TARGETING BIFURCATED METHYLTRANSFERASES TOWARDS USER-DEFINED DNA SEQUENCES

by
Brian Chaikind

A dissertation submitted to Johns Hopkins University in conformity with
the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
February 2014

Abstract

There would be great value in targeting methylation toward user-defined DNA sequences. Directing methylation toward single CpG sites within a genome would provide a means to examine the effects of single epigenetic alterations on cellular phenotype. The spread, erasure, or maintenance of such modifications could be examined in different cellular contexts and at different genomic loci. Further, as aberrant methylation patterns cause or are implicated in many disease states, targeted methylation might be used as a therapeutic.

Many groups have attempted to target methylation toward user-defined sites by fusing a methyltransferase enzyme to a sequence specific DNA binding domain. This strategy biases the methyltransferase toward specific DNA sequences, but the methyltransferase enzyme is active in the absence of the sequence specific DNA binding event. A better strategy would involve linking the DNA binding event of sequence specific proteins to the activity of the methyltransferase enzyme.

The contents of this thesis describe work on an assisted protein assembly strategy for targeting methylation to single CpG sites within a genome. This strategy utilizes naturally or unnaturally bifurcated methyltransferases fused to zinc fingers to affect reassembly over a desired site. The bifurcated methyltransferases are engineered to have reduced affinity for each other and/or for DNA, preventing unassisted

enzymatic reassembly at non-targeted CpG sites. Zinc finger binding to sequences flanking an internal CpG site increase the local concentration of these assembly-deficient, bifurcated methyltransferases, enabling enzymatic reassembly and methylation only over the targeted CpG site.

In Chapter 2, we demonstrate the successful implementation of this strategy for two prokaryotic methyltransferases, M.HhaI and M.SssI. Further, we elucidate design parameters important for constructing active, targeted, bifurcated methyltransferases. In Chapter 3, we describe a novel directed-evolution strategy to quickly identify optimized zinc finger-fused bifurcated methyltransferases. Importantly, we also demonstrate that substitution of bifurcated methyltransferase fragments with new zinc fingers predictably targets methylation toward new zinc finger cognate sequences. Finally, in Chapter 4, we describe successful preliminary studies in human cell lines. We demonstrate the eukaryotic expression of both fragments, targeting specific sites in a mammalian expression vector and methyltransferase activity on chromosomal DNA.

Advisor: Professor Marc Ostermeier

Readers: Professor Sarah Woodson

Professor Jeffrey J. Gray

Acknowledgments

I would like to first sincerely thank my advisor, Dr. Marc Ostermeier for his advice, mentorship, and constant support. I am privileged to have learned from him and to have had such a fantastic example in his mentorship. Marc was always available to meet, always willing to assess my work, and always patient with my thousands of questions. I had not heard of directed evolution or protein engineering before arriving at Johns Hopkins. Whatever I know now about this field, I learned here.

I would also like to thank Dr. Woodson and Dr. Gray for graciously agreeing to read and comment on this thesis. Thanks to the Chemistry-Biology Interface Program, especially Lauren McGhee, former director Dr. Marc Greenberg, and current director Dr. Steve Rokita.

Thanks to current and past lab members; especially Dr. Glenna Meister from whom I inherited this project. I couldn't imagine a better work environment. I met some of my best friends here. Clay Wright and Elad Firnberg were constant sources of advice and support inside and outside the lab. We watched each other get married, brewed beer, and spent long hours joyfully working together.

Baltimore is, of course, the greatest city in America; my family moved here before the Johns Hopkins University was founded, I'm the

first one to attend. The Baltimore Orioles were a great source of inspiration---and frustration.

Thanks to my parents Hinda and Stephen Chaikind who always encouraged my passion for science even when I broke every TV (albeit temporarily) in our house for science's sake. Thanks for helping me with my homework and for constantly encouraging me to drop out of this program to make wine. Thanks to my sister Laurie who was always there to help me put things back into perspective and helped me laugh off the tough days. Thanks Grandma, and to my Aunts and Uncles, who are proud of me, even though they are not sure what it is I do (neither am I). Thanks to my new family, the Marrs, who have always embraced me as one of their own.

Most importantly, thanks to my wife Kendra, who has been a constant source of support, strength and sanity. She has displayed a superhuman patience for my unconventional work schedule and my (poor) sense of humor. I consider myself insanely fortunate to be with her and I honestly could not have accomplished what I have without her.

Also, every time I become frustrated with an experiment, I try to remember that Sir Francis Bacon, father of modern scientific inquiry, died of pneumonia caused by a successful experiment designed to test if a dead chicken could be preserved by refrigeration. 'Hey,' I think to myself, 'My experiment failed, but at least a dead chicken didn't kill me.'

Table of Contents

List of Tables

Table of Figures

1 Introduction

1.1 Biology of methyltransferases

1.1.1 Discovery of methyltransferases

The study of DNA methyltransferases is inexorably linked to our understanding of molecular biology. In 1959, while studying hydroxymethlcytosine incorporation in T-phages, Arthur Kornberg noted that "it would be interesting to look for an enzymatic mechanism for direct methylation of DNA" in eukaryotic tissues [1]. Then, in 1963, Gold and Hurwitz reported partially purified methyltransferases from *E. coli* and showed that they could catalyze methyl transfer onto "acceptor" DNA substrates [2,3]. Tying these discoveries together, Werner Arber and colleagues showed that phage produced in met⁻ strains, grown in the absence of methionine, were not altered by their host [4] and later that adenine methylation conferred the host-specific modification protecting phage DNA from restriction endonuclease degradation [5,6]. Phage "host-specificity" is the result of the methyltransferase activity of the host during phage replication. Phage DNA produced in these hosts is methylated and therefore rendered resistant to host restriction endonucleases, whose activity is blocked by methylated DNA. It is now known that the biological roles of prokaryotic methyltransferases extend beyond the restriction modification (R-M) systems responsible for

bacterial phage immunity. Methyltransferases play important roles in mismatch repair, cell cycle control, and gene expression [7]). Further, as discussed later, groups have challenged the idea that R-M systems function exclusively as an innate immunity against phage infection.

Analysis of genebank data has identified 4990 putative restriction genes and 8080 modification (i.e. methyltransferase) genes [8]. Prokaryotic DNA methyltransferases are divided into two broad classes of S-adenosylmethionine (SAM) dependent enzymes; the first class methylates the C5 position of cytosine and the second class methylates the N4 or N6 exocyclic nitrogen of cytosine or adenine respectively (i.e. N4mC and N6mA methyltransferases) [9]).

1.1.2 M.HhaI and M.SssI: Understanding DNA methyltransferase structure and function

M.HhaI is the archetypal prokaryotic cytosine-5-methyltransferase. Described in the Hamilton Smith's Nobel lecture, M.HhaI recognizes and methylates the internal cytosine of a four base-pair GCGC site [10]. Crystal structures and biochemical studies of this enzyme revealed the general architecture and mechanism of all cytosine-5-methyltransferases. M.HhaI has a bilobal global architecture composed of a large and small domain connected by a hinge [11]. The two lobes create a cleft that binds DNA.
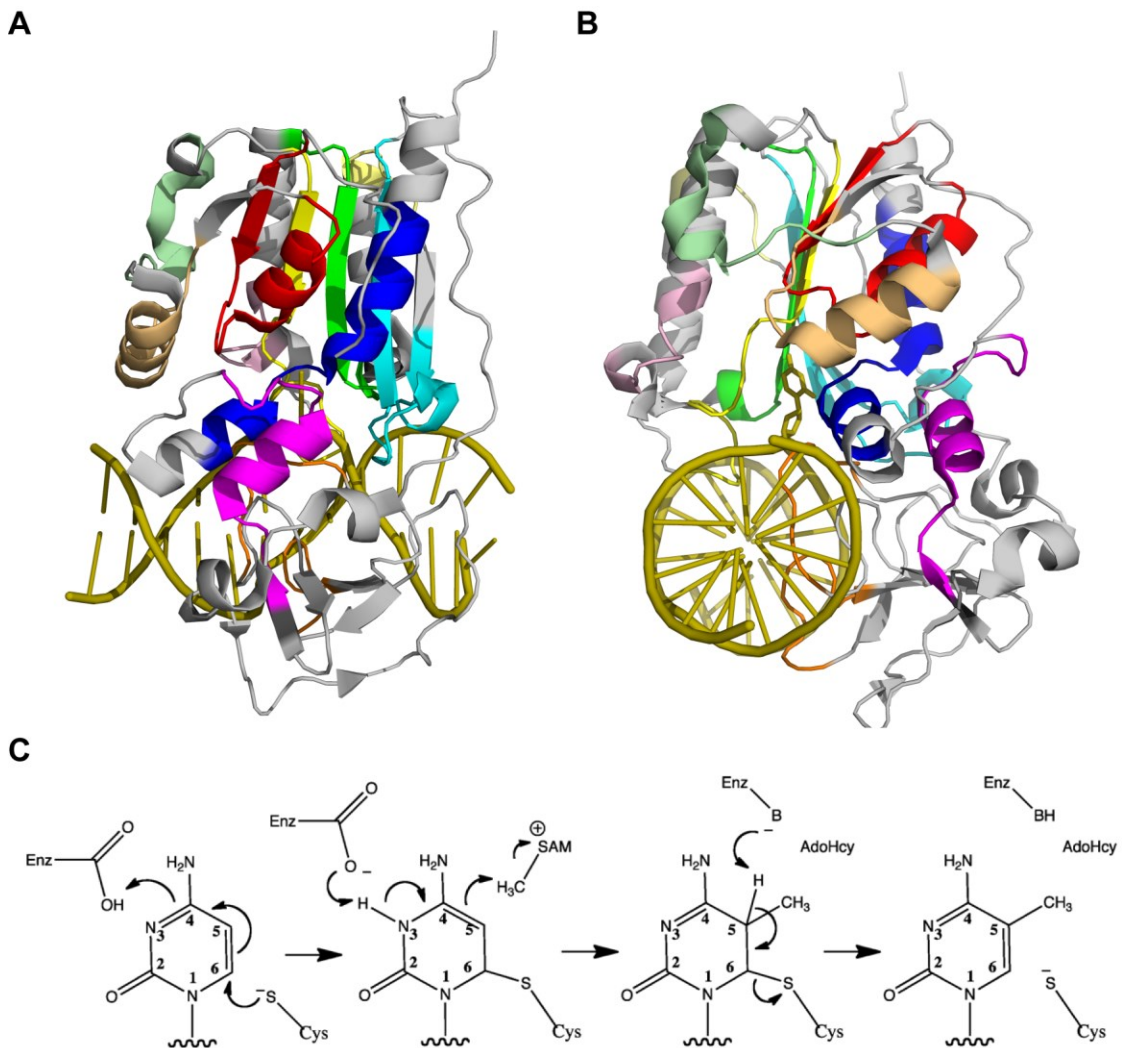
Figure 1.1 Methyltransferase M.HhaI (2HRI) structure and mechanism of action [12]**.** Highly conserved motifs are shown in bright colors, those less conserved are shown in light or pale tint colors. Motif I-red; Motif II-light orange; Motif III-pale green; Motif-IV-yellow; Motif V-light pink; motif VI-green; Motif VII-pale yellow; Motif VIII-cyan; Motif IX-purple; Motif X-blue; target recognition (TRD) loops-orange; and DNA-olive [13]. A) Left: Looking at M.HhaI with the C-terminal α-helix surrounded by highly conserved motifs I, IV, VI, and VIII. B) Right: A view of the methyltransferase looking down the helical axis. Note the position of the TRD loops in the major groove and the cytosine residue flipped 180° out of the helix and into the active site. C) Bottom: Mechanism of action of M.HhaI.

Based on early sequence alignments and crystal structures, ten

conserved motifs could be identified (Fig. 1.1 A and B) [11,13,14].

Domains I-V and domain X are involved in SAM binding [9,15]. Domain IV also contains the catalytically active cysteine residue. Domains VI and VII are involved in binding DNA [9]. Importantly, a relatively non-conserved, glycine rich portion of the enzyme comprises the target recognition domain (TRD) [11].

Later crystal structures revealed the TRD and active site loop make the base-specific contacts with the M.HhaI GCGC recognition site. The enzyme bound to its GCGC recognition site induces a "base-flipping" of the cytosine out of the DNA helix and into the active site of the enzyme (Fig. 1B). Key amino acid residues are shown to stabilize both the unpaired guanine and flipped out cytosine [16,17].

The enzymatic mechanism of cytosine-5-methyltransferases is shown in Figure 1.1C and has been determined by biochemical studies [18,19]. Upon flipping of the cytosine into the active site, the active cysteine attacks the C6 position of the pyrimidine ring. This results in a transient covalent bond between cytosine and the enzyme. It has been proposed, for related enzyme M.HaeIII, that protonation of the N3 facilitates cysteine attack of the C6 and proton removal facilities nucleophilic attack from the C5 position onto the methyl group of SAM [20]. ß-elimination of the C5 hydrogen and covalently bound cysteine restores aromaticity to the ring [21,22].

1.1.2.1   M.SssI: A prokaryotic methyltransferase that recognizes all CpG sites

Like M.HhaI, M.SssI is a prokaryotic CpG methyltransferase. However, upon cloning and characterization, it was discovered M.SssI methylates all CG sites (unlike M.HhaI which methylates GCGC) [23]. This was significant, as it was the first prokaryotic methyltransferase discovered with sequence specificity identical to mammalian methyltransferases. M.SssI was found to contain the same conserved motifs as M.HhaI and a homology model based on M.HhaI crystal structure highlighted many of the potential key interactions between M.SssI and DNA; this study also revealed that the structure and function of these two enzymes are probably very similar [24]. Later mutational studies elucidated important residues affecting the catalytic activity and DNA affinity of M.SssI [25,26].

1.1.3 Natural evolution

The conserved domain architecture and mechanistic similarities of methyltransferases provides evidence for a common evolutionary ancestor. Two, non-mutually exclusive hypotheses: a gene duplication/sequence permutation model and a horizontal gene transfer model, have been proposed to explain the vast diversity and ubiquitous presence of these enzymes in nature. In 1989, Lauster proposed that gene duplications events of a common 12-16 Kda ancestor led to the initial formation of the major classes of methyltransferases. This was based on the observation that adenine and cytosine methyltransferases

have homologous blocks of sequences that are repeated intermolecularly,

and in some cases, intramolecularly [27,28]. Wilson noted that some

motifs on different methyltransferases were permutated with respect to

the linear sequence [29]. Malone *et al.* later identified nine of the ten

structural motifs in the N6mA and N4mC methyltransferases originally

identified in the C5-methyltransferases and subsequently classified these

enzymes by the linear order of their motifs [30].
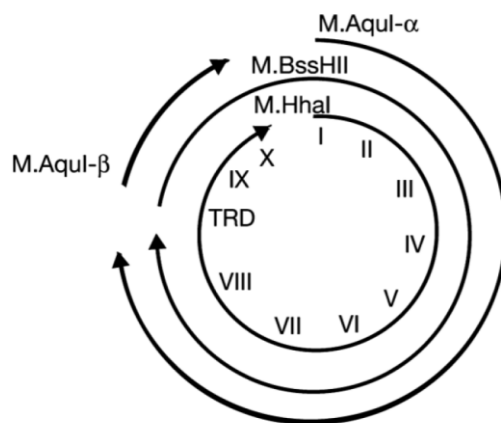
Figure 1.2 Sequence permutations among prokaryotic CpG methyltransferases. An illustration showing how the ten conserved methyltransferase motifs can be naturally circularly permuted (M.BssHII) or bifurcated (M.AquiI) with respect to M.HhaI (republished with permission from [31]).

Sequence permutation appears to have played a role in the diversification

of methyltransferases, functioning either by gene duplication followed by

in-frame fusion, motif shuffling between different methyltransferases or a

combination of the two [32,33]. Regardless of the mechanism, all classes

of methyltransferases show some degree of sequence permutation.

Though rare among C-5-methyltransferases, M.BssHII was shown to be

circularly permuted (Fig. 1.2) [34]. Further, M.AquiI and M.EcoHK31I

were shown to be natural, obligate heterodimers; notably, in both methyltransferases, the open reading frames encoding these heterodimeric fragments overlap [35-37].

Additionally, horizontal gene transfer has been proposed to explain the widespread appearance and diversification of restriction modification systems in prokaryotes. Jeltsch *et. al.* showed a significant correlation between the phylogenetic trees constructed of methyltransferases' innate DNA recognition sequences and these enzymes' amino acid sequences [38]. However, these phylogenies did not match the phylogenetic tree of bacteria from which the methyltransferases were derived [38-40]. In other words, the apparent evolutionary relationships among R-M systems appear different than the evolutionary relationships between the species containing those systems. Jeltsch further went on to show that the codon usage in 29% of endonuclease genes, but to a much lesser extent methyltransferase genes, varied significantly from the codon usage of the parent organism [40]. This might be explained by the sequential nature of the theoretical gene transfer. A prokaryotic organism might first acquire a methyltransferase, in order to protect its' genome, before acquiring a potentially toxic restriction endonuclease. The concept of domain swapping and mutation to obtain new sequence specificities will be discussed in a later section.

Kobayashi *et. al.* proposes that R-M systems may act as selfish, mobile genetic elements at odds with the fitness of the organism [41].

This contrasts with the hypothesis of R-M systems evolving as a means of beneficial immunity against phage or exogenous genetic elements. Among many arguments in this paper, Kobayashi bases this theory on the fact that loss of R-M systems often result in loss of cell viability and that R-M systems are often associated with mobile genetic elements [41]. Both Kobayashi's theory, along with the phage immunity theory, provide explanations for the selective pressures required for the maintenance and diversification of methyltransferases. Restriction endonucleases with novel substrate specificity will degrade chromosomal DNA unless an active methyltransferase is present to methylate DNA and block digestion.

In eukaryotes, there are six subfamilies of DNA methyltransferases (Dnmt1-6) [42]. Diversification within these groups most likely results from gene duplication events and fusion to other regulatory domains, but does not appear to be involved with restriction modification systems [43]. Phylogenetic analysis of Dnmt1, Dnmt2 (an RNA methyltransferase), and Dnmt3 subfamilies suggests that the three families were present in the last common eukaryotic ancestor [42-44].

1.1.4 Mammalian methyltransferases
Cytsosine-5-methyltransferases are the only DNA methyltransferases present in eukaryotes. Dnmt1 and Dnmt3 comprise the two families of human DNA methyltransferases; the motifs of their catalytic domains

show structural homology to M.HhaI [45-47]. Although mechanistically similar, the eukaryotic enzymes are fused to other domains responsible for protein-protein interactions and regulation, adding additional layers of structural and functional complexity relative to their prokaryotic counterparts.

Dnmt1 is responsible for maintenance of methylation due to its innate preference for hemimethylated DNA as well as its recruitment to DNA during replication and DNA repair [48-51]. Dnmt1 functions as part of a large multi-subunit complexes; it localizes to replication foci by binding proliferating cellular nuclear antigen (PCNA) [52] and a protein known as ubiquitin-like, containing PHD and RING finger domains 1 (UHRF1) [53]. Additionally, the methyltransferase has been shown to interact with multiple other proteins (as reviewed in [54]).

Dnmt3a and Dnmt3b are the de novo methyltransferases [55,56]. Dnmt3a binds to DNA as a heterotetramer, comprised of two Dnmt3a enzymes and two molecules of Dnmt3L. Dnmt3L is a catalytically inactive protein essential for maintaining proper regulation and activity of Dnmt3a and potentially Dnmt3b [46,57-59]. Unlike Dnmt1, Dnmt3a and Dnmt3b are highly expressed at different stages in embryonic stem cells, but not ubiquitously in most cell types [55,60].

Dnmt3a and Dnmt3b are also regulated by recruitment to specific histone modifications and proteins involved in chromatin remodeling. Dnmt3L, Dnmt3a and Dnmt3b were shown to bind specifically to the

unmethylated lysine 4 of histone 3 (H3K4) [61-63]. Dnmt3a was shown to interact with the histone 3 trimethylated lysine 36 modification (H3K36me3) [64]. Additionally, Dnmt3 proteins have been shown to interact with the histone methyltransferases and histone associated proteins such as G9a (as reviewed [65]). However, these forms of transcriptional and spatial regulation do not completely explain how methylation patterns are initially established during development.

1.1.5 Role in mammalian development and disease

DNA methylation is the most extensively studied epigenetic modification in eukaryotic cells and is involved in transcriptional repression. CG islands, defined generally as an increase in CpG content at specific regions within the genome, are associated with 72% of gene promoters [66,67]. In mammals, the erasure of DNA methylation is involved in germ cell development and maintenance of stem cell pluripotency. Its reestablishment is important for cellular differentiation, maintenance of haematopoietic stem cell renewal, centromeric formation, inactivation of potentially deleterious repetitive elements, and maintenance of imprinted control regions; DNA methylation is influenced and influences histone posttranslational modifications and protein-DNA interactions (as reviewed in [68]). In many instances it seems that transcriptional repressor binding and/or histone posttranslational modification precedes methylation. However, methylation may play an important role in

stabilizing transcriptional repression and maintaining regions of DNA in a heterochromatin state [69].

The importance of DNA methylation on proper development is evident from numerous knockout studies. In mice, Dnmt3b knockouts result in embryonic lethality and Dnmt3a knockout die shortly after birth [56]. Dnmt1 knockouts also result in embryonic lethality [70]. Interestingly, embryonic stem cell triple knockout mutants of these three enzymes show no signs of chromosomal instability and are capable of normal stem cell renewal [71].

Aberrant methylation patterns are responsible for numerous disease states and are implicated in many more. Imprinting is a difference in methylation of maternal and paternal alleles, leading to differential allelic expression. Mutation and/or loss of proper imprinting causes several diseases such as Prader-Willi and Beckwith-Wiedemann syndromes (as reviewed in [72]). Further mutations in Dnmt3b have been proposed to lead to immunodeficiency, centromeric instability, and facial anomalies (ICF) syndrome [73]. Finally, cancer is characterized by global hypomethylation and hypermethylation at CpG islands (as reviewed in [69]).

1.1.5.1 Epigenetic regulation of intercellular adhesion molecule 1 (ICAM1) in cancer

ICAM1 is a transmembrane protein that plays a role in trans-endothelial cell migration, APC-T-cell communication, and cell signaling (as reviewed

in [74]). The role of ICAM1 expression in cancer progression is not straightforward and may depend on the cancer type and stage of the disease. There is evidence to suggest that increased levels ICAM1 expression in tumors leads to a better disease outcome. Tumor growth in mice was much smaller and showed more leukocyte infiltration if the cancer was transfected with an *ICAM1* gene prior to injection [75]. Among human breast cancer tissues, ICAM1 was negatively correlated with tumor size and metastatic potential [76]. Ovarian cancer tissues showed decreased expression compared to normal tissue, as assessed by immunohistochemistry [77].

In contrast, there is also evidence suggesting that upregulation of ICAM1 facilitates metastasis and tumor growth. The mRNA levels of ICAM1 in cancer cell lines positively correlated with metastatic potential and siRNA ICAM1 gene suppression lead to a decrease in invasion (as assessed by *in vitro* assays) [78]. ICAM1 expression was also associated with metastatic potential in human hepatocellular cancer and expression was shown to increase dramatically upon metastasis in mice models [79]. High levels of ICAM1 in serum have been associated with increased disease potential in numerous cancer lines, presumably by interfering with normal immune cell recognition of the membrane bound form of ICAM1 [78].

Aberrant epigenetic regulation of *ICAM1* may explain its altered expression in disease states. Treatment of ovarian cancer cell lines with

5-aza-2'-deoxycytidine resulted in an increase in expression in all lines tested [77]. ICAM1 promoters were shown to be methylated in 70% of 105 separate bladder cancer tissues removed from patients [80]. Hellebrekers *et al.* showed that ICAM1 transcription in tumor conditioned cells could be upregulated by incubation with Dnmt1 and HDAC inhibitors [81]. Interestingly, however, no difference in methylation was observed before and after incubation with inhibitors. This may be representative of the early stages of epigenetic programming, as changes in histone modifications were correlated with observed increases in expression [81,82]. A similar result was also obtained for highly and moderately methylated ovarian cancer cells [83]. Finally, targeted demethylation within the ICAM1 promoter was correlated to a 2 fold upregulation in gene transcription [84].

1.1.6 Biological implications of site-specific CpG methylation

Though methylation at a single site is not believed to be the main means of epigenetic transcriptional silencing, multiple studies suggest single methylation events can alter expression levels for select genes. *In vitro* methylation of a single CpG site within the *S1000A2* promoter on a reporter plasmid resulted in significant downregulation of gene expression relative to an unmethylated control [85]. Methylated oligonucleotides targeting an intronic region of *peroxisomal membrane protein 4* (*PXMP4 or PMP24*) resulted in a single methylation mark on

chromosomal DNA that down regulated gene expression relative to controls; this corroborated differences observed between normal tissues and tumor cells [86]. Electromobility shift assays show that methylation at a single site impairs the binding of the global insulator CTCF [87].

1.2 Engineering of methyltransferases to target altered or user-defined sequences

Given the biological importance of DNA methylation and its involvement in disease, it would be generally useful to target DNA methylation toward user-defined sequences. In addition to studying effects on transcription, an engineered methyltransferase that methylates a specific single site in a promoter would be generally useful for studying the effects of single aberrant methylation events on the propagation, maintenance, and correction of epigenetic marks. Finally, methyltransferases were recently engineered to more efficiently incorporate the transfer of unnatural alkyl groups donated by S-adenosylmethionine cofactor analogues [88]. This may make it possible to use targeted methyltransferases in order to site-specifically label DNA.

Several groups have used rational design or directed evolution to alter both the specificity and activity of methyltransferases. Strategies to alter the recognition sequences can be divided into three general approaches. The first, I refer to as the *mutant methyltransferase* strategy. This involves mutating important residues in methyltransferases or "swapping" domains responsible for targeting methyltransferases to

specific DNA sequences. The second strategy, which I refer to as the *methyltransferase-DNA binding domain fusion strategy*, involves fusion of a methyltransferase to a sequence specific binding domain. The third, which I refer to as an *assisted protein complementation strategy*, involves using zinc finger binding events to aid the assembly of a heterodimeric methyltransferase.

1.2.1 Mutant methyltransferase strategy

The first strategy involves mutating or exchanging residues involved in targeting specific sequences of DNA. As discussed above, the target recognition domain (TRD) is a region of poor conservation among various methyltransferases and is responsible for most of the residues that make sequence specific contacts with DNA. Balganesh *et. al.* first demonstrated that the targeting specificity of fusion methyltransferases was, in part, due to a variable "non-conserved region" [89]. The variable or target recognition domain of different methyltransferases can be exchanged to alter methyltransferase specificity [90-93]. However, this technique often greatly reduces or eliminates enzymatic activity. Further, this strategy does not allow for the creation of truly new DNA targeting specificities. Groups have also used mutation and *in vitro* selection strategies to relax or alter the specificity of methyltransferases. Most selection strategies rely on methylation-dependent protection from restriction endonuclease digestion to enrich for DNA encoding a methyltransferase with altered

specificity [94-98]. Methyltransferases engineered through these strategies typically do not have activities and specificities comparable to the parent methyltransferase(s). Further, the specificities obtained through these methods are usually palindromic and small (≤ 6 base pairs in length). Thus, the sequence specificities are too small and limited to target specific regions in a genome.

1.2.2 The methyltransferase DNA binding domain fusion strategy

Many groups have engineered methyltransferases that bias methylation towards user-defined DNA sequences. The general strategy, pioneered by Xu and Bestor, involves fusion of a sequence specific DNA binding domain (typically a zinc finger protein) to a methyltransferase enzyme (Fig. 1.3A) [99]. Interaction between the sequence specific DNA binding and its recognition site localizes the fused methyltransferase domain to adjacent CpG sites, resulting in biased methylation. These constructs have been used to affect methylation, *in vitro*, in *E. coli*, and in cancer cell lines [100-104]. Biased methyltransferases have been shown to stably and heritably reduce the expression of *Sox2* and *Maspin* [105]. Siddique *et al.* demonstrated that targeting methylation towards the VEGF-A promoter significantly reduced gene expression in SKOV3 cells [106]. A recent review summarizes much of the literature on the creation and use of these engineered methyltransferases [107].
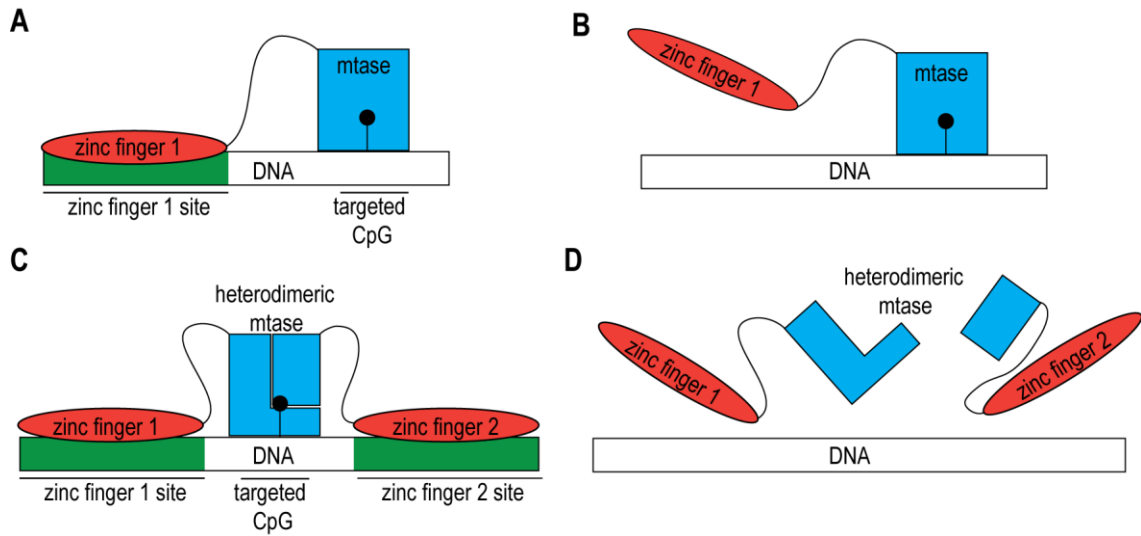
Figure 1.3 A graphical comparison between the methyltransferase DNA binding domain fusion strategy and the assisted protein assembly strategy A) A monomeric methyltransferase-zinc finger fusion protein. DNA–zinc finger interactions bias the methyltransferase activity toward adjacent CpG sites. B) Off-target methylation occurs because the methyltransferase is active even in the absence of zinc finger-DNA interactions. C) Both zinc finger-binding events at sites flanking a targeted CpG site assist heterodimeric methyltransferase reassembly. D) Heterodimeric methyltransferase assembly and activity does not occur without zinc finger-DNA interactions.

The limitation of this strategy is that the methyltransferase catalytic domain is still active and methylates in the absence of zinc finger-DNA interactions (Fig 1.3B). Thus, the methyltransferase fusion construct is still able to methylate non-targeted sequences. Most engineered methyltransferases methylate multiple CpG sites adjacent to the desired target site on the DNA. Despite the successes of these studies in biasing methylation to a particular region, little work has focused on targeting methylation to single CpG sites [108-110].

1.2.3 Assisted protein assembly strategy

Our strategy for achieving single-site, targeted methylation is to make the assembly of a heterodimeric methyltransferase dependent on specific DNA sequences flanking a targeted CpG site (Fig. 1.3C). To accomplish this task, our lab has previously employed a natural heterodimeric DNA methyltransferase and engineered these heterodimers to reduce their ability to reassemble into a functional enzyme [110]. Reducing the ability of the fragments to self assemble in a functional form is necessary as we and others have shown that bifurcated methyltransferases are capable of unassisted reassembly into functional enzymes [31,36,37,111,112]. Using these assembly-defective fragments, reassembly is then assisted by fusion of the fragments to zinc fingers, whose recognition sequences flank the targeted CpG site. The zinc finger domains bind to DNA, increasing the local concentration of the fused methyltransferase fragments over a targeted CpG site. DNA lacking zinc finger binding sites cannot affect functional reassembly, preventing methylation at other sites (Fig 1.3D). Previous work in our lab demonstrated that optimization of the linker length, C-terminal heterodimeric fragment, and target site promotes the functional assembly at the targeted CpG site with little observable activity at other sites [110].

This strategy is similar to protein complementation strategies first demonstrated in the reassembly of dihydrofolate reductase (DHFR). Bifurcated versions of this enzyme were shown to reassemble into a

functional enzyme only if fragments were fused to homodimerizng and heterodimerizing proteins [113]. Reviews of protein complementation assays highlight that many enzymes can be altered through bifurcation and fusion to dimerizing proteins [114,115]. The Ostermeier lab has demonstrated assisted reassembly for bifurcated aminoglycoside phosphotransferase (3')-IIa Neo fused to leucine zippers [116]. Our methyltransferase strategy differs, however, because DNA binding proteins do not dimerize, but rather complex with adjacent sequences, thereby assisting reassembly.

1.2.3.1   Zinc finger nucleases

The assisted protein assembly strategy has been used to develop zinc finger nucleases (ZFN) and TALE-nucleases (TALENs). The wealth of information generated about the function and limitations of ZFNs will help to inform later discussions about targeted methyltransferases. Zinc finger nucleases were first reported by Kim *et. al.* Zinc finger nucleases are composed of two zinc finger proteins fused to restriction endonuclease dimers (typically FokI) [117]. FokI is a type IIs restriction endonuclease with modular endonuclease and DNA recognition domains; importantly, endonuclease domains were shown to require dimerization for activity [118,119]. In a functional ZFN, two zinc finger proteins, each fused to a FokI cleaveage domain, recognize two DNA sequences that flank a DNA site targeted for digestion. Thus, this strategy directs DNA

19

cleavage to the sites specified by the zinc finger proteins. As discussed below, zinc fingers can be designed or evolved to recognize novel sequences. In theory, this strategy provides a means to make user-defined double stranded breaks targeted to any sequence of DNA recognized by zinc finger binding proteins. Induction of a double strand break followed by non-homologuos end joining (NHEJ) or homologous directed repair (HDR) has been used to create genetic knockouts and targeted genetic insertions. The technology has been shown to be widely successful, targeting numerous genes in a number of model organisms including humans (as reviewed in [120]). TALENs, which rely on an alternative sequence specific binding protein, induce double strand breaks using the same principle and have also successfully also been used to affect genetic alterations (as reviewed in [121]).

In spite of the successful widespread application of this technology, zinc finger nuclease technology is still being optimized to reduce off-target nuclease activity. Zinc fingers have been shown to be cytotoxic and this toxicity is believed to be a result of off-target cleavage events [122,123]. Recently, the Liu group was able to use an *in vitro* selection technique to identify large numbers of off-target cleavage events catalyzed by these enzymes [124]. Several factors may cause this off-target cleavage. First, FokI endonucleases can form dimers even if one of the proteins lacks a DNA binding domain [119]. ZFN targeting ability may not be dependent on both the zinc finger-DNA binding interactions and

20

homodimeric formation may cause off-target cleavage events. Several groups have used both structure-guided design and directed evolution techniques to reduce the innate ability of FokI to dimerize, reducing observed cytoxicity [125-128]. Other research has shown that cytoxicity is inversely correlated with the DNA binding specificity of the zinc finger domains; this implicates off-target zinc finger binding as a cause of off-target cleavage events [129]. Finally results in cancer cell lines indicated that the linkers connecting zinc fingers to FokI domains, as well as the spacer length separating zinc finger binding sites greatly affect ZFN selectivity [130]. The combined results of these studies demonstrate that a number of factors will affect the specificity and activity of assisted protein assembly strategies based on obligate heterodimeric-zinc finger fusions proteins.

1.2.3.2   Design of zinc fingers with novel DNA binding specificities
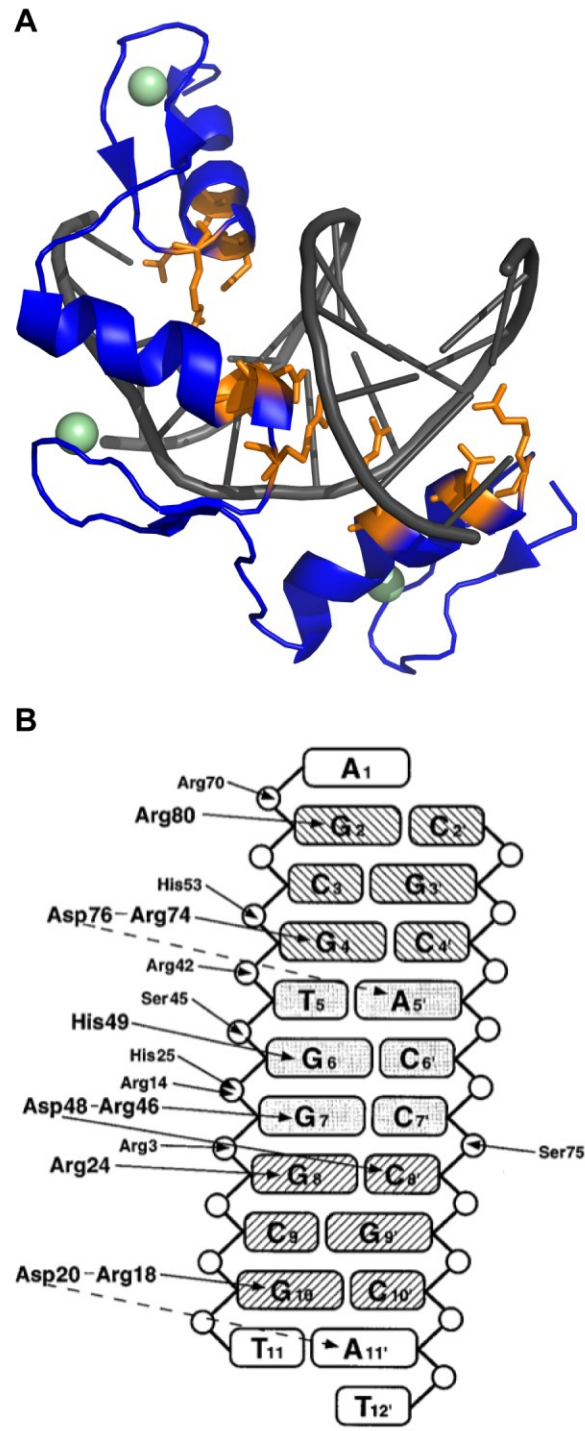
Figure 1.4 Crystal structure and key binding interactions of Zif268 with DNA A) Crystal structure ,1AAY, of DNA (grey) bound by zinc finger 268 (blue). The alpha helices sit in the major groove enabling residues to make base-

specific contacts, here shown as sticks (orange). B) Diagram outlining all residues in zinc finger 268 that make sequence specific contact with DNA (republished with permission from [131]).

Zinc fingers were first identified as a repetitive protein motif in *Xenopus* oocytes. Each motif was found to contain two conserved histidine and cysteine residues coordinated to zinc [132]. There are now known to be many other types of zinc fingers, classified by the composition of the four amino acids responsible for coordinating zinc (as reviewed in [133]). The discussion here will be restricted to a specific class of canonical C_2H_2 zinc fingers that are used in most of the described engineering studies mentioned. Pabo's group solved the crystal structure of murine zinc finger 268, revealing many of the general structural features of these proteins (Fig. 1.4A and B). Zinc fingers are comprised of multiple motifs, each containing two beta sheets, antiparallel to one another, and an alpha helix. This motif structure is stabilized by contacts between hydrophobic residues as well as a coordinated zinc ion [134]. The alpha helix sits within the major groove of DNA and mediates almost all of the base-specific contacts with one strand of DNA. Each zinc finger motif recognizes three bases of DNA. However, residues often contact the bases immediately preceding this three base sequence and each motif may not make contacts with all three bases. The zinc finger binds the three bases of DNA in an 'antiparallel' orientation [131,134].

The periodic nature and regular structure of zinc finger binding proteins has encouraged many groups to design zinc fingers that bind to

new sequences. Theoretically, any sequence of DNA could be targeted if zinc finger motifs were completely modular and if motifs could be identified that recognize each of the 64 possible three-base pair combinations. Pabo *et al.* reviews work attempting to create 'rules' linking zinc finger protein sequence with DNA binding specificity. He concludes that although the protein motif is fairly well conserved, individual zinc finger residue/DNA interactions are not always regular and therefore predictable [135].

Directed evolution techniques have provided a more tractable and successful approach for designing new zinc finger motifs that bind novel sequences. One widely used technique involves creating fusions between randomized zinc finger proteins and the pIII protein of filamentous phage. Expression of these fusion constructs and affinity purification of phage, using novel DNA sequences bound to a solid support, identifies zinc finger motifs with novel sequence specificity [136-138]. This strategy is most successful when one motif of the three-finger protein is modified at a time. Using phage selections, motifs have been selected that recognize 49 of the possible 64 three base pair combinations [139-143]. Available lab programs identify motifs based on input DNA sequences, allowing for facile design of zinc fingers with new DNA binding specificities [144,145]. However, as assessed by a bacterial two-hybrid system, 76% of zinc finger proteins created by assembling previously-selected zinc finger motifs fail to bind their predicted sites [146].

One possible explanation for this observation is that zinc finger motif-DNA interactions are not completely modular, but are instead partly context dependent upon surrounding motifs [147]. To address this, the Joung lab has used several strategies, combining DNA shuffling and bacterial two-hybrid selections to identify 3-motif zinc fingers that recognize DNA *in vivo* [148-150]. These strategies may be more effective at generating functional and specific zinc fingers for *in vivo* applications. Computer programs such as context dependent assembly (CoDA) allow for *in silico* design by combining two-motifs units known to function together. Designing zinc fingers with the context dependent assembly strategy has a greater reported success rate than modular assembly strategies [151].

In spite of the successful application and creation of zinc fingers with novel DNA specificities, there are still limitations to zinc finger design. As discussed above, high levels of off-target activity have been observed for zinc finger-fusion proteins and 15 of the 64 possible three nucleotide DNA binding motifs have yet to be identified. The recent elucidation and use of sequence specific transcription activator-like effectors (TALEs) have proven a promising alternative to zinc fingers. The discovery that a single repeated motif recognizes a single DNA base may make these proteins more modular and less context dependent on surrounding motifs than zinc fingers [152,153]. The discovery of two divariable residues capable of conferring base-protein recognition

enabled the construction of TALEs with predictably altered specificities [154].

Intended to be blank

2 Targeting DNA methylation using and artificially bisected M.HhaI fused to zinc fingers

2.1 Introduction

Many groups have biased methylation to specific DNA sequences by fusing methyltransferase enzymes to sequence-specific DNA binding proteins [99-101,103,155,156]. However, these fusion proteins still methylate away from the desired DNA sequence. This off-target activity occurs because the methyltransferase remains functional in the absence of the DNA-binding protein's association with its cognate DNA sequence. To reduce off-target activity, methyltransferase fusion proteins have been engineered with reduced overall activity, so that a bias in methylation can be observed. However, reducing enzyme activity does not address the fundamental limitation of this strategy, and these fusion constructs still methylate at non-targeted sites. Furthermore, many of these studies assess the level of specificity and activity in eukaryotic cells, which contain endogenous CpG methyltransferases. This therefore limits the ability to conclusively determine the true specificity and activity of these enzymes *in vivo*. A better strategy would make methyltransferase activity contingent upon association of the DNA binding domain with its target DNA sequence. Characterization of these enzymes in *E. coli*, which lack CpG methyltransferases, rather than eukaryotic cells, will allow for the

unambiguous characterization of enzymatic activity and specificity *in vivo*.

Our strategy for designing targeted methyltransferases couples the methyltransferase activity to the DNA binding protein's association with DNA. A monomeric methyltransferase is split into two fragments that are compromised in their ability to assemble into an active heterodimeric enzyme, and each fragment is fused to a different zinc finger. The zinc fingers' DNA binding sites flank a desired CpG site. Thus, zinc finger binding to cognate DNA sites increases the local concentration of the two attached methyltransferase fragments, encouraging the fragments to reassemble only over a desired CpG site. The association of the two fragments into an active enzyme in the absence of the flanking zinc finger binding sites is limited because the two fragments are engineered to have reduced affinity for one another or require each other for proper folding. In this manner, the strategy is akin to a assisted protein assembly or protein complementation assay [157] with a specific DNA sequence mediating assembly of the active methyltransferase.

The Ostermeier laboratory has previously demonstrated this strategy using the naturally split methyltransferase M.EcoHK31I, which methylates the internal cytosine of the 5′-YGGCCR-3′ site. We demonstrated how reduction of the fragment's affinity for each other through truncation of one of the fragments increased the ratio of methylation at the target vs. non-target sites. The optimized construct

exhibited (>50%) methylation at the target site and undetectable methylation at the non-target site under the correct expression conditions [110]. However, off-target methylation could be observed under different expression conditions. Furthermore, the M.EcoHK31I targeted cytosine residue is not a CpG site, and therefore would not be applicable for CpG methylation studies in mammalian cells.

Here we demonstrate our strategy using an artificially split M.HhaI, a CpG methyltransferase derived from M.HhaI fragments previously identified in our lab [31]. Using modeling and experimentation, we show how proper geometric configuration of the M.HhaI fragments and the zinc fingers is important for the bias and activity observed at the target site. With the proper fusion configuration of M.HhaI fragments and zinc finger proteins, we show how bias towards the target site can be increased through mutations rationally designed to reduce the association of the two fragments, through optimization of the linkers connecting the M.HhaI fragments to the zinc fingers, and through optimization of the distance between the zinc finger binding sites and the targeted methylation site. Optimization resulted in an engineered methyltransferase that methylated 50-60% of a desired the target site in *E. coli* cells with minimal levels of methylation at a non-target M.HhaI site.

2.2 Materials and methods

2.2.1 Modeling

The structural model for M.HhaI methyltransferase was obtained from the crystal structure of the M.HhaI/DNA complex (PDB 2HR1) [12]. For target DNA sequences, straight B-DNA structures containing all the three binding sites (one M.HhaI target site and two zinc-finger binding sites) were built using the model.it web server [158].

For zinc fingers HS1 and HS2, homology models were constructed using the Rosetta comparative modeling algorithm employing zinc finger Zif268 (1AAY) as the template [131,159]. Comparative modeling involves 1) copying coordinates from regions aligned with the template sequence, 2) a centroid pseudo-atom side-chain low-resolution building of the unaligned regions using a fragment based loop modeling protocol [160], and 3) a final all-atom high-resolution phase refinement with small backbone perturbations followed by gradient-based minimization and side-chain packing. One thousand models were generated for each of the zinc fingers and the top ranked structures based on the Rosetta standard energy function were selected. Kinks were observed in the C-terminal α-helices when these zinc finger models were superimposed on the template structure, as zinc atoms were not included during modeling. These kinks were fixed by threading the backbone of α-helices over the corresponding C-terminal α-helix from the template structure.

The final complex including zinc fingers and M.HhaI bound to the respective target sites was then assembled. The orientation of the zinc-fingers and M.HhaI at their respective binding sites was determined by aligning target DNA sequences from M.HhaI/DNA complex (2HR1) and Zif268/DNA complex (1AAY) with the straight B-DNA model.

Finally, the linker regions connecting the N-terminal and C-terminal fragments of M.HhaI to the zinc fingers were built using Rosetta kinematic closure (KIC) loop modeling algorithm [161]. The algorithm couples KIC calculations with 1) a low-resolution stage involving loop backbone minimization with side chains represented as centroids, and 2) an all-atom high-resolution stage with Monte Carlo-plus-minimization of side-chain and loop backbone dihedral angles. The calculations in the paper were carried out using Rosetta's developer revision number 46351. All algorithms are also available in Rosetta's release version 3.4. The Rosetta command-line arguments and scores used for the calculations are as follows.

(a) Modeling the zinc fingers

```
minirosetta.<exe> –database <path_to_rosetta_database>
        –run:protocol threading
        –in:file:template_pdb 1AAY.pdb
        –in:file:alignment 1hs1_.aln
        –cm:aln_format general
        –frag3 aa1hs1_03_05.200_v1_3
```

–frag9 aa1hs1_09_05.200_v1_3

–in:file:fasta 1hs1_.fasta

–loops:frag_sizes 9 3 1

–loops:frag_files aa1hs1_09_05.200_v1_3

aa1hs1_03_05.200_v1_3 none

–loops:extended

–loops:build_initial

–loops:remodel quick_ccd

–loops:refine refine_ccd

–out:file:fullatom

–out:nstruct 2000

–out:file:scorefile 1hs1_model.fasc

where 1AAY.pdb is the template structure and 1hs1_.aln is the

sequence alignment of target and template

(b) Constructing the linker regions (loop building)

loopmodel.<exe> –database <path_to_rosetta_database>

–loops:input_pdb fnl1.pdb

–loops:loop_file fnl1.loop

–loops:remodel perturb_kic

–loops:max_kic_build_attempts 500

–in:file:fullatom

–out:nstruct 100

–out:file:scorefile fnl1.fasc

where fnl1.pdb is the input PDB and fnl1.loop defines the range of loop residues

2.2.2 General methods, reagents, and bacterial strains

Restriction enzymes, T4 ligase, and M.HhaI were purchased from New England Biolabs (Ipswich, MA, USA) and were used according to manufacturers instructions. Oligos were purchased from Invitrogen (Carlsbad, CA, USA) and Integrated DNA Technologies (Coralville, IA, USA). Platinum® Pfx DNA Polymerase was purchased from Invitrogen (Carlsbad, CA, USA). dNTPs were purchased from Thermo scientific (Rockford, IL, USA). Agarose gel electrophoresis and PCR were preformed essentially as described previously [162]. *Escherichia coli* K-12 strain ER2267 [*F′ proA+B+ lacIq Δ(lacZ)M15 zzf::mini-Tn10 (KanR)/ Δ(argF-lacZ)U169glnV44 e14-(McrA-) rfbD1? recA1 relA1? endA1 spoT1? thi-1 Δ(mcrC-mrr)114::IS10*] was acquired from New England Biolabs (Ipswich, MA, USA) and was used for cloning and methylation protection assays.

2.2.3 Plasmid and gene construction and design

Plasmid pDIMN8 was derived from pDIMN7 MeND/MeCD [109]. An FspI restriction site was silently mutated within AmpR. Zinc finger genes were fused to M.HhaI methyltransferase gene fragments via desired length linkers using overlap extension PCR. Test sites for methylation (site 1 and site 2) were designed with an internal M.HhaI recognition site (5'-GCGC-3') nested within an FspI restriction site (5'-TGCGCA-3'). These

sites were flanked on either side by HS1 and HS2 zinc finger binding sites [163] or control DNA sequences as desired. Zinc finger recognition sites were separated from the FspI restriction site by 0, 1, 2 or 3 bp.

To facilitate changing the DNA sequences at these sites, site 1 was flanked by XmaI and EcoRI restriction sites and site 2 was flanked by AflIII and BglII sites. The BglII site was created by inserting three bp 66 base pairs downstream from the ColE1 origin of replication. The DNA at sites 1 and 2 were altered by annealing complimentary oligonucleotides encoding the desired DNA sequences. The oligonucleotides were designed such that the annealed product possessed overhangs that complemented the restriction site overhangs produced by digestion at the flanking restriction enzyme sites. Phosphorylation of the annealed oligonucleotides followed by ligation into digested vectors was used to change the sequence at sites 1 and 2.

2.2.4 Methylation protection assays and quantification.

In vivo protection assays were preformed in *E. coli* strain ER2267. Frozen stocks were prepared by inoculating 10 mL of lysogeny broth, supplemented with 100 μg/μl ampicillin and 0.2% w/v glucose, with cells from a single colony. After 12-16 hrs of incubation at 37˚ C, 800 μl of cell culture was mixed with 200 μl of 50% v/v glycerol to create glycerol stocks, which were stored at -80°C.

To perform methylation assays, 5 μl of thawed glycerol stocks were used to inoculate 10 ml of lysogeny broth supplemented with 100 μg/μl ampicillin salt. To repress the *lac* promoter, 0.2% w/v glucose was added. To induce the *lac* and *pBAD* promoters, cultures were supplemented with 1.0 mM of IPTG and 0.0167% w/v arabinose, respectively. Experiments carried out to optimize methylation indicated that inoculation into media containing glucose, IPTG, and arabinose resulted in the highest levels of observed methylation activity. Thus, cultures contained 0.2% glucose, 1.0 mM of IPTG and 0.0167% w/v arabinose unless otherwise indicated. After 12-14 hours of incubation at 250 rpm and 37° C, plasmid DNA was isolated from the cells using QIAprep Spin Miniprep Kit (Qiagen, Valencia, CA).

To ascertain the methylation status at sites 1 and 2 of the plasmid, plasmid DNA (500 ng) was incubated with 2.5 units of FspI and 10 units of NcoI-HF in buffer NEB4 at 37° C for 2 hours. After digestion, the DNA was electrophoresed in a 1.2% w/v agarose gel in TAE buffer at 90 V for 80 minutes at room temperature. Images were captured using the Molecular Imager XRS Gel Doc system with Quantity One software.

To quantify the percentage of plasmids methylated at each site, plasmid DNA (500 ng) was digested with 10 units of NcoI-HF and 2.5 units of FspI in buffer NEB4 at 37° C for 2 hours and half of each digested sample (250 ng) was electrophoresed in a 1.2% w/v gel for 2 hours at 90 V. Images were captured using the Gel Logic 112 Imaging

System. The intensities of each of the four largest bands were determined using Carestream Molecular Imaging Software and corrected to be on a mol basis using the expected length of each DNA fragment. Percentages of methylation are based on the intensity of a given band relative to the total intensity in the lane. Each construct was tested using ≥3 independent cultures. The mean percentage is reported and the error bar represents the standard deviation (n≥3).

2.2.5 Bisulfite sequencing

Unmethylated pDIMN8 plasmid was obtained by inoculating cells in 10 mL of lysogeny broth under conditions that repress gene expression, as described above. Methylated controls were obtained by incubating 8 μg of the unmethylated plasmid at 37°C for 1 hour with 50 units of M.HhaI (New England Biolabs, Ipswich, MA, USA) in 1X HhaI methylase reaction buffer supplemented with 32 mM S-adenosylmethionine. DNA was purified using the Zymo Clean and Concentrator kit according to manufacturers instructions (Zymo Research Corporation, Irvine, CA, USA).

For the experimental samples, ER2267 cells containing pDIMN8 plasmids encoding for methyltransferase fragments, were grown under conditions that induce expression, in triplicate, as described above. DNA was isolated from cells using QIAprep Spin Miniprep Kit and digested with 20-40 units of NcoI at 37° C for 1-2 hours (Qiagen, Valencia, CA).

Linearized DNA was purified using the Zymo Clean and Concentrator-5 kit according to manufacturers instructions. DNA was then treated with bisulfite reagent and purified using the EZ DNA Methylation-Gold Kit according to manufacturers instructions (Zymo Research Corporation, Irvine, CA, USA). Individual strands from site 1 and site 2 were then amplified with a set of unique primers, using One Taq Hot Start DNA Polymerase (New England Biolabs, Ipswich, MA, USA). PCR amplified DNA was purified using the Zymo Clean and Concentrator kit. Purified, PCR amplified DNA, 10 ng, was directly sequenced by Genewiz (South Plainfield, NJ, USA). The heights of trace files in Figure 2.6 were adjusted to aid in direct comparison.

2.2.6 HhaI restriction assay

To ascertain the total methylation status of all 36 HhaI recognition sites, ER2267 cells containing methyltransferase fusion constructs were used to inoculate 10 mL of lysogeny broth medium supplemented with 100 μg/mL ampicillin salt. Cells were inoculated in conditions shown to repress or induce methyltransferase gene expression (0.2% w/v glucose for gene repression or 0.2% glucose, 1.0 mM of IPTG and 0.0167% w/v arabinose for gene induction). After 12-14 hours of incubation at 250 rpm and 37° C, plasmid DNA was isolated from the cells using QIAprep spin miniprep kit (Qiagen, Valencia, CA, USA).

Plasmid DNA (500 ng) was incubated with 10 units of HhaI in 1X NEB4 and 1 µg/mL BSA at 37° C for 2 hours. After digestion, the DNA was electrophoresed in a 2.5% w/v agarose gel in TAE buffer at 90 V for 50 minutes at room temperature. Images were captured using the Molecular Imager XRS Gel Doc system with Quantity One software.

2.2.7 Chromosomal restriction assay

ER2267 cells containing variants with X=3, Y=1, and Z=0, a C-terminal truncation of 6 or 4 amino acids, and zinc finger binding sequences at site 1, were grown under conditions known to induce or repress expression of the heterodimeric methyltransferase (see Materials and Methods). Chromosomal DNA was isolated using the Sigma's GenElute Bacterial Genomic Kit according to the manufacturer's instructions (Sigma-Aldrich, St. Louis, Missouri , USA). Chromosomal DNA was electrophoresed on a 0.8% w/v agarose gel in TAE buffer at 90V for 35 min. to separate genomic DNA from plasmid DNA. Chromosomal DNA was isolated from the gel and purified using Invitrogen's PureLink Quick Gel Extraction Kit (Invitrogen Carlsbad, CA, USA). DNA was further purified by ethanol precipitation as described in Sambrook and Russel [162]. The global level of methylation on the chromosome was assessed by incubating 250 ng of DNA with 5 U of FspI in 1X NEB4 for 1 hr at 37°C. A methylated control was prepared by incubating 250 ng of chromosomal DNA at 37°C for 1 hour with 12.5 U of M.HhaI in 1X HhaI

methylase reaction buffer and 80 μM S-adenosyl-methionine, heat killing the reaction at 65°C for 20 minutes and then incubating the reaction with 5 U of FspI in 1X NEB4 for 1 hr at 37°C. (New England Biolabs, Ipswich, MA, USA). DNA was electrophoresed on a 1% w/v agarose gel in TAE at 90V for 70 minutes. Images were captured using the Molecular Imager XRS Gel Doc system with Quantity One software.

2.3 Results and discussion

Our group previously identified two M.HhaI fragments that could assemble into a functional methyltransferase enzyme in an unassisted fashion [31]. The N-terminal fragment is comprised of amino acids M.HhaI[1-240], and the C-terminal fragment is composed of amino acids M.HhaI[210-326]. Each fragment shared a common internal 30 amino acids, M.HhaI[210-240], referred to as the overlapping region. This overlapping region is analogous to the region where some natural methyltransferases are split or circularly permuted [33].

Nomura and Barbas reported that fusion of one zinc finger to the N-terminus of M.HhaI[1-240] and a second zinc finger to the C-terminus of M.HhaI[210-240] resulted in a targeted methyltransferase [108]. However, our analysis of their engineered enzyme using more definitive assays showed that it methylates target and non-target sites with the same low efficiency [109]. Nevertheless, we imagined that our fragments might be converted to a targeted methyltransferase if we (1) fused the

fragments to zinc fingers in the correct orientation relative to the target DNA sequence, (2) reduced the fragments ability to assemble in an unassisted fashion through mutations designed to reduce the fragments' affinity for each other, (3) optimized the linkers connecting the fragments to the zinc fingers, and (4) optimized the number of bases separating the zinc finger binding sites from the M.HhaI recognition site.

2.3.1 Initial studies

In principle, each methyltransferase fragment could be fused to a zinc finger at the fragment's N- or C-terminus (Fig. 2.1A); combining these fusion variants creates four distinct zinc finger/methyltransferase fragment fusion topologies. We have previously shown that a particular ZN/CZ fusion pair (see Fig. 2.1A for nomenclature) designed by Nomura and Barbas [108] exhibits low-level, non-specific methylation of M.HhaI DNA sites *in vivo* [109]. In contrast, our initial tests of the NZ/ZC fusion pair displayed some bias towards a target flanked by the zinc finger binding sites.

We next desired to model all four combinations of fusion pairs to predict the optimal combination for fragment reassembly at the target site and to estimate the linker lengths that would be required to connect the zinc finger and the methyltransferase fragments. However, the presence of the 30 amino acid overlapping region on both fragments complicated the modeling. We wondered if this region could be removed

from one of the two fragments without compromising activity. Using the NZ/ZC construct, we conducted a set of experiments designed to probe the importance of the common 30 amino acids present in both the N-terminal fragment and C-terminal fragment. These experiments revealed that when the fragments are fused to zinc fingers, the 30 amino acids could be removed from the N-terminal fragment (but not from the C-terminal fragment) without reducing methyltransferase activity. The fragment pair M.HhaI[1-209] and M.HhaI[210-326], which lacks any overlap in sequence, formed the basis for all subsequent experiments.
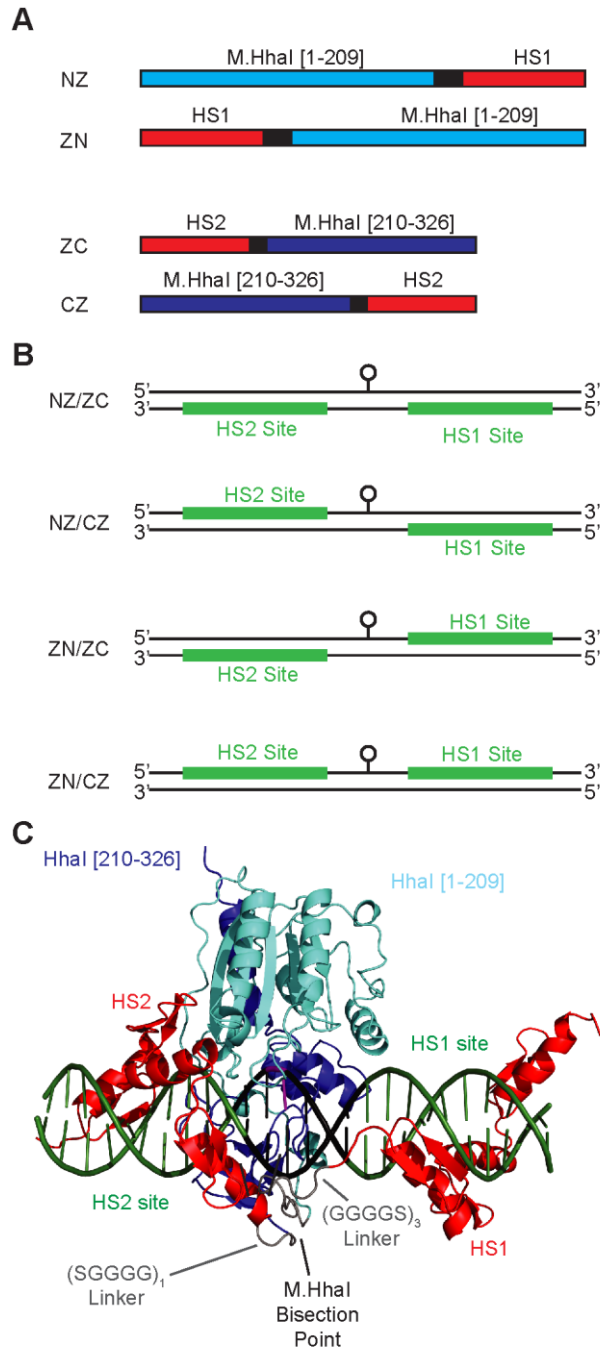
Figure 2.1 Schematic depictions of sequences and nomenclature of modeled protein/DNA complexes (A) Sequences of zinc fingers fused to fragments of M.HhaI methyltransferase. Numbers in brackets correspond to the amino acid numbers. Black segments correspond to linker sequences. (B) The orientation of the zinc finger binding sites relative to the intended methylation target site (the circle). The orientations depicted are the ones that would position the indicated protein pairs over the targeted CpG site. (C)

Molecular model of a particular NZ/CZ construct containing the indicated linkers. The base to be methylated is indicated in purple. Models of all complexes without linkers can be found in Figure 2.2. Krishna Praneeth Kilambi created the pymol model used to make this image.

2.3.2 *In silico* modeling illustrates the spatial constraints of a functional M.HhaI heterodimeric/zinc finger fusion protein with targeted activity

We used in silico modeling to predict the structures of the four possible combinations of M.HhaI fragments and zinc fingers bound to DNA. Each of the four pairs of fusion combinations required a particular placement of the zinc finger binding sites relative to the internal CpG site (Fig. 2.1B). Other orientations of the zinc finger binding sites relative to the internal CpG site would present one or both methyltransferase fragments away from this targeted cytosine. For each of the four configurations depicted in Figure 2.1B, we produced two models in which the methyltransferase was positioned to methylate either the top or the bottom strand relative to the bound zinc fingers (Fig. 2.2). Modeling assumed straight B-DNA structure and thus does not capture any distortions of the DNA that may or may not be induced by the binding of the fusion proteins to DNA.

Modeling predicted that fusion of the zinc fingers to the fragments at the bisection site (i.e. configuration NZ/ZC) would best position the fragments in an orientation capable of reassembling and therefore methylating a targeted CpG site (Fig. 2.1C). We judged this pair as optimal because it required the shortest linkers connecting the zinc

fingers to the methyltransferase fragments. All configurations other than NZ/ZC required linkers that would need to circumvent long distances around the DNA or methyltransferase domains and connect residues separated by at least 40 Å (Fig. 2.2). Although one could conceivably use very long, flexible linkers to traverse these long distances, we reasoned that such constructs would do a poorer job of increasing the local concentration of the two fragments at the target site.

The NZ/ZC model indicated that the linker connecting the N-terminal fragment and its respective zinc finger would need to be longer than that connecting the C-terminal fragment and its zinc finger. The models were consistent with our initial experimental results and provided a rationale for why the NZ/ZC fusion pair, but not the ZN/CZ pair, exhibited some bias for methylating the target site. The models also supported our hypothesis that methylation could be biased towards a target site via a DNA-targeted reassembly method that works by increasing the local concentration of the fragments at the target site.
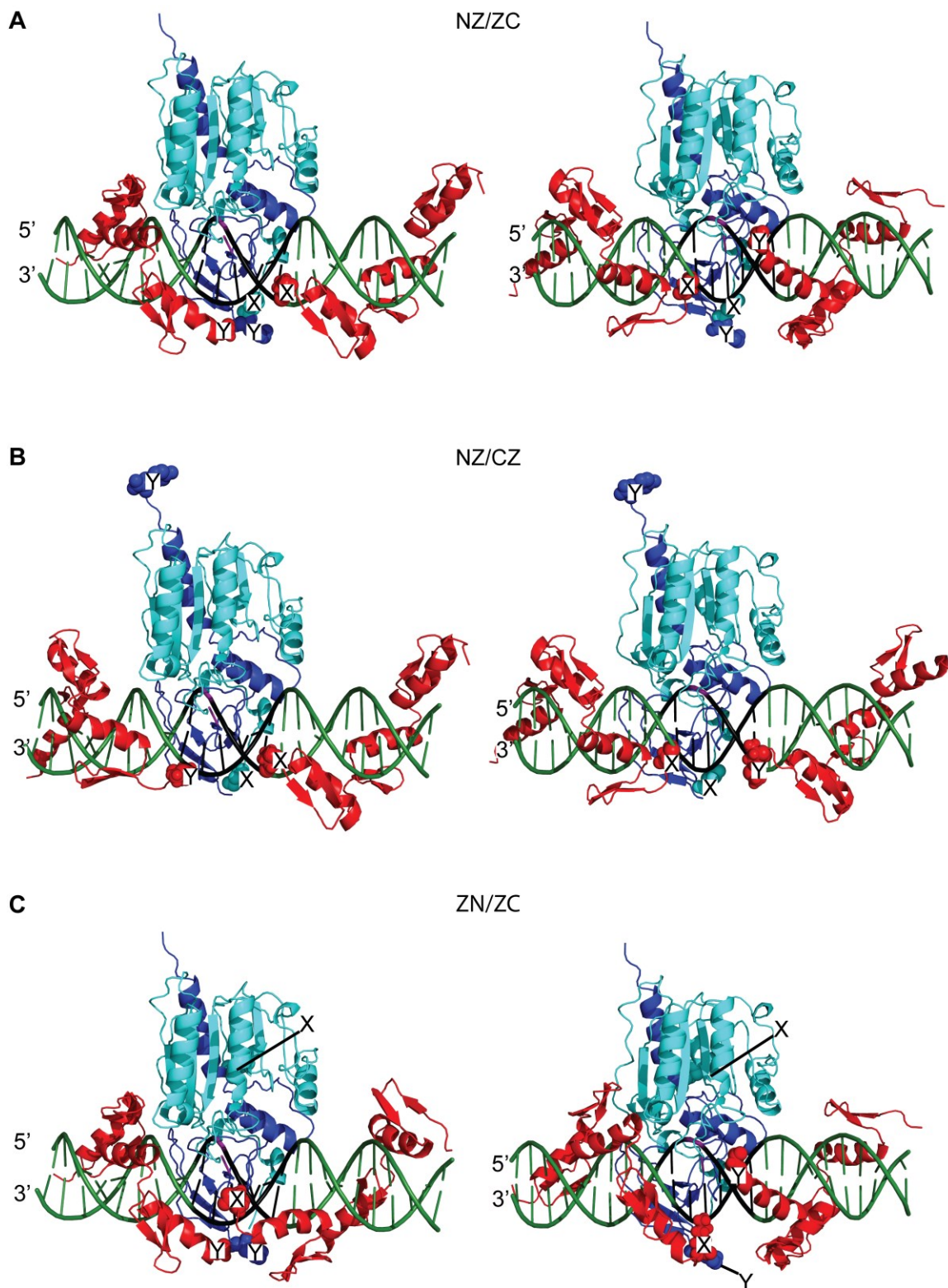
Figure 2.2 Models for four possible fusion combinations of M.HhaI fragments and zinc fingers bound to DNA. Continued on next page.

D

ZN/CZ
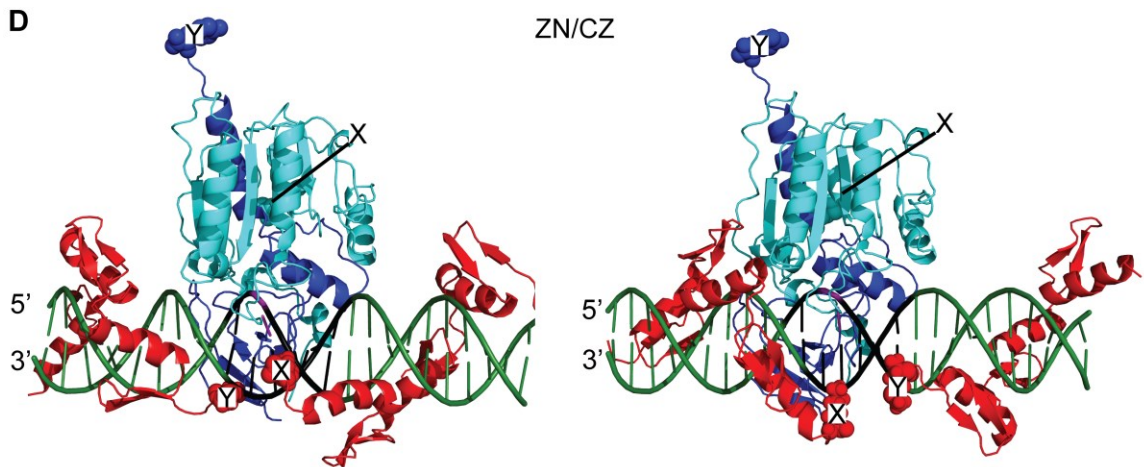
Figure 2.2 continued. Models for four possible fusion combinations of M.HhaI fragments and zinc fingers bound to DNA. For each of the four configurations, models were constructed with the methyltransferase positioned to bind the top or the bottom strand relative to the bound zinc fingers (columns 1 and 2). The 'X' and 'Y' labels indicate the location of the zinc finger and methyltransferase termini that need to be connected via a peptide linker in order for the zinc finger and M.HhaI domains to be bound to DNA as labeled. The 'X' label present on zinc finger HS1 termini should be fused to the 'X' label on the termini of M.HhaI [1–209]. The 'Y' label present on zinc finger HS2 termini should be fused to the 'Y' label on M.HhaI [210–326]. Krishna Praneeth Kilambi created the pymol models used to construct this image.

2.3.3 Plasmid and restriction enzyme protection assay design

We placed the genes encoding the NZ/ZC fragment pairs in a plasmid under separate inducible promoters (Fig. 2.3A). These genes also encoded different length peptide linkers connecting the zinc fingers and the methyltransferase fragments. For assessing methylation levels, the plasmid also contained two M.HhaI test sites (5'-G**C**GC-3') that were nested within FspI sites (5'-TG**C**GCA-3'). FspI digestion is blocked by ^{5m}C methylation at the first cytosine in the recognition sequence [164]. The plasmid also contained a unique NcoI site, so that linearization by NcoI, along with incubation with FspI and agarose gel electrophoresis could be used to distinguish between the four possible methylation states of these

47

two sites (Fig. 2.3 B and C). The two test sites were flanked by HS1 and HS2 zinc finger binding sites (Fig. 2.3 D), control sequences (Fig. 2.3 E), or combinations thereof. Various length spacer nucleotides separated the FspI site and these sequences. The *in vivo* methyltransferase activity assay was preformed by culturing ER2267 cells containing these plasmids in the presence or absence of the inducers for methyltransferase fragment expression.
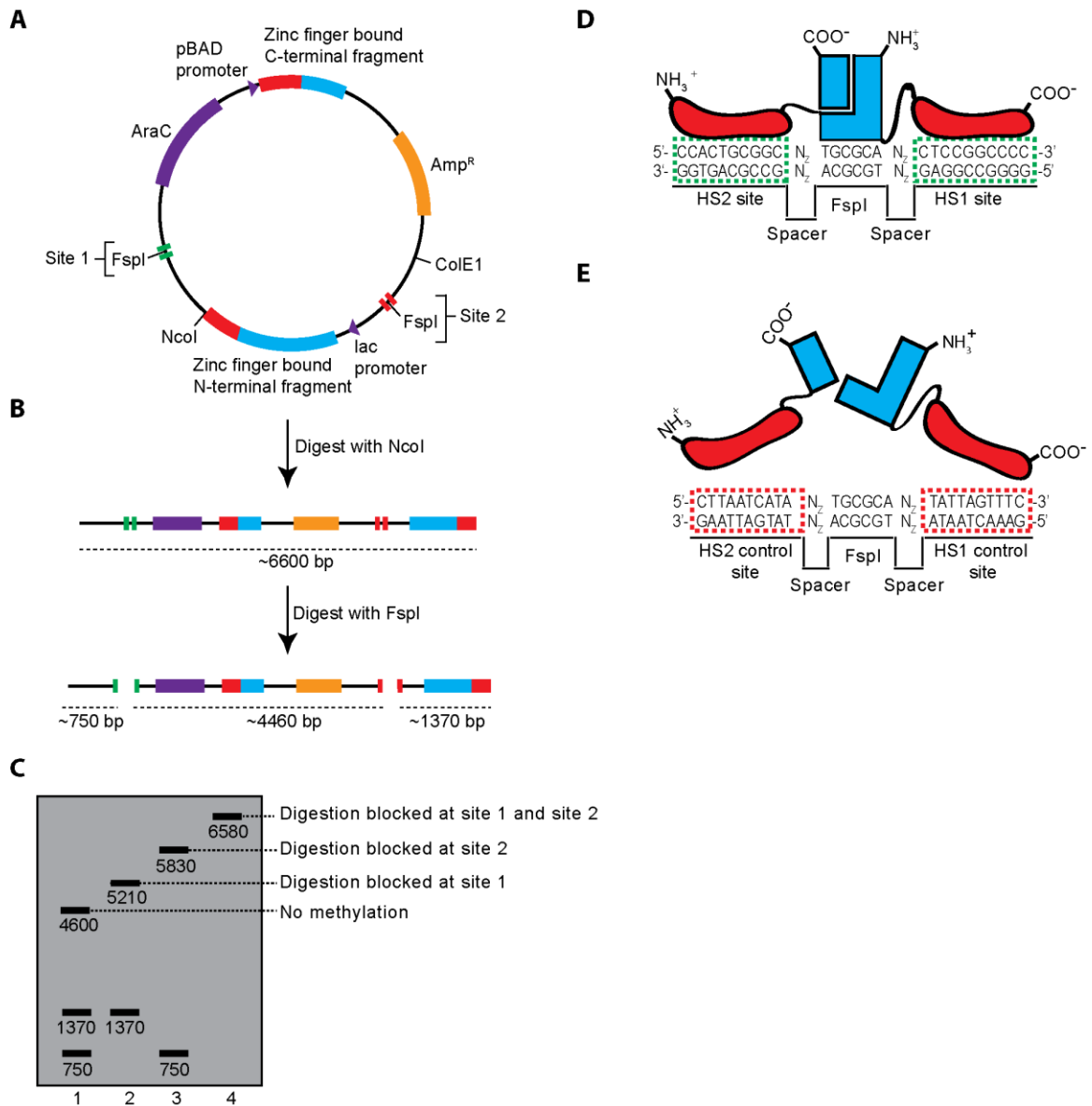
Figure 2.3 A schematic of the restriction enzyme protection assay for targeted methylation (A) A single plasmid, pDIMN8, encodes genes for both methyltransferase fragment-zinc finger fusion proteins, as well as two sites for assessing the degree of targeted methyltransferase activity. Expression of both protein fragments was induced in ER2267 cells and plasmid DNA was isolated. (B) Plasmid DNA was linearized by NcoI-HF digestion and incubated with FspI, an endonuclease whose activity is blocked by methylation. In the absence of methylation, the plasmid is digested twice by FspI and once by NcoI-HF as shown. (C) Methylation at one or both of the FspI containing sites creates unique digestion patterns as assessed by agarose gel electrophoresis. Unique bands are diagnostic of no methylation (~4600 bp), methylation at site 1 (~5210 bp), methylation at site 2 (~5830 bp), or methylation at both sites (~6580 bp). (D) A schematic of the functional methyltransferase at a target site. Zinc finger/DNA recognition mediates methyltransferase assembly. (E) This

assembly is designed not to occur at the non-target control site, which lacks zinc finger binding sites.

2.3.4 Reduction of off-target activity through serial truncation of the C-terminal fragment

We hypothesized that methylation at non-targeted sites resulted from the reassembly of the M.HhaI[1-209] and M.HhaI[210-326] fragments in the absence of the zinc finger binding sites, much like the M.HhaI[1-240] and M.HhaI[210-326] fragments that can assemble in an unassisted fashion [31]. We attempted to improve the bias for the target site through mutations designed to reduce the affinity of the two methyltransferase fragments for one another.

The C-terminal α-helix of M.HhaI is located on the C-terminal fragment and interacts with a set of β-strands located on the N-terminal fragment. Together, the helix and β-strands comprise part of the Rossmann-like fold in M.HhaI [11]. We hypothesized that truncation of the C-terminal α-helix might disrupt this interaction by either reducing the overall stability of the C-terminal fragment or by simply reducing the surface area of the protein-protein interface. Thus, truncation of the C-terminal helix was designed to prevent fragment reassembly when zinc fingers were not bound to their target sites, reducing off-target methylation. Zinc finger binding would facilitate the two fragments' assembly at the target site.
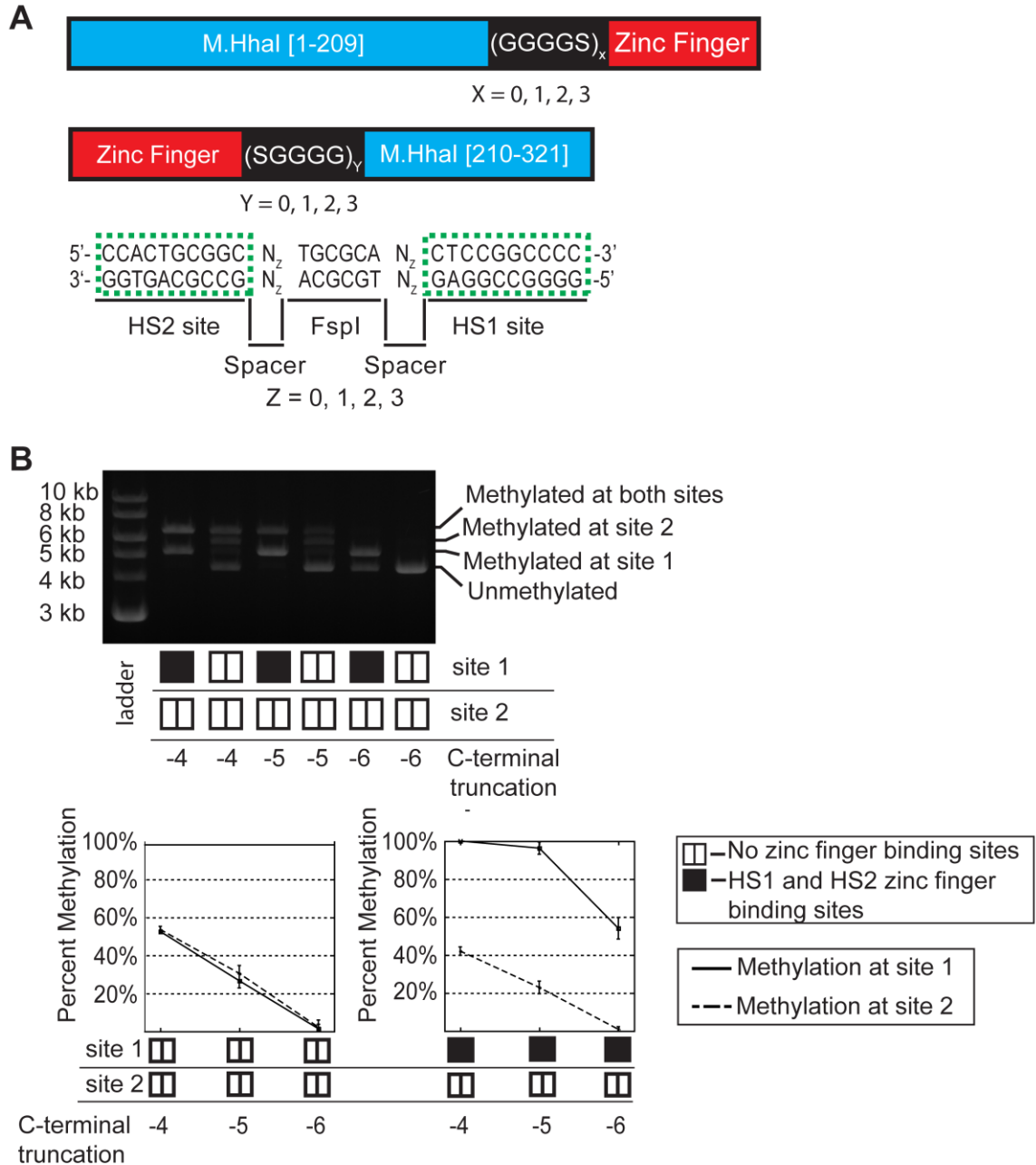
Figure 2.4 The effect of C-terminal truncation, linker lengths, and target site spacing on methyltransferase activity. Continued on next page.

C

Methylated at both sites
Methylated at site 1
Unmethylated

8 kb
5 kb
4 kb
3 kb

ladder

X=	0	0	0	0	1	1	1	1	2	2	2	2	3	3	3	3
Y=	0	1	2	3	0	1	2	3	0	1	2	3	0	1	2	3
Z=	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

% Methylation (80%, 60%, 40%, 20%)

X=	0	0	0	0	1	1	1	1	2	2	2	2	3	3	3	3
Y=	0	1	2	3	0	1	2	3	0	1	2	3	0	1	2	3
Z=	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

D

Methylated at both sites
Methylated at site 1
Unmethylated

8 kb
5 kb
4 kb
3 kb

ladder

X=	1	1	1	1	2	2	2	2	3	3	3	3
Y=	3	3	3	3	2	2	2	2	1	1	1	1
Z=	0	1	2	3	0	1	2	3	0	1	2	3

% Methylation (80%, 60%, 40%, 20%)

X=	1	1	1	1	2	2	2	2	3	3	3	3
Y=	3	3	3	3	2	2	2	2	1	1	1	1
Z=	0	1	2	3	0	1	2	3	0	1	2	3

Figure 2.4 continued. The effect of C-terminal truncation, linker lengths, and target site spacing on methyltransferase activity (A) A schematic of the protein fusions and target DNA sequences indicating the variability in linker length and DNA spacing tested. The linkers connecting the zinc fingers to the N- and C-terminal fragments were varied in 5 amino acid increments (from 0 to 15 amino acids), and combined iteratively. The bases separating the FspI site from the zinc finger binding sites were also varied (0,1,2,3 bases on each side). (B) Truncation of the C-terminus of the C-terminal fragment (indicated in units of amino acids) decreases off-target activity at the methyltransferase. In this experiment X=3, Y=1 and Z=0. The nature of the DNA at site 1 and site 2 (whether a target or non-target site) is depicted at the bottom of the figure and graph. Constructs in which the C-terminus of M.HhaI was truncated by 6 amino acids were used to determine the effect of (C) linker length and (D) target site spacing on methyltransferase activity at the target site. The percent

52

methylation at the target site are indicated in the graphs. All graphs show the mean and the error bar represents the standard deviation of the analysis of plasmid DNA from n ≥ 3 independent cultures.

A

								100 bp ladder	1 kb ladder
Off target methylation									
Methylated at site 1									

									100 bp ladder	1 kb ladder
site 1	▢	▢	■	■	■	■	■	■		
site 2	▢	▢	▢	▢	▢	▢	▢	▢		
Repression	+		+		+		+			
Induction		+		+		+		+		
C-terminal truncation	-6	-6	-6	-6	-5	-5	-4	-4		

▢ —No zinc finger binding sites
■ —HS1 and HS2 zinc finger binding sites

B

										λ HindIII ladder
Intact chromosomal DNA										23,130 bp / 9,416 bp / 6,557bp / 4,361 bp / 2,322 bp / 2,027bp

										λ HindIII ladder
Methylated *in vitro*								+		
FspI	+		+		+		+	+		
Repression			+	+			+	+	+	
Induction	+	+			+	+				
C-terminal truncation	-4	-4	-4	-4	-6	-6	-6	-6	-6	

Figure 2.5 HhaI protection assay of C-terminal truncation variants shown in figure 2.4A (A) Analysis of plasmid DNA. HhaI endonuclease activity is blocked by methylation and one band is indicative of methylation and protection at the target site (site 1). Other, larger bands are indicative of off-target methylation. There are 36 HhaI recognition sites on pDIMN8. Therefore, this assay cannot detect all of the off-target methylation as some bands indicative of off-target methylation may be obscured by other bands in the same lane and some may be too small to observe by this method. (B) Analysis of genomic DNA using FspI digestion. Chromosomal DNA was isolated from cells containing engineered M.HhaI constructs with a 6 or 4 amino acid C-terminal truncation where X = 3, Y = 1, Z = 0. The cells were grown under conditions known to repress or induce methyltransferase fragment expression (see Materials and Methods Fig. 2.4). The K12 chromosome has over 2000 FspI

restriction sites; thus, individual digestion products are not resolvable. For the 6 amino acid-deletion variant this digestion pattern for chromosomal DNA isolated from cells with induced or repressed expression of the engineered methyltransferase is indistinguishable, indicating little to no methylation. However, for the 4 amino acid-deletion variant, induction of the engineered methyltransferase, shifts the digestion pattern toward higher molecular weight bands, which is indicative of some chromosomal methylation. As a control, chromosomal DNA treated with M.HhaI in vitro is protected from FspI digestion. The results show that our targeted methyltransferase (with the 6 amino acid truncation) causes little to no methylation of the chromosome.

Based on the model of NZ/ZC, we used a long linker to connect the N-terminal fragment with HS1 and a short linker to connect HS2 with the C-terminal fragment (i.e. X=3 and Y=1 in Fig. 2.4A). No spacer nucleotides were placed between the FspI site and the HS1/HS2 binding sites (i.e. Z=0 in Fig. 2.4A).

As shown in Figure 2.4B, the progressive deletion of 4 to 6 amino acids from the C-terminus of the C-terminal fragment resulted in the maintenance of a relatively high level of methylation at the target site (>50%) but a severe reduction in methylation at the non-target site. With 6 amino acids deleted, we observed 53±3% methylation of the target site and 1.4±2.4% methylation at the non-target site (Fig. 2.4B) Methylation was not apparent at any other M.HhaI site based on restriction digest protection assays with HhaI (Fig. 2.5A), though the assay is not as sensitive for methylation as the assay with FspI at the non-target site due to the large number of HhaI sites. Similarly, HhaI digestion of genomic DNA failed to provide any evidence of off-target methylation with our optimal construct (Fig. 2.5B), though significant

off-target methylation would need to occur for this assay to detect methylation on the chromosome.

Bisulfite sequencing confirmed that methylation at the target site caused the observed protection from restriction enzyme digestion. Methylation predominantly occurred on one strand on the order of 50% (Fig. 2.6A). No methylation could be detected at the non-target site by bisulfite sequencing (Fig. 2.6B).

A
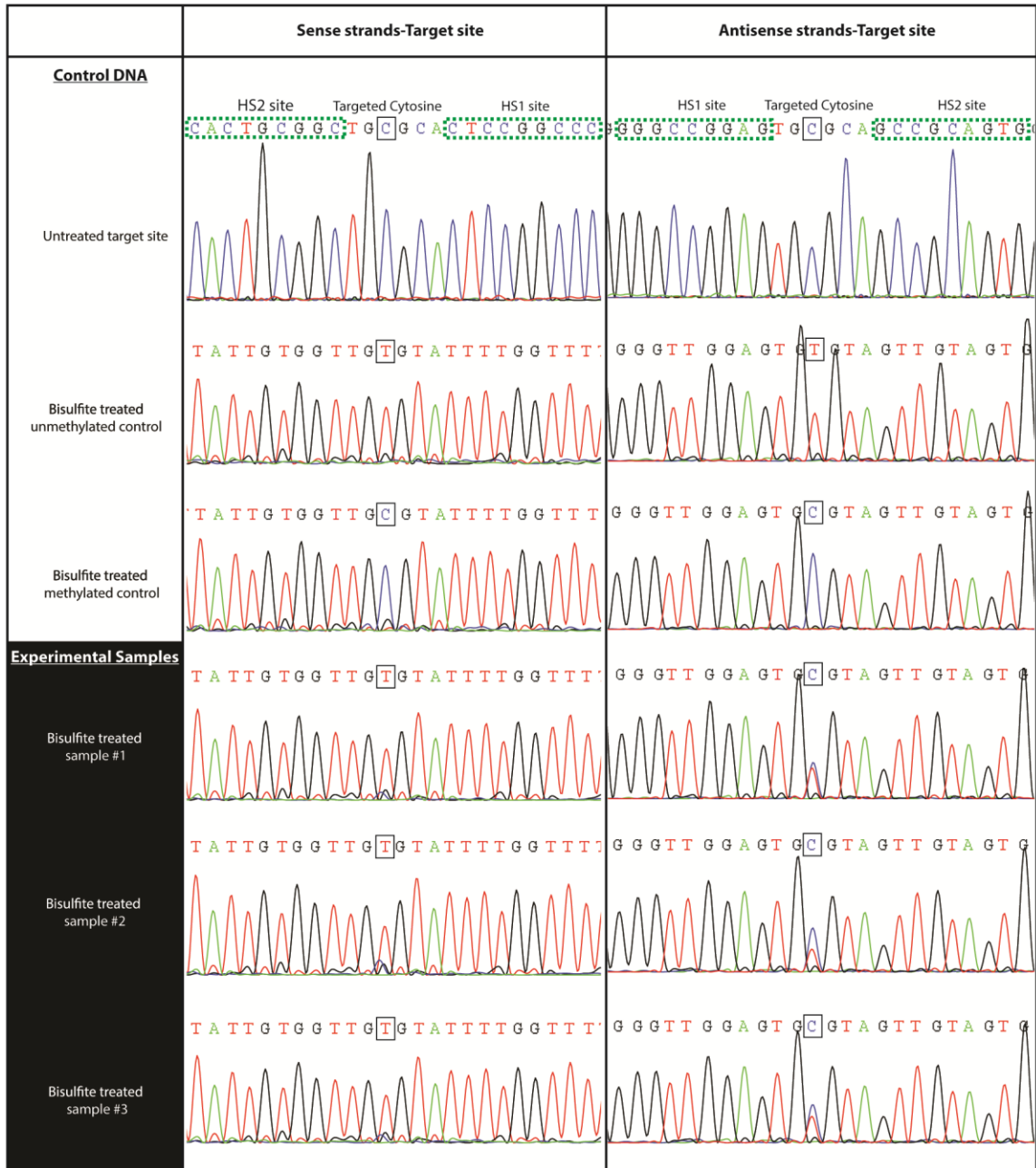
Figure 2.6 Bisulfite analysis of both strands. **Continued on next page.**
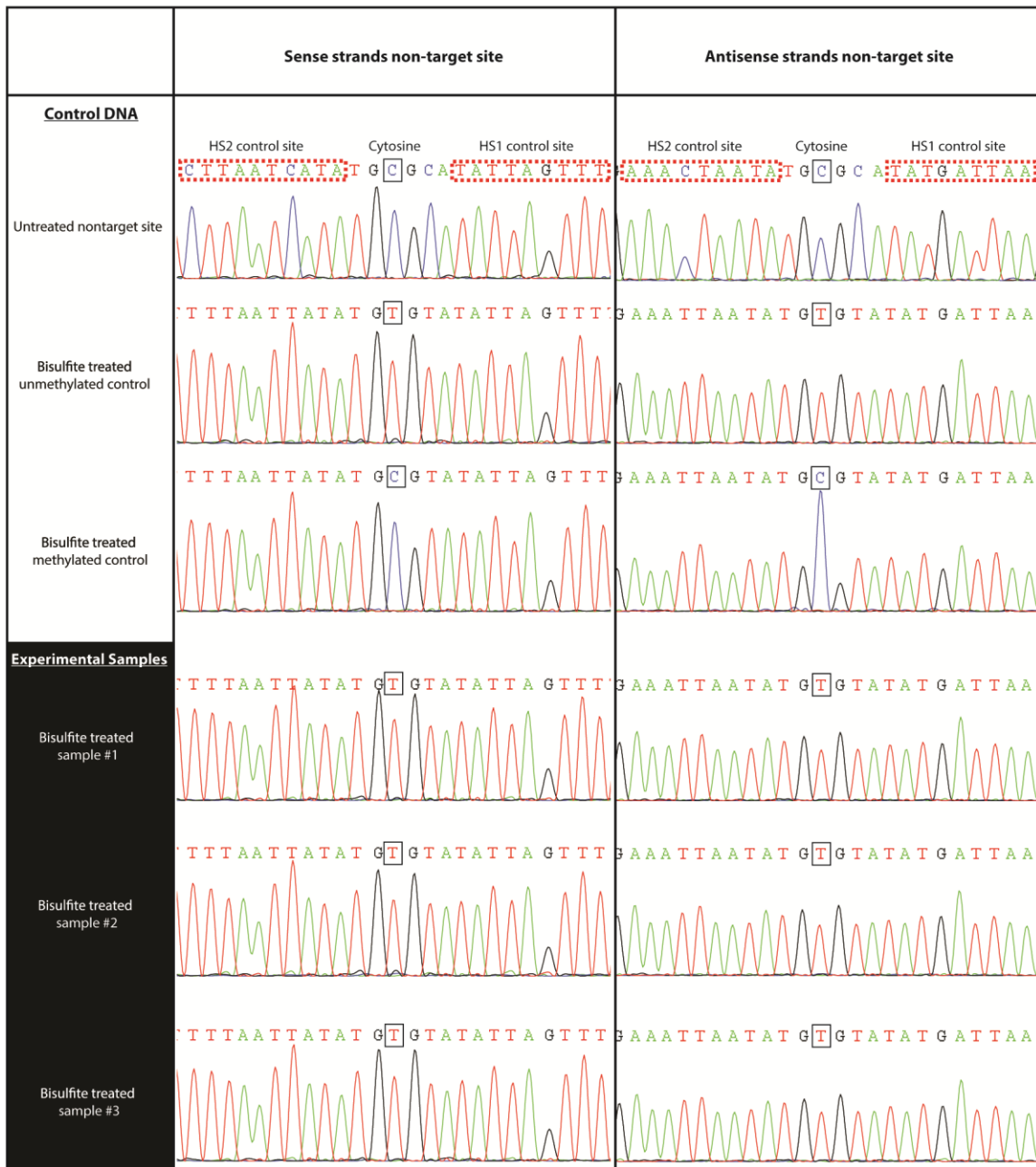
Figure 2.6 Bisulfite analysis of both strands at (A) the target site and (B) the non-target site. Bisulfite treatment followed by PCR amplification converts unmethylated cytosine bases to thymidine bases. Methylated cytosine residues are protected from such a conversion. The sense strand is defined as the top strand of the target and non-target sites shown in Figure 2.3D and E; the antisense strand is the bottom strand in these figures. Sequenced plasmid DNA, which was not bisulfite-treated is shown in column 1 row 1 of each panel. For both panels, the chromatogram of the antisense strand (column 2 row 1) is the computer-generated reverse complement of the chromatogram in column 1 row 1. DNA in rows 2–6, was treated with the bisulfite reagent, amplified and sequenced as described in Methods S1. The plasmid tested was X = 15, Y = 5, Z

= 0 (see Figure 2.4). Rows 2 and 3 show sequencing results for bisulfite-treated unmethylated and methylated control DNA. Rows 4–6 show sequencing results for bisulfite-treated plasmid DNA from three independent cultures. The chromatograms for the following samples were converted to the reverse complement to simplify the comparison (target site, column 1, rows 4–6).

2.3.5 The linker length's effect on methylation at the target site is consistent with the model

We next sought to investigate and optimize the length of the amino acid linkers connecting the M.HhaI fragments to their respective zinc fingers. Our previous work with targeted split methyltransferases indicated that linker length can affect enzymatic activity at the target site [110]. Using overlap extension PCR, we created N-terminal and C-terminal fragments that were fused to zinc fingers by linkers of 0, 5 10 and 15 amino acids (Fig. 2.4A). In all constructs, the C-terminal M.HhaI fragment had its last 6 amino acids removed. All N-terminal fragment linker variants were then crossed with all C-terminal linker variants and tested for methylation activity at the target and non-target site (Fig. 2.4C). The target site lacked spacer nucleotides (Z=0).

All constructs retained bias for methylation at the target site. We observed a reduction in methylation at the target site for shorter amino acid linkers connecting the N-terminal fragment with its respective zinc finger protein. Conversely, in the context of long N-terminal linker, an increase in the length of the linker connecting the C-terminal fragment and its respective zinc finger resulted in a decrease in methylation at the target site.

The relationship between linker length and activity at the target site can be explained by our model of the NZ/ZC/DNA complex. As shown in Figure 2.1C, the optimal 15 amino acid linker, connecting the HS1 zinc finger to the N-terminal fragment, is found to wrap around the DNA backbone. A longer linker is required because the N-termini of the zinc finger and the bisection point of the N-terminal methyltransferase fragment are located on opposite sides of the DNA. Shortening the linker reduces the probability of interaction between the N-terminal M.HhaI fragment with the C-terminal M.HhaI fragment upon HS1 zinc finger binding, thereby reducing methylation at the target site. On the other hand, the close proximity of the N-terminus of the C-terminal M.HhaI fragment and HS2 zinc finger indicates that a short linker would be sufficient between these two domains. However, it would be less entropically favorable for a longer flexible linker between these domains to assist in the assembly of an active enzyme. An overly long linker enables the C-terminal fragment to explore more space upon zinc finger binding, compromising the increase in local concentration of the C-terminal fragment gained by zinc finger binding.

2.3.6 The orientation of the zinc finger binding sites relative to the methylation site modulates targeted methylation

We next sought to characterize the effect of adding bases between the zinc finger binding sites and the FspI site (i.e. varying Z in Figure 2.4A). The addition of bases both increases the distance in the DNA sequence

and rotates the zinc finger binding sequence around the DNA with respect to the central CpG site. Due to this rotation, the addition of bases can potentially increase or decrease the required length of the linker joining the methyltransferase fragment and the zinc finger.

To test this idea, we used three sets of linker variants in which the sum of the number of linker residues was kept constant (X=1/Y=3, X=2/Y=2, and X=3/Y=1). A constant sum total of linker residues helps illustrate that total linker length does not determine enzymatic activity at the target site. For each set of linker variants, we added 0, 1, 2 or 3 bp to both sides of the FspI site and tested methyltransferase activity as before. All constructs retained some methylation at the target site and minimal methylation at the non-target site; however, the length of the spacer DNA modulated activity in a complex manner (Fig. 2.4D).
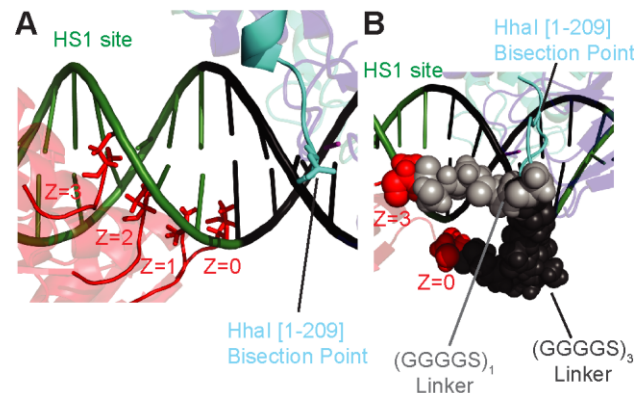
Figure 2.7 Molecular modeling explains how an increase in target spacing can reduced the required protein linker length (A) HS1 zinc fingers are bound to DNA with a target site spacing of Z=0,1,2, or 3. Note that at Z=3, the zinc finger is actually closer to the bisection point of the N-termini than at Z=0. (B) A model demonstrating that a five amino acid linker is sufficient to connect zinc finger HS1 bound at Z=3. In contrast, a longer amino acid linker is required to circumvent the DNA backbone at Z=0. This model provides an explanation for the pattern of target site methylation observed in Figure 3D. Krishna Praneeth Kilambi created the pymol models used to make this image.

The NZ/ZC/DNA model provides a rationale for the observed behavior. For X=1/Y=3/Z=0, methylation at the target site remains low because the linker between the N-terminal M.HhaI fragment and its respective zinc finger is too short. However, methylation at the target site increases to 58% at the highest spacer length (Z=3) because the linker between these domains no longer needs to wrap around the DNA backbone (Fig. 2.7A). When Z=3, modeling indicates that a five amino acid linker (X=1) between the two domains is sufficient for fragment reassembly at the target site (Fig. 2.7B).

For linker combinations that possess high target site methylation with Z=0 (i.e. X=2/Y=2 and X=3/Y=1), the addition of 1 or 2 bp reduces methylation at the target site. However, the addition of 3 bases restores target site methylation to their Z=0 levels (Fig. 2.4D). The initial reduction of activity with the addition of 1 or 2 bp can be explained by the rotation of the zinc fingers further around the DNA such that the linkers have to span even longer distances for reassembly to occur. However, modeling predicts that with the addition of 3 bp, the linker no longer needs to wrap around DNA and the distance can now be spanned by a 5 amino acid linker connecting the N-terminal fragment and HS1 (Fig. 2.7B).

2.3.7 Zinc finger mediated localization of both M.HhaI fragments has a synergistic effect on methylation targeting

We desire targeted methyltransferases that require the binding of both

zinc finger domains for methyltransferase activity.
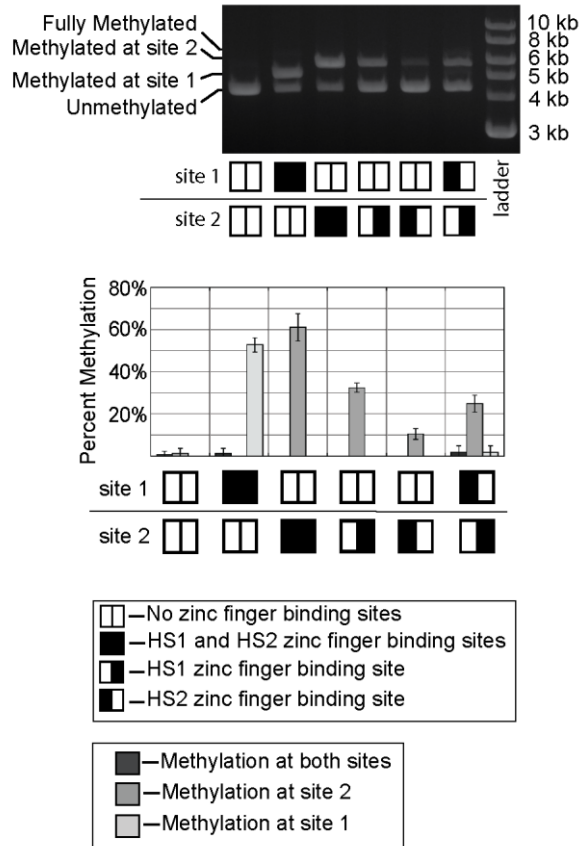
Figure 2.8 The contribution of each zinc finger binding site toward observed, targeted DNA methylation Methylation was assessed as in Figure 2. In this experiment, the C-terminal fragment of M.HhaI is truncated by 6 amino acids, X=3, Y=1, and Z=0. Methyltransferase activity was assessed with and without target sites present. Moving the target site from site 1 to site 2 did not have a large effect on activity. Target half sites (in which either the HS1 or HS2 binding sites were removed) allowed assessment of the contribution of each zinc finger on methylation activity at the target site. The sum of the methylation observed on each half site (43±5%) was less than methylation at the full target site (61±6%). The methylation observed with two distal half sites (<30%) was also less than that observed with the complete target site.

However, all linker length and spacer DNA variants tested (Fig. 2.4 C and

D) retained some bias for methylating the target site, despite our models'

prediction that some variants have insufficient length linkers to allow

target site reassembly. This suggests that some of the methylation observed at the target site may occur without binding of both zinc fingers. In other words, the bias for methylation at the target site may occur in part through localization of only one of the two fragments via its zinc finger domain, followed by a reassembly of M.HhaI that is independent of a second zinc finger-binding event.

To test this hypothesis, target "half sites" were constructed with either the HS1 or the HS2 zinc finger binding site (Fig. 2.8). These experiments were conducted with a construct containing a high degree of specificity and activity for the full HS1/HS2 site (i.e. X=3, Y=1, Z=0). The amount of methylation at the target and non-target sites was assessed as before. Removal of either (but not both) of the zinc finger binding sites reduced, but did not eliminate methylation at the target site (Fig. 2.8). Removal of the HS1 binding site was more detrimental to methylation activity at the target site than removing the HS2 site, indicating that localizing only the N-terminal M.HhaI fragment via zinc fingers was more effective for targeting methylation than localizing only the C-terminal fragment via zinc fingers. This result may be explained by the fact that the target recognition domain (TRD) is present on the C-terminal fragment. Thus, the C-terminal fragment likely possesses greater inherent affinity for the methylation target site than the N-terminal fragment. In other words, the N-terminal fragment has a greater need for fusion to the zinc finger in order to localize it to the target site. This

result is unlikely to be explained by differences in the DNA binding affinity of the two zinc finger proteins. HS1 and HS2 have similar dissociation constants for their target sites (35 nM and 25 nM, respectively) [163].

Although this experiment revealed a shortcoming of our current optimized, split M.HhaI, it also provides evidence for the advantages of targeting methyltransferases using our split enzyme strategy. The level of target site methylation observed at a CpG site flanked by both zinc finger binding sites (61±6%) exceeds the sum of the methylation observed at the half sites (43±5%)(Fig. 2.8). This synergy (i.e. the observed activity at the intact target site is greater than the sum of activity observed at the individual binding sites), is caused by the proximity of zinc finger binding sites and is precisely what our split enzyme system was designed to achieve. We also confirmed that placing the two zinc finger sites at distant locations on the same plasmid cannot provide the same level of targeted methylation observed by placing both zinc finger sites at one target site (Fig. 2.8).

2.3.8 M.SssI can be converted into a heterodimeric/zinc finger fusion enzyme, whose activity is biased towards a desired target site

We were interested in assessing whether other monomeric methyltransferases could be bisected and fused to zinc finger proteins to create targeted methyltransferase. Specifically, we were interested in

bisecting M.SssI, a prokaryotic methyltransferase that recognizes and methylates the cytosine of any 5'-CG-3' site. A targeted methyltransferase derived from M.SssI would, in theory, make it possible to target any CpG site, rather than just 1/16 of the possible CpG sites that could be methylated by our engineered M.HhaI enzyme (which recognizes 5'-GCGC-3').

We used a CLUSTALW alignment to identify a site within M.SssI that was similar to the split site of M.HhaI [165,166]. We fused the zinc finger proteins HS1 and HS2 in the same NZ/ZC orientation as our targeted M.HhaI fusion proteins using 15 and 10 amino acid linkers, respectively (Fig. 2.9A). These constructs were tested for methylation specificity in an analogous fashion to that illustrated in Figure 2.3. Methylation activity was assessed under conditions shown to either induce or repress expression of the methyltransferase fragments. Upon induction, the fusion constructs were very active and, although some bias towards the target site was apparent, the high activity prevented the observation of the extent of this bias (Fig. 2.9B). Slaska-Kiss *et al* very recently demonstrated that M.SssI is amenable to protein fragment complementation at select sites; however, targeted methylation was not demonstrated [112].

We used site-specific mutagenesis to reduce methyltransferase activity in order to reveal the inherent bias of the construct (Fig. 2.9B). Mutating the active site cysteine, C141S, has been shown to reduce the

activity of M.SssI enough to reveal biased methylation activity upon

M.SssI fusion to triple helix forming nucleotides [104]; however, in the

context of our bisected enzyme, this mutation completely eliminated

activity *in vivo* (Fig.

2.9B).
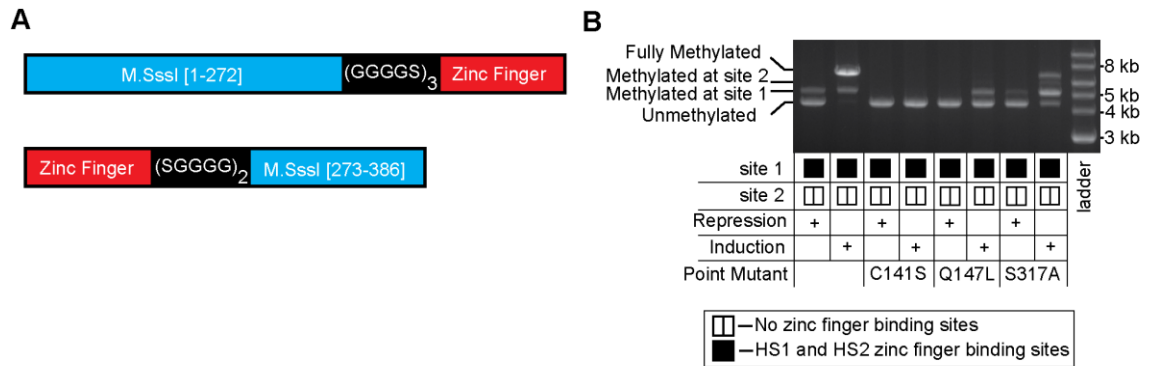
Figure 2.9 M.SssI can be converted into a targeted heterodimeric methyltransferase (A) A schematic showing the sequence of the M.SssI fragments fused to zinc fingers via flexible linkers. (B) A restriction enzyme protection assay showing the split enzyme constructs possess a bias for methylation at the target site. Plasmids were isolated from strains grown under conditions that either repress or induce expression of the two fragments. Plasmid DNA was assayed for methylation as in Figure 2. The activity of these fusion heterodimers was attenuated by the indicated point mutations known to decrease enzyme activity in wild-type M.SssI.

On the other hand, both the Q147L and S317A mutations, which

are known to reduce M.SssI's DNA binding affinity by 12-fold and 3-fold

respectively [26], reduced but did not eliminate the activity of our

bisected M.SssI, revealing the extent of our enzyme's methylation bias

(Fig. 2.9B). The relative activity of the two mutants was consistent with

the reported relative effect of the mutations on M.SssI affinity for DNA.

2.4 Conclusions

We have demonstrated that bisected M.HhaI and M.SssI enzymes, when fused to zinc fingers in the proper orientation, can target methylation to a desired sequence flanked by the respective zinc finger binding sites. Our modeling and experiments have elucidated some of the design principles for constructing a targeted methyltransferase using this strategy. The orientation of methyltransferase fragments relative to each other and to DNA affect the activity at the target site. Mutations designed to reduce the interaction between fragments can improve targeting of the methyltransferase. With the proper linker length, spacing between zinc finger binding and methylation sites, and expression conditions, such constructs can methylate a desired target with high efficiency (50-60%) with levels of non-target site methylation at or below the limit of detection. Part of the targeting arises from the synergistic effect of localizing both fragments to the desired site, which supports our hypothesis of how bisected enzymes could better target methylation.

However, some of the bias for methylation at the target site likely arises from zinc finger mediated localization of only one of the two fragments. Thus, although binding of both zinc finger domains increases target site methylation, such methylation does not yet *require* binding of both zinc finger domains. We believe this limitation arises because the individual N- and C-terminal fragments (particularly the C-terminal fragment) retain some affinity for the 5'-GCGC-3' site and, perhaps,

67

retain sufficient affinity for each other. We next intend to test these hypotheses experimentally in a manner that is guided by our computational model.

2.5 Acknowledgements

I would like to thank Dr. Jeff Gray and Krishna Praneeth Kilambi for their work modeling zinc finger-fused bifurcated M.HhaI and in editing the manuscript. Krishna wrote and executed Rosetta scripts for modeling zinc fingers, modeled DNA target sites, docked and refined structures, and built loops between fused domains. He also wrote the materials and methods section concerning the modeling.

Intended to be blank

3 Directed evolution of improved zinc finger methyltransferases

3.1 Introduction

CpG methylation is one of the most extensively studied epigenetic modifications and broadly regulates or maintains transcriptional activity. As discussed in the introductory chapter, methylation of DNA is involved in proper cellular differentiation, heterochromatin formation and chromosomal stability [68]. Further, aberrant methylation patterns cause or are observed in numerous diseases. Imprinting defects lead to disorders such as Prader-Willi and Angelman syndromes [167]. Notably, aberrant methylation of CpG islands (CGIs) is a hallmark of cancer [69]. Though much has been learned about how methylation patterns are established and erased, the causes of aberrant methylation and the reestablishment of methylation patterns during development remain active areas of research. To study the effects and dynamics of DNA methylation, it would be generally useful to target methylation toward specific, user-defined sequences.

In the previous chapter, we described how a split version of M.SssI DNA methyltransferase fused to zinc finger proteins demonstrates bias for methylating a M.SssI site located between the two zinc finger binding sites [168]. The bifurcation point in M.SssI was chosen based on a CLUSTALW alignment to a site in a similarly engineered M.HhaI enzyme

[168]. However, site specific mutations Q147L or S317A in the M.SssI domain, introduced to reduce the enzyme's DNA binding affinity and activity, were necessary to observe significant bias towards methylation at a target site [168]. Here we present a selection strategy to improve the targeting of methyltransferases and have used this strategy to optimize our M.SssI fusion construct. We performed a negative selection against off-target methylation and a positive selection for methylation at a target site, *in vitro*. This strategy allowed us to quickly identify variants with improved targeting ability and activity *in vivo*. We also demonstrate the modularity of our constructs by altering the zinc finger domains to redirect methylation toward a new target site.

3.2 Materials and methods

3.2.1 Enzymes, oligonucleotides and bacterial strains

Restriction enzymes, T4 ligase, T4 kinase, and Phusion High Fidelity PCR MMX were purchased from New England Biolabs (Ipswich, MA). BoxI was purchased from ThermoFisher Scientific (Waltham, MA). Platinum Pfx DNA polymerase was purchased from Life Technologies (Carlsbad, CA). PfuTurbo Cx Hotstart DNA polymerase was purchase from Agilent Technologies (Santa Clara, CA). Plasmid-Safe-ATP-dependent DNAse was purchased from Epicentre (Madison, WI). pDIM-N8 and pAR plasmids have been previously described [110,168]. All oligonucleotides and gBlocks were synthesized by Invitrogen (Carlsbad, CA) or Integrated DNA

Technologies (Coralville, IA). Gel electrophoresis and PCR were preformed essentially as previously described [162]. Plasmids were isolated using QIAprep Spin Miniprep Kit (Qiagen, Valencia, CA). DNA fragments were purified from agarose gels using QIAquick Gel Extraction Kit (Qiagen, Valencia, CA) or PureLink Quick Gel Extraction Kit (Invitrogen, Carlsbad,CA, USA) and further concentrated using DNA Clean & Concentrator-5 (Zymo Research, Irvine, CA).

Escherichia coli K-12 strain ER2267 [*F′ proA+B+ lacIq Δ(lacZ)M15 zzf::mini-Tn10 (KanR)/ Δ(argF-lacZ)U169glnV44 e14-(McrA-) rfbD1? recA1 relA1? endA1 spoT1? thi-1 Δ(mcrC-mrr)114::IS10*] was acquired from New England Biolabs (Ipswich, MA) and was used in selections, methylation assays and cloning.

3.2.2 Plasmid creation

pDIMN8, was used for library creation and testing of library variants [168]. pDIMN9 was constructed as follows for use in golden gate cloning. Plasmid pDIMN8 was altered by silently mutating a BsaI site in the AmpR gene via pFunkel mutagenesis [169]. PCR, digestion and cloning removed a BbsI restriction site to create vector pDIMN9. Golden gate cloning was used to fuse new zinc finger proteins to methyltransferase fragments. For the creation of plasmids used in golden gate cloning, regions encoding zinc finger proteins were replaced with BbsI sites. pDIM-N9 contained a M.SssI[1-272]-BbsI construct for the addition of zinc fingers to the N-

terminal fragment. pAR contained BbsI-M.SssI[273-386] construct for the addition of new zinc fingers to the C-terminal fragments [110]. gBlocks encoding zinc fingers and BbsI sites were purchased from IDT. Golden gate cloning to fuse zinc finger encoding gBlocks to the above plasmids was performed essentially as described [170]. Zinc finger CD54a was designed using the zinc finger tools website and previously identified zinc finger domains [139,143,145]. Individual C-terminal and N-terminal zinc finger-fused constructs were digested with EcoRI and SpeI as previously described to place these constructs on the same plasmid for characterization in *E. coli* [110]. Site 1 and site 2 were altered as previously described to vary the sequences flanking different CpG sites [168].

3.2.3 Construction of cassette mutagenesis library

An NNK codon mutagenesis library of M.SssI[273-386] was constructed by overlap extension PCR. PCR was carried out using an oligonucleotide degenerate for a five amino acid region in the C-terminal fragment corresponding to amino acids 297-301 in the wild type enzyme. Fragments were digested with AgeI-HF and SpeI and ligated into pDIMN8 containing HS2 and the complete N-terminal fragment-HS1 fusion. Site 1 (i.e. the target site in Fig. 3.1C) contained an FspI site flanked by HS1 and HS2 zinc finger recognition sites. The plasmid also possessed a non-target site that lacked zinc finger binding sites, but contained an internal

SnaBI restriction site (red site in Fig. 3.2A). Ligations were transformed into ER2267 electrocompetent cells and plated onto agarose plates containing 100 μg/ml ampicillin and 2% w/v glucose. Plates were incubated overnight at 37°C. The naive library contained 2 x 10^5 transformants.

3.2.4 Library selection

Plated library variants were recovered from the plate in lysogeny broth supplemented with 15% v/v glycerol and 2% w/v glucose and stored at -80°C. Aliquots were thawed and used to inoculate 10 ml of lysogeny broth supplemented with 100 μg/ml ampicillin salt, 0.2% w/v glucose, 1 mM IPTG, and 0.0167% w/v arabinose. These cultures were incubated overnight at 37°C and 250 rpm. Plasmid DNA was isolated via QIAprep Spin Miniprep Kit and digested for 3 hours at 37°C with McrBC (10 units/μg DNA), FspI (2.5-5 units/μg DNA) in 1X NEBuffer 2 supplemented with 100 μg/ml BSA and 1mM GTP. Reactions were halted by incubation at 65°C for over 20 min minutes to which ExoIII (30 units/μg DNA) was added and the solution incubated at 37°C for 60 min. ExoIII digestion was halted by incubation at 80°C for over 30 minutes and the DNA was desalted using Zymo Clean and Concentrator-5 kits per manufacturer's instructions. DNA was transformed into ER2267 electrocompetent cells and plated on agar supplemented with 2% w/v glucose and 100 μg/ml ampicillin salt.

Cells were recovered from the plate as before and plasmid DNA was isolated using the QIAprep Spin Miniprep Kit. The DNA was digested with FspI (2-2.8 units/μg DNA) in 1X NEBuffer 4 and linear DNA was isolated via gel electrophoresis. PCR was used to amplify the portion of the linear plasmid containing genes encoding for the N-terminal and C-terminal fragments fused to zinc fingers. Purified PCR products were subcloned into the selection plasmid for an additional round of selection.

3.2.5 Restriction endonuclease protection assays

Cultures from colonies were incubated overnight at 37°C and 250 rpm and stored as glycerol stocks. Glycerol stocks were used to inoculate 10 ml of lysogeny broth supplemented with 100 μg/ml ampicillin salt, 0.2% w/v glucose, 1 mM IPTG, and 0.0167% w/v arabinose. After growth overnight at 37°C, plasmid DNA was purified from the cultures by QIAprep Spin Miniprep Kit. Plasmid DNA (500 ng) was digested with NcoI-HF (10 units) and either FspI (2.5 units) or SnaBI (2.5 units) in 1X NEBuffer 4 for over one hour at 37°C. SnaBI digests were supplemented with 100 μg/ml BSA. Half of each digested sample was loaded onto agarose gels (1.2% w/v in TAE) and electrophoresed at 90 V for 105-120 minutes. Bands were quantified as described [168].

3.3 Results and discussion

3.3.1 Design of the selection system

M.SssI naturally methylates CpG sites [23]. Our previously described, bifurcated M.SssI DNA methyltransferase zinc finger fusions (Fig. 3.1B) biased methylation toward a targeted M.SssI site flanked by the cognate zinc finger binding sequences. However, active variants also methylated other M.SssI sites [168]. We sought to reduce this off-target methylation while maintaining high levels of methylation at the targeted M.SssI site. Here, we describe an *in vitro* selection system that preferentially enriches variants possessing the ability to methylate the target site, but lacking the ability to methylate other non-targeted M.SssI sites on the plasmid (Fig. 3.1D).

In vitro selection strategies have been used to enrich for methyltransferases with relaxed or altered specificity. Most strategies rely on methylation-dependent protection from restriction endonuclease digestion to positively select for DNA encoding a methyltransferase with altered specificity [95-98]. Our selection scheme differs from previous studies as it additionally employs McrBC as a negative selection against unwanted methylation activity. In our system for altering methyltransferase specificity, a single plasmid contains both genes encoding the zinc finger-fused M.SssI fragments as well as a targeted M.SssI site nested within an FspI restriction site and flanked by zinc finger binding sites (Fig. 3.1A and C).

76
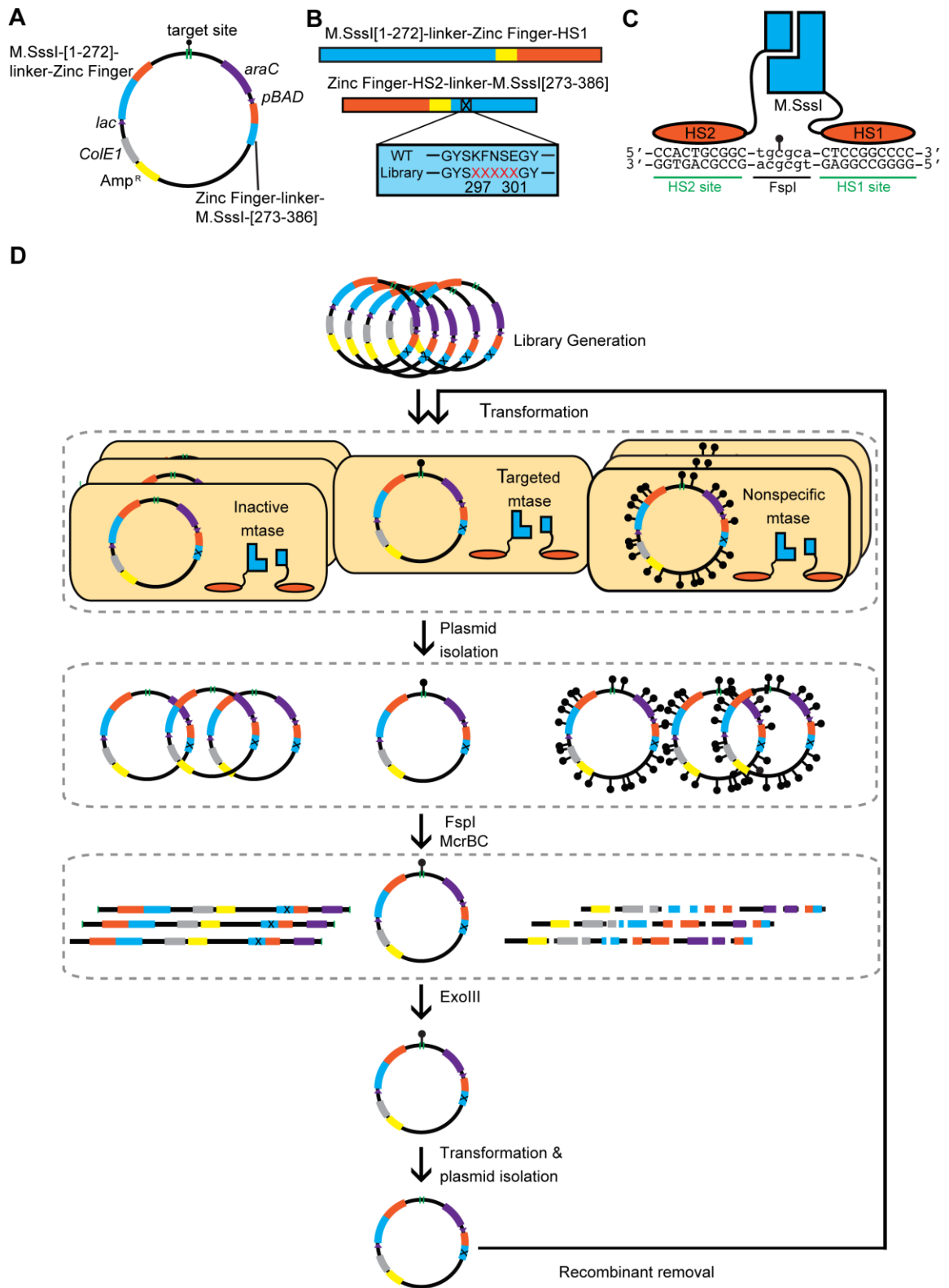
Figure 3.1 Schematics of the vector, library, proteins, and selection used in these experiments (A) The vector used in selections. The vector encodes for

both heterodimeric fragments fused to zinc fingers under the control of separate inducible arabinose (pBAD) and IPTG (lac) promoters, a target site, and the araC gene. (B) A scheme of the zinc finger-fused, bifurcated M.SssI and the mutagenized codons used in library construction. Residues 297-301 of M.SssI (located in the C-terminal fragment) were randomized. Numbering scheme is that of the wildtype M.SssI. (C) An assembled zinc finger-fused heterodimeric M.SssI methyltransferase assembled at the target site. (D) An overview of the selections used in this experiment. The schematic illustrates the fates of plasmids encoding an inactive methyltransferase (left), the desired targeting methyltransferase methylating the target site (middle) and a nonspecific methyltransferase methylating multiple M.SssI (i.e. CpG) sites.

The plasmid also has over 400 other M.SssI (i.e. CpG) sites. Once

transformed into *E. coli*, the methyltransferase fragments encoded by the

plasmid are expressed, resulting in methylation of the same plasmid. The

plasmid DNA is isolated and subjected to *in vitro* digestions with

endonucleases FspI and McrBC (Fig. 3.1D). Since FspI digestion is

blocked by methylation at the target site, FspI digestion serves to select

for methylation at the targeted CpG site. McrBC is an endonuclease that

recognizes and cleaves DNA with two distal methylated sites [171,172].

McrBC will not digest a single site that is methylated or hemimethylated

unless there is a second methylated site on the same DNA within about

40-3000 bp [173]. We therefore expect that most plasmids methylated at

multiple M.SssI sites will be digested by McrBC. Thus, McrBC digestion

selects against off-target methylation. The DNA is then incubated with

ExoIII to degrade any plasmid that is digested at least once, ideally

leaving the plasmid DNA encoding a highly specific methyltransferase

intact for the subsequent transformation.

Initial proof of principal selections demonstrated that McrBC, FspI

and ExoIII treatment of unmethylated plasmid DNA, followed by

transformation resulted in a 99.85% decrease in the number of transformants relative to untreated DNA. Similarly, McrBC, FspI and ExoIII treatment of a highly methylated plasmid reduced transformants by 99.95% relative to untreated control.

3.3.2 Design of the library

We constructed a library of M.SssI C-terminal fragment variants randomized at residues 297-301 (Fig. 3.1B). We hypothesized that mutations to these residues might reduce the ability of the split methyltransferase to methylate non-targeted CpG sites by reducing the fragment's inherent affinity for double-stranded DNA. Early studies indicated that M.SssI interacts with DNA, irrespective of the presence of CpG sites and subsequently methylates processively [174]. Further, a homology model of M.SssI suggested that residues 297 and 299 form contacts with the ribose phosphate backbone on the CpG bases complementary to the methylated CpG site [24]. Mutational studies showed that for monomeric M.SssI, K297A or N299A mutations did not appreciably affect either the catalytic activity or the dissociation constant of a CpG containing oligonucleotide [26]. Mutating these residues, we hypothesized, might eliminate the innate affinity of our fragments for DNA without affecting the catalytic activity of the enzyme.

Additionally, the homology model indicated the amide backbone of serine residue at position 300 made base-specific contacts with the

cytosine and guanine bases complementary to the methylated strand. This model initially implicated serine's conserved and catalytically important role for stabilizing the complementary strand during base flipping and methylation [24]. However, the S300P mutation resulted in only a three-fold increase in a dissociation constant and no significant change in initial rate of reaction [25].

3.3.3 Library selections

Initial selection experiments on this library resulted primarily in the isolation of plasmid DNA with a deleted FspI restriction site, presumably formed by a recombination event. This false positive was a trivial, albeit frequently observed solution for plasmid survival in our devised scheme. Thus, we subjected the plasmid DNA from the resulting transformants to additional steps to enrich for those plasmids that survived our selection and retained their FspI site. In these additional steps, the plasmid DNA was transformed into ER2267 cells and the cells were plated under conditions known to repress the promoters controlling methyltransferase fragment expression. We digested plasmid DNA from these cells with FspI and purified the linear, FspI-digested DNA away from undigested plasmid DNA by agarose gel electrophoresis. The portion of the plasmid encoding the zinc fingers and methyltransferase genes was PCR amplified, ligated back into the same plasmid backbone. This entire procedure was

repeated, subjecting variants to an additional round of selection. Selected variants were then analyzed.

3.3.4 Analysis of library variants that survived the selection

We assayed 47 variants for methylation activity at both the target and non-target site and determined the variants' sequences. The best variants (e.g. PFCSY, CFESY, and SYSSS, which are named for the sequence at residues 297-301) methylated 65-80% of the plasmids at the target site with minimal methylation (0-8%) at the non-target site (Fig. 3.2).

A

SnaBl (non-target)   Fspl (target)

~1560 bp   ~4480 bp   ~750 bp

B

```
5'-CCACTGCGGC-tgcgca-CTCCGGCCCC-3'
3'-GGTGACGCCG-acgcgt-GAGGCCGGGG-5'
     HS2 site    Fspl    HS1 site
```

C

```
5'-CTTAATCATA-tacgta-TATTAGTTTC-3'
3'-GAATTAGTAT-atgcat-ATAATCAAAG-5'
     control site   SnaBl   control site
```

D

methylated — / unmethylated —

10 kb / 8 kb / 6 kb / 5 kb / 4 kb — 1kb ladder

	RYVSL		AWDCS		PFCSY		YFLSE		NYRAL		NFRAD		FWWVG		PYTSN		SYSSY		KFNSE		KFNSE		KFNSE	
Fspl (target)		X		X		X		X		X		X		X		X		X		X		X		X
SnaBl (non-target)	X		X		X		X		X		X		X		X		X		X		X		X	
Ncol	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
--[297-301]--	RYVSL		AWDCS		PFCSY		YFLSE		NYRAL		NFRAD		FWWVG		PYTSN		SYSSY		KFNSE		KFNSE		KFNSE	
point mutant																					C141S		Q147L	

representative selected library variants | "WT" split M.SssI | inactive site mutant | known biased variant

Figure 3.2 Methylation assay for selected variants (A) Relative locations of the target site and non-target site on a plasmid linearized by NcoI digestion. (B) The target site is comprised of the HS1 and HS2 zinc finger recognition sites flanking an internal FspI restriction site. The targeted CpG site is nested within this FspI restriction site (C) The non-target site lacks the HS1 and HS2 recognition sequences, but contains a SnaBI restriction site with a nested CpG site for the assessment of off-target methylation (D) The restriction

81

endonuclease protection assay for methylation at the target and non-target site uses digestion with NcoI and either FspI or SnaBI for assessment of target and non-target methylation, respectively. FspI and SnaBI cannot digest a methylated site. Shown are results from select variants as well as the 'wildtype' heterodimeric enzyme (i.e. no mutations to residues 297-301) with or without a catalytically inactivating (C141S), or a catalytically compromised (Q147L) mutation.

Most variants displayed biased methyltransferase activity toward the targeted site. A complete list of sequenced variants can be found in Table 1. A comparison of the sequences of active variants, using weblogo 3.3, indicated that a functional heterodimeric methyltransferase strongly preferred certain residues at positions 298 and 300 (Fig. 3.3) [175,176]. Position 298 (wildtype phenylalanine) was almost exclusively composed of aromatic residues. Position 300 (wildtype serine) was almost exclusively composed of small residues (defined as an amino acid with an R side chain containing 1-3 heavy atoms). The observed conservation at these residues is consistent with sequence alignments showing these two residues are relatively well-conserved among methyltransferases of different species [24].

In contrast, positions 297, 299 and 301 exhibited little preference for specific amino acids. This finding is consistent with the mutational studies discussed above, as these residues were not found to be important for the catalytic activity of the monomeric enzyme[26]. Our study reveals that there are numerous solutions for improving the specificity of our zinc finger-fused, bifurcated methyltransferases. To further characterize some of these variants, we cloned library fragments into plasmids containing either a control non-target site (lacking both zinc finger binding sites) or a half-site (lacking one of the zinc finger sites) adjacent to the FspI restriction site.

As with our previously described split M.HhaI constructs, these split M.SssI constructs did not require the presence of both zinc finger binding sites for methylation activity (data not shown) [168]. However, CFESY and SYSSS exhibited a synergistic activity caused when both zinc finger recognition sites flanked the targeted CpG site. In other words, the observed activity at the full site was greater than the additive effects of each individual half site.

A

Wild type	**-K**	**F**	**N**	**S**	**E-**
	297	298	299	300	301

B
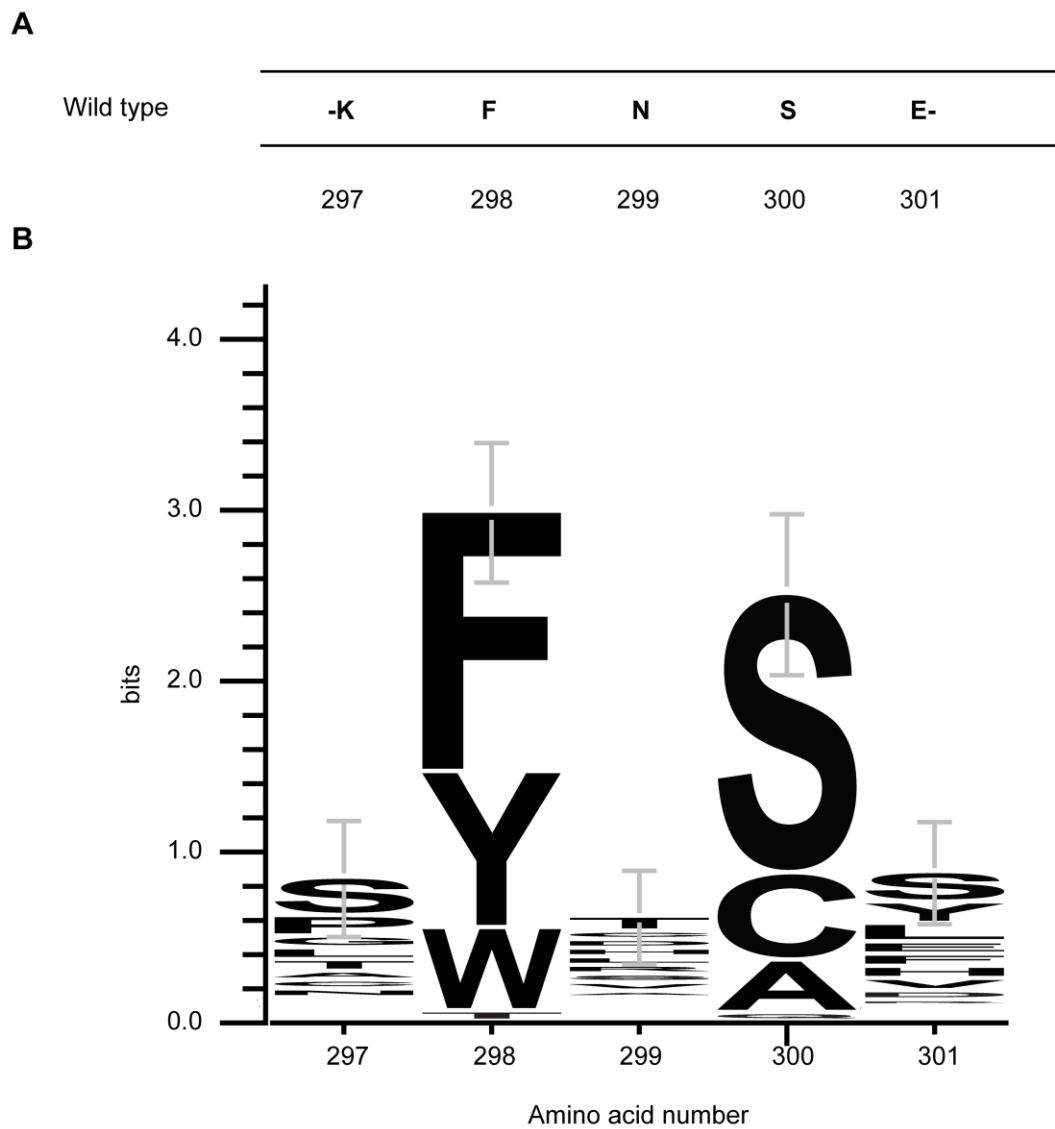
Figure 3.3 Sequence conservation at residues 297-301 of all catalytically active selected variants (A) The wild type sequence for residues 297-301 (B) A sequence logo of active variants.

Colony #	Amino acid at position					assayed	active	notes
	K297	F298	N299	S300	E301			
1	T	F	T	A	H	x	x	
5	P	Y	C	S	F	x	x	
6	G	W	H	S	Y	x	x	
7	C	F	E	S	Y	x	x	
8	V	F	M	*	L			
9	R	F	D	S	L	x	x	
12	S	F	R	C	D	x	x	
13	Y	L	N	G	I	x		
14	S	W	L	S	S	x	x	
15	C	F	A	S	S	x	x	
16	R	R	I	L	*			
18	L	F	L	S	A	x	x	
21	R	W	A	S	*			
22	S	Y	S	S	S	x	x	
23	K	F	N	S	E	x		WT
24	L	W	N	A	S	x	x	
25	H	F	T	S	S	x	x	
26	G	F	E	S	F	x	x	
29	S	F	T	A	R	x	x	
30	S	F	V	S	T	x	x	
31	K	F	N	S	E			WT
32	S	Y	H	S	V	x	x	
34	G	Y	K	C	R	x	x	
35	P	F	F	C	H	x	x	
37	L	K	C	G	G	x		
41	C	F	A	S	S		x	duplicate
42	L	Y	Y	C	E	x	x	
43	L	W	A	S	L	x	x	
45	S	Y	S	C	Y	x	x	
46	R	Y	V	S	L	x	x	
47	S	Y	A	*	M			
49	L	Y	R	*	E			
50	A	W	D	C	S	x	x	
54	P	F	C	S	Y	x	x	
56	Y	F	L	S	E	x	x	
58	L	F	T	A	Y			
59	N	Y	R	A	L	x	x	
62	P	Y	C	S	F		x	duplicate
63	S	F	R	C	D		x	duplicate
64	N	F	R	A	D	x	x	
66	F	W	W	V	G	x		
67	P	Y	T	S	N	x		
69	S	Y	S	S	Y	x	x	
70	L	*	*	Y	P			
71	T	F	T	A	H		x	duplicate

75	S	Y	H	S	V		x	duplicate
76	P	F	V	S	H	x	x	
79	Q	F	M	S	E	x	x	S296N mutation outside cassette; colony not included in weblogo
88	N	F	P	S	F	x	x	
92	A	W	T	S	V	x	x	
95	S	Y	D	S	L	x	x	
98	T	F	N	C	E	x	x	
99	S	Y	H	S	V		x	duplicate
100	F	W	S	S	Q	x	x	
101	P	F	C	S	Y		x	duplicate on amino acid level; different codon usage than colony 54
102	A	F	D	S	S	x	x	
103	I	Y	L	Q	E	x	x	
105	S	Y	V	S	L	x	x	
107	G	T	P	C	T	x	x	
109	V	F	G	C	P	x	x	
110	P	F	T	S	Y	x	x	
114	T	W	F	S	S	x	x	
116	H	F	T	S	S		x	duplicate

Table 1: Complete list of variants sequenced, assayed and confirmed to have methyltransferase activity Yellow indicates aromatic residues. Blue indicates "small" residues as defined in the text.

3.3.5 The targeted heterodimeric methyltransferases are modular

To test whether or not our targeted M.SssI methyltransferases are modular with respect to the zinc finger domains, we replaced zinc fingers HS1 and HS2 with two zinc fingers designed to target a specific site in the promoter of intercellular adhesion molecule 1 (ICAM1). The previously designed zinc finger CD54-31Opt [177] is adjacent to a CpG site in this promoter. To generate a pair of zinc fingers capable of flanking this CpG site, we designed a second zinc finger, CD54a, to bind downstream from the recognition sequence of CD54-31Opt and adjacent

to the targeted CpG site. The two zinc fingers were fused to fragments comprising non-optimized bifurcated M.SssI fragments (residues KFNSE at positions 297-301) and to two selected variants (CFESY and SYSSS at positions 297-301), replacing the HS1 and HS2 zinc fingers (Fig. 3.4A). These two optimized variants were chosen because methylation at the target site (containing both zinc finger binding sites) was greater than the additive amount of methylation levels observed at half sites, as discussed above.

We assessed the methyltransferase activity and specificity of these constructs in *E. coli* with a restriction endonuclease protection assay (Fig. 3.4C and D). Although all three constructs biased methylation to the target site from the ICAM1 promoter, only the CFESY and SYSSS constructs targeted methylation to the desired site with little to no observable methylation at the non-target site (Fig. 3.4D). Notably, the 'non-target' site in this experiment contained the zinc finger sites recognized by HS1 and HS2 (Fig. 3.4B).

The CD54-31Opt was chosen because it was shown to effectively target the ICAM1 promoter, altering transcription levels when bound to transcriptional activators or repressors [83,177]. Additionally, fusion of CD54-31Opt to Ten-Eleven Translocation gene 2 resulted in a small, observable amount of demethylation around the target site, correlating with a 2-fold upregulation in ICAM1 transcription [84]. Our construct

may potentially enable assessment of the biological effects of targeted

methylation at this site.
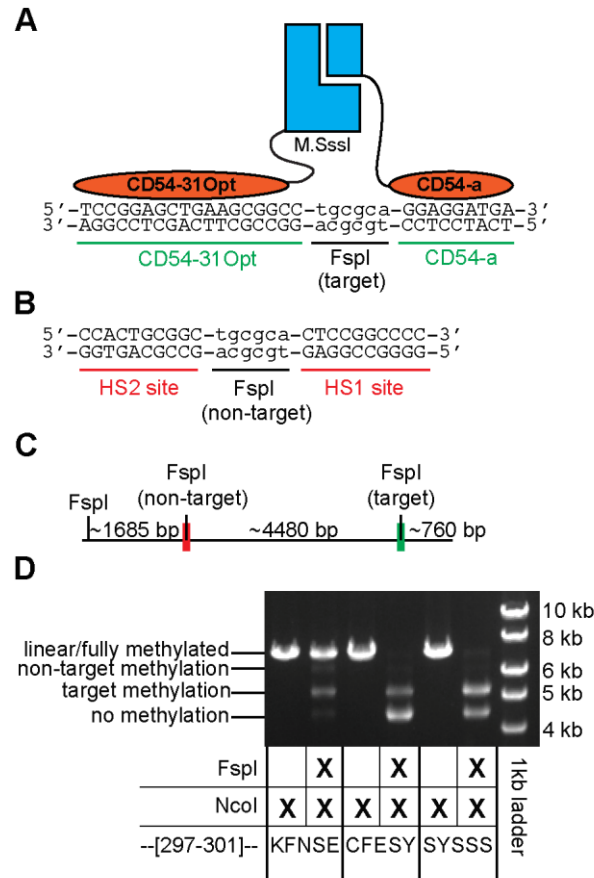
Figure 3.4 Substitution of new zinc fingers targets methylation towards a new site (A) A schematic of the designed methyltransferase is shown assembled over the new targeted CpG site. New cognate zinc finger recognition sequences flank a CpG site nested within an FspI site. Zinc fingers CD54-31Opt and CD54a have replaced the HS1 and HS2 zinc fingers. (B) The non-target site contains the HS1 and HS2 zinc finger recognition sites flanking a CpG site nested within a FspI restriction site (i.e. this was the target site used in the experiments shown in Figure 2). (C) The relative locations of the target site and non-target sites are shown on a plasmid linearized by NcoI digestion. (D) The restriction endonuclease protection assay for methylation at the target and non-target site for the 'wildtype' heterodimeric enzyme (KFNSE) and two selected variants with mutations in the region 297-301.

3.4 Acknowledgments

Intended to be blank

4 Initial tests in eukaryotic cells

4.1 Introduction

Our targeted methyltransferases may be useful as tools to study the phenotypic effects of site-specific methylation at different promoters. These enzymes would also enable us to observe how epigenetic alterations spread or are corrected in human cell lines. Finally, if a hypomethylated region of DNA is responsible for a disease state, a targeted construct might act as a potential therapeutic. However, before a targeted methyltransferase can be utilized in these contexts, it will be necessary to show that zinc finger-fused heterodimeric methyltransferases are functional in eukaryotic cells. In *E. coli*, we have shown that expression levels of zinc finger-fused methyltransferase fragments affect the activity and function of these enzymes [110]. Thus, in eukaryotic cells, the stability, expression levels, toxicity, and nuclear transport may also affect the function of these enzymes. In addition, DNA within the chromosome may be less accessible than plasmid DNA. As discussed in Chapter 1, other biased methyltransferases have successfully methylated eukaryotic chromosomal DNA. However, it is unknown whether our heterodimeric methyltransferases will be able to methylate the chromosome as well.

The goals of this research were to demonstrate that 1) methyltransferase fragments could be expressed in human cell lines, 2)

methylation could be targeted to a specific site in a mammalian expression vector and 3) heterodimeric methyltransferases could methylate chromosomal DNA in human cell lines. Achieving these goals would provide the proofs of principle necessary for larger-scale studies attempting to target the methyltransferase to specific CpG sites within eukaryotic chromosomal DNA.

4.2 Materials and methods

4.2.1 Reagents and bacterial strains

Restriction endonucleases, T4 ligase, T4 PNK, and Phusion High Fidelity MMX were purchased from New England Biolabs (Ipswich, MA). Oligonucleotides and gBlocks were purchased from Integrated DNA Technologies (Coralville, IA). For cloning and sequencing, plasmids were isolated using QIAprep Spin Miniprep Kit (Qiagen,Valencia, CA). DNA fragments and PCR products were purified from agarose gels using PureLink Quick Gel Extraction Kit (Life Technologies, Carlsbad,CA, USA) and further concentrated using DNA Clean and Concentrator-5 (Zymo Research, Irvine, CA) according to manufacturers instructions.

Escherichia coli K-12 strain ER2267 *[F proA$^+$B$^+$ lacI^q D(lacZ)M15 zzf::mini-Tn10 (Kan^R)/ D(argF-lacZ)U169 glnV44 e14$^-$(McrA$^-$) rfbD1? recA1 relA1? endA1 spoT1? thi-1 D(mcrC-mrr)114::IS10]* was acquired from New England Biolabs (Ipswich, MA) and was used for cloning. NEB 10-beta Competent *E. coli* (High Efficiency) *[Δ(ara-leu) 7697 araD139 fhuA*

ΔlacX74 galK16 galE15 e14- φ80dlacZΔM15 recA1 relA1 endA1 nupG

rpsL (StrR) rph spoT1 Δ(mrr-hsdRMS-mcrBC)] and NEB 5-alpha

Competent *E. coli* (High Efficiency) *[fhuA2D(argF-lacZ)U169 phoA glnV44*

φ80Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17] were also used

for cloning and purchased from New England Biolabs (Ipswich, MA).

4.2.2 Plasmid construction

Genes containing M.SssI heterodimeric fragments have been described

elsewhere [168](Chapter 3). Genes encoding zinc finger-fused M.SssI

heterodimeric fragments were cloned into mammalian expression vector

pBUDCE4.1. The C-terminal fragment zinc finger fusion gene was placed

under the control of the CMV immediate-early promoter. The N-terminal

fragment zinc finger fusion gene was placed under the control of the EF-

1α promoter. Oligonucleotides encoding the SV40-NLS and a FLAG-tag

were annealed to their reverse complement sequence by incubating at

over 95°C for over 2 minutes and cooling to room temperature. Annealed

oligonucleotides contained overhangs complementary to cut sites at

either the N-termini or C-termini of the zinc fingers. Double stranded

DNA was phosphorylated and ligated to fuse these DNA sequences to

zinc finger genes, creating the constructs shown in Figure 4.1B.

The region between the origin of replication and CMV promoter was

removed; we cloned various target sites in its place. These target sites

were created by annealing complementary, phosphorylated

oligonucleotides as above. Oligonucleotides encoded the desired target site and, when annealed to each other, created double stranded sequences of DNA with overhangs complementary for restriction sites in the pBUD plasmid. This DNA was then ligated into pBUD plasmids.

We modified the above plasmid with two BsmBI sites in order clone and test optimized variants that were identified through *E. coli* selections in Chapter 3. A gBlock of the CD54a-fused-C-terminal M.SssI fragment was designed; within this gBlock, two adjacent BsmBI sites separated by an internal sequence replaced the region encoding amino acids [297-301]. This gBlock was then cloned into pBUD and replaced the zinc finger-fused-C-terminal M.SssI fragment in this vector. The internal sequence between the two BsmBI sites was later also altered to remove an unwanted DNA sequence. The final construct is shown in Figure 4.1B.

The above plasmid was used to construct optimized C-terminal constructs, following a golden gate procedure preformed essentially as described [170]. In order to insert novel DNA sequences in the region encoding wildtype residues 297-301, variant sequences were created by designing two complementary oligonucleotides, annealed as above. These oligonucleotides contained sequences encoding novel amino acids flanked by regions complementary to BsmBI cut sites in the plasmid. BsmBI sites were then placed outside of these complementary regions. Digestion of BsmBI in the presence of the plasmid, the annealed

oligonucleotides and T7 ligase allowed for the rapid creation of optimized C-terminal fragments into the pBUD mammalian vectors. These were transformed into *E. coli* and sequenced.

4.2.3 Cell culture

HEK293 cells (a gift from Jim Stivers' lab) were grown in RPMI 1640 with glutamine (Cat #11875-093, Life Technologies, Carlsbad, CA) supplemented with 10% FBS (Hyclone Cat #SH30088.03, Thermo Scientific, Waltham, MA). RKO cells were obtained from the American Type Culture Collection (Manassas, VA). Cells were grown in Minimal Essential Media with Earles (E-MEM) balanced salts and glutamine (Cat#112-018-101, Quality Biologicals, Gaithersburg, MD) supplemented with 10% FBS. Cells were grown at 5% CO_2 and at 37°C. Cells were split by washing with DPBS (Cat #14190-250, Life Technologies, Carlsbad, CA), adding 1-2 mL 0.25% Trypsin-EDTA Cat #25-053-Cl (MediaTech, Herndon, VA) and diluting in appropriate media. Cells were frozen by trypsinizing, diluting in complete media and adding 5% DMSO before storage o/n at -80°C. Cells were then transferred and stored in liquid nitrogen.

4.2.4 Transfection into HEK293 and RKO cells

Cells were transfected with Lipofectamine 2000 Transfection Reagent (Life Technologies, Carlsbad, CA). DNA used for transient transfections was isolated from *E. coli* cultured in low salt media at pH 7.5,

supplemented with 50 µg/ml zeocin (Life Technologies, Carlsbad, CA). Plasmid was isolated with the PureYield Plasmid Miniprep Sytem (Promega, Madison, WI) according to the large culture volume protocol.

The day before transfection, HEK293 cells were seeded into 6-well plates (6x10^{-5} cells/well) or 10 cm dishes (3x10^{-6} cells/dish) to achieve cultures of 90-95% confluency on the day of transfection. For transfections in 6-well plates, 5 µg of DNA was incubated in 625 µl Opti-MEM media (Life Technologies, Carlsbad, CA) for five minutes and combined with 12.5 µl lipofectamine in 625 µl Opti-MEM, which was then incubated for at least 20 minutes at room temperature. RPMI complete media (RPMI+10% FBS) was removed and replaced with 1250 µl Opti-MEM media. The DNA, lipofectamine/Opti-MEM solution was added to cells and incubated for 24 hours at 5% CO_2 and 37°C. This protocol was scaled up six-fold for transfections in 10 cm plates.

For transient transfections of RKO cells, 5x10^{-4} cells/well were seeded into 6-well plates and grown for several days until they achieved 40-60% confluency. A mixture of 2 µg of DNA in 100 µl of E-MEM was incubated for five minutes and mixed with 6 µl of lipofectamine in 100 µl of E-MEM. DNA in E-MEM was combined with lipofectamine in E-MEM and incubated at room temperature for over 20 minutes. Fresh complete media (E-MEM + 10% FBS) (0.8 µl) was added to each well before transfection. The DNA/lipofectamine/E-MEM mixture (200 µl) was added

to each well in a dropwise fashion and incubated for 24 hours at 5% CO_2 and 37°C.

For both RKO and HEK293 cells, after a 24-hour incubation of the transfection reagent and DNA transfection mixture was replaced with 2 ml of the appropriate complete media (per well of a 6-well plate). Media was replaced, if necessary, at 24-hour intervals and the cells were harvested 72 hours after the initial addition of the transfection reagent.

4.2.5 Plasmid digestion assays

Isolation of plasmid DNA was preformed essentially as described [178]. Briefly, for 6-well plates, cells were disrupted mechanically or with trypsin and washed several times in DPBS. Cells were spun at 1500xg, resuspended in residual DPBS and lysed by the addition of 250 μl Hirt lysis buffer (0.6% w/v SDS and 10 mM EDTA). After lysis at room temperature for 20 minutes, 100 μl of ice cold 5M NaCl was added and the mixture was incubated at 4°C overnight. The mixture was spun at 14,000xg for 15 minutes. A phenol chloroform extraction and ethanol precipitation were preformed essentially as described [162]. Phenol:Chloroform extraction of the aqueous layer was preformed at least twice and mixtures were back extracted with TE buffer. Aqueous layers were combined and extracted with an equal volume of chloroform. The aqueous layer was supplemented with 40 mM $MgCl_2$ and 2 μl pellet paint co-precipitant (EMD Millipore, Billerica, MA) per 500 μl of aqueous

97

solution. Three volumes of ethanol (-20°C) per one volume of aqueous layer was added and incubated overnight at -20°C. The solution was centrifuged at 14,000xg and at 0°C for 30 minutes or more. The pellet was washed once in 70% w/v ethanol and redissolved in water. The protocol was scaled 6x and slightly modified for larger 10 cm dish transfection experiments.

Isolated DNA was purified with a Zymo Clean and Concentrator-5 columns essentially as recommended by the manufacturer. Depending on size of the transfection experiment (6-well or 10 cm dish), DNA was incubated with 5 or 15 units of Plasmid-Safe-ATP-Dependent DNAse (Epicentre, Madison, WI) and 5 or 15 μg of DNAse and protease free RNAse (ThermoScientific, Waltham, MA), supplemented with 1mM ATP and 1X Plasmid-Safe reaction buffer. Reactions were incubated for at least 1 hr at 37°C and heat killed at over 70°C for at least 20 minutes. Reactions were divided into three equal aliquots and incubated with SnaBI (2.5 units) supplemented with BSA, FspI (2.5 units), or no enzyme at 37°C for 1 hour. Digestions were analyzed on a 1.2% w/v agarose gel in TAE run at 90 volts for 40 minutes. Images were captured using a Gel Logic 112 Imaging System.

4.2.6 Bisulfite sequencing

RKO cells, transfected with plasmid DNA, were harvested 72 hours after transfection via trypsinization and washed in DPBS. Chromosomal DNA

98

was isolated using a Genomic DNA Extraction PureLink Kit (Life Technologies, Carlsbad, CA) per manufacturers instructions. Isolated DNA was treated with bisulfite DNA reagent using and EZ DNA Methylation-Gold Kit (Zymo Research, Irvine, CA). PfuTurbo Cx Hotstart DNA polymerase (Agilent Technologies, Santa Clara, CA) was used to amplify bisulfite converted DNA. Touch down PCR was used to amplify only the correct region associated with the ICAM1 promoter and was modified from [179]. An initial cycle of 95°C for 3 min was followed by a touchdown PCR (95°C for 1 min, annealing temperature for 1 minute, 72°C for 1 minute). The annealing temperature started at 64°C and was dropped 2°C degrees after two cycles and then decreased 1°C after every other cycle until the annealing temperature reached 57°C. After the touchdown PCR, an additional 40 cycles were carried out with the parameters above and the annealing temperature of 56°C.

Amplified PCR products were purified, ligated into pDIMN plasmids and transformed into NEB5 alpha or NEB10 beta cells. Colony PCR identified colonies containing the insert and these colonies were sent for sequencing. The sense strand was amplified with primers 5'-TAG TGA GCG GCC GCT AAG TTG GAG AGG GAG GAT TTG A-3' (Fw) and 5'-TAG TTT GAA TTC CAT AAA CAA CTA CCT AAA CAT ACA TAA CCT AAC C-3'(Rev). The anti-sense strand was amplified with primers 5'-TGA GTG CGG CCG CAT AAA ATA AAC ACA ATA ACA ATC TCC ACT CTC-3'(Fw)

and 5'-TTG TAT GAA TTC AGG TTG TAA TTT TGA GTA GTA GAG GAG

TTT AG-3' (Rev).

4.2.7 Cell lysis and western blot analysis

At 72 hours after transfection, HEK293 cells in 6-well plates were

washed in ice cold DPBS and lysed in 50 µl ice cold RIPA lysis buffer (per

well) supplemented with 1X protease inhibitor cocktail P8340 (Sigma

Aldrich, St. Louis, MO). Lysates were vortexed intermittently and

incubated on ice for 30 minutes before the soluble fraction was recovered

by centrifugation. A 26 µl aliquot of soluble fraction was mixed with 10 µl

of 4x NuPage LDS Sample Buffer (Life Technologies, Carlsbad, CA) and 4

µl DTT (0.5 M) and incubated at over 70°C for 10 minutes. Samples were

loaded on a 4-12% bis-tris gel and run in MES running buffer

supplemented with 500 µl NuPAGE Antioxidant (Life Technologies,

Carlsbad, CA) at 190 volts for 40 minutes.

Proteins were transferred to PVDF membranes using a Trans-Blot

SD Semi-Dry Electrophoretic Transfer Cell (Biorad, Hercules, CA) in

transfer buffer (10 ml of 20X NuPAGE transfer buffer, 100 µl NuPAGE

antioxidant, 10 ml methanol in 100 ml) at 15 V for 30 minutes. The

membrane was incubated with anti-FLAG monoclonal antibody (cat

#0420 Lifetein, South Plainfield, NJ) diluted 2000-fold in blocking buffer

(5% w/v milk in TBST) overnight at 4°C. The membrane was washed

several times in TBST and incubated at room temperature for 30 min

with a goat anti-mouse-HRP conjugate (cat#170-5047, Biorad, Hercules, CA) diluted 6000-fold in blocking buffer (0.4% w/v dry milk in TBST) in a SNAP I.D. system (Millipore, Billerica, MA). After washing the membrane in TBST, the membrane was developed using the Immun-Star WesternC Chemiluminescence Kit (Biorad, Hercules, CA). Images were taken using the Molecular Imager XRS Gel Doc system and analyzed with Quantity One software.

4.3 Results and discussion

4.3.1 Heterodimeric methyltransferase-fusion proteins target methylation toward specific sites and are expressed in HEK293 cells

We first attempted to demonstrate that methyltransferase fragments can be expressed and can target methylation in HEK293 cells. Each zinc finger methyltransferase fusion protein was cloned under the control of a separate constitutive promoter in the pBUD plasmid (Fig. 4.1A). In these experiments, HS1 and HS2 zinc fingers were fused to N-terminal and C-terminal M.SssI fragments as described in chapters 2 and 3. Additionally, sequences encoding the SV40 NLS and FLAG-tag were fused to the terminal ends of each zinc finger (Fig. 4.1B). Finally, we added a targeted CpG site, nested within an FspI restriction site and flanked by HS1 and HS2 recognition sequences, to the pBUD plasmid (Fig. 4.1C). Transient transfection of pBUD plasmid containing an unrelated gene, Haps59-EGFP fusion [180], demonstrated that under the

101

conditions used to transfect our methyltransferase variants, 75-80% of

the Haps59-EGFP transfected cells were fluorescent 72 hours post-

transfection.
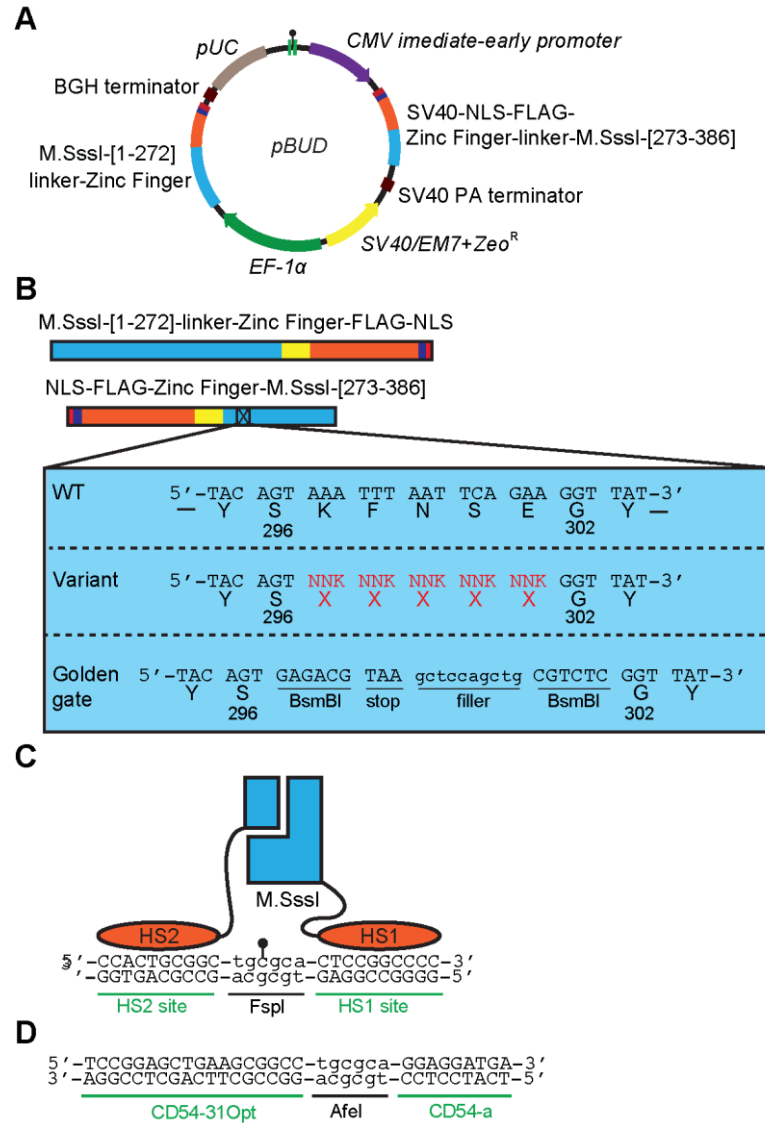
Figure 4.1 Constructs for eukaryotic expression vectors A) The pBUD mammalian expression vector with relevant gene sequences, promoters, resistance marker, and origin of replication. B) A graphical representation of the zinc finger (orange) fused methyltransferase fragments (cyan). FLAG-tags (purple) and NLS-SV40 (red) sequences are attached to each zinc finger. Below the C-terminal fragment, an enlarged area illustrates changes made to amino acid residues 295-303. The 'wildtype' heterodimeric methyltransferase, a generic library variant, or a construct designed to enable golden gate cloning of optimized constructs are shown. Note that the amino acid numbering

corresponds to the monomeric wildtype M.SssI construct. C) A schematic of a zinc finger-fused heterodimeric methyltransferase binding to its' target site. D) The target site for N-terminal and C-terminal heterodimeric methyltransferase fragments fused to CD54-31opt and CD54a, respectively.

The plasmids expressing methyltransferase fragments were

isolated 72 hours after transfection. Transfected plasmids and non-

transfected plasmids were assayed for their sensitivity to endonucleases

whose activity is blocked by CpG methylation. Similar to the *E. coli*

expression tests in Chapter 3, the targeted CpG site is nested within an

FspI site. A SnaBI restriction site present in the CMV-promoter is not

flanked by these zinc finger binding recognition sequences and is

considered a non-target site. Thus, nicked or supercoiled plasmid in FspI

or SnaBI digestion lanes indicate methylation-dependent protection at

the target or non-target sites, respectively.
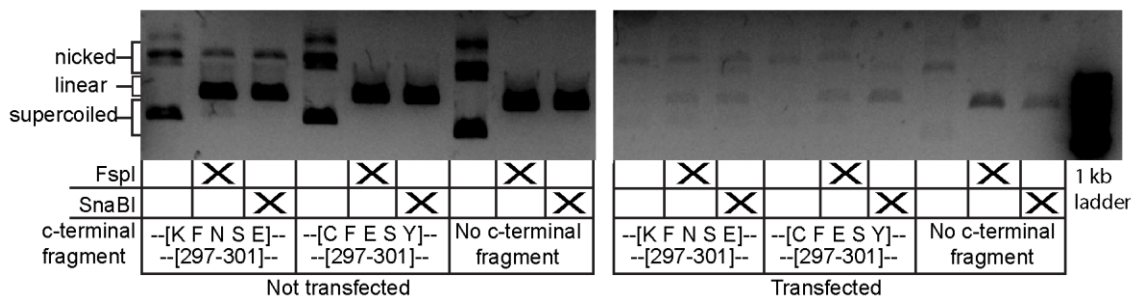
Figure 4.2 Restriction digest assays of the 'wildtype', optimized and inactive variants. Inactive variants lack the zinc finger-fused C-terminal fragment. Variants are digested with no enzyme, FspI or SnaBI. Panel 1 depicts plasmid DNA prior to transfection. In panel 2, plasmid DNA was recovered from transfected HEK293 cells. Top (nicked) and bottom (supercoiled) bands are indicative of methylation-dependent protection from endonuclease digestion. Pixels of control DNA and ladder were saturated. The image was inverted and image contrast proportionally altered to enable visualization of transfected plasmids.

Results demonstrated that the plasmid DNA, prior to transfection,

was sensitive to SnaBI and FspI digestion. This is expected because the

pBUD plasmid lacks promoters recognized by native *E. coli* transcription machinery; methyltransferase fragments, therefore, should not be actively expressed in the *E. coli* from which the plasmid DNA was prepared. However, plasmid DNA encoding 'wildtype' (i.e. no mutations to residues 297-301) methyltransferase fragments appear to be partially protected from digestion prior to transfection (as indicated by nicked DNA in Fig 4.2 panel 1). This may be due to low-level, leaky transcription, and subsequent non-specific methylation of these highly active methyltransferase fragments in *E. coli.* Regardless, the ratio of protected DNA to digested DNA was so low that this was not expected to alter the interpretation of the protection assays in transfected plasmids. Undigested, non-transfected plasmids were present in nicked and supercoiled forms. In this case, the high levels of nicked DNA may result from the isolation procedure or from the use of zeocin, a DNA damaging agent, as a selectable marker during preparation in *E. coli* [181].

For plasmid isolated from transfected cells, the 'wild-type' heterodimeric methyltransferase fusion protein (KFNSE in the region corresponding to 297-301) methylates equally at the target and non-target site, as indicated by the increased presence of nicked DNA relative to linear DNA (Fig 4.2 panel 2). The lack of specificity for the target site over non-target site in HEK293 cells mirrors the lack of specificity observed in *E. coli* [168]. Similar to our *in vivo E. coli* experiments, in HEK293 cells, the optimized variant (residues CFESY in the region

corresponding to 297-301) appears only methylated at the target site.

This result is indicated by the presence of nicked band in the FspI

digested, but not the SnaBI digested lanes (Fig 4.2 panel 2). As expected,

plasmid lacking one of the two obligate heterodimeric fragments shows

no nicked or supercoiled DNA when digested with either FspI or SnaBI.

However, unlike our results in *E. coli,* we observed large amount of

unprotected plasmid DNA in our transfected 'wildtype' constructs. This

may be due to inefficient transcription or translation of the

methyltransferase fragments in our transfected cells. Further, incomplete

methylation may also be due to a limited number of plasmids present in

the nucleus compared to the cytoplasm [182].
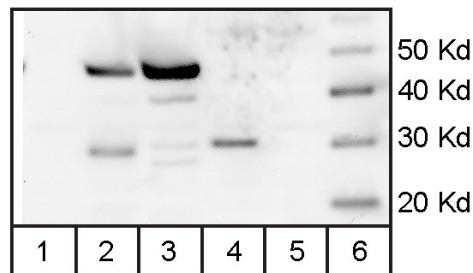
Figure 4.3 Western blot of transiently transfected HEK293 cells Lane 1:
Empty pBUD.CE.4.1; lane 2: pBUD expressing zinc finger-fused N-terminal and
C-terminal 'wild type' fragments; lane 3: pBUD expressing only the zinc finger-
fused N-terminal fragment; lane 4 pBUD expressing FLAG-EGFP-Haps59
fusion; lane 5: empty; lane 6: MagicMark XP Western Protein Standard.

To further demonstrate that both fragments were expressed in at

least some population of HEK293 cells, transiently transfected cells were

lysed 72 hours after transfection. A western blot of the lysates using

anti-FLAG antibodies revealed that cells transfected with the 'wildtype' N-

terminal and C-terminal methyltransferase-zinc finger fusion fragments produced two bands of the expected sizes (45 Kd and 25.8Kd respectively) (Fig. 4.3). Cells transfected with plasmid encoding only the N-terminal fragment expressed only one band (45Kd) of the expected size.

4.3.2 'Wildtype' heterodimeric zinc finger fusion proteins methylate chromosomal DNA.

It would be significant to show that a heterodimeric methyltransferase is active on chromosomal DNA. Studies have shown that zinc fingers known to interact with plasmid DNA may not be able to access the same sequences within the chromosome due to the DNA's inaccessibility within the chromatin structure [183].

To demonstrate that our heterodimeric, zinc finger-fused methyltransferases are active on the chromosome, we transfected pBUD plasmids containing zinc finger methyltransferase fusion proteins into RKO cells. In these experiments, the N-terminal construct was fused to CD54-31Opt and the C-terminal constructs were fused to CD54a (see Chapter 3). A target site with cognate zinc finger binding sequences flanking an internal AfeI site was also cloned into these vectors (Fig. 4.1D). We used these constructs because they encode zinc fingers that, in *E. coli*, efficiently targeted methylation to a region of DNA matching one found in the promoter of the *Intercellular Cell Adhesion Molecule 1* (ICAM1) gene. Further, the promoter of *ICAM1* was found to be

hypomethylated in RKO cells [184]. Preliminary bisulfite analysis confirmed this.

Bisulfite sequencing of the antisense strand (relative to the top strand in Fig. 4.2D) reliably covers 29 CpG sites. When we analyzed 8 clones from bisulfite treated CFESY optimized variant, we observed one methylated site present in one of the 8 clones. This site was not the CpG site flanked by the zinc finger recognition sequences. When we assessed chromosome isolated from cells transected with the 'wildtype variant,' 4 of 15 clones had methylation on at least two sites. One clone was methylated at 16 of the possible 29 sites assessed. Only one sequence appeared methylated at the target site.

The results are the first evidence to suggest that these heterodimeric methyltransferases methylate chromosomal DNA. The transfection efficiency was estimated qualitatively to be 30-40% based on the fluorescence observed in RKO cells transfected with a pBUD Haps59-EGFP construct (the same conditions used to transfect the methyltransferase-containing constructs). Assuming the transfection efficiency of the active 'wildtype' methyltransferase construct is the same, then all successfully transfected cells showed some degree of methylation.

Given the level of activity observed in the plasmid digestion assays, one would expect to observe a higher level of methylation at the desired target site. This might be explained by a decrease in the accessibility of

chromosomal DNA at this site. A lack of observed methylation on the chromosomal target site for both the 'wildtype' and optimized variants might also be explained by the presence of the target site on the pBUD vector. In transfected cells, there would be many more equivalents of target site sequences on the plasmid than on the chromosome. Assuming the levels of activity are relatively low compared to *E. coli* assays, the plasmid's target site would essentially mask any targeting ability of the methyltransferase. Alternatively, the methyltransferase may be methylating the target site on only the sense strand, rather than the antisense strand. A preference for one strand has been observed in our M.HhaI methyltransferase zinc finger fusion proteins [168]. Sequencing data of the sense strand however, was poor, due presumably to the highly repetitive sequences found on the bisulfite-converted sequence.

To test other variants, we used golden gate cloning to quickly construct mammalian expression vectors with optimized C-terminal constructs identified in chapter 3. We also removed the ICAM1 target site from these plasmids and replaced it with the site shown in Figure 4.1C. However, the transient transfection efficiency of these experiments was so low, (10-30%) that little methylation could be detected in the bisulfite analysis.

4.3.3 Conclusion

We demonstrate that both heterodimeric methyltransferase-zinc finger fusion proteins are expressed in HEK293 cells. Further, we provide the first evidence that these constructs bias methylation toward a desired target site in transiently transfected cells. In RKO cells, we demonstrate that very active variants methylate chromosomal DNA. More optimization is needed to increase the transfection efficiency and expression levels in RKO cells before any statements can be made about the ability of these constructs to target methylation on the chromosome.

4.4 Acknowledgments

Intended to be blank

5 Conclusions and future directions

5.1 Introduction

We have engineered targeted methyltransferases by fusing zinc fingers to various bifurcated methyltransferases [110,168]. Zinc finger recognition sequences flank a targeted CpG site; zinc finger binding events assist the reassembly of bifurcated methyltransferase fragments over this targeted CpG site. Targeted methylation is only possible when the bifurcated fragments are impaired in their innate ability to reassemble into a functional enzyme or in their innate affinity for DNA. Deletions to either the presumed fragment-fragment interface or mutations of residues responsible for innate DNA affinity have proven necessary to reduce the off-target methylation of our zinc finger-fused heterodimeric methyltransferases [110,168](Chapter 3).

We have shown that several parameters affect the ability of the fragments to functionally reassemble at a target site and methylate DNA. Important factors include the length of the linkers connecting zinc fingers to their methyltransferase fragment partners as well as the DNA spacers between a targeted CpG site and zinc finger recognition sequences [110,168]. The topology of the fusion between the zinc finger protein and the methyltransferase fragment will also facilitate or impair enzymatic activity at a target site. We believe these factors affect observed methylation at a targeted CpG site because they alter how

fragments are oriented with respect to each other and DNA. Proper orientation, we believe, facilitates fragment reassembly and methylation at a targeted CpG site [110,168].

Directed evolution provides a means to select for enzymes with a desired activity. We successfully devised a positive and negative selection scheme to couple the targeted methylation activity of engineered methyltransferases to the *in vitro* survival of the DNA encoding those enzymes. To select against off-target methylation, we utilized methylation-dependent restriction endonuclease, McrBC [171]. To select for methylation at a target site, we utilized an endonuclease, FspI, whose activity is blocked by methylation.

Finally, we have provided initial evidence to suggest that these zinc finger-fused bifurcated methyltransferases function in mammalian cell lines. However, further work must be done to demonstrate that optimized constructs can target specific sequences within the chromosome.

5.2 Directed evolution: selecting for enzymatic activity regulated by DNA sequence recognition of both zinc finger-binding events.

Our heterodimeric constructs methylate at half sites. Methyltransferase activity is observed if a CpG site is flanked by only one of the zinc finger recognition sequences. We attribute this activity to the residual affinity of the fragments for each other and/or for DNA. However, the presence of two zinc finger cognate sites which flank a CpG site results in a synergistic effect on observed methylation [168]. In other words, the

additive effect observed at both half sites is less than the activity

observed at an intact target site.
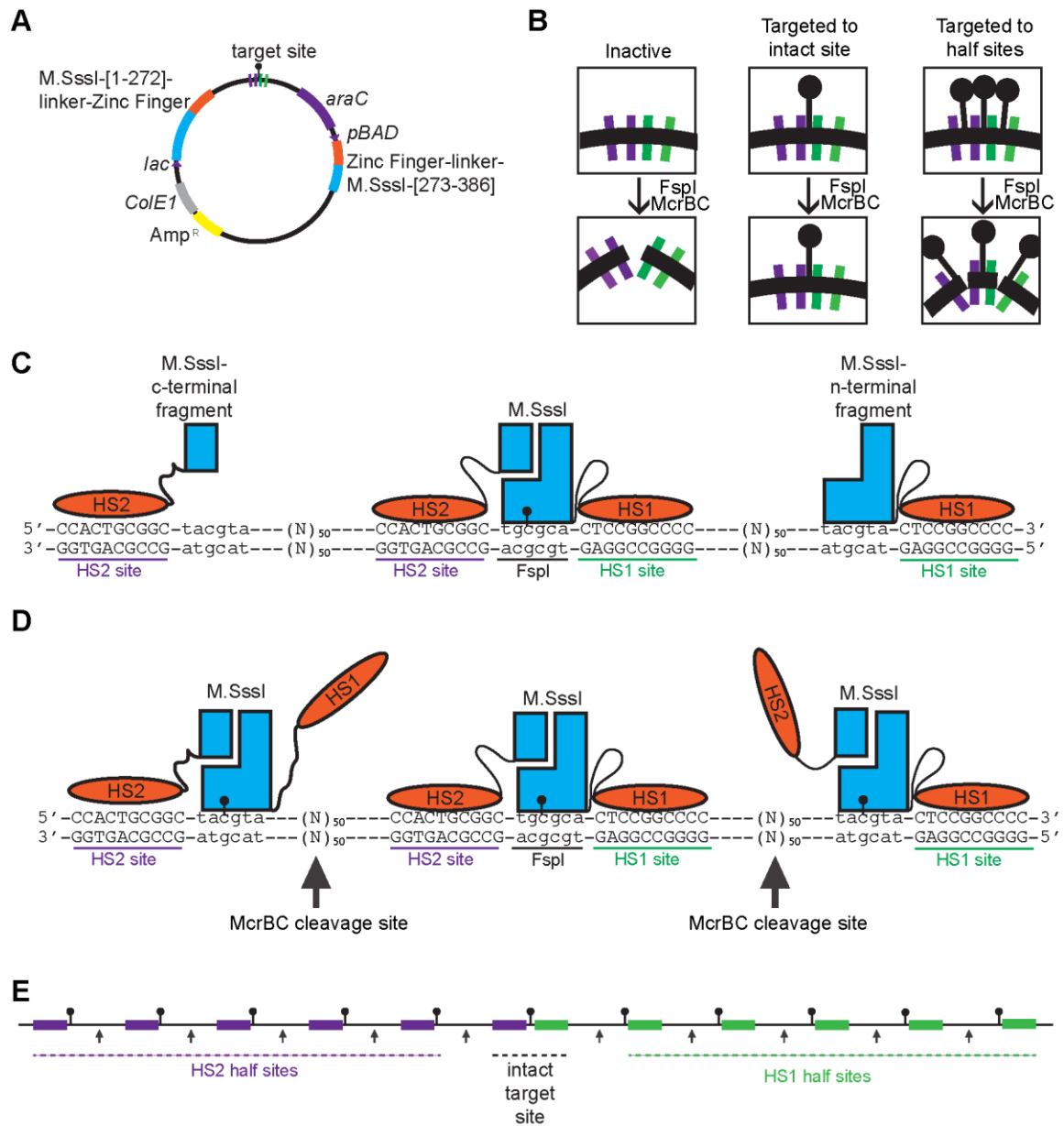
Figure 5.1 Selecting for heterodimeric methyltransferases that require both zinc finger-binding events The selection is preformed as in chapter 3. (A) As in chapter 2 and 3, the vector encodes both heterodimeric fragments fused to zinc fingers under the control of separate, inducible arabinose (*pBAD*) and IPTG (*lac*) inducible promoters; the plasmid also contains an *araC* gene. The target site has been modified to contain two half sites flanking an intact target site; a larger representation is shown in Fig 5.1C. (B) The fate of different methylation patterns in our selection is shown. Unmethylated DNA is digested by FspI. DNA methylated at the target site is protected from FspI digestion but

113

is not a substrate for McrBC digestion. DNA methylated at the target site and half sites is protected from FspI digestion, but is a substrate for McrBC digestion. (C) A target site flanked on either side by two half sites is shown. The ideal zinc finger-fused heterodimeric methyltransferase methylates at the intact target site, but not the half sites. (D) A zinc finger-fused heterodimeric methyltransferase is shown that is not dependent on both zinc finger-cognate DNA sequence interactions. DNA is subject to McrBC digestion (represented by arrows). (E) An illustration demonstrates how the stringency of the selection against half-site activity may be increased by adding tandem half sites. Increasing the number of half sites increases the likelihood that two half sites may be methylated, resulting in subsequent DNA digestion by McrBC.

By modifying the selection preformed in Chapter 3, it may be possible to engineer methyltransferases that are completely dependent upon both zinc finger protein-binding events. This may be accomplished by selecting against the activity at half sites. Figure 5.1 illustrates how modifications can be made to the selection scheme originally outlined in chapter 3.

As in Figure 3.1, the intact target site contains both zinc finger recognition sequences that flank an internal CpG, which is itself nested within an FspI restriction site. However in Figure 5.1A, this target is further flanked by two half sites. Because McrBC endonuclease activity requires two CpG sites, activity occurring at the target site and either of the two half sites will result in endonuclease digestion of the plasmid. Plasmid encoding inactive methyltransferases will be digested with FspI. Methylation at the target site and only the target site will be protected from FspI and McrBC digestion (Fig. 5.1B).

Figure 5.1C and D illustrates the DNA sequence of these new sites demonstrating the activities of our ideal (C) or non-optimized (D) methyltransferase constructs. Finally, the stringency of the selection may

be increased by adding more half sites. Assuming the intact target site sequence is methylated, any other methylation event should result in the destruction of the plasmid DNA.

5.3 Fusion of heterodimeric methyltransferase fragments to TAL effectors

Zinc fingers have been used to create user-defined targeted methyltransferases, nucleases, and transcription factors. However, as noted in the introductory chapter, zinc finger motifs are not completely modular. Zinc finger motifs may bind DNA bases outside their three base pair registers [147]. Attempts at modular assembly of these motifs have high failure rates as assessed by bacterial two-hybrid assays [146]. Though not discussed in this thesis, we designed several other zinc finger methyltransferase fusion constructs that failed to methylate any CpG sites *in vivo*. Recently, zinc finger nucleases were shown to catalyze numerous off-target cleavage events *in vitro* and in cancer cell lines [124]. These recent reports speak to the limitations of zinc finger-mediated protein reassembly strategies.

TAL effectors (TALEs) motifs have been shown to be more modular than zinc finger motifs. Because a single motif binds to a single base, the design of TALEs with new binding specificity has also proven much simpler. Given the ease of design, it would be advantageous to use modular TALEs to construct targeted methyltransferases. Initial experiments not reported in this thesis demonstrated that TALE-

methyltransferase fusion proteins were inactive. However, TALE variants that are expressed well in eukaryotic cells may not express well in *E. coli* (David R. Liu lab personal communication).

TALEs may not be a panacea for those attempting to target specific DNA sequences. Repetitive addition of the same base-specific, repeat-variable di-residues (RVDs) often fail to bind DNA, as assessed by a plant reporter assay [185]. Further, engineered TALEs may not be able to distinguish every base in every sequence context with equal efficiency [186]. RVDs, like zinc fingers, may not be perfectly modular. Further, several variables, including the sequence of non-repetitive TALE N-terminal and C-terminal domains, the topology of the fusion constructs, linkers connecting domains, and base pair spacing may have to be re-optimized to enable proper activity of TALE methyltransferases.

5.4 Targeting methylation toward a human chromosome

Our proof of principle studies demonstrate that targeted methyltransferases are expressed in HEK293 cells, methylate a targeted site on plasmid DNA in HEK293 cells, and will methylate chromosomal DNA in RKO cells. However, we have yet to show that optimized heterodimeric fragments will target a specific site within chromosomal DNA.

There are several complicating factors that must be overcome to enable one to target biologically relevant CpG sites. First, two multi-

domain zinc finger proteins must be designed to target each CpG site. As discussed, zinc finger motifs do not exist for every three base-pair site and modular construction of known motifs may not always result in new zinc finger proteins capable of binding novel DNA sequences. The success of modular assembly in particular was shown to drastically decrease with the decrease in GNN sequences present in a target sequence [146]. Thus, certain sequences flanking a desired CpG site may not be amenable to zinc finger construction. However, zinc fingers need not always be designed *de novo* for every site. Many researchers have designed zinc fingers that bind to gene promoters (as reviewed in [187]). A literature search for a desired target may reveal previously constructed and characterized zinc fingers for a desired region. Such an approach identified CD54-31Opt (Chapters 3 and 4).

Secondly, the human cell line must be hypomethylated at the desired site. HEK293 cells are hypermethylated in the ICAM1 promoter (data not shown). Thus, these cells were not appropriate models for testing our designed ICAM1 targeting methyltransferases.

Finally, the efficiency of transient transfection must be high enough to allow for adequate assessment of the targeted methyltransferase. Inefficient transfection may prevent the observation of the targeted methylation. Assuming the activity of optimized variants in mammalian cells is similar to that observed in *E. coli* (often 40-60% methylation at the target site), then a low (~10%) transfection efficiency

will result in 4-6% methylation at a targeted CpG site. Other groups have overcome this issue by co-transfecting with plasmids expressing LNGFR and GFP. Magnetic activated cell sorting can then be used to enrich for transfected cells and the percentage of enriched transiently transfected cells can then be assessed using fluorescence-activated cell sorting (FACS), creating a more accurate assessment of enzymatic activity [106].

Finally, demonstrating site-specific methylation in human cell lines is complicated by the presence of endogenous human methyltransferases. Unless engineered constructs are transfected and assayed in DNMT knockouts, it will be difficult to assess whether observed methylation is a result of the engineered or endogenous enzymes. Observed off-target methylation around a targeted CpG site may result from the recruitment of endogenous cellular machinery to the original targeted methylated CpG site, encouraging the methylation of surrounding CpG sites.

Assuming these criteria are met and constructs can successfully methylate a specific CpG site within the chromosome, a set of experiments could be designed to probe the effects of site-specific methylation. Quantitative PCR has been used to measure differences in expression caused by biased methyltransferases [105,106]. Chromatin immunoprecipitation of histone modifications might elucidate any epigenetic alterations caused by targeted methylation. Further, long term culture of transiently transfected constructs might allow one to assess

how aberrant epigenetic traits are corrected, maintained or spread over

days and months. A similar experiment was recently performed to first

target H3K9 methylation and then to observe the maintenance or loss of

histone modifications [188].

Intended to be blank

Bibliography

1.   Kornberg A, Zimmerman SB, Kornberg SR, Josse J (1959) Enzymatic synthesis of deoxyribonucleic acid. Influence of bacteriophage T2 on the synthetic pathway in host cells. Proc Natl Acad Sci U S A 45: 772–785.

2.   Gold M, Hurwitz J, Anders M (1963) The enzymatic methylation of RNA and DNA. I. Biochem Biophys Res Commun 11: 107–114. doi:10.1016/0006-291X(63)90075-5.

3.   Gold M, Hurwitz J, Anders M (1963) The enzymatic methylation of RNA and DNA, II. On the species specificity of the methylation enzymes. Proc Natl Acad Sci U S A 50: 164-169.

4.   Arber W (1965) Host specificity of DNA produced by Escherichia coli. V. The role of methionine in the production of host specificity. J Mol Biol 11: 247–256. doi:10.1016/S0022-2836(65)80055-9.

5.   Kühnlein U, Linn S, Arber W (1969) Host specificity of DNA produced by Escherichia coli. XI. In vitro modification of phage fd replicative form. Proc Natl Acad Sci U S A 63: 556–562.

6.   Kühnlein U, Arber W (1972) Host specificity of DNA produced by Escherichia coli. XV. The role of nucleotide methylation in in vitro B-specific modification. J Mol Biol 63: 9–19.

7.   Wion D, Casadesús J (2006) N6-methyl-adenine: an epigenetic signal for DNA–protein interactions. Nat Rev Microbiol 4: 183–192. doi:10.1038/nrmicro1350.

8.   Roberts RJ, Vincze T, Posfai J, Macelis D (2009) REBASE--a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res 38: D234–D236. doi:10.1093/nar/gkp874.

9.   Cheng X (1995) Structure and function of DNA methyltransferases. Annu Rev Biophys Biomol Struct 24: 293–318. doi:10.1146/annurev.bb.24.060195.001453.

10.   Smith HO (1979) Nucleotide sequence specificity of restriction endonucleases. Science 205: 455–462.

11.   Cheng X, Kumar S, Posfai J, Pflugrath JW, Roberts RJ (1993)

Crystal structure of the HhaI DNA methyltransferase complexed with S-adenosyl-L-methionine. Cell 74: 299–307.

12. Shieh F, Youngblood B, Reich NO (2006) The role of Arg165 towards base flipping, base stabilization and catalysis in M.HhaI. J Mol Biol 362: 516–527. doi:10.1016/j.jmb.2006.07.030.

13. Kumar S, Cheng X, Klimasauskas S, Mi S, Posfai J, et al. (1994) The DNA (cytosine-5) methyltransferases. Nucleic Acids Res 22: 1–10.

14. Posfai J, Bhagwat AS, Pósfai G, Roberts RJ (1989) Predictive motifs derived from cytosine methyltransferases. Nucleic Acids Res 17: 2421–2435.

15. Sankpal UT, Rao DN (2002) Structure, function, and mechanism of HhaI DNA methyltransferases. Crit Rev Biochem Mol Biol 37: 167–197. doi:10.1080/10409230290771492.

16. Klimasauskas S, Kumar S, Roberts RJ, Cheng X (1994) HhaI methyltransferase flips its target base out of the DNA helix. Cell 76: 357–369.

17. Gerasimaite R, Merkiene E, Klimasauskas S (2011) Direct observation of cytosine flipping and covalent catalysis in a DNA methyltransferase. Nucleic Acids Res 39: 3771–3780. doi:10.1093/nar/gkq1329.

18. Wu JC, Santi DV (1987) Kinetic and catalytic mechanism of HhaI methyltransferase. J Biol Chem 262: 4778–4786.

19. Osterman DG, DePillis GD, Wu JC, Matsuda A, Santi DV (1988) 5-Fluorocytosine in DNA is a mechanism-based inhibitor of HhaI methylase. Biochemistry 27: 5204–5210.

20. Erlanson DA, Chen L, Verdine GL (1993) DNA methylation through a locally unpaired intermediate. J Am Chem Soc 115: 12583–12584. doi:10.1021/ja00079a047.

21. Chen L, MacMillan AM, Chang W, Ezaz-Nikpay K, Lane WS, et al. (1991) Direct identification of the active-site nucleophile in a DNA (cytosine-5)-methyltransferase. Biochemistry 30: 11018–11025. doi:10.1021/bi00110a002.

22. Chen L, MacMillan A, Verdine GL (1993) Mutational separation of DNA binding from catalysis in a DNA cytosine methyltransferase. J Am Chem Soc 115: 5318–5319.

23.   Renbaum P, Abrahamove D, Fainsod A, Wilson GG, Rottem S, et al. (1990) Cloning, characterization, and expression in Escherichia coli of the gene coding for the CpG DNA methylase from Spiroplasma sp. strain MQ1(M.SssI). Nucleic Acids Res 18: 1145–1152. doi:10.1093/nar/18.5.1145.

24.   Koudan E, Bujnicki J, Gromova E (2004) Homology modeling of the CG-specific DNA methyltransferase SssI and its complexes with DNA and AdoHcy. J Biomol Struct Dyn 22: 339–345.

25.   Darii MV, Kirsanova OV, Drutsa VL, Kochetkov SN, Gromova ES (2007) Isolation and site-directed mutagenesis of DNA methyltransferase SssI. Mol Biol 41: 110–117. doi:10.1134/S0026893307010153.

26.   Darii MV, Cherepanova NA, Subach OM, Kirsanova OV, Raskó T, et al. (2009) Mutational analysis of the CG recognizing DNA methyltransferase SssI: Insight into enzyme–DNA interactions. Biochim Biophys Acta 1794: 1654–1662. Available: http://linkinghub.elsevier.com/retrieve/pii/S1570963909001824 .

27.   Lauster R (1988) Duplication and variation as a phylogenetic principle of type-II DNA methyltransferases. Gene 74: 243.

28.   Lauster R (1989) Evolution of type II DNA methyltransferases. A gene duplication model. J Mol Biol 206: 313–321.

29.   Wilson GG (1992) Amino acid sequence arrangements of DNA-methyltransferases. Methods Enzymol 216: 259–279.

30.   Malone T, Blumenthal RM, Cheng X (1995) Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. J Mol Biol 253: 618–632. doi:10.1006/jmbi.1995.0577.

31.   Choe W, Chandrasegaran S, Ostermeier M (2005) Protein fragment complementation in M.HhaI DNA methyltransferase. Biochem Biophys Res Commun 334: 1233–1240. doi:10.1016/j.bbrc.2005.07.017.

32.   Jeltsch A (1999) Circular permutations in the molecular evolution of DNA methyltransferases. J Mol Evol 49: 161–164. doi:10.1007/PL00006529.

33.   Bujnicki JM (2002) Sequence permutations in the molecular

evolution of DNA methyltransferases. BMC Evol Biol 2: 3.

34. Xu S, Xiao J, Posfai J, Maunus R, Benner J (1997) Cloning of the BssHII restriction-modification system in Escherichia coli : BssHII methyltransferase contains circularly permuted cytosine-5 methyltransferase motifs. Nucleic Acids Res 25: 3991–3994.

35. Karreman C, de Waard A (1990) Agmenellum quadruplicatum M.AquI, a novel modification methylase. J Bacteriol 172: 266–272.

36. Pinarbasi H, Pinarbasi E, Hornby D (2002) Recombinant alpha and beta subunits of M.AquI constitute an active DNA methyltransferase. J Biochem Mol Biol 35: 348–351.

37. Lee KF, Kam KM, Shaw PC (1995) A bacterial methyltransferase M.EcoHK311 requires two proteins for in vitro methylation. Nucleic Acids Res 23: 103–108.

38. Jeltsch A, Kröger M, Pingoud A (1995) Evidence for an evolutionary relationship among type-II restriction endonucleases. Gene 160: 7–16.

39. Olsen GJ, Woese CR, Overbeek R (1994) The winds of (evolutionary) change: breathing new life into microbiology. J Bacteriol 176: 1–6.

40. Jeltsch A, Pingoud A (1996) Horizontal gene transfer contributes to the wide distribution and evolution of type II restriction-modification systems. J Mol Evol 42: 91–96. doi:10.1007/BF02198833.

41. Kobayashi I (2001) Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. Nucleic Acids Res 29: 3742–3756. doi:10.1093/nar/29.18.3742.

42. Ponger L, Li W-H (2005) Evolutionary diversification of DNA methyltransferases in eukaryotic genomes. Mol Biol Evol 22: 1119–1128. doi:10.1093/molbev/msi098.

43. Colot V, Rossignol J-L (1999) Eukaryotic DNA methylation as an evolutionary device. Bioessays 21: 402–411. doi:10.1002/(SICI)1521-1878(199905)21:5<402::AID-BIES7>3.0.CO;2-B.

44. Jurkowski TP, Jeltsch A (2011) On the evolutionary origin of eukaryotic DNA methyltransferases and Dnmt2. PLoS ONE 6: e28104. doi:10.1371/journal.pone.0028104.

45. Bestor TH (1988) Cloning of a mammalian DNA methyltransferase. Gene 74: 9–12. doi:10.1016/0378-1119(88)90238-7.

46. Jia D, Jurkowska RZ, Zhang X, Jeltsch A, Cheng X (2007) Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature 449: 248–251. doi:10.1038/nature06146.

47. Song J, Rechkoblit O, Bestor TH, Patel DJ (2011) Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. Science 331: 1036–1040. doi:10.1126/science.1195380.

48. Gruenbaum Y, Cedar H, Razin A (1982) Substrate and sequence specificity of a eukaryotic DNA methylase. Nature 295: 620–622. doi:10.1038/295620a0.

49. Leonhardt H, Page AW, Weier HU, Bestor TH (1992) A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell 71: 865–873.

50. Goyal R, Reinhardt R, Jeltsch A (2006) Accuracy of DNA methylation pattern preservation by the Dnmt1 methyltransferase. Nucleic Acids Res 34: 1182–1188. doi:10.1093/nar/gkl002.

51. Mortusewicz O, Schermelleh L, Walter J, Cardoso MC, Leonhardt H (2005) Recruitment of DNA methyltransferase I to DNA repair sites. Proc Natl Acad Sci U S A 102: 8905–8909. doi:10.1073/pnas.0501034102.

52. Chuang LS, Ian HI, Koh TW, Ng HH, Xu G, et al. (1997) Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. Science 277: 1996–2000.

53. Sharif J, Muto M, Takebayashi S-I, Suetake I, Iwamatsu A, et al. (2007) The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature 450: 908–912. doi:10.1038/nature06397.

54. Qin W, Leonhardt H, Pichler G (2011) Regulation of DNA methyltransferase 1 by interactions and modifications. Nucleus 2: 392–402. doi:10.4161/nucl.2.5.17928.

55. Okano M, Xie S, Li E (1998) Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat Genet 19: 219–220. doi:10.1038/890.

56. Okano M, Bell DW, Haber DA, Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 99: 247–257.

57. Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH (2001) Dnmt3L and the establishment of maternal genomic imprints. Science 294: 2536–2539. doi:10.1126/science.1065848.

58. Chedin F, Lieber MR, Hsieh C-L (2002) The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. Proc Natl Acad Sci U S A 99: 16916–16921. doi:10.1073/pnas.262443999.

59. Gowher H, Liebert K, Hermann A, Xu G, Jeltsch A (2005) Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L. J Biol Chem 280: 13341–13348. doi:10.1074/jbc.M413412200.

60. Watanabe D, Suetake I, Tada T, Tajima S (2002) Stage- and cell-specific expression of Dnmt3a and Dnmt3b during embryogenesis. Mech Dev 118: 187–190. doi:10.1016/S0925-4773(02)00242-3.

61. Ooi SKT, Qiu C, Bernstein E, Li K, Jia D, et al. (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature 448: 714–717. doi:10.1038/nature05987.

62. Otani J, Nankumo T, Arita K, Inamoto S, Ariyoshi M, et al. (2009) Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. EMBO Rep 10: 1235–1241. doi:10.1038/embor.2009.218.

63. Zhang Y, Jurkowska R, Soeroes S, Rajavelu A, Dhayalan A, et al. (2010) Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. Nucleic Acids Res 38: 4246–4253. doi:10.1093/nar/gkq147.

64. Dhayalan A, Rajavelu A, Rathert P, Tamas R, Jurkowska RZ, et al. (2010) The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. J Biol Chem 285: 26114–26120. doi:10.1074/jbc.M109.089433.

65. Cedar H, Bergman Y (2009) Linking DNA methylation and histone modification: patterns and paradigms. Nat Rev Genet 10: 295–304. doi:10.1038/nrg2540.

66. Gardiner-Garden M, Frommer M (1987) CpG islands in vertebrate

genomes. J Mol Biol 196: 261–282.

67. Deaton AM, Bird A (2011) CpG islands and the regulation of transcription. Genes Dev 25: 1010–1022. doi:10.1101/gad.2037511.

68. Smith ZD, Meissner A (2013) DNA methylation: roles in mammalian development. Nat Rev Genet 14: 204–220. doi:doi:10.1038/nrg3354.

69. Bergman Y, Cedar H (2013) DNA methylation dynamics in health and disease. Nat Struct Mol Biol 20: 274–281. doi:10.1038/nsmb.2518.

70. Li E, Bestor TH, Jaenisch R (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell 69: 915–926.

71. Tsumura A, Hayakawa T, Kumaki Y, Takebayashi S-I, Sakaue M, et al. (2006) Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. Genes Cells 11: 805–814. doi:10.1111/j.1365-2443.2006.00984.x.

72. Robertson KD (2005) DNA methylation and human disease. Nat Rev Genet 6: 597–610. doi:10.1038/nrg1655.

73. Hansen RS, Wijmenga C, Luo P, Stanek AM, Canfield TK, et al. (1999) The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. Proc Natl Acad Sci U S A 96: 14412–14417.

74. Lawson C, Wolf S (2009) ICAM-1 signaling in endothelial cells. Pharmacol Rep 61: 22–32.

75. Tachimori A, Yamada N, Sakate Y, Yashiro M, Maeda K, et al. (2005) Up regulation of ICAM-1 gene expression inhibits tumour growth and liver metastasis in colorectal carcinoma. Eur J Cancer 41: 1802–1810. doi:10.1016/j.ejca.2005.04.036.

76. Ogawa Y, Hirakawa K, Nakata B, Fujihara T, Sawada T, et al. (1998) Expression of intercellular adhesion molecule-1 in invasive breast cancer reflects low growth potential, negative lymph node involvement, and good prognosis. Clin Cancer Res 4: 31–36.

77. Arnold JM, Cummings M, Purdie D, Chenevix-Trench G (2001) Reduced expression of intercellular adhesion molecule-1 in

ovarian adenocarcinomas. Br J Cancer 85: 1351–1358. doi:10.1054/bjoc.2001.2075.

78.   Rosette C (2005) Role of ICAM1 in invasion of human breast cancer cells. Carcinogenesis 26: 943–950. doi:10.1093/carcin/bgi070.

79.   Sun JJ, Zhou XD, Liu YK, Tang ZY, Feng JX, et al. (1999) Invasion and metastasis of liver cancer: expression of intercellular adhesion molecule 1. J Cancer Res Clin Oncol 125: 28–34. doi:10.1007/s004320050238.

80.   Friedrich MG, Chandrasoma S, Siegmund KD, Weisenberger DJ, Cheng JC, et al. (2005) Prognostic relevance of methylation markers in patients with non-muscle invasive bladder carcinoma. Eur J Cancer 41: 2769–2778. doi:10.1016/j.ejca.2005.07.019.

81.   Hellebrekers DMEI, Castermans K, Viré E, Dings RPM, Hoebers NTH, et al. (2006) Epigenetic regulation of tumor endothelial cell anergy: silencing of intercellular adhesion molecule-1 by histone modifications. Cancer Res 66: 10770–10777. doi:10.1158/0008-5472.CAN-06-1609.

82.   Hellebrekers DMEI, Melotte V, Viré E, Langenkamp E, Molema G, et al. (2007) Identification of epigenetically silenced genes in tumor endothelial cells. Cancer Res 67: 4138–4148. doi:10.1158/0008-5472.CAN-06-3032.

83.   de Groote ML, Kazemier HG, Huisman C, van der Gun BTF, Faas MM, et al. (2013) Upregulation of endogenous ICAM-1 reduces ovarian cancer cell growth in the absence of immune cells. Int J Cancer 134: 280–290. doi:10.1002/ijc.28375.

84.   Chen H, Kazemier HG, de Groote ML, Ruiters MHJ, Xu G-L, et al. (2013) Induced DNA demethylation by targeting Ten-Eleven Translocation 2 to the human ICAM-1 promoter. Nucleic Acids Res [Epub ahead of print]: 1–12. doi:10.1093/nar/gkt1019.

85.   Wicki R, Franz C, Scholl FA, Heizmann CW, Schäfer BW (1997) Repression of the candidate tumor suppressor gene S100A2 in breast cancer is mediated by site-specific hypermethylation. Cell Calcium 22: 243–254. doi:10.1016/S0143-4160(97)90063-4.

86.   Zhang X, Wu M, Xiao H, Lee M-T, Levin L, et al. (2010) Methylation of a single intronic CpG mediates expression silencing of the PMP24 gene in prostate cancer. Prostate 70: 765–776. doi:10.1002/pros.21109.

87.	Renda M, Baglivo I, Burgess-Beusse B, Esposito S, Fattorusso R, et al. (2007) Critical DNA binding interactions of the insulator protein CTCF: a small number of zinc fingers mediate strong binding, and a single finger-DNA interaction controls binding at imprinted loci. J Biol Chem 282: 33336–33345. doi:10.1074/jbc.M706213200.

88.	Lukinavicius G, Lapinaite A, Urbanaviciute G, Gerasimaite R, Klimasauskas S (2012) Engineering the DNA cytosine-5 methyltransferase reaction for sequence-specific labeling of DNA. Nucleic Acids Res 40: 11594–11602. doi:10.1093/nar/gks914.

89.	Balganesh TS, Reiners L, Lauster R, Noyer-Weidner M, Wilke K, et al. (1987) Construction and use of chimeric SPR/Φ3T DNA methyltransferases in the definition of sequence recognizing enzyme regions. EMBO J 6: 3543–3549.

90.	Klimasauskas S, Nelson JL, Roberts RJ (1991) The sequence specificity domain of cytosine-C5 methylases. Nucleic Acids Res 19: 6183–6190.

91.	Gubler M, Braguglia D, Meyer J, Piekarowicz A, Bickle TA (1992) Recombination of constant and variable modules alters DNA sequence recognition by type IC restriction-modification enzymes. EMBO J 11: 233–240.

92.	Walter J, Trautner TA, Noyer-Weidner M (1992) High plasticity of multispecific DNA methyltransferases in the region carrying DNA target recognizing enzyme modules. EMBO J 11: 4445–4450.

93.	Lange C, Wild C, Trautner TA (1996) Identification of a subdomain within DNA-(cytosine-C5)-methyltransferases responsible for the recognition of the 5' part of their DNA target. EMBO J 15: 1443–1450.

94.	Cohen HM, Tawfik DS, Griffiths AD (2002) Promiscuous methylation of non-canonical DNA sites by HaeIII methyltransferase. Nucleic Acids Res 30: 3880–3885. doi:10.1093/nar/gkf507.

95.	Cohen HM, Tawfik DS, Griffiths AD (2004) Altering the sequence specificity of HaeIII methyltransferase by directed evolution using in vitro compartmentalization. Protein Eng Des Sel 17: 3–11. doi:10.1093/protein/gzh001.

96.	Tímár E, Groma G, Kiss A, Venetianer P (2004) Changing the recognition specificity of a DNA-methyltransferase by in vitro

evolution. Nucleic Acids Res 32: 3898–3903. doi:10.1093/nar/gkh724.

97.   Gerasimaite R, Vilkaitis G, Klimasauskas S (2009) A directed evolution design of a GCG-specific DNA hemimethylase. Nucleic Acids Res 37: 7332–7341. doi:10.1093/nar/gkp772.

98.   Rockah-Shmuel L, Tawfik DS (2012) Evolutionary transitions to new DNA methyltransferases through target site expansion and shrinkage. Nucleic Acids Res 40: 11627–11637. doi:10.1093/nar/gks944.

99.   Xu GL, Bestor TH (1997) Cytosine methylation targetted to pre-determined sequences. Nat Genet 17: 376–378. doi:10.1038/ng1297-376.

100.   McNamara AR, Hurd PJ, Smith AEF, Ford KG (2002) Characterisation of site-biased DNA methyltransferases: specificity, affinity and subsite relationships. Nucleic Acids Res 30: 3818–3830.

101.   Carvin CD, Parr RD, Kladde MP (2003) Site-selective in vivo targeting of cytosine-5 DNA methylation by zinc-finger proteins. Nucleic Acids Res 31: 6493–6501.

102.   Carvin CD, Dhasarathy A, Friesenhahn LB, Jessen WJ, Kladde MP (2003) Targeted cytosine methylation for in vivo detection of protein-DNA interactions. Proc Natl Acad Sci U S A 100: 7743–7748. doi:10.1073/pnas.1332672100.

103.   Li F, Papworth M, Minczuk M, Rohde C, Zhang Y, et al. (2007) Chimeric DNA methyltransferases target DNA methylation to specific DNA sequences and repress expression of target genes. Nucleic Acids Res 35: 100–112. doi:10.1093/nar/gkl1035.

104.   van der Gun BTF, Maluszynska-Hoffman M, Kiss A, Arendzen AJ, Ruiters MHJ, et al. (2010) Targeted DNA methylation by a DNA methyltransferase coupled to a triple helix forming oligonucleotide to down-regulate the epithelial cell adhesion molecule. Bioconjug Chem 21: 1239–1245. doi:10.1021/bc1000388.

105.   Rivenbark AG, Stolzenburg S, Beltran AS, Yuan X, Rots MG, et al. (2012) Epigenetic reprogramming of cancer cells via targeted DNA methylation. Epigenetics 7: 350–360. doi:10.4161/epi.19507.

106.   Siddique AN, Nunna S, Rajavelu A, Zhang Y, Jurkowska RZ, et al. (2013) Targeted methylation and gene silencing of VEGF-A in

human cells by using a designed Dnmt3a-Dnmt3L single-chain fusion protein with increased DNA methylation activity. J Mol Biol 425: 479–491. doi:10.1016/j.jmb.2012.11.038.

107. de Groote ML, Verschure PJ, Rots MG (2012) Epigenetic Editing: targeted rewriting of epigenetic marks to modulate expression of selected target genes. Nucleic Acids Res 40: 10596–10613. doi:10.1093/nar/gks863.

108. Nomura W, Barbas CF (2007) In vivo site-specific DNA methylation with a designed sequence-enabled DNA methylase. J Am Chem Soc 129: 8676–8677. doi:10.1021/ja0705588.

109. Meister GE, Chandrasegaran S, Ostermeier M (2008) An engineered split M.HhaI-zinc finger fusion lacks the intended methyltransferase specificity. Biochem Biophys Res Commun 377: 226–230. doi:10.1016/j.bbrc.2008.09.099.

110. Meister GE, Chandrasegaran S, Ostermeier M (2010) Heterodimeric DNA methyltransferases as a platform for creating designer zinc finger methyltransferases for targeted DNA methylation in cells. Nucleic Acids Res 38: 1749–1759. doi:10.1093/nar/gkp1126.

111. Pósfai G, Kim SC, Szilák L, Kovács A, Venetianer P (1991) Complementation by detached parts of GGCC-specific DNA methyltransferases. Nucleic Acids Res 19: 4843–4847.

112. Ślaska-Kiss K, Tímár E, Kiss A (2012) Complementation between inactive fragments of SssI DNA methyltransferase. BMC Mol Biol 13: 17. doi:10.1186/1471-2199-13-17.

113. Pelletier JN, Campbell-Valois FX, Michnick SW (1998) Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally designed fragments. Proc Natl Acad Sci U S A 95: 12141–12146. doi:10.1073/pnas.95.21.12141.

114. Paschon DE, Ostermeier M (2004) Construction of protein fragment complementation libraries using incremental truncation. Methods Enzymol 388: 103–116. doi:10.1016/S0076-6879(04)88010-8.

115. Morell M, Ventura S, Avilés FX (2009) Protein complementation assays: approaches for the in vivo analysis of protein interactions. FEBS Lett 583: 1684–1691. doi:10.1016/j.febslet.2009.03.002.

116. Paschon DE, Patel ZS, Ostermeier M (2005) Enhanced catalytic efficiency of aminoglycoside phosphotransferase (3 ')-IIa achieved through protein fragmentation and reassembly. J Mol Biol 353: 26–37. doi:10.1016/j.jmb.2005.08.026.

117. Kim YG, Cha J, Chandrasegaran S (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc Natl Acad Sci U S A 93: 1156–1160.

118. Li L, Wu LP, Chandrasegaran S (1992) Functional domains in Fok I restriction endonuclease. Proc Natl Acad Sci U S A 89: 4275–4279.

119. Bitinaite J, Wah DA, Aggarwal AK, Schildkraut I (1998) FokI dimerization is required for DNA cleavage. Proc Natl Acad Sci U S A 95: 10570–10575.

120. Carroll D (2011) Genome engineering with zinc-finger nucleases. Genetics 188: 773–782. doi:10.1534/genetics.111.131433.

121. Joung JK, Sander JD (2013) TALENs: a widely applicable technology for targeted genome editing. Nat Rev Mol Cell Biol 14: 49–55. doi:10.1038/nrm3486.

122. Alwin S, Gere MB, Guhl E, Effertz K, Barbas CF, et al. (2005) Custom zinc-finger nucleases for use in human cells. Mol Ther 12: 610–617. doi:10.1016/j.ymthe.2005.06.094.

123. Pruett-Miller SM, Connelly JP, Maeder ML, Joung JK, Porteus MH (2008) Comparison of zinc finger nucleases for use in gene targeting in mammalian cells. Mol Ther 16: 707–717. doi:10.1038/mt.2008.20.

124. Pattanayak V, Ramirez CL, Joung JK, Liu DR (2011) Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. Nat Methods 8: 765–770. doi:10.1038/nmeth.1670.

125. Szczepek M, Brondani V, Büchel J, Serrano L, Segal DJ, et al. (2007) Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. Nat Biotechnol 25: 786–793. doi:10.1038/nbt1317.

126. Miller JC, Holmes MC, Wang J, Guschin DY, Lee Y-L, et al. (2007) An improved zinc-finger nuclease architecture for highly specific genome editing. Nat Biotechnol 25: 778–785. doi:10.1038/nbt1319.

127. Doyon Y, Vo TD, Mendel MC, Greenberg SG, Wang J, et al. (2011) Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. Nat Methods 8: 74–79. doi:doi:10.1038/nmeth.1539.

128. Ramalingam S, Kandavelou K, Rajenderan R, Chandrasegaran S (2011) Creating designed zinc-finger nucleases with minimal cytotoxicity. J Mol Biol 405: 630–641. doi:10.1016/j.jmb.2010.10.043.

129. Cornu TI, Thibodeau-Beganny S, Guhl E, Alwin S, Eichtinger M, et al. (2008) DNA-binding specificity is a major determinant of the activity and toxicity of zinc-finger nucleases. Mol Ther 16: 352–358. doi:10.1038/sj.mt.6300357.

130. Händel E-M, Alwin S, Cathomen T (2009) Expanding or restricting the target site repertoire of zinc-finger nucleases: the inter-domain linker as a major determinant of target site selectivity. Mol Ther 17: 104–111. doi:10.1038/mt.2008.233.

131. Elrod-Erickson M, Rould MA, Nekludova L, Pabo CO (1996) Zif268 protein-DNA complex refined at 1.6 angstrom: A model system for understanding zinc finger-DNA interactions. Structure 4: 1171–1180. doi:10.1016/S0969-2126(96)00125-6.

132. Miller J, McLachlan AD, Klug A (1985) Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. EMBO J 4: 1609–1614.

133. Iuchi S (2001) Three classes of C2H2 zinc finger proteins. Cell Mol Life Sci 58: 625–635.

134. Pavletich NP, Pabo CO (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. Science 252: 809–817.

135. Pabo CO, Peisach E, Grant RA (2001) Design and selection of novel Cys2His2 zinc finger proteins. Annu Rev Biochem 70: 313–340. doi:10.1146/annurev.biochem.70.1.313.

136. Choo Y, Klug A (1994) Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions. Proc Natl Acad Sci U S A 91: 11168–11172.

137. Rebar EJ, Pabo CO (1994) Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. Science 263: 671–673. doi:10.1126/science.8303274.

138. Jamieson AC, Kim SH, Wells JA (1994) In vitro selection of zinc fingers with altered DNA-binding specificity. Biochemistry 33: 5689–5695.

139. Segal DJ, Dreier B, Beerli RR, Barbas CF (1999) Toward controlling gene expression at will: Selection and design of zinc finger domains recognizing each of the 5"-GNN-3" DNA target sequences. Proc Natl Acad Sci U S A 96: 2758–2763. doi:10.1073/pnas.96.6.2758.

140. Dreier B, Segal DJ, Barbas CF III (2000) Insights into the molecular recognition of the 5′-GNN-3′ family of DNA sequences by zinc finger domains. J Mol Biol 303: 489–502. doi:10.1006/jmbi.2000.4133.

141. Dreier B, Beerli RR, Segal DJ, Flippin JD, Barbas CF (2001) Development of zinc finger domains for recognition of the 5"-ANN-3" family of DNA sequences and their use in the construction of artificial transcription factors. J Biol Chem 276: 29466–29478. doi:10.1074/jbc.M102604200.

142. Dreier B, Fuller RP, Segal DJ, Lund CV, Blancafort P, et al. (2005) Development of zinc finger domains for recognition of the 5"-CNN-3" family DNA sequences and their use in the construction of artificial transcription factors. J Biol Chem 280: 35588–35597. doi:10.1074/jbc.M506654200.

143. Blancafort P, Magnenat L, Barbas CF (2003) Scanning the human genome with combinatorial transcription factor libraries. Nat Biotechnol 21: 269–274. doi:10.1038/nbt794.

144. Sander JD, Zaback P, Joung JK, Voytas DF, Dobbs D (2007) Zinc Finger Targeter (ZiFiT): an engineered zinc finger/target site design tool. Nucleic Acids Res 35: W599–W605. doi:10.1093/nar/gkm349.

145. Mandell JG, Barbas CF (2006) Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases. Nucleic Acids Res 34: W516–W523. doi:10.1093/nar/gkl209.

146. Ramirez CL, Foley JE, Wright DA, Müller-Lerch F, Rahman SH, et al. (2008) Unexpected failure rates for modular assembly of engineered zinc fingers. Nat Methods 5: 374–375. doi:10.1038/nmeth0508-374.

147. Isalan M, Choo Y, Klug A (1997) Synergy between adjacent zinc fingers in sequence-specific DNA recognition. Proc Natl Acad Sci U

S A 94: 5617–5621.

148. Hurt JA, Thibodeau SA, Hirsh AS, Pabo CO, Joung JK (2003) Highly specific zinc finger proteins obtained by directed domain shuffling and cell-based selection. Proc Natl Acad Sci U S A 100: 12271–12276. doi:10.1073/pnas.2135381100.

149. Maeder ML, Thibodeau-Beganny S, Osiak A, Wright DA, Anthony RM, et al. (2008) Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. Mol Cell 31: 294–301. doi:10.1016/j.molcel.2008.06.016.

150. Maeder ML, Thibodeau-Beganny S, Sander JD, Voytas DF, Joung JK (2009) Oligomerized pool engineering (OPEN): an "open-source" protocol for making customized zinc-finger arrays. Nat Protoc 4: 1471–1501. doi:10.1038/nprot.2009.98.

151. Sander JD, Dahlborg EJ, Goodwin MJ, Cade L, Zhang F, et al. (2011) Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). Nat Methods 8: 67–69. doi:10.1038/nmeth.1542.

152. Moscou MJ, Bogdanove AJ (2009) A Simple Cipher Governs DNA Recognition by TAL Effectors. Science 326: 1501–1501. doi:10.1126/science.1178817.

153. Scholze H, Boch J (2010) TAL effector-DNA specificity. Virulence 1: 428–432. doi:10.4161/viru.1.5.12863.

154. Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, et al. (2009) Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. Science 326: 1509–1512. doi:10.1126/science.1178811.

155. Smith AE, Hurd PJ, Bannister AJ, Kouzarides T, Ford KG (2008) Heritable gene repression through the action of a directed DNA methyltransferase at a chromosomal locus. J Biol Chem 283: 9878–9885. doi:10.1074/jbc.M710393200.

156. Minczuk M, Papworth MA, Kolasinska P, Murphy MP, Klug A (2006) Sequence-specific modification of mitochondrial DNA using a chimeric zinc finger methylase. Proc Natl Acad Sci U S A 103: 19689–19694. doi:10.1073/pnas.0609502103.

157. Michnick SW, Remy I, Campbell-Valois FX, Vallée-Bélisle A, Pelletier JN (2000) Detection of protein-protein interactions by protein fragment complementation strategies. Methods Enzymol 328: 208–230.

158. Vlahovicek K, Pongor S (2000) Model.it: building three dimensional DNA models from sequence data. Bioinformatics 16: 1044–1045. doi:10.1093/bioinformatics/16.11.1044.

159. Qian B, Raman S, Das R, Bradley P, McCoy AJ, et al. (2007) High-resolution structure prediction and the crystallographic phase problem. Nature 450: 259–264. doi:10.1038/nature06249.

160. Wang C, Bradley P, Baker D (2007) Protein-protein docking with backbone flexibility. J Mol Biol 373: 503–519. doi:10.1016/j.jmb.2007.07.050.

161. Mandell DJ, Coutsias EA, Kortemme T (2009) Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. Nat Methods 6: 551–552. doi:10.1038/nmeth0809-551.

162. Sambrook J, Russell D (2001.) Molecular Cloning: A Laboratory Manual. 3rd ed. Argentine J, Irwin N, Janssen KA, Curtis S, Zierler M, et al., editors Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press. 722; 5.4–5.17; 8.21–8.22 pp.

163. Beerli RR, Segal DJ, Dreier B, Barbas CF (1998) Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. Proc Natl Acad Sci U S A 95: 14628–14633.

164. Nelson M, McClelland, M (1989) Effect of site-specific methylation on DNA modification methyltransferases and restriction endonucleases. Nucleic Acids Res 17: r389-r415.

165. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947–2948. doi:10.1093/bioinformatics/btm404.

166. Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, et al. (2010) A new bioinformatics analysis tools framework at EMBL-EBI. Nucleic Acids Res 38: W695–W699. doi:10.1093/nar/gkq313.

167. Horsthemke B, Buiting K (2008) Genomic imprinting and imprinting defects in humans. Adv Genet 61: 225–246. doi:10.1016/S0065-2660(07)00008-9.

168. Chaikind B, Kilambi KP, Gray JJ, Ostermeier M (2012) Targeted DNA methylation using an artificially bisected M.HhaI fused to

zinc fingers. PLoS ONE 7: e44852. doi:10.1371/journal.pone.0044852.

169. Firnberg E, Ostermeier M (2012) PFunkel: efficient, expansive, user-defined mutagenesis. PLoS ONE 7: e52031. doi:10.1371/journal.pone.0052031.

170. Sanjana NE, Cong L, Zhou Y, Cunniff MM, Feng G, et al. (2012) A transcription activator-like effector toolbox for genome engineering. Nat Protoc 7: 171–192. doi:10.1038/nprot.2011.431.

171. Sutherland E, Coe L, Raleigh EA (1992) McrBC: a multisubunit GTP-dependent restriction endonuclease. J Mol Biol 225: 327–348. doi:10.1016/0022-2836(92)90925-A.

172. Dryden DT, Murray NE, Rao DN (2001) Nucleoside triphosphate-dependent restriction enzymes. Nucleic Acids Res 29: 3728–3741. doi:10.1093/nar/29.18.3728.

173. Stewart FJF, Raleigh EAE (1998) Dependence of McrBC cleavage on distance between recognition elements. Biol Chem 379: 611–616.

174. Renbaum P, Razin A (1992) Mode of action of the Spiroplasma CpG methylase M.SssI. FEBS Lett 313: 243–247.

175. Schneider TD, Stephens RM (1990) Sequence logos: a new way to display consensus sequences. Nucleic Acids Res 18: 6097–6100.

176. Crooks GE, Hon G, Chandonia J-M, Brenner SE (2004) WebLogo: a sequence logo generator. Genome Res 14: 1188–1190. doi:10.1101/gr.849004.

177. Magnenat L, Blancafort P, Barbas CF (2004) In vivo selection of combinatorial libraries and designed affinity maturation of polydactyl zinc finger transcription factors for ICAM-1 provides new insights into gene regulation. J Mol Biol 341: 635–649. doi:10.1016/j.jmb.2004.06.030.

178. Hirt B (1967) Selective extraction of polyoma DNA from infected mouse cell cultures. J Mol Biol 26: 365–369.

179. Jolly CJ, Neuberger MS (2001) Somatic hypermutation of immunoglobulin kappa transgenes: Association of mutability with demethylation. Immunol Cell Biol 79: 18–22. doi:10.1046/j.1440-1711.2001.00968.x.

180. Wright CM, Wright RC, Eshleman JR, Ostermeier M (2011) A protein therapeutic modality founded on molecular regulation. Proc Natl Acad Sci U S A 108: 16206–16211. doi:10.1073/pnas.1102803108.

181. Oliva-Trastoy M, Trastoy MO, Defais M, Larminat F (2005) Resistance to the antibiotic Zeocin by stable expression of the Sh ble gene does not fully suppress Zeocin-induced DNA cleavage in human cells. Mutagenesis 20: 111–114. doi:10.1093/mutage/gei016.

182. Lechardeur D, Lukacs GL (2006) Nucleocytoplasmic transport of plasmid DNA: a perilous journey from the cytoplasm to the nucleus. Hum Gene Ther 17: 882–889. doi:10.1089/hum.2006.17.882.

183. Zhang L, Spratt SK, Liu Q, Johnstone B, Qi H, et al. (2000) Synthetic zinc finger transcription factor action at an endogenous chromosomal site. Activation of the human erythropoietin gene. J Biol Chem 275: 33850–33860. doi:10.1074/jbc.M005341200.

184. Easwaran HP, Van Neste L, Cope L, Sen S, Mohammad HP, et al. (2010) Aberrant silencing of cancer-related genes by CpG hypermethylation occurs independently of their spatial organization in the nucleus. Cancer Res 70: 8015–8024. doi:10.1158/0008-5472.CAN-10-0765.

185. Streubel J, Blücher C, Landgraf A, Boch J (2012) TAL effector RVD specificities and efficiencies. Nat Biotechnol 30: 593–595. doi:10.1038/nbt.2304.

186. Meckler JF, Bhakta MS, Kim M-S, Ovadia R, Habrian CH, et al. (2013) Quantitative analysis of TALE-DNA interactions suggests polarity effects. Nucleic Acids Res 41: 4118–4128. doi:10.1093/nar/gkt085.

187. Sera T (2009) Zinc-finger-based artificial transcription factors and their applications. Adv Drug Deliv Rev 61: 513–526. doi:10.1016/j.addr.2009.03.012.

188. Hathaway NA, Bell O, Hodges C, Miller EL, Neel DS, et al. (2012) Dynamics and memory of heterochromatin in living cells. Cell 149: 1447–1460. doi:10.1016/j.cell.2012.03.052.

Intended to be blank

Curriculum Vitae

BRIAN CHAIKIND
PhD: Chemical Biology
Phone: (301) 346-4004
Email: bchaiki1@gmail.com

BIRTH
Washington, District of Columbia
06-06-1985

EDUCATION

Johns Hopkins University-Chemistry-Biology Interface Program..............2008-2014
Baltimore, MD
- PhD candidate *degree expected in Winter 2014*
- Masters in Chemical Biology obtained June 2010

Northwestern University..2003-2007
Evanston, IL
- Bachelor of Arts in Environmental Science, Minor in Chemistry
- Cumulative GPA: 3.80
- Cum Laude
- Departmental Honors
- Senior Thesis: "Solving the supply problem associated with marine natural products chemistry"

LABORATORY EXPERIENCE

Protein Engineering Laboratory-Dr. Marc Ostermeier...............................2008-2014
Johns Hopkins University, Baltimore, MD
- Rational design of zinc finger targeted methyltransferase proteins
- Developed novel selection strategy for the directed evolution of targeted methyltransferases
- Worked on proof of principle studies testing engineered methyltransferases in human cell lines
- Responsible for training graduate and undergraduate students

Marine Natural Products Laboratory-Dr. William Gerwick.........................2007-2008
Scripps Institution of Oceanography, San Diego, CA
- Employed as a Staff Research Associate II
- Responsible for extracting and fractionating algal samples
- Responsible for training laboratory volunteers

Analytical Laboratory Course..2007
Northwestern University, Evanston, IL
- Trained to use and interpret HPLC and GC-MS equipment; trained in the theory of chromatography.

Natural Products Laboratory-Dr. Lyndon West...2006
University of Georgia, Athens, GA

- Awarded Northwestern's Undergraduate Research Grant to create a quantitative analysis of natural products under Dr. Lyndon West. In this facility, I became proficient with HPLC, NMR, and UV spec. equipment.
- Collected samples while scuba diving and snorkeling

Plant Systematics Course...2006
Northwestern University, Evanston, IL
- Trained to identify and collect plant species.

Biology laboratory for genetics, biochemistry, and molecular biology......2005-2006
Northwestern University, Evanston, IL
- Trained to plate and grow bacteria using sterile techniques, used affinity chromatography to purify chimeric proteins

Independent Research-Dr. Richard Silverman's Group..............................2004-2005
Northwestern University, Evanston, IL
- Synthesized selective nNOS inhibitors, ran butyl lithium reactions, trained to run NMR spectrometers, experienced with column chromatography

Organic Synthesis Lab Course for Chemistry Majors.......................................2004
Northwestern University, Evanston, IL
- Successfully designed a procedure to oxidize a benzylaldehyde using green chemistry,
  trained to interpret H-NMR, C-NMR, GC/MS, and FT-IR, trained to use FT-IR equipment

Organic Chemistry Lab...2003-2004
Northwestern University, Evanston, IL
- Trained in basic synthesis techniques, synthesized N,N Diethyl-m-toulamide from toulic acid thionyl chloride, and diethylamine, experienced in thin layer chromatography

PUBLICATIONS

Chaikind, B., Ostermeier, M. "Directed evolution of improved zinc finger methyltransferases" [submitted]

Chaikind, B., Kilambi, K. P., Gray, J. J. & Ostermeier, M. "Targeted DNA Methylation Using an Artificially Bisected M.HhaI Fused to Zinc Fingers." *PLoS ONE* 7, e44852 (2012).

CONFERENCES

B. Chaikind, K. Praneeth Kilambi, J.J. Gray, and M. Ostermeier. "Targeted DNA Methylation Using a Bisected M.Hhal Fused to Zinc Fingers." AIChe Annual Meeting (Talk). Pittsburgh, PA. October 29, 2012

B. Chaikind and M. Ostermeier. "Toward a Site-Specific CpG Methyltransferase" (Poster). Gordon Research
Conference. Epigenetics: Mechanisms, Development and Disease. Stonehill College, Easton, MA. August 9, 2011.

B. Chaikind. "Solving the Supply Side Problem of Marine Natural Product Chemistry: A Directed Approach" (Poster). Chicago Undergraduate Research Symposium. Loyola University, Chicago, IL. April 14, 2007.

RELEVANT EXPERIENCE

Editorial Assistant- Journal of Natural Products......................................2007-2008
Scripps Institution of Oceanography, La Jolla, CA
- Worked under Dr. William Gerwick to assist with the editorial process

Shedd Aquarium Workshop Reporting...2007
Shedd Aquarium, Chicago, IL
- Compiled a white paper based on the Shedd Workshop "Managing Cetaceans for Optimal Health." The workshop brought together an interdisciplinary group of researchers to discuss the best husbandry practices for marine mammals.

HONORS/AWARDS

- Voted by faculty to give only student lecture at Johns Hopkins Chemistry-Biology Interface Symposium-Fall 2012
- Award for best poster Johns Hopkins University Chemistry-Biology Interface Symposium- Fall 2011
- Honorable Mention National Science Foundation Graduate Research Fellowship Program
- Northwestern Undergraduate Research Grant-studying natural products chemistry in University of Georgia
- Deans List (3.7 GPA or higher) 7 out of 11 possible quarters at Northwestern University
- Invited to apply to Northwestern Medical School's accelerated 7 year program