57's interaction with SFG in the co-crystal structure (Fig. 8). This disfavors methylation and provides an indicates that base-flipping is capable in the absence of cofactor.

We also explored a model with WT CcrM and single-stranded DNA. CcrM can uniquely methylate ssDNA as well as dsDNA, therefore we wanted to understand CcrM base-flipping of ssDNA to better understand this process (Fig. 9). The WT/ssDNA model omits the DNA strand-separation step, but base-flipping is the slowest step in the pathway and rate-

determining for catalysis (Fig. 9). We opted for a 5-step model with ssDNA to account for the truncated pathway that CcrM undergoes with a substrate that does not require DNA strand-separation. The results with ssDNA can be useful to the study of other enzymes that modify single-stranded duplexes and rely on a base-flipping mechanism.

Global fitting was essential to derive the forward and reverse rate constants for the individual steps in all models (Table 6.1). Because of the series of steps with comparable rate constants, conventional equation-based data fitting is invalid. Confidence contour analysis shows that the models are well constrained by the data (Fig. 6.11). All models reveal that base-flipping is the rate-determining step in catalysis, although it should be noted that the rate of DNA methylation is not well defined because it appears to follow immediately after base-flipping. The mechanism of base-flipping in the context of DNA strand separation by CcrM could be insightful to understanding other enzymes that recognize double-stranded DNA and modify a single nucleotide.

Enzyme specificity is a function of all steps leading up to the first step that is kinetically irreversible, which can be identified by the highest overall barrier on free energ profile. In this case base flipping is the final specificity-determing step (k_4) since it is rapidly followed by methylation so that base-flipping is effectively irreversible. Accordingly, the specificity constant is defined approximately by $k_{cat}/K_m = K_1K_2K_3k_4$. CcrM achieves such high specificity through three unfavorable equilibirum steps involving DNA binding, enzyme conformational change, strand separation, with the final specificity step defined the rate of base flipping. Since each step preceeding base flipping is unfavorable, DNA that fails to reach the final state can be readily released. Base flipping represents the final checkpoint for DNA methlyation and it

depends on proper interactions of the ssDNA with the enzyme. Although bases-flipping is also rate limiting, that only defines how fast the reaction proceeded. The role of base-flipping as the final checkpoint in specificity is far more important.

VI: References

6.1. Bellamy,S.R.W., Krusong,K., and Baldwin,G.S. (2007) A rapid reaction analysis of uracil DNA glycosylase indicates an active mechanism of base flipping. *Nucleic Acids Research*, 35,5,1478-1487.

6.2. Matje,D.M., Zhou,H., Smith,D.A., Neely,R.K., Dryden,D.T.F, Jones,A.C., Dahlquist,F.W., and Reich, N.O. (2013) Enzyme-promoted base flipping controls DNA methylation fidelity. *Biochemistry*, 52, 1677-1685.

6.3. Su.T., Tock,M.R., Egelhaaf,S.U., Poon,W.C.K., and Dryden,D.T.F. (2005) DNA bending by M.EcoKI methyltransferase is coupled to nucleotide flipping. *Nucleic Acids Research,* 33,10, 3235-3244.

6.4. Allan,B.W., Reich,N.O., and Beechem,J.M. (1999) Measurement of the absolute temporal coupling between DNA binding and base flipping. *Biochemistry*, 38, 5308-5314.

6.5. Mashhoon,N. and Reich,N.O. (1993) Presteady state kinetics of an s-adenosylmethionine-dependent enzyme. *The Journal of Biological Chemistry*, 268, 13, 9191-9193.

6.6. Youngblood,B. and Reich,N.O. (2006) Conformational transitions as determinants of specificity for the DNA methyltransferase EcoRI*. The Journal of Biological Chemistry, 281, 37, 26821-26831.

6.7. Allan,B.W., Garcia,R., Maegley,K., Mort,J., Wong,D., Lindstrom,W., Beechem,J.M. and Reich,N.O. (1999) DNA bending by EcoRI DNA methyltransferase accelerates base flipping but compromises specificity. The Journal of Biological Chemistry, 274, 27, 19269-19275.

6.8. Neely,R.K., Daujotyte,D., Grazulis,S., Magennis,S.W., Dryden,D.T.F., Klimašauskas,S., and Jones,A.C. (2005) Time-resolved fluorescence of 2-aminopurine as a probe of base flipping in M.HhaI-DNA complexes. *Nucleic Acids Research*, 33, 22, 6953-6960.

6.9. Klimasauskas,S., Kumar,S., Roberts,R.J. and Cheng,X.D. (1994) Hhal methyltransferase flips its target base out of the DNA helix. Cell, 76, 357–369.

6.10. Reinisch,K.M., Chen,L., Verdine,G.L. and Lipscomb,W.N. (1995) The crystal structure of HaeIII methyltransferase covalently complexed to DNA: an extrahelical cytosine and rearranged base pairing. *Cel*l, 82, 143–153.

6.11. Goedecke,K., Pignot,M., Goody,R.S., Scheidig,A.J. and Weinhold,E. (2001) Structure of the N6-adenine DNA methyltransferase M.TaqI in complex with DNA and a cofactor analog. *Nature Struct. Biol.*, 8, 121–125.

6.12. Stivers,J.T. and Jiang,Y.L. (2003) A mechanistic perspective on the chemistry of DNA repair glycosylases. *Chemical Rev.*, 103, 2729–2759.

6.13. Yang.K., Matsika,S., and Stanley,R.J. (2007) 6MAP, a fluorescent adenine analogue, is a probe of base flipping by DNA photolyase. *Journal of Physical Chemistry B*, 111,35,10615-10625.

6.14. Hawkins,M.E., Pfleiderer,W., Balis.F.M., Porter,D., and Knutson.J.R. (1997) Fluorescence properties of pteridine nucleoside analogs as monomers and incorporated into oligonucleotides. *Analytical Biochemistry*, 244,1,86-95.

6.15. Hawkins,M.E., Pfleiderer,W., Jungmann,O., and Balis,F.M. (2001) Synthesis and fluorescence characterization of pteridine adenosine nucleoside analogs for DNA incorporation. *Analytical Biochemistry*, 298,2, 231-240.

6.16. Merkiene,E. and Klimasauskas,S. (2005) Probing a rate-limiting step by mutational perturbation of AdoMet binding in the HhaI methyltransferase. *Nucleic Acids Research*, 33,1,307-315.

6.17. Vilkaitis,G., Merkiene,E., Serva,S., Weinhold,E., and Klimasauskas,S. (2001) The mechanism of DNA cytosine-5 methylation Kinetic and mutational dissection of HhaI methyltransferase. The *Journal of Biological Chemistry*, 276,24,20924-20934.

6.18. Lindstrom,W.M., Jr, Flynn,J. and Reich,N.O. (2000) Reconciling structure and function in HhaI DNA cytosine-C-5 methyltransferase. The *Journal of Biological Chemistry*, 275, 4912–4919.

Chapter VII: The conserved and essential C-terminal domain, Loop-2B, and Loop-45 are found broadly in N^4/N^6-adenine DNA methyltransferases including human or animal pathogens

VII: Abstract

We have identified three CcrM protein moieties that are essential for DNA strand separation, substrate discrimination, and efficient catalysis: the C-terminal domain, Loop-2B, and Loop-45. Bioinformatics and BLAST searches within the UniProt database resulted in identification of a vast variety of proteins that contain some or all of these structural moieties. Interestingly, all of the identified proteins recognize GANTC. The C-terminal domain, Loop-2B, and Loop-45 are found broadly in N^4/N^6-adenine DNA methyltransferases, some of which are human or animal pathogens, across three Proteobacteria classes, three other phyla, and in *Thermoplasma acidophilum*, an Archaea. Several of the identified proteins have a putative or cognate restriction endonuclease. CcrM and orphan ortholog (BabI) stand out mechanistically from orthologs from RM systems (HinfI and M.Linc). CcrM and BabI have higher level of substrate discrimination than HinfI and M.Linc suggesting that enzymes from RM systems may have less stringint requirements for selectivity resulting in host protection than orphan enzymes that are resposible for gene expression and cell-cycle regulation. The results presented here provide a possible correlation between the unique DNA strand separation mechanism and high-fidelity recognition with the conserved protein moities that are found in other enzymes.

VII: Introduction

We have shown that CcrM from *Caulobacter crescentus* stands out from other characterized DNA Mtases in many ways. Regarding sequence, CcrM has a unique 83-amino acid C-terminal domain. Structurally, CcrM relies on a DNA strand separation of several bases within the recognition site rather than "base flipping" of solely the target base. Mechanistically, CcrM displays an unprecedented level of sequence discrimination and performs highly efficient methylation during processive catalysis. We resolved that DNA strand separation is the dominant discrimination step. Finally, target adenine base flipping follows DNA strand separation and is the rate-determining step rather than product release. We wanted to understand how the protein moieties and structural elements that we determined were essential for strand separation were conserved in other enzymes.

CcrM ortholog from *Brucella abortus,* as well as the β-class DNA MTase M.HinfI, both of which have the 83-residue C-terminal domain and methylate both ssDNA and dsDNA in GANTC sites are capable of DNA strand separation (Fig. 3.2)[2.3]. M.HhaII, which recognizes GANTC, lacks the C-terminal segment, and which shows no activity with single stranded DNA does not DNA strand separate (Fig. 3.2)[2.2,3.0]. Thus, CcrM orthologs are functionally and structurally like CcrM[2.2,3.0].

DNA strand separation is also referred to as strand-displacement in other enzyme systems. CRISPR-Cas9 gene editing enzymes rely on extensive strand-displacement in which the pairing between the target DNA and crRNA guide sequence requires initial recognition of a short protospacer adjacent motif (eg., 5'NGG3')[4.2,4.3]. Similarly to CcrM, the onset of strand displacement by CRISPR is endergonic and does not require ATP[3.20]. In contrast to CcrM,

CRISPR-Cas9 recognition is mediated through RNA/DNA hybridization rather than specific protein residues[3.20,4.2,4.3].

DNA strand separation is also observed at the initiation of transcription by the RNA polymerase sigma subunit[7.1,7.2]. The sigma subunit recognizes 5'TATAAT3' in promotor region coding strands and base-flips the underlined A and T out of the DNA duplex causing melting of all bases in the recognition site except for the 5'T which maintains Watson-Crick base pairing[7.1,7.2]. The flipped A and T are recognized by sigma subunit protein residues via base-specific interactions in a lock and key mechanism[7.1,7.2]. Similarly to CcrM, recognition and strand separation are coupled events that do not require ATP[4.1,7.1,7.2]. In contrast to CcrM, the flipped bases by the sigma subunit are recognized in the coding strand, which leaves the template strand in a single-stranded conformation that facilitates RNA transcription[7.1,7.2]. CcrM, recognizes and modifies only the target strand[2.5].

A newly described human ß class adenine methyltransferase (MettL3-MettL14 complex) methylates single stranded DNA and unpaired regions of double stranded DNA with lesions or mismatches with reduced activity and may rely on a strand separation mechanism[3.7,4.15].

Although the actual strand separation mechanisms may differ, the motifs and mechanisms that govern DNA strand-separation and substrate fidelity are relevant to other proteins such as CRISPR-Cas9 and RNA polymerase sigma factor[5.21,5.22,7.1,7.2]. The mechanisms described for CcrM, CRISPR, and sigma factor, in which an enzyme strand-separates double-stranded DNA, represent a new DNA recognition mechanism. The results presented here for CcrM can be useful for understanding substrate discrimination in other

systems such as CRISPR/Cas9 where gene editing tools require overcoming barriers of selectivity between enzymes and DNA substrates.

In this chapter, we identify a vast variety of proteins that contain some or all of the structural moieties that we have shown are essential to CcrM's mechanism of DNA strand separation and catalysis.

VII: Materials and methods

BLAST searches and multiple sequence alignments

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.

Phylogenetic tree

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), neutral (purple).

Moraxella Lincolni

The *Moraxella Lincolni* plasmid was constructed by Twist Biosciences. The plasmid was transformed and the protien was purified as previously described for WT CcrM.

Specificity experiments

Substrate specificity were perfomed for CcrM and orthologs via radiochemical methods. All reactions consisted of CcrM[2.5μM], DNA substrate[1.0μM], SAM[15 μM], in CcrM reaction buffer and one 120-minute timepoint was spotted in triplicate.

Burst experiments

Radiochemical steady-state burst experiments with CcrM orthologs consisted of CcrM[200nM], DNA[3 μM], SAM[15 μM], in CcrM reaction buffer and three independent reactions were averaged. Error bars represent standard error and linear regression models were fit in Graphpad Prism 10.0.2.

VII: Results

The C-terminal domain is conserved across human and animal pathogens

The multiple sequence alignment from the 83 amino acid C-terminus of CcrM reveals conservation amongst a variety of N^4/N^6-DNA methyltransferases from alphaproteobacteria

(Fig. 7.1A, top). The organisms listed have diverse biological capabilities and some are human or animal pathogens. The variety include *Rhodobacter massiliensis*[7.3], *Pannonibacter pragmitetus*[7.4], *Inquilinus limosus*[7.5], *Haematospirillum jordaniae*[7.6], and *Methylocapsa palsarum*[7.7], a bacteria that contributes to biogas production *Rhodopseudomonas faecalis*[7.8], a plant-growth promoting bacteria *Azospirillum sp. RU38E*[7.9] two methanotrophs *Methylocapsa palsarum*[7.10] and *Methylocella silvestris*[7.11], and *Tepidicaulis Marinus,* which reduces *nitrate*[7.12]. Also *highlighted are a nitrogen fixing* alphaproteobacterium *Hartmannibacter diazotrophicus*[7.13] and *Martelella endophytica*, which exhibits inhibitory activity against fungal plant pathogens[7.14].

The multiple sequence alignments using the 80 amino acid C-terminal segment of M.HinfI (Figure 7.1A, bottom) reveals conservation amongst a variety of DNA methyltransferases in the N^4/N^6-methyltransferase family including the highlighted human pathogens *Mycoplasma girerdii*[7.15], *Bartonella bacilliformis*[7.16], *Bartonella tamaie*[7.17], *Capnocytophaga canimorsus*[7.18], *Helicobacter pylori*[7.19], *Brucella abortus*[7.20], and *Haemophilus influenzae*[7.21], animal pathogens *Mycoplasma nasistruthionis*[7.22], *Ureaplasma diversum*[7.23], *Campylobacter sputorum biovar sputorum*[7.24], *Brachyspira catarrhinii*[7.25], *Mycoplasma californicum*[7.26], *Moraxella lincolnii*[7.27], *Mycoplasma falconis*[7.28], *Moraxella macacae*[7.29] and *Brucella melitensis*[8.16,8.17],. 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, and Tenericutes) as well as an archaea. These organsms are highlighted in the M.HinfI phylogenetic tree (SI Fig. $8^{3.0}$).

The logo shows that the residues we mutated (E280, G305, S315, H317, N330, W332, and R350) are highly conserved amongst CcrM and M.HinfI (Fig. 7B). 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 (Fig. 3.5).

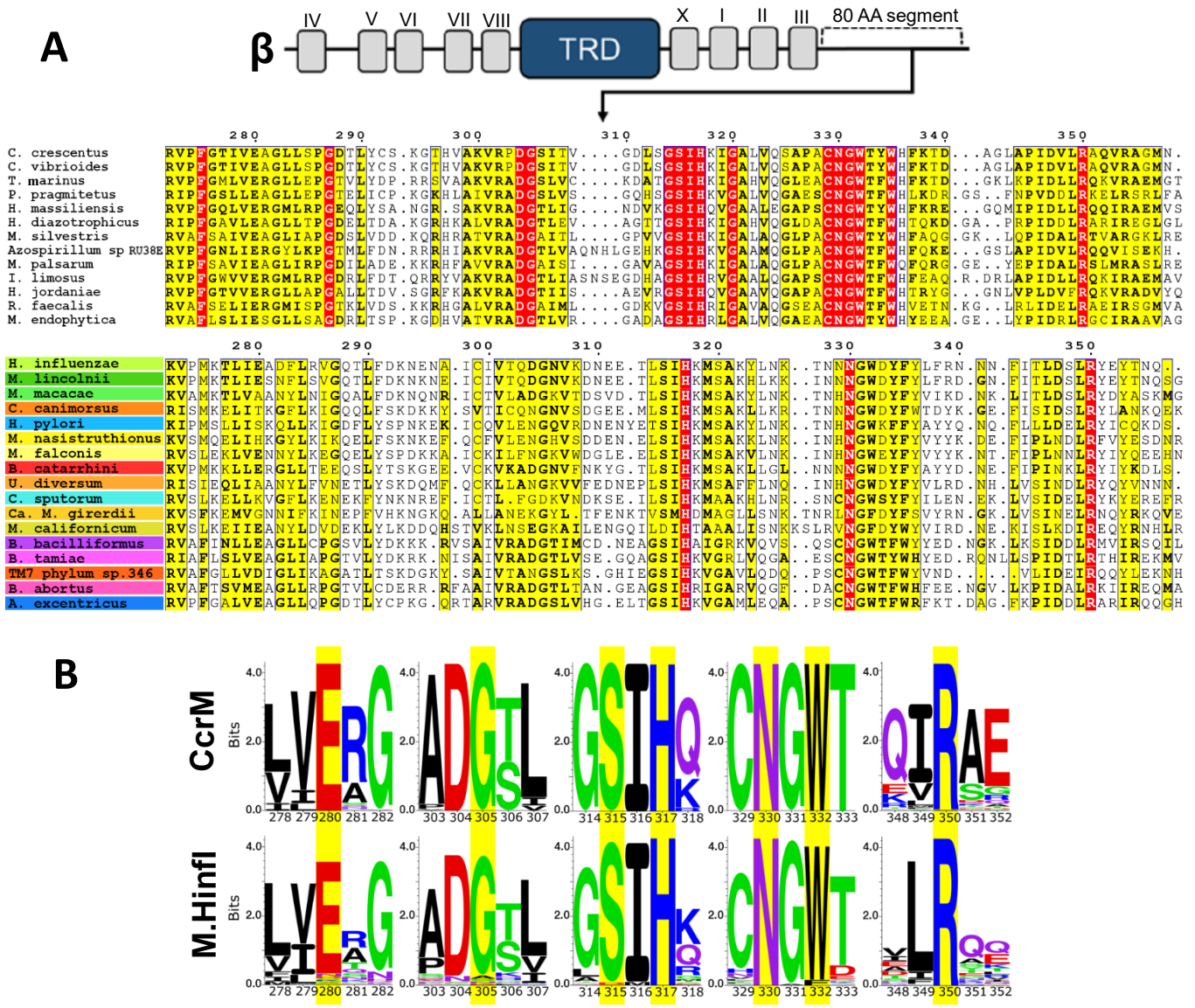


Figure 7.1. The 80 amino acid C-terminal segment is widespread amongst N4 and N6-DNA methyltransferases. A. The motifs of β-class methyltransferases are depicted in relation to the C-terminal 80-residue segment[2.1]. Proteins were collected through UniProt BLAST searches of the 80 amino acid segment (dashed line) from β–class CcrM (*C. crescentus,* top) and ɣ-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 amongst N^4/N^6-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 (SI Fig. $8^{3.0}$). 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), *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), *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 5 residues are shown. Full sequence logos are provided in SI Fig. $6^{3.0}$. The colors represent the chemical properties of each residue; polar residues (green), acidic (red), basic (blue), hydrophobic (black), neutral (purple). Yellow highlighted residues were mutated in this study. The logos were made with SeqLogo.

Loop-2B and Loop-45 are conserved across human and animal pathogens

We sought to further understand the importance of these structural elements that regulate strand separation (Loop-2B and Loop-45) by looking at these motifs in CcrM orthologs identified from the C-term BLAST search. Loop-2B and Loop-45 are inserted within the separated DNA strands in the cocrystal structure[2.5]. These residues and loops are conserved across DNA methyltransferases in human and animal pathogens, including *Mycoplasma falconis, Mycoplasma nasistruthionis, Campylobacter sputorum, Moraxella lincolnii, Haemophilus influenzae, Moraxella macacae, Brachyspira catarrhinii, Helicobacter pylori,*

Capnocytophaga canimorsus, Ureaplasma diversum, Mycoplasma californicum, Brucella melitensis, Brucella abortus, Bartonella tamiae, Bartonella bacilliformis and certainly others (Fig. 7.2). All the orthologs have the CcrM 83 amino acid C-terminal domain and multiple sequence alignment reveals the high conservation of Loop-2B and Loop-45 (Fig. 7.3). 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). Organisms that contain a CcrM ortholog with a C-term, Loop-2B, and Loop-45 are found across many orders of bacteria, as well as some eukaryotes and some archaea (Fig. 7.3). The variation of these loops is shown via a SeqLogo (Fig. 7.4).

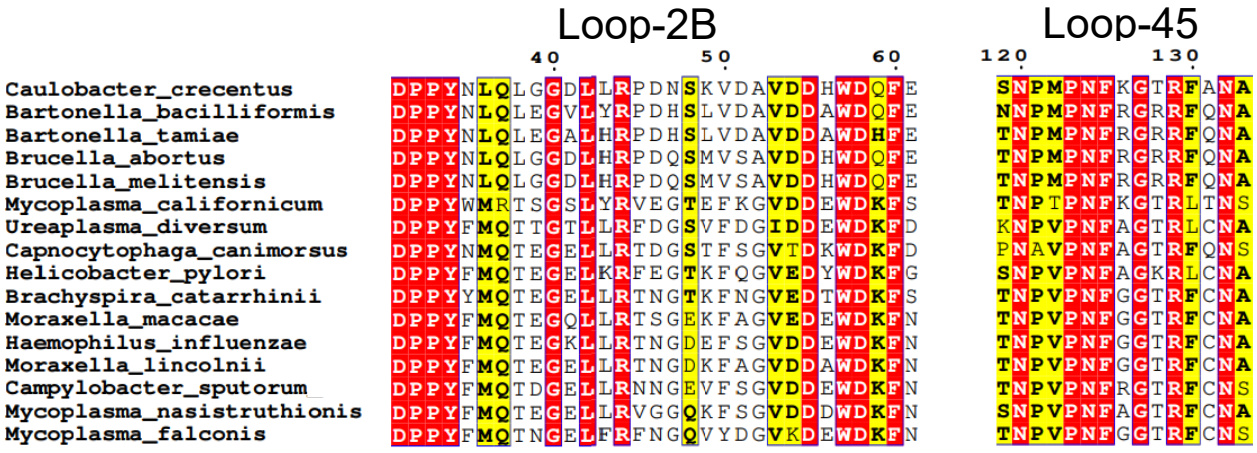


Figure 7.2. Loop-2B and Loop-45 are highly conserved in CcrM orthologs across pathogenic orthologs. The C-terminal domain was used as the search seed for this BLAST. Multiple sequence alignment was done using Clustal Omega and the MSA was visualized using ESPript 3.0 with a similarity color scheme global score of 0.7. Numbering is based on the CcrM sequence from *Caulobacter crescentus*.

Orphan orthologs are more discriminating than orthologs from restriction-modification systems

We also sought to understand how orthologs within different biological contexts (orphans or part of RM systems) differed mechanistically. We carried out specificity analysis with a CcrM orphan ortholog and related RM enzymes (Fig. 7.5a). BabI is an orphan ortholog, while HinfI and *Moraxella lincolnii* (M.Linc) are orthologs within RM systems. CcrM and BabI are more discriminating against non-cognate substrates than HinfI and M.Linc (Fig. 7.5a). CcrM and BabI do not display a steady-state burst, indicating that a step preceeding methylation is rate-limiting (Fig. 7.5), unlike M.Linc and HinfI which show the typical burst, showing that product release is likely rate-limiting (Fig. 7.5). The comparison between orphan enzymes and enzymes from RM systems supports our hypothesis that enzymes from RM systems may have less stringint requirements for selectivity than orphan enzymes that are resposible for gene expression and sophisticated cell-cycle regulation.

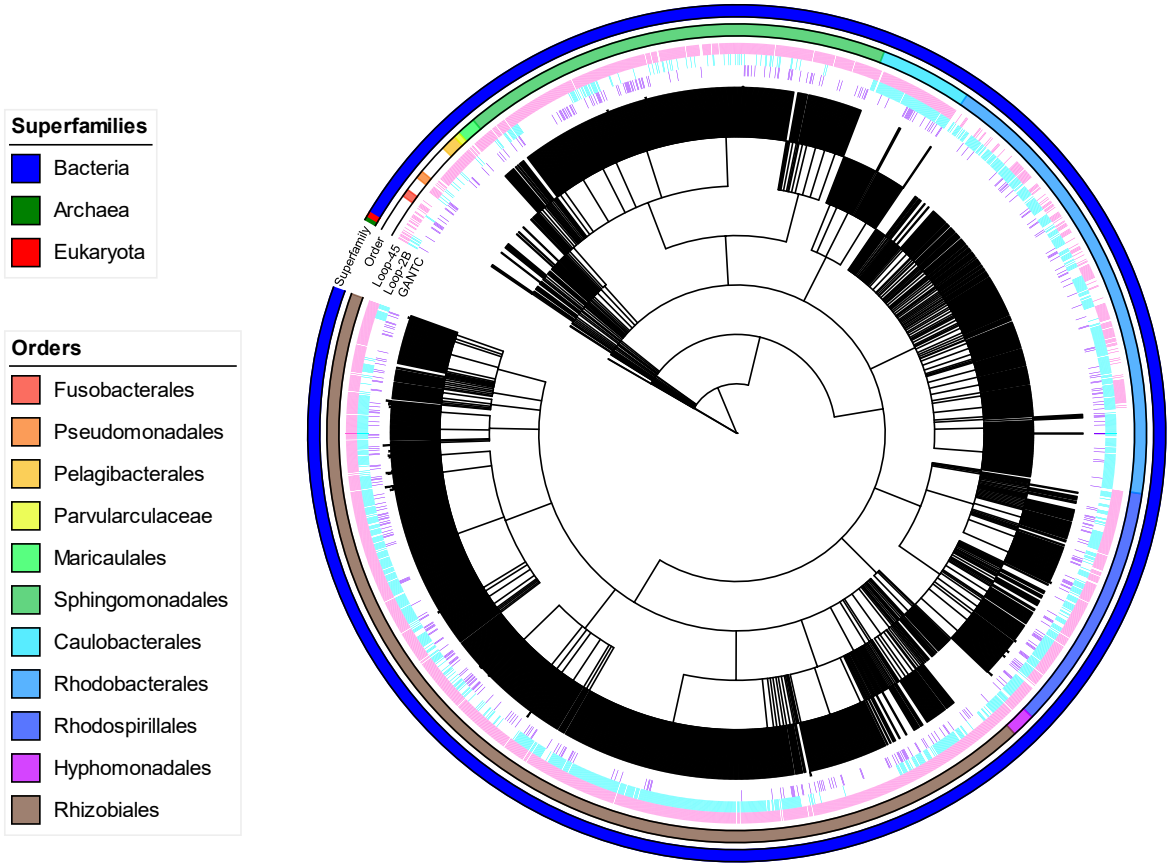


Figure 7.3. Presence of Loop Motif Sequences in Proteins with 80 AA consensus sequence. This Phylogenetic tree displays the organisms with proteins that appeared in a ScanProsite search using the 83 AA consensus sequence: . [RK]-[VI]-[PAS]-[FML]-x(4)-[EDNS]-x-[GNDH]-x(4)-[GN]-x-x-[LVFI]-x(8,12)-[ACV]-x(4,6)-[DNGS]-[GAS]-x(9,16)-S-I-H-x(12,14)-N-G-[WF]-x(14,16)-[DN]-x-x-R. The Superfamilies of the organisms are indicated in the outermost ring of annotations. The Orders containing more than 5 organisms are indicated by the second ring. The presence of Loop-45 is indicated by a pink line in the third ring, and the presence of Loop 2B is indicated by a cyan line in the fourth ring, and confirmed GANTC recognizers (REBASE) are indicated by a violet line in the innermost ring of annotations. The Tree was created using NCBI commontree and visualized with iTOL. All nodes are collapsed to the species level. All of the methyltransferases with known recognition sequences in this tree recognize GANTC. This annotation was provided by REBASE. The unlabeled methyltransferases have unidentified recognition sequences and are not annotated in REBASE. There are no other recognition sequences that we have identified other than GANTC.

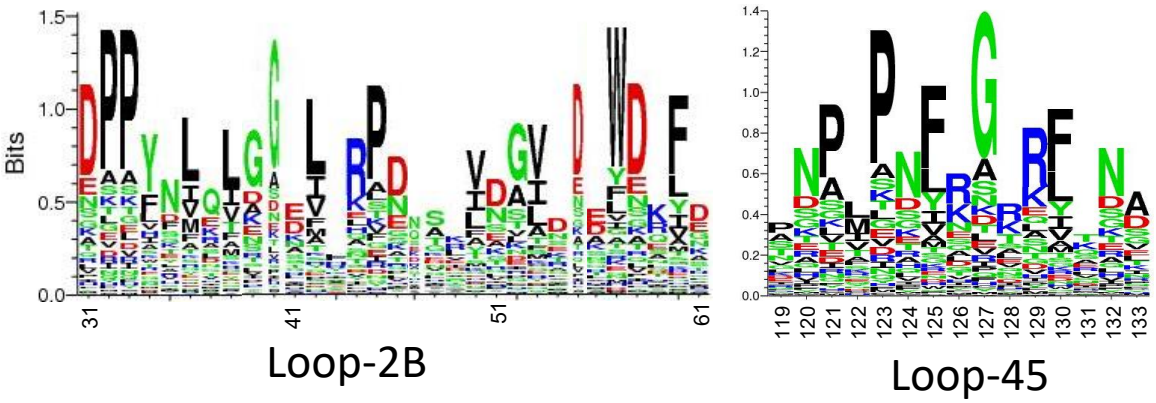


Figure 7.4. Motif organization and sequence logos of Loop-2B and Loop-45. The logos were constructed from the methyltransferases in the phylogenetic tree. Residue labeling is based off the CcrM sequence. The sequences were aligned with Clustal Omega and the logos were constructed with Seq2Logo.

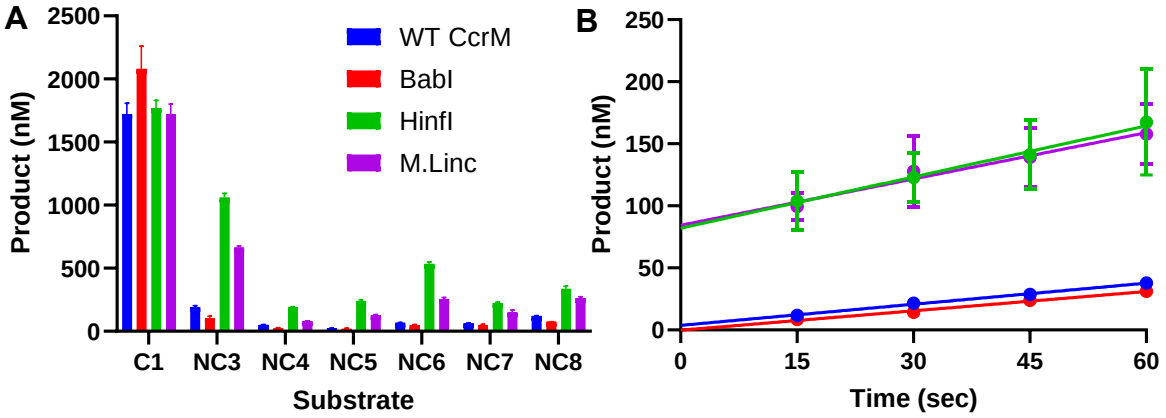


Figure 7.5. CcrM orphan orthologs are more discriminating than orthologs from RM systems, and orthologs from RM systems display a steady-state burst. A. CcrM and BabI are orphan DNA methyltransferases and are highly discriminating against non-cognate substrates. RM enzymes, HinfI and M.Linc, are less discriminating. Each reaction consisted of CcrM[2.5μM], DNA substrate[1.0μM], SAM[15 μM] and one 120-minute timepoint was spotted in triplicate. **B.** Steady-state burst experiment with CcrM orthologs. Radiochemical steady-state burst experiments with CcrM orthologs consisted of CcrM[200nM], DNA[3 μM], SAM[15 μM], in CcrM reaction buffer and three independent reactions were averaged.

VII: Discussion

Using the CcrM and M.HinfI protein sequences, we searched for other organisms which have a protein with similar C-terminal segments and found N^4/N^6-DNA MTases with this segment in broadly distributed organisms. The residues we investigated via mutational analysis in CcrM (Figure 7.1) show a high degree of conversation across this entire group which spans multiple Phyla (SI Fig. $8^{3.0}$). The segment found in M.HinfI is observed in diverse bacteria, including one Archaea example (SI Fig. $8^{3.0}$). 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 (SI Fig. $8^{3.0}$). 477/500 of the proteins in the M.HinfI BLAST are β-class DNA methyltransferases based on the organisation of conserved motifs[2.1]. Three proteins could not be assigned due to problems in identifying the catalytic domains.

We previously showed that the C-terminal domain of CcrM is essential for binding and DNA strand separation[3.0]. The C-terminal domain is dispersed across CcrM orthologs in several orders of bacteria, some eukaryotes and some archaea (Fig. 7.1). In addition to the C-term, Loop-2B and Loop-45 are also essential for strand separation (Fig. 5.4, 5.6, 5.7, 5.8, 5.9) and these loops are highly conserved across a variety of bacteria (Fig. 7.2).

Organisms that contain a CcrM ortholog with a C-term, Loop-2B, and Loop-45 are found across many orders of bacteria, as well as some eukaryotes and some archaea (Fig. 7.3). Together, the C-terminal domain, Loop-2B, and Loop-45 are essential for generating strand-separation, stabilizing the strand-separated intermediate, and in these steps regulating CcrM's extreme substrate fidelity. This unique substrate discrimination mechanism is complex and

useful for understanding how enzymes have evolved unique mechanisms for regulating cellular processes.

CcrM and orphan ortholog (BabI) stand out mechanistically from orthologs from RM systems (HinfI and M.Linc). CcrM and BabI's higher level of substrate discrimination and lack of a steady-state burst (Fig. 7.5) indicate that MTases which display a slow step prior to product release can display steps contributing to discrimination which impact turnover. Typically, a steady-state burst indicates rapid enzyme-bound product formation followed by product release which is rate-limting. HinfI and M.Linc's steady state burst suggests that the less stringent discriminatory steps are either faster than CcrM or that product release is significantly slower for enzymes from RM systems. Thus, enzymes from RM systems may have less stringint requirements for selectivity resulting in host protection than orphan enzymes that are resposible for gene expression and cell-cycle regulation.

VII: References

7.1. Liu,X., Bushnell,D.A., and Komberg,R.D. (2011) Lock and key to transcription: σ-DNA Interaction, *Cell*, 147, 6, 1218-1219.

7.2. Feklistov,A. and Darst,S.A. (2011) Structural basis for promoter-10 element recognition by the bacterial RNA polymerase σ subunit. *Cell*, 147, 6, 1257-1269.

7.3. Greub, G., Raoult, D., (2003) Rhodobacter massiliensissp. nov., a new amoebae-resistant speciesisolated from the nose of a patient, *Research in Microbiology*, 154: 631-635.

7.4. Wang, M., Zhang, X., Jiang, T., Hu, S., Yi, Z., Zhou, Y., Ming, D., Chen, S., (2017) Liver abscess caused by Pannonibacter phragmitetus: case report and literature review, 4, 48, 1-4.

7.5. McHugh, K.E., Rhoads, D.D., Wilson, D.A., Highland, K.B., Richter, S.S., Procop, G.W., (2016) Inquilinus limosusin pulmonary disease: case report and review of the literature, *Diagnostic Microbiology and Infectious Disease*, 86: 446-449.

7.6. Hovan, G., Hollinger, A., (2018) Clinical Isolation and Identification of Haematospirillum jordaniae, *Emerging Infectious Diseases* 24, 10, 1955-1956.

7.7. Benítez-Páez et al. (2020) Breast-milk microbiota linked to celiac disease development in children: a pilot study from the PreventCD cohort, *Frontiers in Microbiology* 11, 1335, 1-12.

7.8. Abendroth, C., Latorre-Pérez, A., Porcar, M., Simeonov, C., Luschnig, O., Vilanova, C., Pascual, J., (2019) Shedding light on biogas: a transparent reactor triggers the development of a biofilm dominated by Rhodopseudomonas faecalis that holds potential for improved biogas production, *BioRxiv*.

7.9. Casán, F., Diaz-Zorita, M., (2016) Azospirillum sp. in current agriculture: From the laboratory to the field, Elsevier, *Soil Biology and Biochemistry,* 103: 117-130.

7.10. Miroshnikov et al. (2020) Draft Genome Sequence of Methylocapsa palsarum NE2T, an Obligate Methanotroph from Subarctic Soil, *American Society for Microbiology Prokaryotes* 5, 24, 1-2.

7.11. Dedysh, S.N., Knief, C., Dunfield, P.F., (2005) Methylocella Species Are Facultatively Methanotrophic, *Journal of Bacteriology* 187, 13, 4665-4670.

7.12. Takeuchi et al. (2015) Tepidicaulis marinus gen. nov., sp. nov., a marine bacterium that reduces nitrate to nitrous oxide under strictly microaerobic conditions, *International Journal of Systematic Microbiology*, 65, 1749-1754.

7.13. Suarez, C., Ratering, S., Geissler-Plaum, R., 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, *International Journal of Systematic and Evolutionary Microbiology* 64, 3160–3167.

7.14. Bibi, F., Chung, E.J., Khan, A., Jeon, C.O., Chung, Y.R., (2013) Martelella endophytica sp. nov., an antifungal bacterium associated with a halophyte, *International Journal of Systematic and Evolutionary Microbiology* 63, 2914–2919.

7.15. Costello, E.K., Sun, C.L., Carlisle, E.M., Morowitz, M.J., Banfield, J.F., Relman, D.A., 2017. *Candidatus Mycoplasma girerdii* replicates, diversifies, and cooccurs with Trichomonas vaginalis in the oral cavity of a premature infant, *Scientific Reports* 7: 3764.

7.16. Gomes, C., Martínez-Puchol, S., Ruiz-Roldán, L., Pons, M.J., del Valle Mendoza, J., Ruiz, J. (2016) Development and characterization of highly antibiotic resistant *Bartonella bacilliformis* mutants, *Scientific Reports* 6:33584.

7.17. Kosoy et. al., (2008) *Bartonella tamiae* sp. nov., a Newly Recognized Pathogen Isolated from Three Human Patients from Thailand, *Journal of Clinical Microbiology* 46, 2, 772-775.

7.18. Mader, N., Lührs, R., Langenbeck, M., Herget-Rosenthal, H., (2020) *Capnocytophaga canimorsus* – a potent pathogen in immunocompetent humans – systematic review and retrospective observational study of case reports, *Infectious Diseases* 52,2, 65-74.

7.19. Beydoun, M.A., Beydoun, H.A., Weiss, J., Hossain, S., El-Hajj, Z.W., 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, *Molecular Psychiatry*.

7.20. Atluri, V.L., Xavier, M.N., de Jong, M.F., den Hartigh, A.B., Tsolis, R.M., (2011) Interactions of the human pathogenic *Brucella* Species with their hosts, *Annual Review of Microbiology*, 65:523-541.

7.21. Price, E.P., Sarovich, D.S., Nosworthy, E., Beissbarth, J., Marsh, R.L., Pickering, J., Kirkham, L.S., Keil, A.D., Chang, A.B., Smith-Vaughan, H.C., (2015) *Haemophilus influenzae*: using comparative genomics to accurately identify a highly recombinogenic human pathogen, *BMC Genomics*, 16:641.

7.22. Spergser, J., Botes, A., Nel, T., Ruppitsch, W., Lepuschitz, S., Langer, S., Ries, S., Dinhopl, N., Szostak, M., Loncaric, I., Busse, H., (2019) *Mycoplasma nasistruthionis* sp. nov. and Mycoplasma struthionis sp.nov. isolated from ostriches with respiratory disease, *Systematic and Applied Microbiology* 43, 126047.

7.23. Andrade et al. (2020) Multilocus sequence typing characterizes diversity of *Ureaplasma diversum* strains, and intra-species variability induces different immune response profiles, *BMC Veterinary Research*, 16:163.

7.24. Miller, W.G., Yee, E., Chapman, M.H., Bono, J.L., (2017) Comparative Genomics of All Three *Campylobacter sputorum Biovars* and a Novel Cattle-Associated *C. sputorum* Clade, *Genome Biol. Evol.* 9(6):1513-1518

7.25. Phillips, N.D., La, D., 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, *Anaerobe* 59: 8-13.

7.26. Hewicker-Trautwein, M., Feldmann, M., Kehler, W., Schmidt, R., Thiede, S., Seeliger, F., Wohlsein, P., Ball, HJ., Buchenau, I., Spergser, J., Rosengarten, R. (2002) Outbreak of pneumonia and arthritis in beef calves associated with *Mycoplasma bovis* and *Mycoplasma californicum Veterinary Record* 151, 699-703.

7.27. 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, International Journal of Systematic Bacteriolo*gy 43, 3, 474-481.

7.28. Lierz, M., Hagen, N., Lueschow, D., 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 Diseases* 52,1, 94-99.

7.29. Whitehouse, C.A., Chase, K., Embers, M.E., Kulesh, D.A., Ladner, J.T., Palacios, G.F., 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.

Chapter VIII: Small molecule screening from MMV libraries for CcrM inhibition: potency, target selectivity, mechanism of action and bacterial cell studies.

VIII. Abstract

High profile human pathogens are becoming increasingly multi-resistant to a variety of strong antibiotics which pose a serious health risk to humans worldwide. Therefore, identification and development of novel antibiotic therapeutics is in high demand. DNA methyltransferases that regulate cell cycle progression are validated targets for small molecules because several have been shown to be essential for bacterial viability[8.1]. Here, CcrM was used as a target for small molecule inhibition from several MMV (Medicines for Malaria Venture) libraries. The MMV libraries screened for CcrM inhibition were the Global Health Priority Box, the Pathogen Box V2, and the Covid Box. The MMV libraries were screened manually via radiochemical methods and inhibitors were identified when <20% methylation was observed compared to uninhibited CcrM. 29 hits were identified out of 621 compounds resulting in a 4.7% hit rate. The 29 hits were investigated further and were prioritized based on potency and selectivity. Suramin was identified as the most potent CcrM inhibitor with an IC_{50} of 1.8 μM. Suramin was not selective for CcrM over human DNMT3A; Suramin also has an IC_{50} of 1.8 μM for DNMT3A. This study serves as a basis for the identification and repurposing of small molecules that could be modified for improved potency and selectivity of inhibition of CcrM for novel antibiotic development.

VIII. Intro

Antibiotics typically have a bacterial-selective mechanism of action. For example, Penicillins have been shown to inhibit bacterial cell wall synthesis by interacting with penicillin binding proteins which results in bacterial cell lysis and death[8.2]. Another mechanism of action used by Kanamycin, Streptomycin, Tetracyclines, and other antibiotics inhibit the ribosomal 30S subunit which disrupts protein synthesis[8.3,8.4]. Both examples rely on inhibition of bacteria-specific cellular targets, therefore, limiting the off-target human toxicity. However, bacterial pathogens are becoming increasingly resistant to these classes of traditional antibiotics[.5,8.6]. Therefore, development of selective antibiotics with novel mechanisms of action is increasingly pertinent. N^6-adenine DNA methyltransferases that regulate cell cycle genes and are essential to pathogenesis and cell viability are validated targets for novel antibiotics, however, the challenge of selectivity for bacterial DNA Mtases over human DNA MTases is a major barrier to drug development[8.7,8.8,8.9,8.10].

Most DNA methyltransferases and protein methyltransferases catalyze the transfer of a methyl group from methyl-donor SAM (S-adenosylmethionine) to DNA or protein substrates, respectively (Fig. 8.1)[8.11,8.12]. Historically, SAM-dependent DNA methyltransferase inhibitors have been SAM analogs that compete with SAM in the active site, thus disabling SAM binding and subsequent methyl-transfer[8.10,8.13,8.14]. Human DNA methyltransferases are known targets for anti-cancer therapeutics and a variety of SAM-analogs have been developed[8.15]. Often, these molecules have off-target specificity because of a conserved SAM-binding fold[8.12]. Therefore, development of non-SAM-analog inhibitors with high selectivity for a single target is of high interest.

In this study, we use CcrM from *Caulobacter crescentus* as a model for preliminary discovery of small molecule inhibitors. *Caulobacter crescentus* is not a pathogen, however, our bioinformatic efforts (Chapter 7) identified pathogens that contain a CcrM homolog. CcrM from *Caulobacter crescentus* is structurally homologous to CcrM from *Brucella melitensis* (Fig. 8.2). *Brucella melitensis* is a human and animal pathogen and CcrM has been shown to be critical for *Brucella* viability[8.15,8.16].

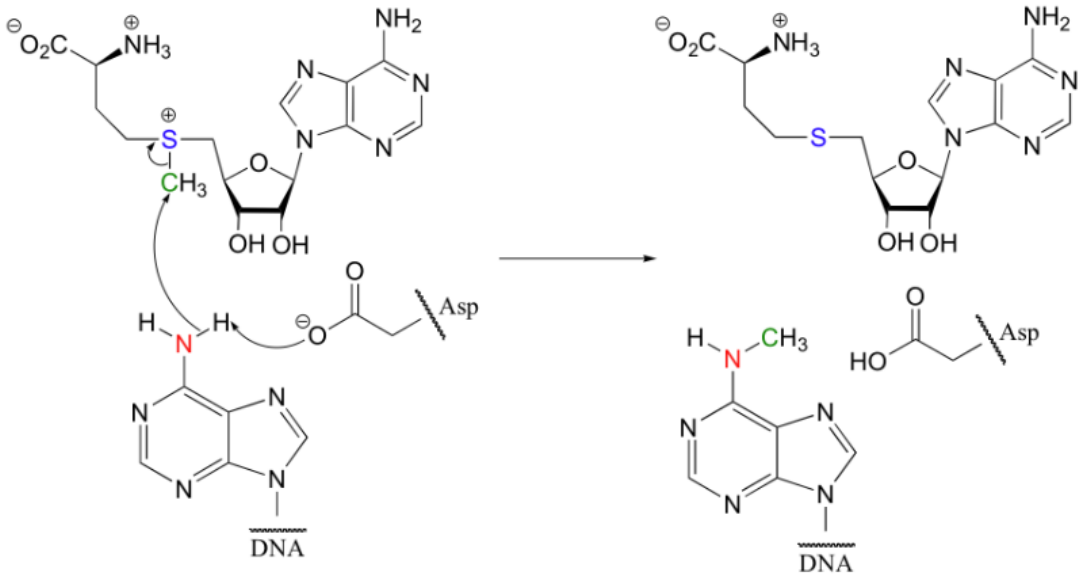


Figure 8.1. Methylation of N^6-adenine from SAM is an enzyme-catalyzed S_N2 mechanism. A basic protein residue in the active site of DNA methyltransferases deprotonates N^6 resulting in increased nucleophilicity of the adenine towards the methyl-thiol in SAM.

Small molecule libraries with high chemical diversity are a great starting point for drug-target discovery and validation. Small molecules have a near-infinite chemical space, while protein folds have a finite chemical space. Therefore, it has become common practice in the drug design industry to repurpose drugs by screening diverse libraries of small molecules for

novel targets. The MMV libraries contain drugs that were previously developed for malaria, tuberculosis, dengue, HIV, tumors, leukemia, fungi, and several other disease sets.

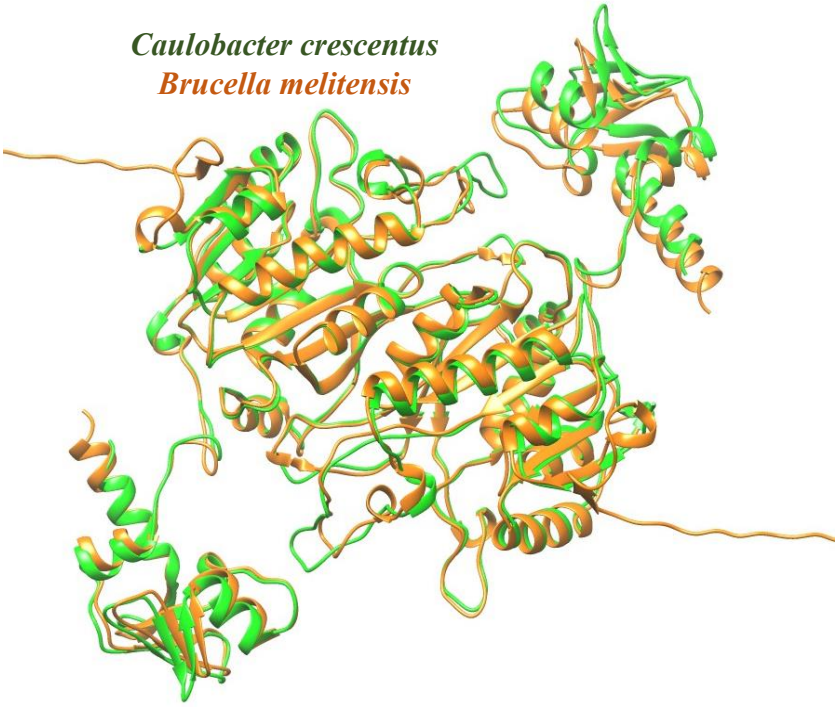


Figure 8.2. Superimposition of CcrM from *Caulobacter crescentus* and *Brucella melitensis* reveals structural homology. Alphafold was used to predict the Apo structure of CcrM from *Caulobacter crescentus* (green) and *Brucella melitensis* (orange). Structural images were made and in UCSF Chimera.

Here, we screened MMV small molecule libraries against CcrM in search of a potent, selective, non-SAM-analog inhibitor. Our criteria for molecule selection were high potency, high selectivity, not a DNA intercalator, and not a SAM-analog. This small molecule screening study of CcrM is a useful basis for continued efforts. We discovered some potent compounds but did not find a compound that met the criteria of selectivity. Therefore, further studies of screening larger libraries or synthesis of analogs could yield a meaningful antibiotic through means of homologous CcrM inhibition in a high profile and antibiotic resistant pathogen.

VIII. Methods

Radiochemical screening

Hits were determined by a single turnover radiochemical assay. Reactions were done in [100]nM dsC1, [250]nM WT CcrM,[15] µM (1:50) SAM, [150]µM compound, 2mM DTT, 0.2mg/ml BSA, reaction buffer (1 M HEPES, 10 mM EDTA, 200 mM NaCl, pH 8) with a final DMSO concentration of 1.5%. Background measurements with 1.5% DMSO withheld DNA and were subtracted from each data point. A standard control reaction contained 1.5% DMSO and no compound. Background and standard measurements were collected in triplicates and reactions with compound were collected as single data points, all spotted at a singular 30-minute time point with a 5ul spot size. Hits were identified when >60% inhibition of CcrM was observed.

IC_{50} determination

IC_{50} values were determined for hit compounds via radiochemical initial velocity methods. Conditions for CcrM IC_{50} assay were [50]nM WT CcrM, [1]µM dsC1, [15]µM SAM, [0-X]µM compound (X = concentration of compound where 100% inhibition). 5µL were spotted in triplicate at a single 20 minute timepoint.Conditions for DNMT3A IC_{50} assays were [150]nM DNMT3A, [10]µM poly dI-dC, [15]µM SAM, [0-X]µM compound (X = concentration of compound where 100% inhibition) . Backgrounds contained in 1.5% DMSO and witheld DNA. 15uL spot sizes in duplicate at a one hour time point. Data were fit to a sigmoidal dose-response model in GraphPad Prism 10.0.2 to determine the IC_{50} values; Y=Bottom + (Top-Bottom)/(1+10^((LogIC50-X)*HillSlope)).

Anisotropy

Conditions consisted of [0, 150, 500]nM WT CcrM, 10nM FAM-DNA, 50µM compound, 15µM SAH, in CcrM reaction buffer with 1.5% DMSO. 100µl samples were loaded onto Costar 96-well flat bottom black plates and scanned on a Tecan SPARK microplate reader equipped with polarizing filters. The excitation and emission wavelengths were 485nm and 535nm, respectively with 20nm bandwidths.

SAM vs Suramin competition assay

Multiple-turnover conditions consisted of 50nM CcrM, 1µM dsC1, [0.05, 0.1, 0.5, 1.0, 2.5, 5.0, 10, 15]µM SAM, [0, 0.5, 1.0, 3.0, 5.0]µM Suramin, and 2.5% DMSO in CcrM reaction buffer. One timepoint was collected in triplicate for each sample and all samples were background subtracted. Data were transformed to double-reciprocol plots and then fit to linear regression models in GraphPad Prism 10.0.2 to determine the point of intersection.

Pseudo in-vivo inhibition of CcrM with suramin (Ecoli cell studies)

5mLs of LB/Kanamycin (30ug/ml) was inoculated with glycerol stock stab for WT and W332A growths. The *Ecoli* DE3 expression cells contained Pet-28a plasmids with the CcrM gene. Liquid cultures grew for 3 hrs and then the entire 5mLs of cells were pelleted and the plasmid DNA was purified using an Agilent miniprep kit. Purified plasmids were digested with HinfI to challenge the methylation state of GANTC sites within the pet-28a plasmid. Digested products were run on a precast 2.5% agarose gel with GelRed. The NEB DNA broad range

marker (ladder), undigested plasmids, digested plasmids, and controls were run for 1 hour at 100 V in TBE buffer, and imaged on an EZ Gel Doc imager. After assay validation, suramin [150µM] was added to the liquid culture 3 hr growth and controls were grown with 2.5% DMSO.

VIII. Results

Initial screening of MMV libraries resulted in 29 hits out of 621 compounds resulting in a 4.7% hit-rate. Hits were identified via radiochemical methods as compounds that inhibited 80% or more of CcrM methylation activity. An initial screen of the MMV Pathogen Box V2 Plate B displays an example where we identified 4 compounds as hits that showed 10-25% product turnover compared to the DMSO control (Fig. 8.3).

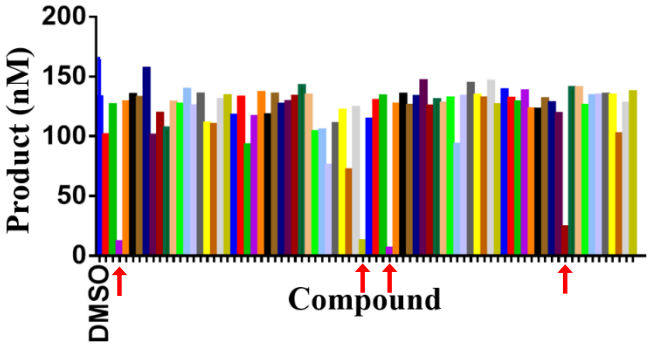
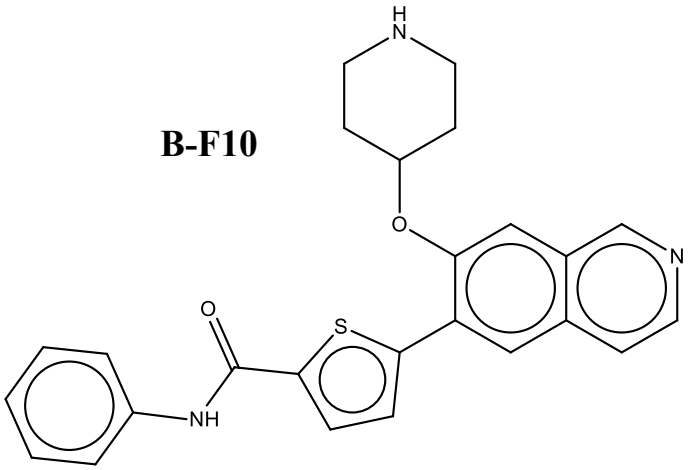


Figure 8.3. Initial screen of MMV Pathogen Box V2 Plate B shows 4 small molecules that were identified as CcrM inhibitors via radiochemical screening. Red arrows point to the hit compounds. Hit compounds showed <20% product formation compared to the 1.5% DMSO control.

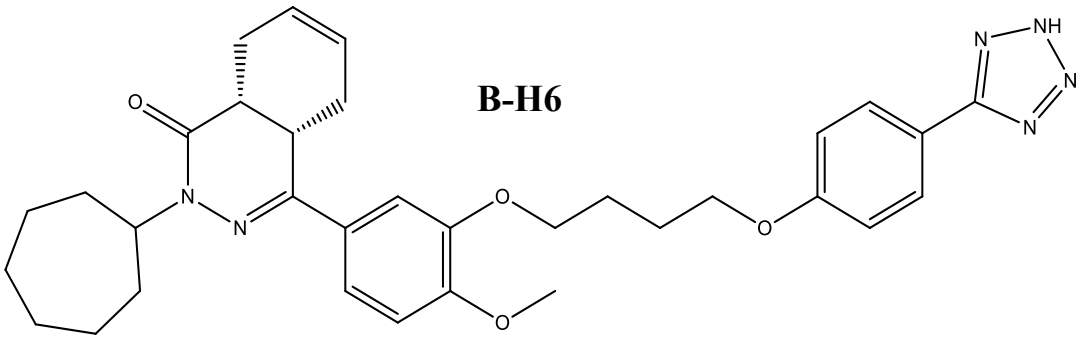
The MMV Pathogen Box V2 consisted of plates A-E and 6 compounds were identified as hits: B-F10, B-H6, C-C7, C-H9, B-D8 (Suramin), and E-F7 (Fig. 8.4). The compounds identified as hits are structurally dissimilar (Fig. 8.4). Hit compounds from the MMV Pathogen

Box V2 show dose-dependent inhibition on CcrM methylation (Fig. 8.5) and these compounds were investigated for mechanism and potency. We first wanted to rule out any DNA intercalators from this list because we wanted to identify CcrM-specific inhibitors. Equilibrium binding anisotropy was used to possibly determine if the compounds interfered with CcrM/DNA binding either through inhibition of DNA or protein (Fig. 8.6). Compounds B-D8, B-H6, and C-C7 resulted in no change in anisotropy indicating either prtoein or DNA inhibition (Fig. 8.6). Compounds B-F10, C-H9 and E-F7 show increases in anisotropy indicating that are less than the DMSO control suggesting that complex association is occuring but is disrupted (Fig. 8.6).

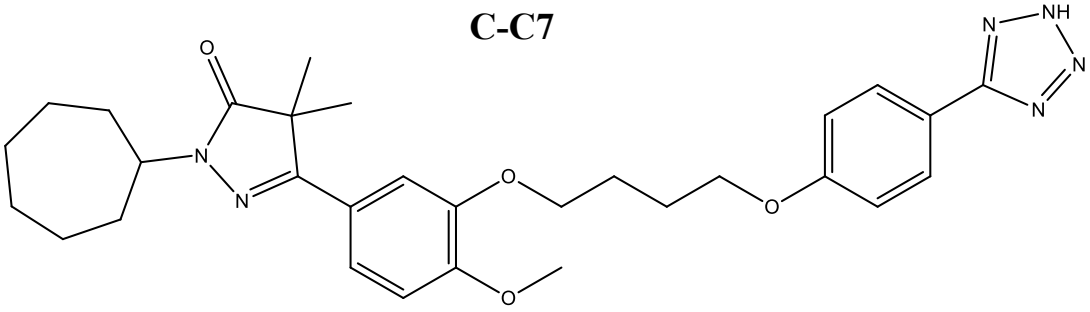
B-F10



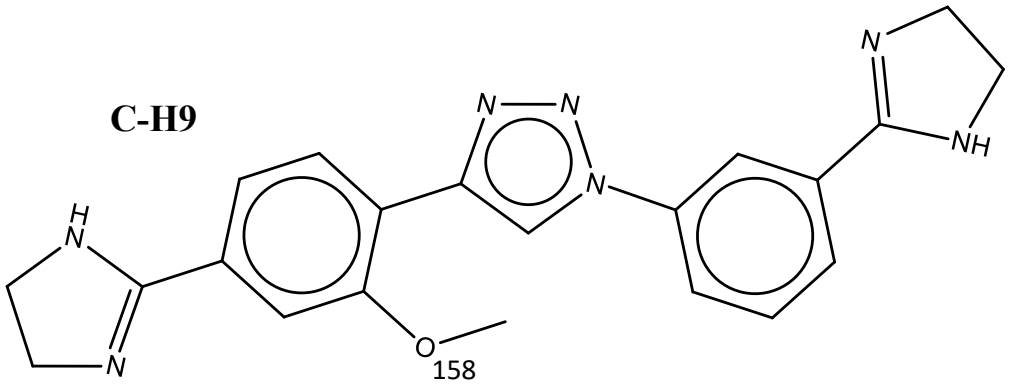
B-H6



C-C7



C-H9



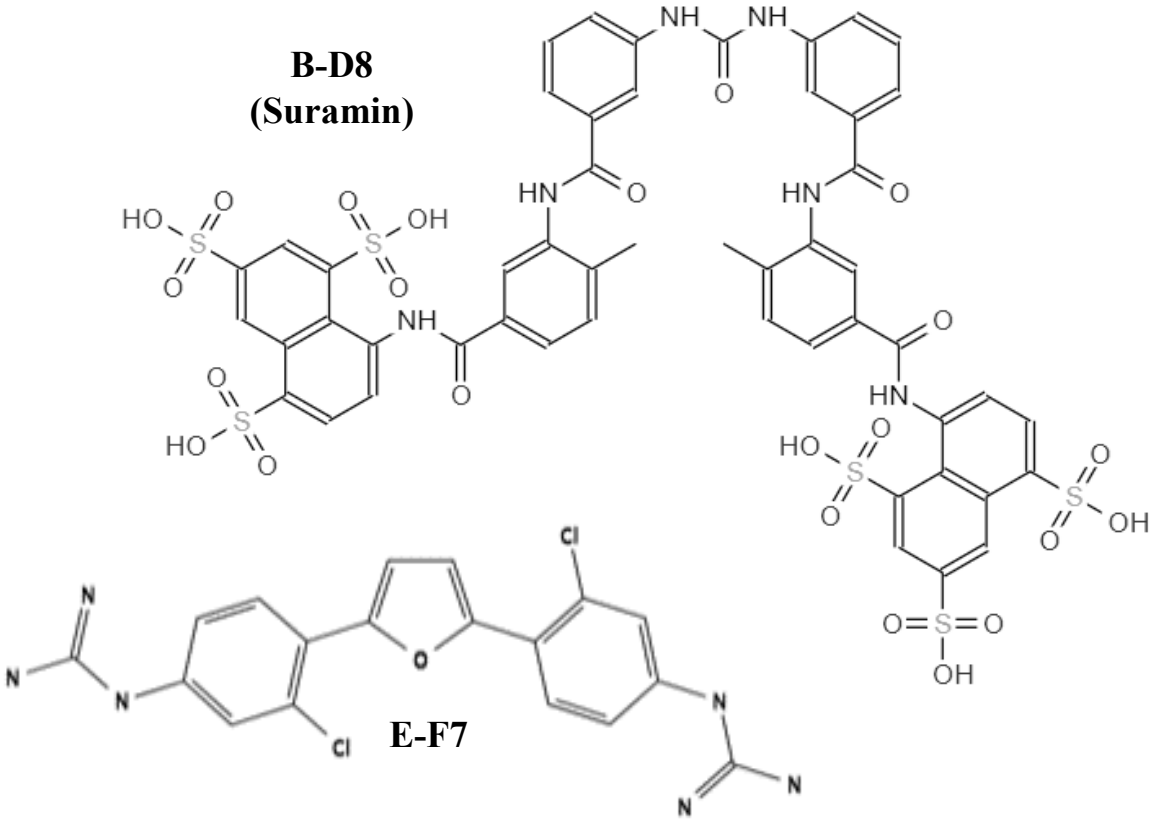


Figure 8.4. Structures of small molecules identified as inhibitors of CcrM via screening of the MMV Pathogen Box V2. The compound ID's indicate "Plate-Well position" from the Pathogen Box V2.

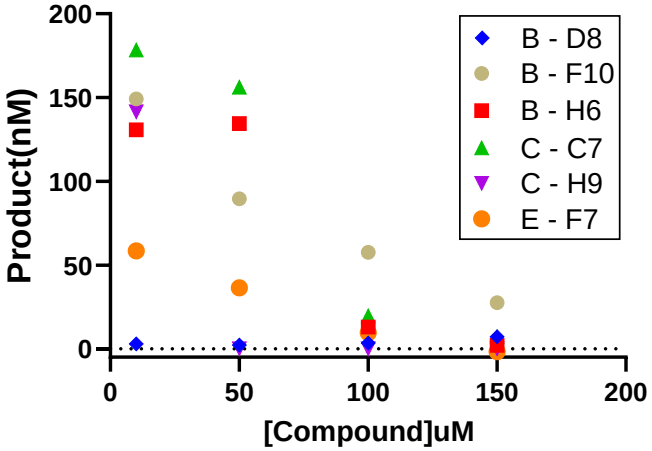


Figure 8.5. Hit compounds from the MMV Pathogen Box V2 that show dose-dependent inhibition on CcrM methylation. Conditions consisted of 100nM dsC1, 250nM WT CcrM, 15µM (1:50 Hot:Cold) SAM, [10, 50, 100, 150]µM compound, 2mM DTT in CcrM reaction buffer with 1.5% DMSO.

The Pathogen Box V2 hit compounds were also tested for selectivity against CcrM orthologs (BabI, AtucI, M.Linc, HinfI, HhaII) and human DNMT3A (Fig. 8.7). Suramin inhibits BabI and DNMT3A but is not selective for HinfI and M.Linc, and less potent for HhaII (Fig. 8.7). IC_{50}s were determined for Suramin with CcrM and each enzyme (Fig. 8.8, Table 8.1). Suramin has an IC_{50} for CcrM = 1.8µM, BabI = 1.1µM, AtucI = 2.7µM, HinfI = 149.4µM, HhaII = 3.8µM, M.Linc = 176.7µM, and DNMT3A = 1.8µM (Table 8.1).

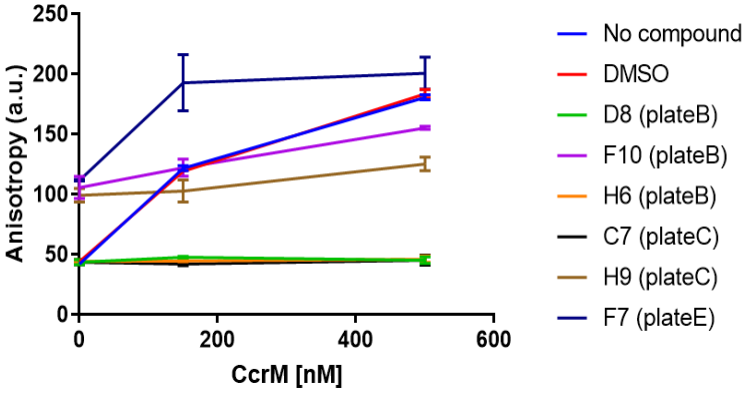


Figure 8.6. Anisotropy reveals that compounds from the MMV Pathogen Box V2 affect the stability of the CcrM/DNA complex. Conditions consisted of [0, 150, 500]nM WT CcrM, 10nM FAM-DNA, 50µM compound, 15µM SAH, in CcrM reaction buffer with 1.5% DMSO. D8, H6, and C7 either inhibit CcrM and prevent DNA binding or bind to DNA preventing complex association.

Mechanism studies were conducted to determine the mechanism of inhibition of Suramin for CcrM. Suramin vs SAM competition assay suggests that Suramin does not bind at the SAM binding site and that Suramin has a mixed-type mode of inhibition for CcrM (Fig. 8.9). CcrM has a SAM- and Suramin-concentration dependence on velocity (Fig. 8.9a). A double-reciprocal plot (1/V vs 1/[SAM]) with uninhibited (DMSO) and inhibited (Suramin 3.0µM) does not intersect at the y- or x-axis suggesting mixed-type inhibition of Suramin for CcrM (Fig. 8.9b). Compound E-F7 was also investigated for mechanism of inhibition via SAM

vs Suramin competition assay. The double-reciprocal plot for E-F7 shows an intersection in the negative x-and y-axis suggesting non-competitive inhibition of E-F7 for CcrM (Fig. 8.10).

CcrM inhibitors were also identified from screening the MMV Global Health Priority Box. 9 inhibitors (MB2-A10, MB2-C2, MB2-C7 MB2-C8, ZND-A8, ZND-B3, ZND-B5, ZND-G8, and ZND-H2) were identified which are structurally dissimilar (Fig. 8.11). Preliminary IC_{50}s were determined to estimate the potency of these hit-compounds, which ranged from ~39-~125μM (Table 8.2). MB2-A10 was the most potent with an IC_{50} of ~39μM (Fig. 8.12).

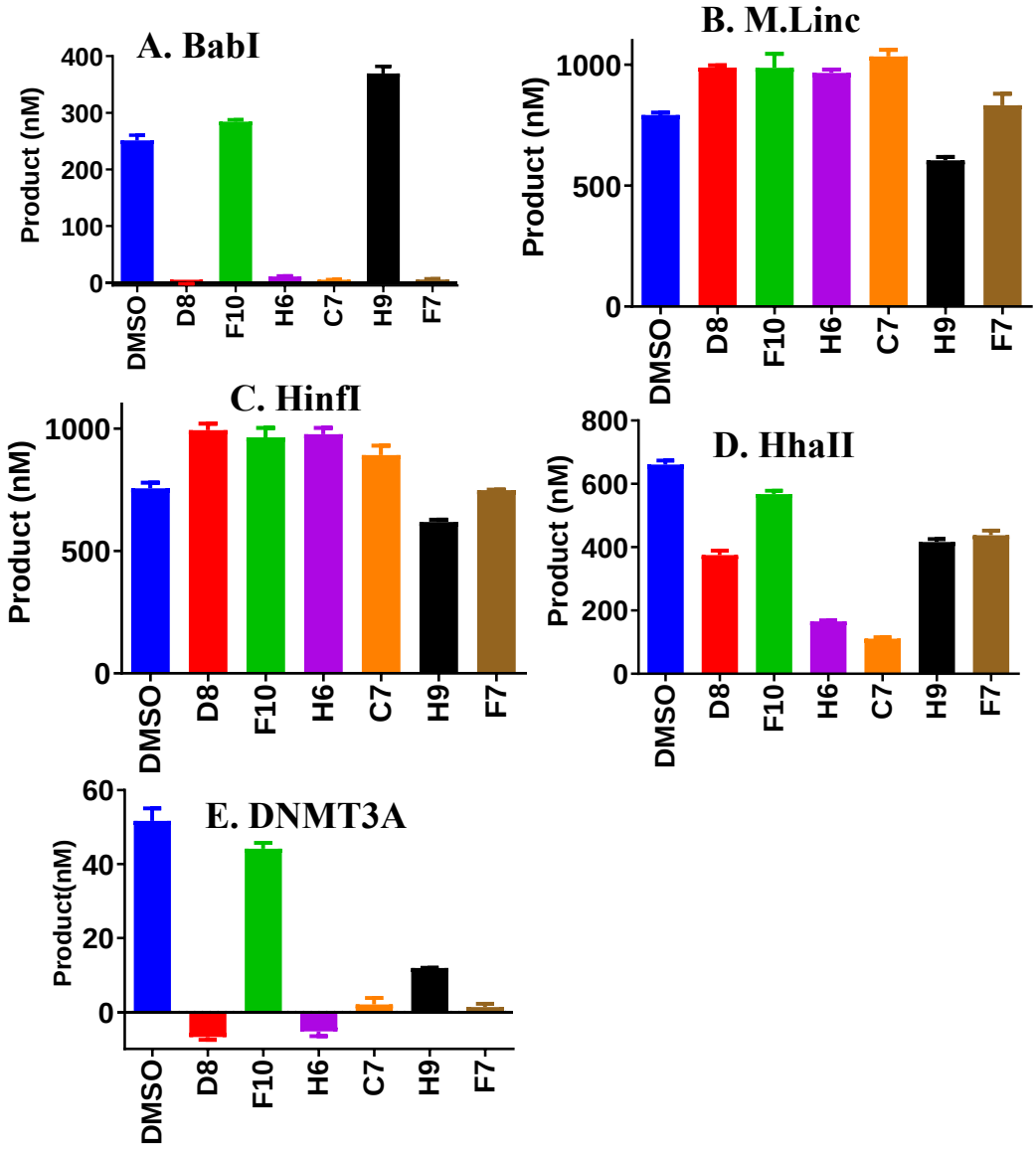


Figure 8.7. Selectivity for compounds from the MMV Pathogen Box V2 against CcrM orthologs and human DNMT3A. **A.** BabI is an orphan CcrM ortholog from *Brucella Abortus*. **B.** M.Linc is a CcrM ortholog that is part of an R/M system from *Moraxella Licolnii.* **C.** HinfI is a CcrM ortholog that is part of an R/M system from *Haemophilus Influenzae*. **D.** HhaII is a CcrM ortholog lacking a C-terminal domain and part of a restriction modification system from *Haemophilus parahaemolyticus*. **E.** DNMT3A is a C^5-cytosine human DNA methyltransferase. Conditions for BabI, M.Linc, HinfI, and HhaII consisted of 50nM enzyme, 1µM dsC1, 15µM SAM, 10µM D8 (Suramin), and 100µM F10,H6,C7,H9,F7. Conditions for DNMT3A consisted of 150nM DNMT3A, 10µM poly dI-dC DNA, 15µM SAM and the same compound concentrations.

The MMV Covid Box yielded 4 hits (A-C3, A-A7, B-A5, and A-H11) that were structurally similar (Fig. 8.13). A-H11 and B-A5 had high potency with IC_{50} values of 17.9µM and 6.5µM, respectively, (Fig. 8.14). A-H11 and B-A5 were not prioritized for further investigation because they are characterized DNA intercalator drugs named Pyronaridine and Doxorubicin, respectively. A-C3 and A-A7 were not investigated further because of low potency with IC_{50}s of 96.9µM and 85.0µM, respectively (Fig. 8.14).

We developed a pseudo *in-vivo* bacterial cell-based CcrM inhibition assay (Fig. 8.15). The cell-based assay relied on the gram negative Nico21 (DE3) expression cells containing a Pet-28a-CcrM vector. The *E.coli* cells contained a low level of CcrM expression due to a leaky promoter and CcrM expression was not induced with IPTG. CcrM is able to methylate the 12 GANTC sites within the pet-28a plasmid, resulting in protection from HinfI digestion. HinfI endonuclease recognizes and cuts the internal adenine in 5'G^ANTC'3 sites. The WT CcrM plasmid purified from a 3 hour outgrowth with no inhibitor shows that the cellular CcrM mehtylated and protected all GANTC sites from HinfI digestion (Fig. 8.16, lane 4). A control for this assay is the W332A variant of CcrM. W332A is catalytically dead (ref Clay) and therefore the *E.coli* containing the pet-28a-W332A CcrM gene does not methylate the plasmid. The lack of methylation and protection by W332A results in full HinfI digestion with 9 distict bands (Fig. 8.16. lane 5).

This preliminary *in-vivo* result is the basis for determining cellular inhibition of CcrM for compounds that inhibit CcrM *in-vitro*. Compounds that inhibit WT CcrM in E.coli would have the same gel banding phenotype as the W332A gel banding phenotype due to inhibited CcrM methylation activity.

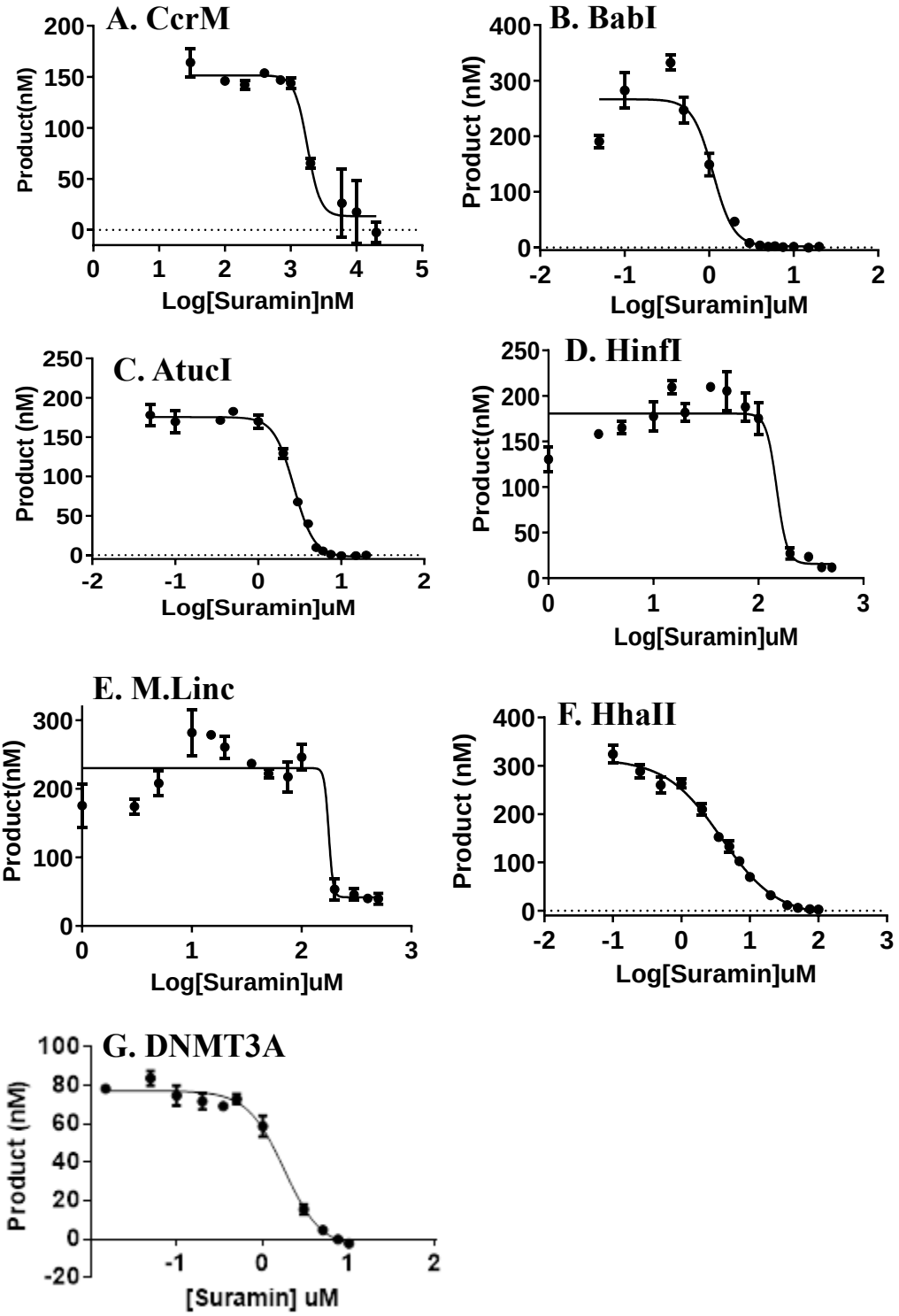


Figure 8.8. IC_{50} curves for suramin against CcrM, CcrM orthologs, and DNMT3A. IC_{50}'s were determined via radiochemical initial velocity assay. **A.** CcrM conditions were 50nM CcrM,

1µM dsC1, 15µM SAM, [0-10] µM suramin. **B.** BabI: 50nM BabI, 1µM dsC1, 15µM SAM, [0-20] µM suramin. C. AtucI: 50nM AtucI, 1µM dsC1, 15µM SAM, [0-20] µM suramin. D. HinfI: 25nM HinfI, 1µM dsC1, 15µM SAM, [0-500] µM suramin. E. M.Linc: 25nM M.Linc, 1µM dsC1, 15µM SAM, [0-500] µM suramin. F. HhaII: 50nM HhaII, 1µM dsC1, 15µM SAM, [0-100]µM suramin. **G.** DNMT3A: 150nM DNMT3A, 10µM poly dI-dC DNA, 15µM SAM, 1.5% DMSO, [0-10]µM suramin. Backgrounds done in 1.5% DMSO and witheld DNA. Spotted 5µL in duplicate triplicate at a single 20 minute timepoint.

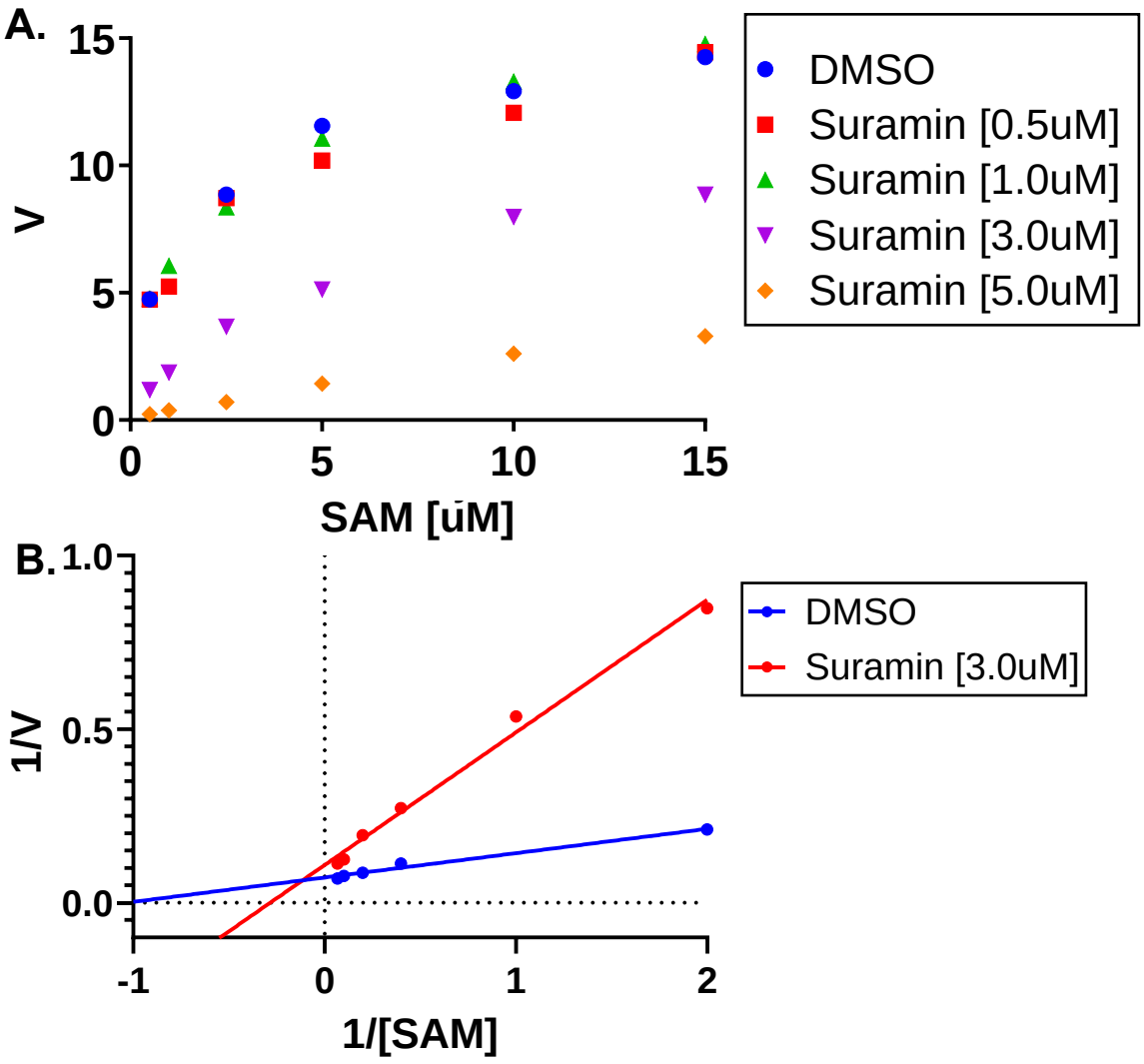


Figure 8.9. Suramin vs SAM competition assay reveals a mixed-type or noncompetitive mechanism of inhibiiton. A. [SAM]µM vs V plot shows SAM and suramin concentration dependences on velocity. B. Double-reciprocal plot of uninhibited (DMSO) and inhibited

(suramin 3.0µM) reveal non-intersecting linear regressions which is best-fit to a mixed-typed or noncompetitive mechanism of suramin inhibition of CcrM. Conditions consisted of 50nM CcrM, 1µM dsC1, [0.05, 0.1, 0.5, 1.0, 2.5, 5.0, 10, 15]µM SAM, [0, 0.5, 1.0, 3.0, 5.0]µM Suramin, and 2.5% DMSO in CcrM reaction buffer.

We used our most potent CcrM inhibitor (Suramin) identified from the screening for the cell-based inhibition assay to see if Suramin could enter a bacterial cell wall (Fig. 8.17). *E.coli* containing the Pet-28a-WTCcrM plasmid were grown in LB/Kan with 200µM Suramin did not result in a banding-pattern (Fig. 8.17, lane 3). This result indicates that WT CcrM in E.coli was capable of methylating and protecting its Pet-28a-WTCcrM plasmid and therefore not inhibited.

Enzyme	**Suramin IC_{50}**
CcrM	1.8 µM
BabI	1.1 µM
AtucI	2.7 µM
HinfI	149.4 µM
HhaII	3.8 µM
M.Linc	176.7 µM
DNMT3A	1.8 µM

Table 8.1. Suramin is not selective for CcrM. IC_{50}s for Suramin inhibition compare the selectivity against CcrM, CcrM orthologs, and human DNMT3A. Suramin is a potent inhibitor of CcrM, BabI, AtucI, HhaII, and DNMT3A, while non-potent for HinfI and M.Linc.

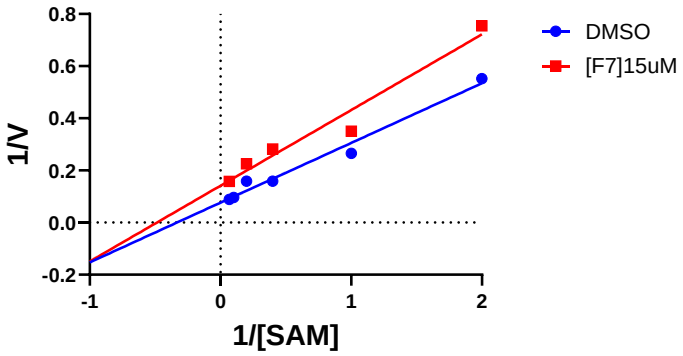


Figure 8.10. F7 vs SAM competition assay reveals non-competitive inhibition of CcrM. Conditions consisted of 50nM CcrM, 1µM dsC1, [0.05, 0.1, 0.5, 1.0, 2.5, 5.0, 10, and 15µM] SAM, 15µM F7, and 2.5% DMSO in CcrM reaction buffer. The lack of intersection on the y-axis is suggestive of a non-competitive mechanism of F7 for CcrM.

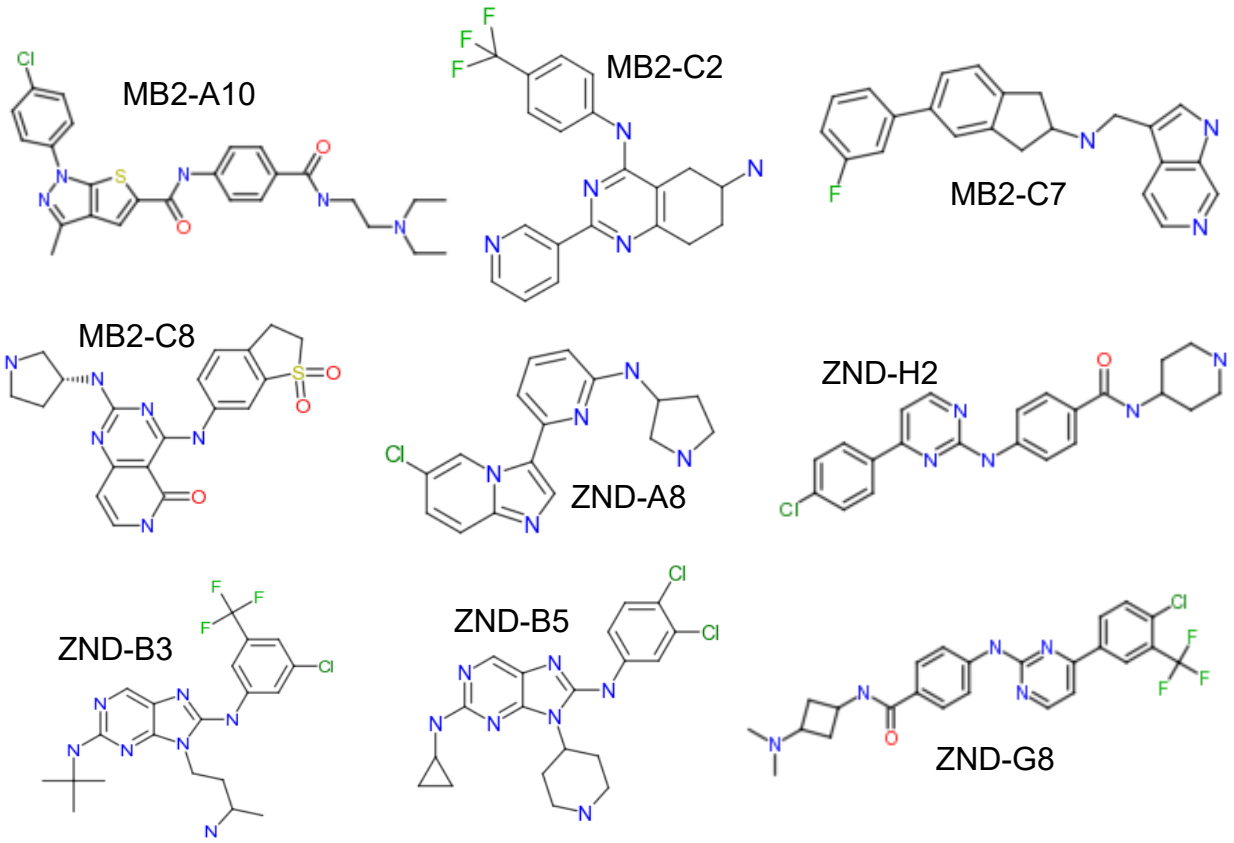


Figure 8.11. CcrM inhibitors identified from the MMV Global Health Priority Box. Compound identifiers denote "plate name-well position". Conditions consisted of 250nM WT CcrM, 100nM dsC1, 15uM SAM, [0-150]uM Compound. Single 5uL spot at 30 minute time point.

Global Health Priority Box Compound	IC_{50}
MB2-A10	~39µM
MB2-C2	~75µM
MB2-C7	~125µM
MB2-C8	~70µM
ZND-A8	~125µM
ZND-B3	~125µM
ZND-B5	~125µM
ZND-G8	~125µM
ZND-H2	~125µM

Table 8.2. Preliminary IC50s of hit-compounds from the Global Health Priority Box for CcrM.

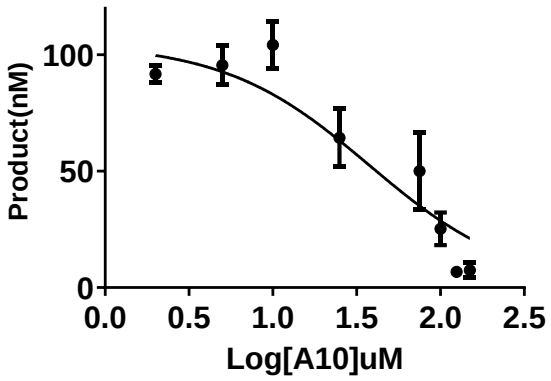


Figure 8.12. IC_{50} for A10 is 39µM. Conditions consisted of 250nM CcrM, 100nM dsC1, 15µM SAM, [0-150]µM MB2-A10. 5µL spot taken in triplicate at a 30 minute time point**.**

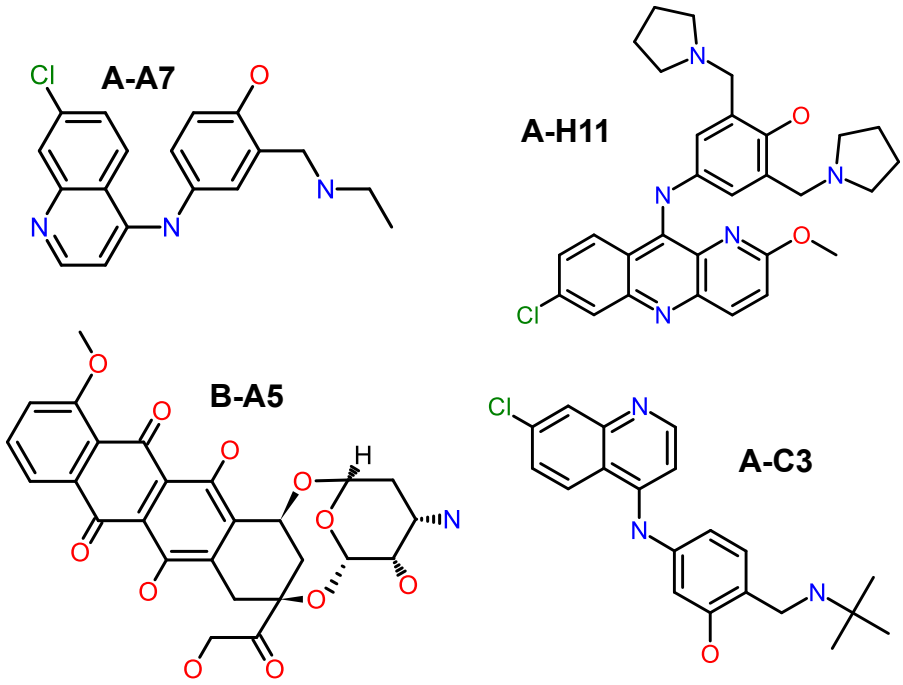


Figure 8.13. Hits identified from the MMV Covid Box screen. 4 small molecules were identified as CcrM inhibitors from the MMV Covid Box: A-A7, A-H11, A-C3, and B-A5. Compound identifiers denote "Plate-Well". A-H11 and B-A5 are known DNA intercalator drugs named Pyronaridine and Doxorubicin, respectively.

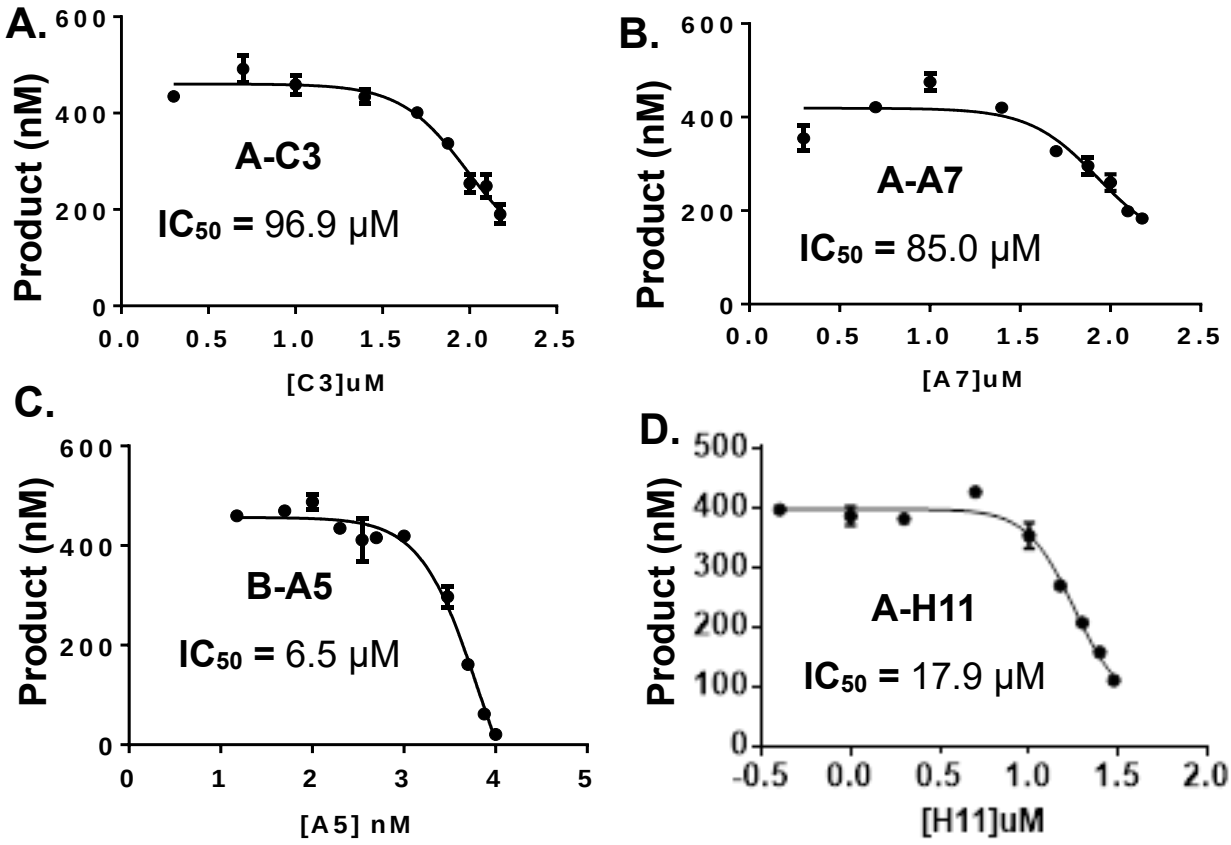


Figure 8.14. IC_{50} curves from the MMV Covid Box hits for CcrM. **A.** A-C3 has an IC_{50} = 96.9µM. **B.** A-A7 has an IC_{50} = 85.0µM. **C.** B-A5 has an IC_{50} = 6.5µM. **D.** A-H11 has an IC_{50} = 17.9µM. Conditions consisted of 50nM CcrM, 1µM dsC1, 15µM SAM, [0-150] µM A-A7 and A-C3, [0-25] µM A-H11, [0-10] µM B-A5.

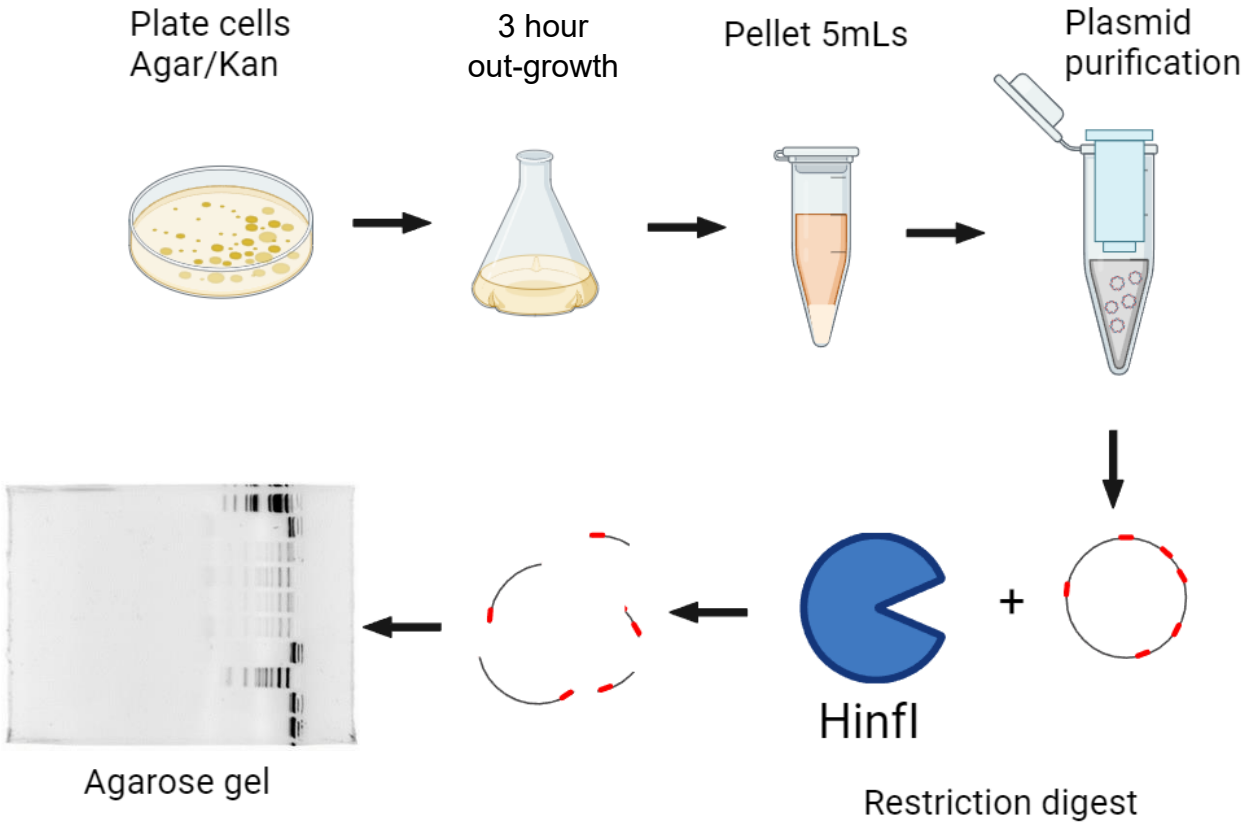


Figure 8.15. Cell study assay development for *in-vivo* CcrM inhibition in *E.coli*. Nico21 (DE3) cells containing a pet-28a vector with the CcrM gene were grown on LB Agar/Kanamycin plates. A single colony was selected for inoculation into liquid LB/Kan containing 2.5% DMSO (5mLs). The 5mLs of growth were pelleted by centrifugation and the plasmids were purified using an Agilent Miniprep kit. The purified plasmids were digested with HinfI to probe the methylation status of the pet-28a-CcrM plasmid's GANTC sites. Digested and undigested plasmids were loaded onto a 2.5% Agarose GelRed precast gel. Gels were imaged on an EZ Doc Gel Imager.

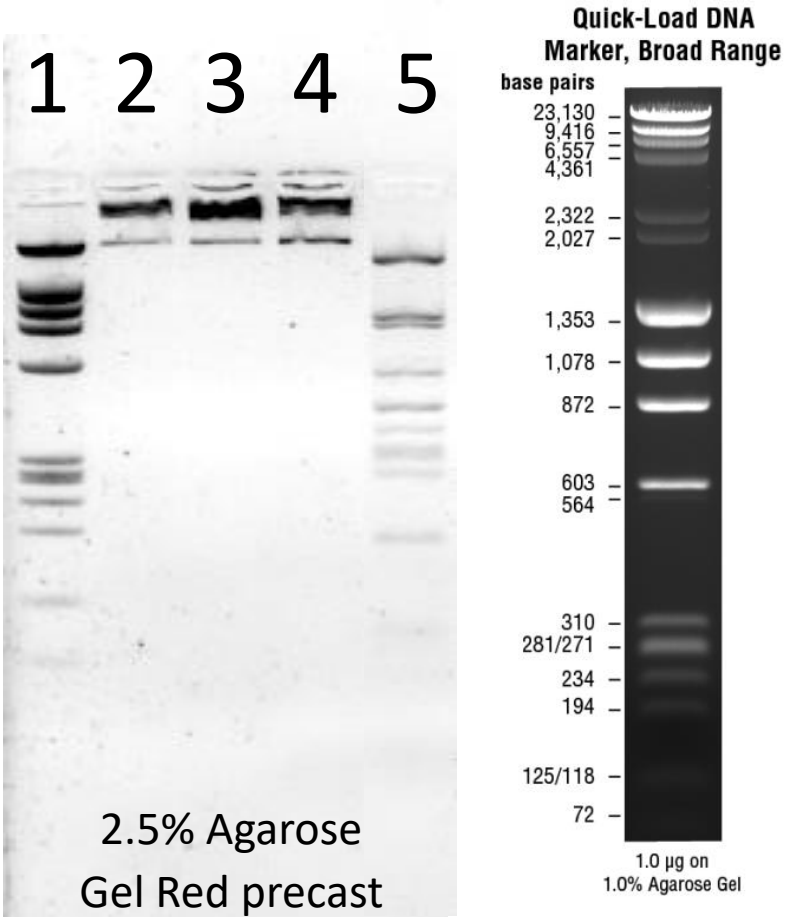


Figure 8.16. Validation of *in-vivo* CcrM inhibition assay. Gel Lane 1: DNA Broad range marker (NEB), 2: WT plasmid (undigested), 3: W332A plasmid (undigested), 4: WT plasmid (digested), 5: W332A plasmid (digested).

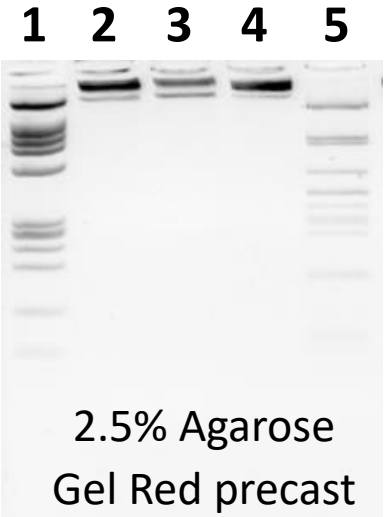


Figure 8.17. *In-vivo* CcrM inhibition of suramin does not reveal that suramin is capable of inhibiting CcrM in *E.coli*. Gel Lane 1: DNA Broad range marker (NEB), 2: WT + DMSO (undigested), 3: WT + Suramin 200µM (digested), 4: W332A + DMSO (undigested), 5: W332A + DMSO (digested).

VIII. Discussion

The CcrM compound screening efforts serve as a basis for the development of novel antibiotics that inhibit DNA methyltransferases which are essential to bacterial cellular viability. Three MMV libraries (Pathogen Box V2, Covid Box, Global Health Priority Box) were screened for inhibition of CcrM methylation activity. 29 compounds were determined to be hits out of the 621 compounds resulting in a 4.7% hit rate. The compounds in the MMV libraries are compiled sets of FDA-approved drugs and the goal of this study was to repurpose a drug for CcrM as a novel target and develop an antibiotic.

The results presented here identified compounds that met one of a few important criteria but failed to meet all criteria. Criteria were high potency, not a DNA intercalator, and selective for only bacterial DNA methyltransferases while not human DNA methyltransferases. DNA intercalators are small molecules containing a heterocyclic aromatic flat moiety that inserts itself within nucleic acids of DNA, which causes distortions in the DNA structure[8.17].

Suramin is our best candidate from the MMV screening efforts. One possible interpretation of the lack of Suramin cellular inhibition of CcrM is that 200uM Suramin is far below the cellular IC_{50} and does not result in observed inhibition. Another explanation is that Suramin degraded during the 3 hour out-growth by water-catalyzed hydrolysis and enough CcrM was expressed to perform efficient methylation of the plasmid. A third interpretation is due to Suramin's lack of selectivity, it could have an affinity for other enzymes therefore decreasing the amount of Suramin available to inhibit CcrM. A final interpretation is that Suramin is a large charged polysulfonated molecule that is incapable of getting through the E.coli bacterial cell wall.

Suramin is a polysulfonated naphthylurea and is a well-studied and known inhibitor of several targets with a variety of therapeutic effects[8.18,8.19]. For example, Suramin inhibits a heparanase enzyme endo-beta-D-glucuronidase that is important for metastatic melanoma cell invasion[8.20]. Suramin has an ID_{50} of 46 μM for heparinase inhibition which correlated to significant lack of melanoma cell growth[8.20]. Suramin has also been shown to be a potent inhibitor of the reverse transcriptase (RNA-directed DNA polymerase) of retroviruses for the treatment of AIDS where it was also shown to be highly nonspecific but was potent at concentrations that were non-toxic for host cells[8.21,8.22]. Suramin has been used as an inhibitor for a variety of targets for over 100 years and remains a candidate for more-selective and less-toxic analogs for treating disease[8.19].

The compounds were provided by MMV for no cost and the libraries are no longer available. Follow up cell-studies would require significantly more ngs of compounds and small quantities of these compounds were very costly, limiting the accessibility to perform these

experiments with low chances of interesting results. Further studies would benefit from collaboration with organic chemists to synthesize high volumes of compounds, analogs for structure-activity-relationship (SAR) determination. Collaboration with bacteriologists would elevate cell studies and determine pharmacokinetic parameters for drug characterization.

8. References

8.1. Robertson, G.T., Reisenauer, A., Wright, R., Jensen, R.B., Jensen, A., Shapiro, L, Roop II, R.M. (2000) The Brucella abortus CcrM DNA Methyltransferase is essential for viability, and it's overexpression attenuates intracellular replication in murine macrophages. *Journal of Bacteriology,* 182, 12.

8.2. Ghooi,R.B. and Thatte,S.M. (1995) Inhibition of cell wall synthesis – is this the mechanism of action of penicillins? *Medical Hypotheses*, 44, 127-131.

8.3. Weisblum,B. and Davies,J. (1968) Antibiotic inhibitors of the bacterial ribosome. *Bacteriological Reviews*, 32, 4, 493-528.

8.4. McCoy,L.S., Xie,Y., and Tor,Y. Antibiotics that target protein synthesis, John Wiley & Sons, Ltd. 2010, WIREs RNA, 2, 2.

8.5. Demerec,M. (1948) Origin of bacterial resistance to antibiotics. *Journal of Bacteriology,* 56, 63-74.

8.6. Lambert,P.A. (2005) Bacterial resistance to antibiotics: modified target sites. *Advanced Drug Delivery Reviews*, 57, 1471-1485.

8.7. Omeershffudin, U.N.M. and Kumar, S. (2019) Bacterial DNA Adenine Methyltransferase as a novel drug target of antibiotics: Current status and future drug discovery challenges. *Int.J.Curr.Microbiol.App.Sci,* 8, 4, 2494-2504.

8.8. Mashhon, N., Pruss, C., Carrol, M., Johnson, P.H., and Reich, N.O. (2006) Selective inhibitors of bacterial DNA adenine methyltransferases. *Journal of Biomolecular Screening,* 11, 5, 497-510.

8.9. Mashoon.N., Pruss,C., Carroll,M., Johnson,P.H., and Reich,N.O. (2006) Selective inhibitors of bacterial DNA adenine methyltransferases. *Journal of Biomolecular Screening*, 11, 5, 497-510.

8.10. Zhou,J. Deng,Y., Iyamu,I.D., Horton.J.R., Yu,D. Hajian,T., Vedadi,M., Rotili,D., Mai,A., Blumenthal,R.M., Zhang,X., Huang,R., and Cheng,X. (2023) Comparative study of adenosine analogs as inhibitors of protein arginine methyltransferases and a Clostridioides difficile-specific DNA adenine methyltransferase. *ACS Chemical Biology*, 18, 4, 734-745.

8.11. Copeland,R.A., Solomon,M.E., and Richon,V.M. (2009) Protein methyltransferases as a target class for drug discovery. *Nature Reviews Drug Discovery*, 8, 724-732.

8.12. Martin,J.L. and McMillan,F.M. (2002) SAM (dependent) I AM: the s-adenosylmethionine-dependent methyltransferase fold. *Current Opinion in Structural Biology,* 12, 6, 783-793.

8.13. Ferreiera de Freitas,R., Ivanochko,D., and Schapira,M. (2019) Methyltransferase inhibitors: competing with, or exploiting the bound cofactor. *Molecules*, 24, 24, 4492.

8.14. Zhang.Z., Wang,G., Li,Y., Lei,D., Xiang,J., Ouyang,L., Wang,Y., and Yang,J. (2022) Recent progress in DNA methyltransferase inhibitors as anticancer agents. *Frontiers in Pharmacology*, 13, 1-27.

8.15. Moreno,E. and Moriyon,I. (2002) Brucella melitensis: A nasty bug with hidden credentials for virulence. *PNAS*, 99, 1, 1-3.

8.16. Awadh,H., Hammed,Z.N., Hamzah,S.S., Saleh,T.H., and AL-Rubaii,B.A.L. (2022) Molecular identification of intracellular survival related Brucella melitensis virulence factors. *Biomedicine*, 42, 4, 761-765.

8.17. Baliaeva,T.N. and Faddeeva,M.D. (1991) DNA intercalators: their interaction with DNA and other cell components and their use in biological research. *Tsitologiia*, 33, 10, 3-31.

8.18. Zhang, Y., Keng, Y., Zhao, Y., Wu, L., and Zhang, Z. (1998) Suramin is an active site-directed, reversible, and tight binding inhibitor of protein-tyrosine phosphatase. *Journal of Biological Chemistry*, 273, 20, 12281-12287.

8.19. Wiedemar, N., Hauser, D.A., and Maser, P. (2020) 100 Years of Suramin. *ASM Journals Antimicrobial Agents and Chemotherapy,* 64, 3, 1-14.

8.20. Nakajima, M., DeChavigny, A., Johnson, C.E., Hamada, J., Stein, C.A., and Nicolson, G.L. (1991) Suramin. A potent inhibitor of melanoma heparinase and invasion. *Journal of Biological Chemistry*, 266, 15, 9661-9666.

8.21. De Clercq, E. (1987) Suramin in the treatments of AIDS: Mechanism of action. *Antiviral Research*, 7, 1, 1-10.

8.22. De Clercq, E. (1979) Suramin: A potent inhibitor of the reverse transcriptase of RNA tumor viruses. *Cancer Letters*, 8, 1, 9-22.