2021

Investigation of a Type II-C CRISPR-Cas9 on Its Sensitivity to the Epigenetic State of DNA Substrates

Ethan Wickline

FLORIDA STATE UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

INVESTIGATION OF A TYPE II-C CRISPR-CAS9 ON ITS SENSITIVITY TO THE EPIGENETIC STATE OF DNA SUBSTRATES

By

ETHAN SAMUEL WICKLINE

A Thesis submitted to the
Department of Chemistry and Biochemistry
in partial fulfillment of the
requirements for the degree of
Master of Science

2021

Ethan Samuel Wickline defended this thesis on July 1, 2021.
The members of the supervisory committee were:

Hong Li
Professor Directing Thesis

Beth Stroupe
Committee Member

Scott Stagg
Committee Member

Wei Yang
Committee Member

The Graduate School has verified and approved the above-named committee members, and certifies that the thesis has been approved in accordance with university requirements.

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

CRISPR Cas9 is a bacterial immune system which has been found to be extremely useful for the purposes of genomic editing and DNA detection methods. While it is possible to "program" Cas9 proteins for specific target sites, any site must also contain a match for the Cas9 protein's PAM sequence otherwise the protein will not bind. Preliminary experiments reveal AceCas9, one variant of CRISPR Cas9 isolated from *Acidothermus cellulolyticus* bacteria with a PAM sequence of 5'-NNNCC-3', is sensitive to the presence of methylation on the fourth-position cytosine in its PAM sequence meaning it will not successfully cleave any target site possessing 5-methylcytosine at this position. DNA methylation is an important epigenetic change which can influence gene expression but current methods of detecting DNA methylation either require harsh reactions involving sodium bisulfite that can degrade sample DNA or methylation-sensitive restriction enzymes which frequently have highly inflexible and non-programmable target sites. This information was used in the creation of a methylation-detection assay which uses AceCas9 protein to "filter" non-methylated DNA by cleaving it, the methylated DNA remaining may then be amplified via PCR, and then analyzed after gel electrophoresis. The detection limits of this methylation assay have been found to consistently detect methylated DNA despite only constituting $1*10^{-5}$% of a given sample. This would allow for the detection of methylation for a single nucleotide to a concentration as low as $1*10^{-5}$% without requiring the harsh reaction conditions involving sodium bisulfite or an inflexible target site required by a methylation-sensitive restriction enzyme.

CHAPTER 1

INTRODUCTION

1.1 CRISPR

Without a suitable defense mechanism against them, bacteria are highly vulnerable to bacteriophages. The various "Clustered Regularly Interspaced Short Palindromic Repeats" (CRISPR) nuclease proteins were originally discovered as part of a bacterial immune system against such bacteriophages (*1-3*). The CRISPR system is capable of incorporating pieces of bacteriophage DNA into a dedicated CRISPR locus as a "spacer" to target in the case of future infection (*1-3*). In the event of such an infection, a "CRISPR-associated" (Cas) protein with the incorporated DNA guiding it (a component of the final Cas protein referred to as its "guide RNA") will then bind to DNA that matches and cleave it. Because the locus may accept different sequences of DNA, vectors encoding CRISPR protein and its respective locus have since been found to be highly useful tools for targeted genome editing (*4, 5, 15*). These vectors may be "programmed" by adding or modifying a spacer in the locus to target a specific site, used to transform cells, the transformed cells will express Cas protein with the newly programmed guide RNA, and the Cas protein will then find any sequences which match its preprogrammed target site before binding and cleaving it (*16*). This will leave the cleaved DNA of the host cell to either remain cleaved, fix itself via non-homologous end joining, or fix itself with template DNA via homology directed repair if provided with such a template (*4*).

Of the various types of CRISPR that exist, CRISPR Cas9 will be the focus of this research. Regardless of exact variant of CRISPR Cas9, each Cas9 protein is composed of several common domains (Figure 1). The HNH and RuvC domains are responsible for cleaving the target and non-target strands of DNA respectively (*5*), the recognition (REC) lobe plays a major role in recognizing and binding to target sites as well as reducing off-target effects (*17, 18*), immediately downstream of its target site is the "Protospacer Adjacent Motif" (PAM) which is a nonprogrammable sequence essential for Cas9 binding which differs based on the exact variant of Cas9 protein (*5-7*), and the "PAM Interacting Domain" (PID) is vital for PAM sequence recognition (*5, 19, 20*). The PID and PAM sequence is of particular interest for the purposes of this research because, while the target site may be "programmed" into a CRISPR

locus easily, any variant of Cas9 still requires its variant-specific PAM sequence downstream of the target site before it will bind (*5, 19, 20*).

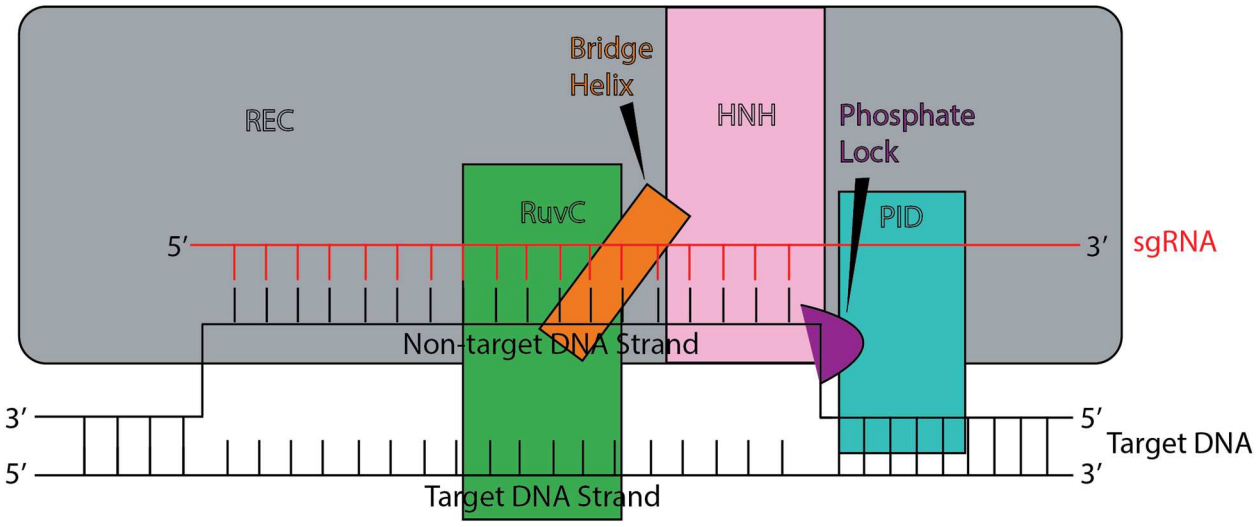


Figure 1. Cartoon representation of AceCas9, one example of a Cas9 protein with several of the common domains including the REC, RuvC, HNH, and PID domains.

1.2 Methylation of DNA

Epigenetic modifications of DNA include several major modifications and may be caused by a variety of outside factors. The methylation of DNA by addition of a methyl group to cytosine or adenine is one such example of these modifications, with pathways demonstrating their creation and maintenance being well characterized (*21*). Methylation of cytosine at its fifth-position carbon to create 5-methylcytosine is of particular interest for the purpose of this research but the creation of N6-methyladenine or N4-methylcytosine are also possible via methylation of adenine at its sixth-position nitrogen or cytosine at its fourth-position nitrogen respectively (*21, 22*). 5-methylcytosine is of particular interest because it appears to be the most frequent form of methylated DNA measured in mammals, occurring almost exclusively where a guanine nucleotide immediately follows a cytosine nucleotide (CpG) (*23, 24*). These CpG sites are methylated at a rate between 70-80% in mammals on average (*14*) and experimental evidence supports their presence as a form of gene regulation by silencing genes (*25-27*). This means an abnormal quantity of methylation may result in genetic diseases or tumorigenesis. For example, hypermethylation (the overabundance of methylated DNA) can inactivate tumor-suppressing genes whereas hypomethylation (the lack of methylated DNA) could activate oncogenes (*26-28*).

This is further supported when examining "CpG islands", sections of DNA at least 200 base pairs in length with a concentration of cytosine and guanine higher than 50% and an observed-to-expected ratio of CpG sites higher than 60% (*29*). It was found that 72% of human promoters contained a CpG island (*30*). Typically, these promoter CpG islands are unmethylated but hypermethylation of these islands have been detected in many tumors (*26, 27, 29*). Silencing genes is not the only function of DNA methylation. Restriction enzymes may bind and cleave DNA at short, specific sequences which is highly useful for eliminating foreign DNA but, without a way to differentiate self from non-self DNA, these enzymes would also target host DNA (*8, 31*). Many restriction enzymes are sensitive to the presence of methylation so host DNA is typically methylated at target sites to prevent self-targeting (*8, 31*).

Due to the value in researching DNA methylation, several assays were developed to determine the presence of methylation at various resolutions. The "gold standard" for this is bisulfite sequencing whereupon DNA of interest is reacted with sodium bisulfite, amplified by PCR, then sequenced to determine a sample's methylation state (*37*). The sodium bisulfite deaminates unmethylated cytosine nucleotides to uracil then the subsequent PCR amplification converts uracil to thymine while leaving methylated cytosine untouched (*37*). When sequenced, the results of bisulfite sequencing should reveal the methylation status of all cytosines between the amplified primers, with the potential for single-nucleotide resolution. Bisulfite sequencing is not without its flaws, however, as the harsh reaction of sodium bisulfite can degrade available DNA by up to 99% after a 16-hour bisulfite treatment reaction, the high salt concentration required for the reaction can result in renaturation of DNA before it is fully converted, and designing primers for amplification requires special consideration due to the decreased prevalence of cytosine (resulting in a decrease of effective primer locations) (*42*). All this together means methods that require bisulfite sequencing, while very useful, are typically better suited for large-scale DNA study instead of individual nucleotides of interest. Other methods like the use of assays involving methylation-sensitive restriction enzymes offer different advantages by treating DNA with restriction enzymes which cannot cleave methylated DNA. These enzymes proceed to digest unmethylated DNA while leaving methylated DNA for study (*42*). A major drawback of these assays however is their need for inflexible target sites, resulting in requiring different restriction enzymes to target different sites (*42*). If a 5-methylcytosine of

interest does not fall within the target site of one of these restriction enzymes then this method would not be usable.

1.3 Impact of Methylation on CRISPR

Restriction enzymes have the potential to be methylation-sensitive in order to help protect host DNA from self-targeting so perhaps there is reason to believe CRISPR Cas9 may also be methylation-sensitive. Restriction Enzymes and CRISPR Cas9 are similar in their role to defend their host from foreign DNA invasion by searching for specific sequences and cleaving them. It is important to note that while a given Cas9 protein may tolerate mismatches in its target site and remain functional (*9-11)*, it will not tolerate mismatches within its PAM sequence (5, *19, 20*). This tolerance also includes methylation as SpCas9 was experimentally discovered to be capable of cleaving its target site despite the presence of 5-methylcytosine in either its target site or its PAM sequence (*32*). SpCas9, however, requires a PAM sequence that matches the nucleotide sequence 5'-NGG-3' (*32-34*) so while it may recognize a PAM sequence containing a cytosine (in the form of 5'-CGG-3'), the cytosine may also be substituted with another nucleotide and retain functionality. This information would therefore support 5-methylcytosine being tolerated in the target site of a Cas9 protein, but further information would be necessary to determine if a PAM sequence requiring cytosine would similarly tolerate the presence of 5-methylcytosine. AceCas9, a variant of Cas9 protein originally isolated from *Acidothermus cellulolyticus*, has a PAM sequence of 5'-NNNCC-3' (*12, 13*), making it a prime candidate for studying cytosine methylation. Because the two cytosines in its PAM sequence are required for AceCas9 activity, both may be targeted for selective methylation to find the impact of methylation on protein activity.

1.4 Preliminary Data

Preliminary experiments were performed which demonstrate AceCas9 is sensitive to methylation on the fourth position cytosine in its PAM sequence (henceforth referred to as "^{5M}C4" and "^{5M}C5" for a 5-methylcytosine at the PAM's fourth or fifth positions respectively). This was determined because, like other Cas9 variants, AceCas9 may be used for the purpose of targeted cleavage of DNA sequences. This allows study of the protein's function and its efficacy by use of DNA cleavage ("in-vitro oligonucleotide cleavage assay") and cell survival assays

("in-vivo cell survival assay") wherein the cleavage of targeted DNA may be measured by a denaturing gel or by the presence of cell colonies respectively.

1.4.1 *In-Vivo Cell Survival Assay Reveals ^{5M}C5 Does Not Impact AceCas9 Activity*

An "in-vivo cell survival assay" requires the transformation and creation of competent cells with a plasmid encoding for the production of ccdB when in the presence of arabinose as well as a resistance to ampicillin (plasmid henceforth referred to as "ccdB-AcPAM"). ccdB acts as a DNA gyrase toxin and will kill any cells growing in an environment containing arabinose (*35*). This plasmid also contains a target site for AceCas9 so if AceCas9 protein is both present and functional the plasmid will be cleaved and the cell will survive. The method is, however, restricted by what methyltransferase enzymes are endogenous to tested cells. The BW25141 *E. coli* cells only contain a *dcm* gene to produce Dcm methyltransferase. This methyltransferase can target for methylation the second cytosine in sequences matching 5'-CCWGG-3' (*36*). The AceCas9 target site in ccdB-AcPAM is 5'-ggtaggatggcaagatcctggtatACACCaag-3', 44 base pairs downstream of its *ccdB* gene (capitalization indicates PAM sequence). The last five nucleotides, 5'-CCaag-3', almost match the target sequence of Dcm methyltransferase so site directed mutagenesis was used to induce a point mutation in order to change it to 5'-CCagg-3'. This change would methylate the fifth position cytosine in the PAM sequence, creating ^{5M}C5. Once the mutated ccdB-AcPAM plasmid was used to transform BW25141 *E. coli* cells, it was necessary to verify the presence of ^{5M}C5. This was accomplished via bisulfite treatment and sequencing of plasmid DNA extracted from cells post-transformation. Results of Sanger sequencing successfully verified the presence of ^{5M}C5 (Figure 2).

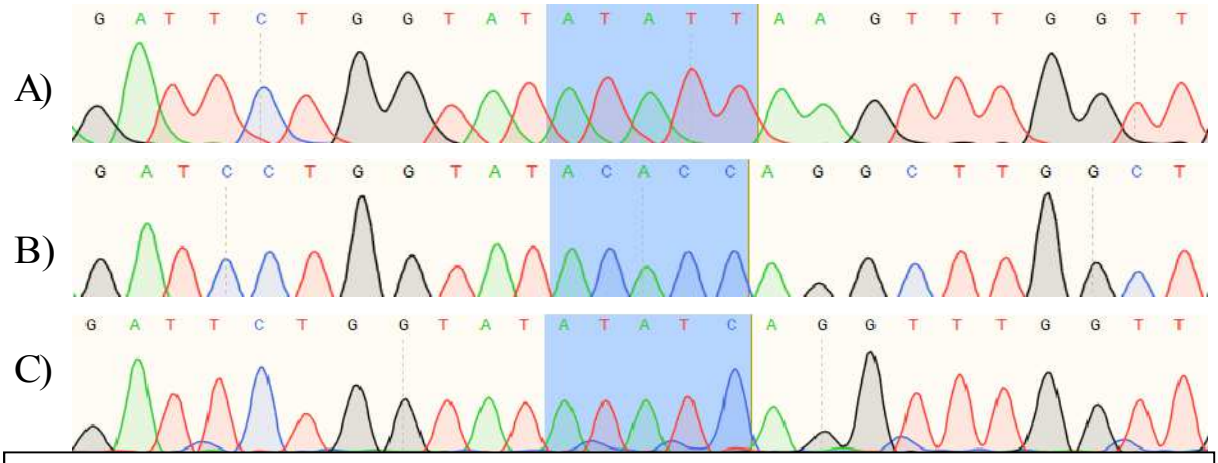


Figure 2. Sanger sequencing results of ccdB-AcPAM PAM sequence (shaded in blue)

Figure 2 – Continued
with varying states of methylation.

A) Depicts DNA after bisulfite treatment while lacking the 5'-CCagg-3' mutation, resulting in the conversion of all cytosines within the PAM sequence to thymine.

B) Depicts DNA before bisulfite treatment while containing the 5'-CCagg-3' mutation.

C) Depicts DNA after bisulfite treatment while containing the 5'-CCagg-3' mutation, resulting in methylation of the fifth position cytosine.

After the presence of ^{5M}C5 was verified via bisulfite sequencing, BW25141 *E. coli* cells containing the ccdB-AcPAM plasmid were transformed again with either Ac9g123 pACYCDuet plasmid (which encodes for AceCas9 and chloramphenicol resistance) or empty pACYCDuet plasmid (which only encodes for chloramphenicol resistance). Cells were then plated and incubated on agar plates containing either only chloramphenicol or chloramphenicol and arabinose. Cell survival on plates containing only chloramphenicol would require a successful transformation of either the Ac9g123 or empty pACYCDuet plasmid because both plasmids encode for chloramphenicol resistance. Cell survival on plates containing both chloramphenicol and arabinose would require the target site on the ccdB-AcPAM to be cleaved due to its role in producing toxic ccdB, therefore requiring both successful transformation of Ac9g123 pACYCDuet plasmid and AceCas9 insensitivity towards ^{5M}C5.

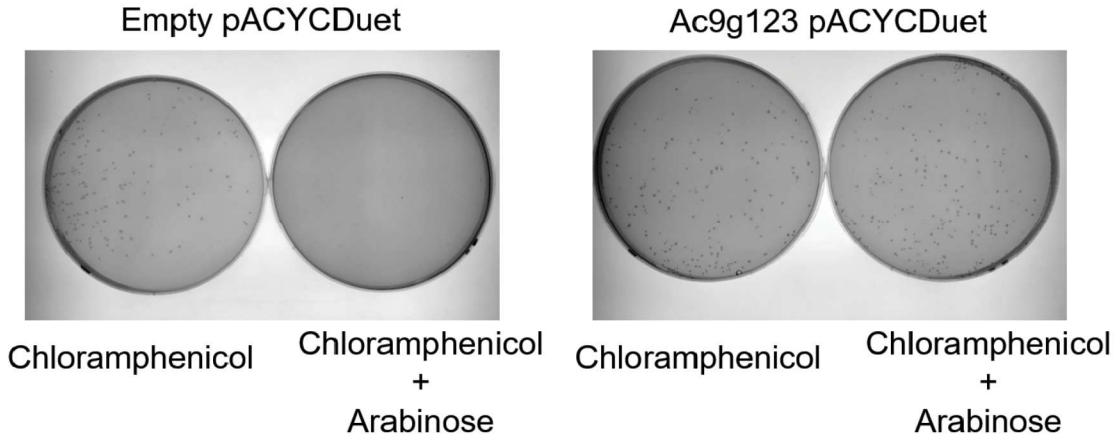


Figure 3. Depicts colonies of cells growing with either empty pACYCDuet plasmid or Ac9g123 pACYCDuet plasmid. Colonies of empty pACYCDuet-containing cells grew in the presence of chloramphenicol but not in the presence of chloramphenicol with arabinose, implying ccdB-AcPAM plasmid was not cleaved. Colonies of Ac9g123 pACYCDuet-containing cells grew in the presence of both chloramphenicol

Figure 3 - Continued
and chloramphenicol with arabinose, implying ccdB-AcPAM plasmid was cleaved.

It was expected for cells containing the empty pACYCDuet plasmid to survive on plates containing only chloramphenicol due to the antibiotic resistance granted by the empty pACYCDuet plasmid but would produce ccdB and die if placed into chloramphenicol and arabinose plates due to the absence of AceCas9 protein to cleave the ccdB-AcPAM plasmid. Cells transformed with Ac9g123 pACYCDuet plasmid instead were expected to survive on both plates if AceCas9 was 5-methylcytosine insensitive but would behave similar to cells transformed with empty pACYCDuet plasmid if AceCas9 was 5-methylcytosine sensitive. This experimental method was the first to provided evidence that AceCas9 was insensitive to ^{5M}C5 (Figure 3).

1.4.2 *In-Vitro Cleavage Assay Shows ^{5M}C4 Inhibits AceCas9 Activity*

An "in-vitro oligonucleotide cleavage assay" utilizes purified AceCas9 protein and oligonucleotides to simulate cell conditions wherein the protein combines with supplied single guide RNA (sgRNA), binds with free-floating annealed oligonucleotides, cleaves them, and the resulting sample is then run on a denaturing gel.

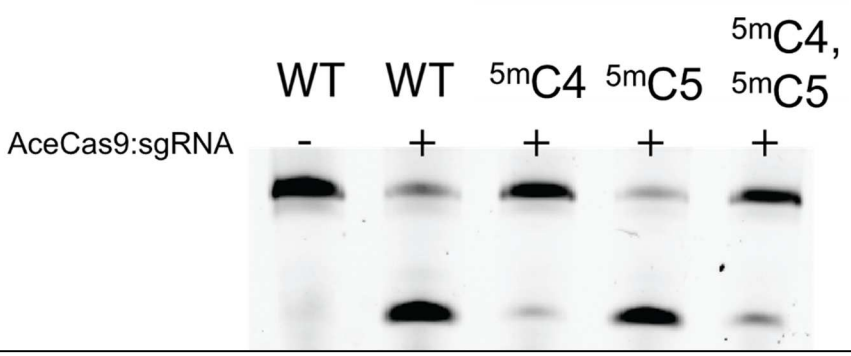


Figure 4. Depicts five samples used for AceCas9 cleavage assay. +/- denotes presence or absence of AceCas9 protein and sgRNA. WT- and WT+ both contain unmethylated annealed oligos but WT- lacks AceCas9 protein. This implies AceCas9 protein cannot cleave its target site when C4 is methylated but can when only C5 is methylated.

Doing so was necessary because, while the BW25141 cells used lacked any endogenous methyltransferases that could easily produce ^{5M}C4, ordering oligonucleotides for an in-vitro reaction allows for them to be pre-methylated. The non-target strand of the annealed

oligonucleotides was also labeled with a 5'-HEX fluorescent tag for visibility on the gel. Four separate tubes were labeled for each sample (WT, ^{5M}C4, ^{5M}C5, ^{5M}C4+^{5M}C5), including a fifth negative control sample wherein WT DNA was used like normal but extra water replaced the AceCas9 protein and sgRNA to ensure no reaction took place. When finished, pictures of the gel taken by a Chemidoc MP Imaging System provided evidence that AceCas9 is insensitive to ^{5M}C5 but sensitive to ^{5M}C4 (Figure 4).

1.5 Development of Methylation Assay and Purpose

Research was performed which used Cas9 protein in an assay to determine the presence of point mutations (*39*). By ensuring the nucleotide of interest fell within the PAM sequence of the Cas9 used, the researchers demonstrated capability to filter large quantities of "healthy" DNA (lacking the deleterious point mutation) via cleaving with Cas9, PCR amplification of the DNA remaining, and to check for results via gel electrophoresis. If any DNA is left after being treated with Cas9, it can then be amplified by PCR which allows for the detection of very dilute quantities of mutated DNA. Because AceCas9 is ^{5M}C4 methylation-sensitive, it could be used in a similar fashion to instead determine the presence of methylation. As covered earlier, there already exists several methods to determine methylation status in DNA but each have a drawback that prevents them from being ideal for every circumstance. Bisulfite sequencing is still the golden standard for detecting methylation in DNA but due to the harsh reaction, potential for partial conversion, and specific requirements for designing primers the technique is less reliable when targeting individual nucleotides. An alternative to this, assays that are based on using methylation-sensitive restriction enzymes, can be used to target individual nucleotides but only when they fall within the inflexible target site of such an enzyme. These target sites are typically restrictive patterns of four or more nucleotides so an alternative that has a more permissive target site could greatly expand useful locations. The preliminary data described prior supports AceCas9 being methylation-sensitive to ^{5M}C4 so it may be used in a similar manner as methylation-sensitive restriction enzymes to determine the presence of methylation. Such a technique would be capable of determining methylation status of individual nucleotides without degradation of DNA (in the case of bisulfite sequencing) and has a significantly more flexible required target sequence (due to its PAM of 5'-NNNCC-3' and programmable target site).

Testing whether AceCas9 could be used as part of a "Methylation Assay" required first an experiment where DNA is known to be methylated and unmethylated. In-vitro conditions were chosen to best ensure DNA being used is methylated or unmethylated. The DNA, ccdB-AcPAM plasmid was chosen instead of oligonucleotides. DNA would first be mutated to change its PAM sequence from 5'-ACACC-3' to 5'-AGGCC-3' so it could be treated with HaeIII methyltransferase. HaeIII methyltransferase is a methyltransferase which targets the first cytosine in all sequences matching 5'-GGCC-3' so it is ideal for creating ^{5M}C4 after the required mutation is made. Once DNA is methylated after treatment with HaeIII methyltransferase, it could be diluted with unmethylated DNA to create a sample with any percent methylated needed then these samples could be "filtered" with an AceCas9 cleavage reaction in much the same way as the method used by Lee et al. to detect point mutations (*39*). In this case, instead of the non-mutated DNA and mutated DNA used by Lee et al, all unmethylated DNA should be cleaved leaving only methylated DNA which could then be amplified via PCR then analyzed with gel electrophoresis. If successful as a methylation detection method, this assay would be useful in the study of epigenetics and diseases caused by abnormal methylation of individual nucleotides and could, in the future, potentially be adapted to use other Cas9 proteins that have different PAM requirements so long as they are also methylation-sensitive.

CHAPTER 2

METHODS

2.1 Determining AceCas9 Methylation Sensitivity

2.1.1 *Preparation of ccdB-AcPAM*

In order to utilize the endogenous Dcm methyltransferase of the tested BW25141 cells, the nucleotide sequence immediately downstream of the PAM region for the target site needed to be changed from 5'-CCaag-3' to 5'-CCagg-3'. NEB's "Q5® Site-Directed Mutagenesis Kit" was used for this purpose following the provided protocol. The forward and reverse primers used in the initial PCR step of this reaction were 5'-atACACCaggcttggctgttttgg-3' and 5'-accaggatcttgccatcct-3' respectively. After making the required point mutation and subsequent transforming of competent cells, the desired mutated plasmid was first extracted using Omega Bio-tek's "E.Z.N.A. ® Plasmid DNA Kit" following the provided protocol.

2.1.2 *Bisulfite Sequencing*

Once the mutated ccdB-AcPAM plasmid was extracted, it was necessary to verify the presence of ^{5M}C5. NEB's "EpiMark ® Bisulfite Conversion Kit" was first used to convert all unmethylated cytosine in the DNA to uracil by following its provided protocol then amplified via PCR using NEB's "EpiMark ® Hot Start *Taq* DNA Polymerase" and its provided protocol to convert those uracil to thymine. The forward and reverse primers used for PCR amplification were created using MethPrimer (*38*) and were 5'-gggaagaagtggttgattttagttat-3' and 5'-ccaaacaaattctattttatcaaacc-3' respectively. The amplified DNA was then sequenced using Sanger sequencing. Sanger sequencing was performed by the FSU DNA Sequencing Facility.

2.1.3 *In-Vivo Survival Assay*

After the presence of ^{5M}C5 was verified via bisulfite sequencing, BW25141 E. coli cells containing the mutant ccdB-AcPAM plasmid were transformed again with either empty pACYCDuet plasmid or Ac9g123 pACYCDuet plasmid. Cells were then plated and incubated at 37 °C for 16 hours on agar plates containing either only chloramphenicol or chloramphenicol and arabinose. This resulted in four plates of interest: One with cells containing empty pACYCDuet plasmid with only chloramphenicol, one with cells containing empty pACYCDuet plasmid with

both chloramphenicol and arabinose, one with cells containing Ac9g123 pACYCDuet plasmid with only chloramphenicol, and one with cells containing Ac9g123 pACYCDuet plasmid with both chloramphenicol and arabinose.

2.1.4 *In-Vitro Cleavage Assay*

In order to test sensitivity of ^{5M}C4 and verify methylation insensitivity of ^{5M}C5, an in-vitro cleavage assay was performed using oligonucleotides ordered from Eurofins Genomics. The non-target strand of the annealed oligos were labeled with a 5'-HEX fluorescent tag for visibility on a gel and had a sequence of 5'-[HEX] gctaaggtggataccaggatcttgccatcc-3'. The other oligos had the following sequences: 5'-ggatggcaagatcctggtatCCACCttagc-3' (WT), 5'-ggatggcaagatcctggtatCCA^{5M}CCttagc-3' (^{5M}C4), 5'-ggatggcaagatcctggtatCCAC^{5M}Cttagc-3' (^{5M}C5), and 5'-ggatggcaagatcctggtatCCA^{5M}C^{5M}C ttagc-3' (^{5M}C4+^{5M}C5). For each sample, HEX-tagged and non-HEX-tagged oligos were combined at a molar ratio of 1:20 and then heated in a water bath to 75 °C for five minutes to anneal before being allowed to cool back to room temperature again. Four tubes were labeled for each sample (WT, ^{5M}C4, ^{5M}C5, ^{5M}C4+^{5M}C5) and each combined 5uL nuclease-free water, 2uL AceCas9 cleavage buffer (composed of 20mM tris buffer at a pH of 7.5, 150mM potassium chloride, 2mM dithiothreitol, 10mM magnesium chloride, and 5% glycerol), 1uL of 10uM appropriate sgRNA (final concentration of 1uM), and 1uL of 10uM AceCas9 protein (final concentration of 1uM) to each tube before being mixed via pipetting then heated in a water bath to 37 °C for 15 minutes. After the 15 minutes each sample was allowed to cool to room temperature before 1uL of their respective annealed oligo was added to each tube. A negative control was also created (7uL nuclease-free water, 2uL AceCas9 cleavage buffer, and 1uL WT annealed oligo) and all five tubes were set in a water bath at 50 °C to incubate for one hour. 10uL 2X formamide gel loading buffer was added to each tube after this period, the five tubes were heated to 90 °C for two minutes, and then loaded into a 15% urea PAGE gel for gel electrophoresis. Pictures of the gel were then taken by a Chemidoc MP Imaging System for analysis.

2.2 Methylation Assay

2.2.1 *Creation of Methylated DNA*

Samples of ccdB-AcPAM containing a PAM sequence of 5'-AGGCC-3' were treated with HaeIII methyltransferase using NEB's "Recommended Protocol for Methylation of Genomic DNA", modified to instead use between 100-150ng of DNA to ensure complete methylation of DNA. In a tube 5uL of the HaeIII methyltransferase kit-supplied methyltransferase reaction buffer, 5uL of S-adenosyl methionine at a concentration of 1600uM, 100-150ng of DNA, 4-25 units (approximately 1uL) of the kit-supplied methyltransferase enzyme, and nuclease-free water up to a total reaction volume of 50uL was mixed by pipetting then incubated in a 37 °C water bath for one hour before being transferred to a 65 °C water bath for 20 minutes to stop the reaction.

2.2.1.1 Verifying Methylation Status

After the methylation reaction, methylation status of the DNA needed to be verified. The following protocol was based on NEB's HaeIII methyltransferase "Control Protection Assay" but modified to use only a portion of the methylated DNA to ensure complete cleavage of DNA if it was unprotected. 40uL of NEBuffer 2.1, 10 units (approximately 1uL) of HaeIII restriction endonuclease, and 9uL of the methylated DNA was mixed via pipetting before incubating in a 37 °C water bath then analyzed on a 1% agarose gel. Unmethylated DNA was used in a separate reaction to act as a control. When there was no evident sign of cleavage in the methylated DNA sample despite cleavage in the unmethylated DNA sample, the methylated DNA was used to create diluted samples of methylated DNA.

2.2.2 *Serial Dilution of Samples*

Samples containing varying percentages of methylated ccdB-AcPAM plasmid were created via serially diluting with unmethylated ccdB-AcPAM. Tested percentages ranged from 100% methylated to $1*10^{-5}$% methylated (decreasing by an order of magnitude with each dilution, so 100% was used to make 10% then the 10% was used to make the 1% and so on) as well as 0% methylated (containing exclusively unmethylated DNA). The concentration of the methylated DNA had become highly diluted initially due to the methylation reaction resulting in a calculated concentration of approximately 3ng/uL whereas the concentration for the

unmethylated DNA was approximately 30ng/uL. This difference in concentration was accounted for when creating the first dilution, 10%, by diluting the methylated DNA by mass instead of volume. Afterwards all less methylated dilutions were based on volume instead. After the dilutions were created, they would be used as the "DNA" component of the AceCas9 Cleavage Assay.

2.2.3 *AceCas9 Cleavage Assay*

The in-vitro cleavage reaction was started by combining into a tube 1uL of 10uM AceCas9 protein (final concentration 1uM), 1uL of 10uM appropriate sgRNA (final concentration 1uM), 2uL of AceCas9 cleavage buffer (same ingredients as before), and 5uL of nuclease-free water then mixing via pipetting before incubating in a 37 °C water bath for 15 minutes before being allowed to cool to room temperature. Then 1uL of DNA respective to the sample being tested was added to each tube for a total combined reaction volume of 10uL before being placed in a 50 °C water bath for two hours to incubate.

2.2.3.1 Cleanup of DNA

DNA was cleaned using NEB's "Monarch® PCR & DNA Cleanup Kit (5 µg)" following the provided protocol. Before adding the cleanup binding buffer each sample was diluted with 15uL nuclease-free water to reach a total volume of 25uL then mixed with 60uL of cleanup binding buffer. At the end of the protocol 12uL of elution buffer was used to elute the cleaned DNA. This DNA was then used as the "template DNA" component during PCR amplification.

2.2.4 *PCR Amplification of DNA*

PCR amplification was performed using NEB's "Q5® High-Fidelity 2X Master Mix for PCR" following the provided protocol. An additional pair of PCR primers were added to the mixture to act as a control for the amplification. The target site forward and reverse primers for this reaction are 5'-cgtgaactttacccggtggt-3' and 5'-agagcgttcaccgacaaaca-3' respectively and were intended to amplify the target site for the AceCas9 cleavage reaction (resulting in amplification only if there is DNA left uncleaved). The additional forward and reverse primers, referred to as the control site primers, were 5'-cagcaacgcggccttttttac-3' and 5'-cgccggctgactttctcaat-3' respectively and were of a region not targeted at all by AceCas9 during

the cleavage reaction (resulting in amplification regardless of cleavage). Initially the PCR reaction for each sample consisted of 12.5uL kit-provided master mix, 1.25uL of each primer at a concentration of 10uM for each, 4uL template DNA, and 3.5uL nuclease-free water for a total volume of 25uL. Due to the faint target site band on the gel, this was later changed to 12.5uL kit-provided master mix, 1.25uL of the forward and reverse control site primers at a concentration of 10uM for each, 2.5uL of the forward and reverse target site primers, 4uL template DNA, and 1uL nuclease-free water for a total volume of 25uL. The conditions for the PCR mix was as follows: 1 cycle at 98 °C for 30 seconds, 35 cycles of 98 °C (10 seconds) followed by 59 °C (30 seconds) followed by 72 °C (20 seconds), 1 cycle at 72 °C (2 minutes), then held at 4 °C until collected.

2.2.5 *Analysis of DNA via Gel Electrophoresis*

The resulting PCR product was then mixed with NEB 6X Purple Gel Loading Dye before 4uL of each sample were added to a 1.5% agarose gel containing ethidium bromide then run until bands separated. Pictures of the gel taken by a BIO-RAD Chemidoc MP Imaging System were then analyzed using BIO-RAD Image Lab Software to quantitatively determine volume, adjusted volume, and band percent for all bands of interest.

CHAPTER 3

RESULTS AND DISCUSSION

3.1 Refinement of Methylation Assay

The methylation assay as described initially had fewer required steps. During the cleavage assay, NEB buffer 3.1 was used instead of AceCas9 cleavage buffer and the incubation period was one hour instead of two. Control primers were not added to the PCR amplification step and the steps involving DNA cleanup and analysis via Image Lab Software were also not present. The first gel derived from this series of steps (Figure 5, left) and second gel derived from the same steps with the addition of the control primers to PCR amplification (Figure 5, right) depicted a perfect cleavage for the unmethylated sample containing no methylated DNA (0%) and all other samples other than $1*10^{-5}$% would depict clear bands without error.

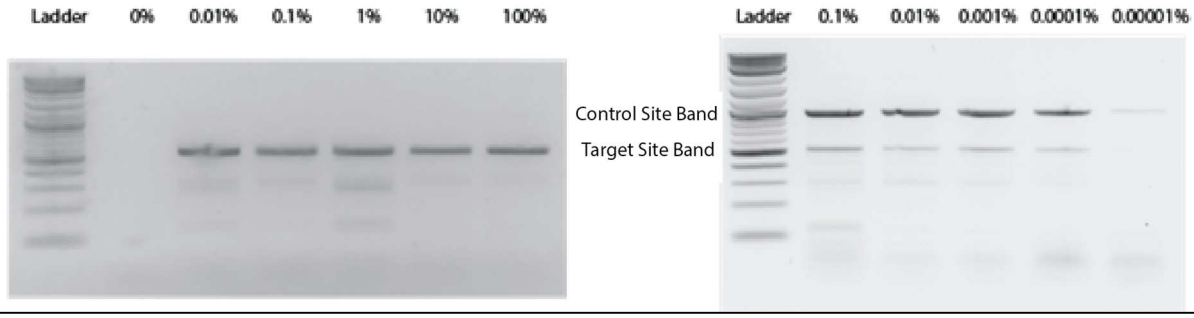


Figure 5. Depicts two agarose gels. There contents are listed from left-to-right. Of note, the top-most band in the right gel (other than the ladder) is the result of the control primers which were added to the protocol after the left gel was made. The band after this, and the top-most band in the left gel, is the band of interest while others below are extraneous.

Left gel: Ladder, 0% methylated, 0.01% methylated ($1*10^{-2}$%), 0.1% methylated ($1*10^{-1}$%), 1% methylated, 10% methylated, and 100% methylated.

Right gel: Ladder, 0.1% methylated ($1*10^{-1}$%), 0.01% methylated ($1*10^{-2}$%), 0.001% methylated ($1*10^{-3}$%), 0.0001% methylated ($1*10^{-4}$%), and 0.00001% methylated ($1*10^{-5}$%).

Replicating these gels, however, had several issues as anytime a gel would be replicated using the same conditions used to create the initial gels several samples seemingly at random would not be cleaved properly or show up on the gel. Eventually this issue was solved by implementing the DNA cleanup step to remove all non-DNA impurities before amplification. This resulted in its own problem as all samples, including the 0%, showed some faint target site

band. Further refinement of the method involved swapping the NEB buffer 3.1 (used originally because this was the buffer Lee et al. used in their experiment) for AceCas9 cleavage buffer and extension of the cleavage assay incubation from one hour to two hours to increase target site band prominence. Twice the original volume of target site primers were added during the PCR amplification step to further increase target site band prominence. Once this was accomplished, the gel could no longer be simply analyzed based on target site band presence because there existed a faint band for even the 0% methylated samples. Instead, BIO-RAD Image Lab Software was used to quantify band intensity for comparison. Despite these changes to the original methylation assay more will still be required. The methylation assay has proven to be highly sensitive in terms of detecting methylation but its limit has not yet been found. Further tests will be required to determine this limit.

3.2 Measuring Band %

Results obtained from Image Lab Software shows the "band %", a measurement of how much the band intensity of one band in a sample contributes to the total band intensity from all bands measured (in this case just the control site band and the target site band). This provides a ratio which can be used to quantitatively measure if the target site band for any given sample on a gel represents methylated DNA (if its target site band % is higher than the target site band % from the 0% sample) or represents the baseline band one could expect from an entirely unmethylated sample (if its target site band % is approximately equivalent to the target site band % from the 0% sample). It is therefore important to compare target site band % of all other samples to the target site band % from the 0% sample to accurately determine the presence or lack of methylated DNA.

3.3 Methylation Assay Can Detect Methylation as Low as $1*10^{-5}$%

After experiments resulting in the refinement of the methylation assay, a working version was derived (described in Chapter 2) and was being used to determine the theoretical lowest concentration of methylated DNA it could detect. The most recent measurements obtained by this assay was a set of three gels which were obtained from the DNA of one cleavage assay that was then PCR amplified three separate times. These gels depicted samples containing as little as $1*10^{-5}$% methylated DNA. One such gel is listed below followed by a graph listing the band %

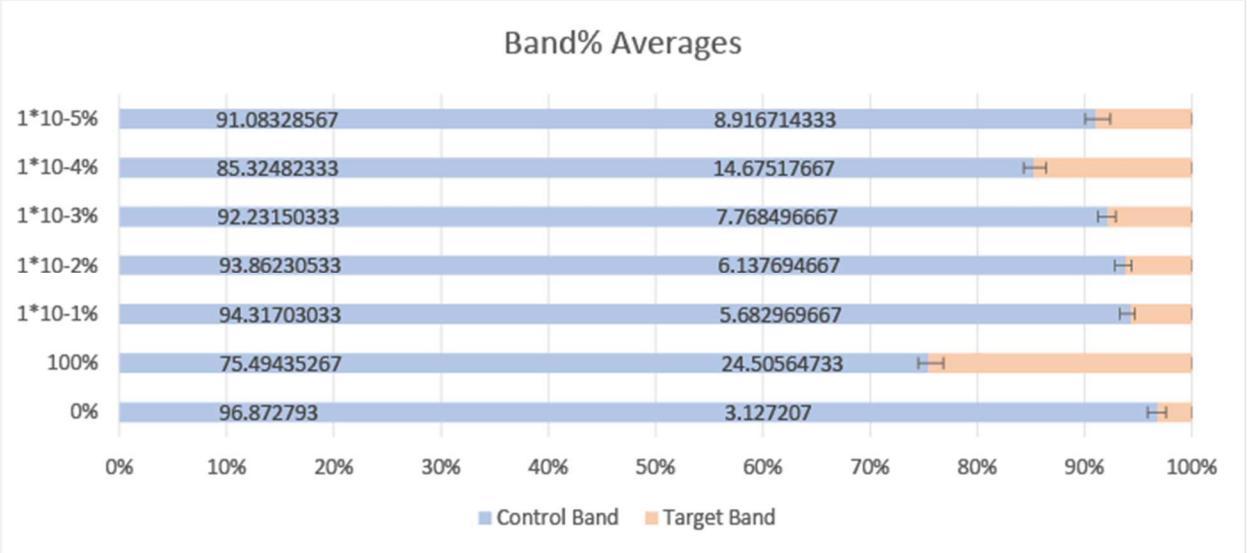
values for each of the samples (Figure 6). Averaging the band % values for all three of the gels and finding their standard deviation shows the target band % values for any given sample is consistently higher than the target band % for the 0% sample, with enough of a difference for random deviation to be unlikely explanation (Figure 7). This provides support for the methylation assay to currently be capable of detecting methylated DNA as low as $1*10^{-5}$% of any given sample. More experimentation will be necessary to determine if this is the detection limit of the assay but, once perfected, this assay could theoretically detect without bisulfite sequencing the presence of methylation on a single targeted nucleotide that accounts for as little as $1*10^{-5}$% (or lower) of tested DNA.

Figure 6. Gel (top) and corresponding graph depicting band % values (bottom). For a sample to be considered too diluted to be detected its target band % needs to approximately match the target band % of the 0% sample (which contains no methylated DNA). The

Figure 6 - Continued
100% sample has a much higher target band % which some samples approach but did not have enough methylated DNA to match.

Table 1. Table listing average band % values and their standard deviation (top) and corresponding graph depicting band % values (bottom). Standard deviation reveals variance between 0% sample and any other sample is too large to reliably be random chance.

Band% Averages			
Sample	Control Band%	Target Band%	Standard Deviation
0%	96.872793	3.127207	0.729723213
100%	75.49435267	24.50564733	1.362130911
$1*10^{-1}$%	94.31703033	5.682969667	0.396866044
$1*10^{-2}$%	93.86230533	6.137694667	0.571810964
$1*10^{-3}$%	92.23150333	7.768496667	0.708786563
$1*10^{-4}$%	85.32482333	14.67517667	1.100304316
$1*10^{-5}$%	91.08328567	8.916714333	1.356474663



3.4 Proposed Experiments and Next Steps

3.4.1 *Expanding Current Methylation Assay Utility*

In addition to refining the protocol involved for the methylation assay, there exists two possible means to expand its utility without modifying the Cas9 protein in use. Several oxidation products of 5-methylcytosine (5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxylcytosine), believed to be important to the demethylation pathway of cytosine, cannot be detected using traditional bisulfite sequencing methods (*42*). When treated with sodium bisulfite, 5-hydroxymethylcytosine is converted into cytosine-5-methylenesulfonate which is read as 5-methylcytosine while both 5-formylcytosine and 5-carboxylcytosine are read as thymine (*42*). Additional testing using DNA containing these oxidation products could determine if AceCas9 is similarly sensitive to them as well. It would be hypothesized AceCas9 would similarly be sensitive to the presence of these products on its fourth position PAM cytosine but not its fifth position PAM cytosine because of the lack of direct interactions between its PID and this cytosine (*41*) (Figure 8, bottom)

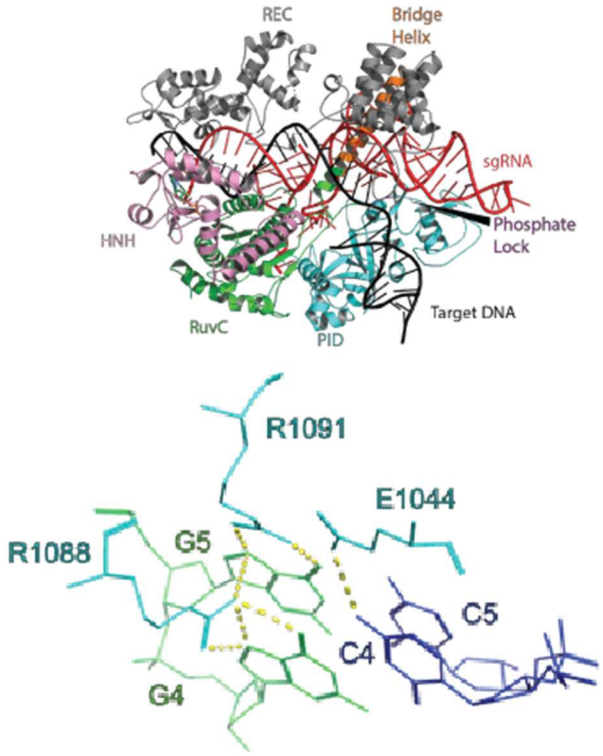


Figure 7. Domains of AceCas9. These figures depict AceCas9 protein in complex with DNA (top) and a close-up (bottom) showing the three amino acids from the PID that directly interact with C4 and C5 in the PAM sequence.

Figure 7 - Continued
In the bottom image, the target strand of DNA is also color-coded blue while the non-target strand is green.

Thus far, only oligonucleotide or plasmid DNA have been used in designing and testing the methylation assay. Determining the presence of methylation in genomic DNA, particularly DNA from human cells, is an essential next step. While CpC methylation (5-methylcytosine immediately followed by an unmethylated cytosine) is uncommon in healthy human DNA, Xiao et al. tested a variety of human cancer cell lines that each contained different CpN methylation including CpC methylation (*43*). SKBR3 cells specifically seemed to have a higher than typical concentration of CpC methylation as opposed to healthy human cells. Recreating conditions analogous to a target site containing CpC methylation for in-vitro testing and, ultimately, in-vivo tests using SKBR3 cells would be important for this goal.

3.4.2 *Engineering AceCas9 Which Recognizes 5'-NNNCG-3' PAM Sequence*

Due to the prevalence of promoters containing a CpG island (*30*) and tendency of CpG being the most frequent site for cytosine methylation in humans (*23, 24*), a version of AceCas9 that could recognize a 5'-NNNCG-3' PAM sequence while still retaining sensitivity to ^{5M}C4 would be useful in studying diseases caused by abnormal methylation levels. Research demonstrates the possibility to use site directed mutagenesis to loosen PAM specificity of SpCas9 with the stated possibility to replicate these results for any Cas9 protein with a known structure (*40*). Walton et al. ultimately implemented several such mutations to SpCas9 in order to loosen its PAM from 5'-NGG-3' to 5'-NGN-3'. A mutated AceCas9 protein which instead recognizes a PAM sequence of 5'-NNNCG-3' or 5'-NNNCN-3' would greatly expand potential target sites and utility of a methylation assay utilizing it if it were to remain methylation-sensitive. A crystal structure of AceCas9 was derived which should allow use of this method (*41*) (Figure 8).

REFERENCES

1. Sorek, R; Lawrence, M. C; Wiedenheft, B.; CRISPR-Mediated Adaptive Immune Systems in Bacteria and Archaea. *Annu. Rev. Biochem.* **2013.** 82, 237-266

2. Garneau, J; Dupuis, M; Villion, M; et al; The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*. **2010.** 468, 67–71

3. Bolotin, A; Quinquis, B; Sorokin, A; Ehrlich, D, S; Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology.* **2005.** 151, 2551-2561

4. Cui, Y; Xu, J; Cheng, M; et al; Review of CRISPR/Cas9 sgRNA Design Tools. *Interdiscip Sci Comput Life Sci.* **2018.** 10, 455–465

5. Jiang, F; Doudna, J, A; CRISPR–Cas9 Structures and Mechanisms. *Annu. Rev. Biophys.* **2017.** 46, 505-529

6. Mojica, F, J, M; Díez-Villaseñor, C; García-Martínez, J; Almendros, C; Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology.* **2009.** 155, 733-740

7. Zhang, F; Development of CRISPR-Cas systems for genome editing and beyond. *Q Rev Biophys.* **2019.** 52, 1-31

8. McClelland, M; The effect of sequence specific DNA methylation on restriction endonuclease cleavage. *Nucleic Acids Res.* **1981.** 9, 22, 5859-5866

9. Maier, J, A, H; Jeltsch A; Design and Application of 6mA-Specific Zinc-Finger Proteins for the Readout of DNA Methylation. *Methods Mol Biol.* **2018.** 1867, 29-41

10. Kungulovski, G; Jeltsch, A; Epigenome Editing: State of the Art, Concepts, and Perspectives. *Trends Genet.* **2016.** 32, 101-113

11. Anderson, E, M; Haupt, A; Schiel, J, A; et al; Systematic analysis of CRISPR–Cas9 mismatch tolerance reveals low levels of off-target activity. *J. Biotechnol.* **2015.** 211, 56-65

12. Tsui, T, K, M; Hand, T, H; Duboy, E, C; Li, H; The Impact of DNA Topology and Guide Length on Target Selection by a Cytosine-Specific Cas9. *ACS Synth Biol.* **2017.** 6, 1103-1113

13. Hand, T, H; Das, A; Li, H; Directed evolution studies of a thermophilic Type II-C Cas9. *Methods in Enzymology.* **2019.** 616, 265-288

14. Jabbari, K; Bernardi, G; Cytosine Methylation and CpG, TpG (CpA) and TpA Frequencies. *Gene.* **2004**. 333, 143-149

15. Fu, Y; Sander, J; Reyon, D; et al; Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol.* **2014.** 32, 279–284

16. Santos, D.P; Kiskinis, E; Eggan, K; Merkle, F.T; Comprehensive protocols for CRISPR/Cas9□based gene editing in human pluripotent stem cells. *Curr. Protoc. Stem Cell Biol.* **2016.** 38: 5B.6.1□ 5B.6.60

17. Sung, K; Park, J; Kim, Y; et al; Target Specificity of Cas9 Nuclease via DNA Rearrangement Regulated by the REC2 Domain. *J. Am. Chem. Soc.* **2018**, 140, 25, 7778-7781

18. Palermo, G; Chen, J, S; Ricci, C, G; et al; Key role of the REC lobe during CRISPR-Cas9 activation by 'sensing', 'regulating', and 'locking' the catalytic HNH domain. *Q Rev Biophys.* **2018.** 51, e9, 1-11

19. Anders, C; Niewoehner, O; Duerst, A; Jinek, M; Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature.* **2014.** 513, 569-573

20. Leenay, R, T; Beisel, C, L; Deciphering, communicating, and engineering the CRISPR PAM. *J Mol Biol.* **2017.** 429, 177-191

21. Kumar, S; Chinnusamy, V; Mohapatra, T; Epigenetics of Modified DNA Bases: 5-Methylcytosine and Beyond. *Front Genet.* **2018.** 9, 640, 1-14

22. Ratel, D; Ravanat, J; Berger, F; Wion, D; N6-methyladenine: the other methylated base of DNA. *Bioessays.* **2006.** 28, 3, 309-315

23. Ehrlich, M; Gama-Sosa, M, A; Huang, L, H; et al; Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. *Nucleic Acids Res.* **1982.** 10, 8, 2709-2721

24. Mikeska, T; Craig, J, M; DNA Methylation Biomarkers: Cancer and Beyond. *Genes (Basel).* **2014.** 5, 3, 821-864

25. Feng, Q; Hawes, S, E; Stern, J, E; et al; DNA Methylation in Tumor and Matched Normal Tissues from Non-Small Cell Lung Cancer Patients. *Cancer Epidemiol Biomarkers Prev.* **2008.** 17, 3, 645-654

26. Esteller, M; CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future. *Oncogene.* **2002.** 21, 5427–5440

27. Ehrlich, M; DNA methylation in cancer: too much, but also too little. *Oncogene.* **2002.** 21, 5400–5413

28. Ehrlich, M; DNA hypomethylation in cancer cells. *Epigenomics.* **2009.** 1, 2, 239-259

29. Gardiner-Garden, M; Frommer, M; CpG islands in vertebrate genomes. *J Mol Biol.* **1987.** 196, 2, 261-82

30. Saxonov, S; Berg, P; Brutlag, D, L; A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci U S A.* **2006.** 103, 5, 1412–1417

31. Kessler, C; Manta, V; Specificity of restriction endonucleases and DNA modification methyltransferases — a review (edition 3). *Gene.* **1990.** 92, 1-2, 1-240

32. Hsu, P, D; Scott, D, A; Weinstein, J, A; et al; DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol.* **2013.** 31, 9, 827–832

33. Zhang, Z; Zhang, Y; Gao, F; et al; CRISPR/Cas9 Genome-Editing System in Human Stem Cells: Current Status and Future Prospects. *Mol. Ther. Nucleic Acids.* **2017.** 9, 230-241

34. Nishimasu, H; Shi, X; Ishiguro S; et al; Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science.* **2018.** 361, 6408,1259-1262

35. Chen, Z; Zhao, H; A highly sensitive selection method for directed evolution of homing endonucleases. *Nucleic Acids Res.* **2005.** 33, e154

36. Hanck, T; Schmidt S; Fritz, H, J; Sequence-specific and mechanism-based crosslinking of Dcm DNA cytosine-C5 methyltransferase of E. coli K-12 to synthetic oligonucleotides containing 5-fluoro-2'-deoxycytidine. *Nucleic Acids Res.* **1993.** 21, 303-309

37. Fraga, M, F; Esteller, M; DNA Methylation: A Profile of Methods and Applications. *BioTechniques.* **2002.** 33, 632-649

38. Li, L, C; Dahiya, R; MethPrimer: designing primers for methylation PCRs. *Bioinformatics.* **2002.** 18, 1427-1431

39. Lee, S, H; Yu, J; Hwang, G-H; et al; CUT-PCR: CRISPR-mediated, ultrasensitive detection of target DNA using PCR. *Oncogene.* **2017.** 36, 6823-6829

40. Walton, R, T; Christie, K, A; Whittaker, M, N; Kleinstiver, B, P; Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science.* **2020.** 368, 290-296

41. Das, A; Hand, T, H; Smith, C, L; et al; The Molecular Basis for Recognition of 5'-NNNCC-3' PAM and Its Methylation State by Acidothermus cellulolyticus Cas9. *Nat. Commun*. **2020**. 11: 6346.

42. Bhattacharjee, R; Moriam, S; Umer, M; Nguyen, N; et al; DNA methylation detection: recent developments in bisulfite free electrochemical and optical approaches. *Analyst.* **2018.** 143(20):4802-4818.

43. Xiao, W; Liu, X; Niu, X; Li, C; et al; The frequency of CpG and non-CpG methylation of Notch3 gene promoter determines its expression levels in breast cancer cells. *Exp. Cell Res.* **2020.** 386, 111743

BIOGRAPHICAL SKETCH

Ethan Samuel Wickline

Education

Thesis Master of Science (Biochemistry) July 2021
Florida State University, Tallahassee, FL Overall GPA: 3.42

Bachelor of Science (Biochemistry) May 2018
Stetson University, Deland, FL Overall GPA: 3.456

Skills

- Creating cell agar plates as well as plating and incubating cell colonies
- Designing primers for use in site directed mutagenesis/DNA sequencing/PCR/bisulfite sequencing
- Site directed mutagenesis
- Polymerase Chain Reaction (PCR)
- Bisulfite sequencing
- Planning and performing cell survival assays
- CRISPR Cas9 in vitro cleavage assay
- Gel electrophoresis (both agarose and polyacrylamide) and assays like western blot and gel mobility shift assay
- Protein purification via Ni-NTA purification and AKTA HPLC system
- Bacterial plasmid transformation via electroporation or chemical transformation
- Preparation of competent cells
- Bradford protein assay
- Chemical titration
- From-scratch creation of pertinent lab reagents, synthesis/separation/purification of compounds
- Thin Layer Chromatography (TLC)
- Reading and interpreting IR/NMR/UV-Vis spectra
- Softwares: Pymole, Snapgene, Mathematica, Adobe Illustrator, Microsoft Office (Certified in Word and Powerpoint)

Experience

Graduate Student Research January 2019-Present
Florida State University, Tallahassee, FL

- Planned and conducted research pertaining to the recognition of methylated cytosine by AceCas9 (acidothermus cellulolyticus Cas9)
- Co-author of "The molecular basis for recognition of 5′-NNNCC-3′ PAM and its methylation state by Acidothermus cellulolyticus Cas9"
    - Paper available from Nature Communications at: https://www.nature.com/articles/s41467-020-20204-1

Teaching Assistant August 2018-Present
Florida State University, Tallahassee, FL

- Assisted with chemistry course lectures and teaching labs
- Responsible for student engagement and safety
- Graded student assignments
- Designed online course for transition to virtual teaching

Undergraduate Student Research January 2015 – May 2018
Stetson University, Deland, FL

- Assisted in research by running Gaussian simulations of 11th row transition metal-bonded NHC complexes and interpreted physical properties from measurements