Methylated DNA Detection using a High-Affinity Methyl-CpG Binding Protein and Photopolymerization-based Amplification

ARCHIVES
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
JUN 22 2015
LIBRARIES

by

Brandon Walter Heimer

B.S. Chemical Engineering
The University of Texas at Austin, 2007

Submitted to the Department of Chemical Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Science in Chemical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2015

© 2015 Massachusetts Institute of Technology
All rights reserved.

Signature of Author . . . . . Signature redacted .

Brandon Walter Heimer
Department of Chemical Engineering
May 18, 2015

Certified by . . . . . Signature redacted .

Hadley D. Sikes
Assistant Professor of Chemical Engineering
Thesis Supervisor

Accepted by . . . . . Signature redacted .

Richard D. Braatz
Professor of Chemical Engineering
Graduate Officer

Methylated DNA Detection using a High-Affinity Methyl-CpG Binding Protein and Photopolymerization-based Amplification

by

Brandon Walter Heimer

Submitted to the Department of Chemical Engineering on May 18, 2015 in Partial Fulfillment of the Requirements for the Degree of Doctor of Science in Chemical Engineering

Abstract

Methylation at the 5 position of the cytosine base, when followed by guanine (CpG) in the promoter region of a protein-coding gene, is an epigenetic modification that has been shown to be involved in gene silencing. While important for regulating many normal, cellular processes, aberrant DNA methylation has been implicated in the development of human diseases such as cancer. Medical research has begun to identify hypermethylated gene promoters that can be used as predictive, prognostic, and diagnostic biomarkers. Current technologies for DNA methylation detection, however, rely on sodium bisulfite conversion of unmethylated cytosine bases to urasil in the target DNA. Bisulfite treatment is time consuming to perform, degrades >90% of the DNA, and is susceptible to incomplete conversion which can lead to false results. As such, a method that eliminates both chemical conversion and PCR amplification of DNA would enable significantly lower-cost and rapid DNA methylation analysis in the clinic.

Methyl binding domain (MBD) family proteins specifically bind double-stranded, methylated DNA which makes them excellent transducers for DNA methylation in biosensing applications. I displayed three of the core members MBD1, MBD2, and MBD4 on the surface of *Saccharomyces cerevisiae* yeast cells. Using the yeast display platform, I determined the equilibrium dissociation constant of human MBD2 (hMBD2) to be 5.9±1.3 nM for binding to singly methylated DNA. I further used the yeast display platform to evolve the hMBD2 protein for improved binding affinity. Affecting five amino acid substitutions doubled the affinity of the wild-type protein to 3.1±1.0 nM. Further, concatenating the high-affinity MBD variant from 1x to 3x and expressing it in *Escherichia coli* as a green fluorescent protein (GFP) fusion improved affinity 6-fold for interfacial binding.

After optimizing the expression and purification of engineered MBD-GFP proteins in *E. coli*, I developed a simple method for detecting methylated DNA fragments from the *MGMT* gene promoter. A defining feature is that target oligonucleotides from the test sample hybridize directly to capture probes printed in 300 μm diameter spots on an inexpensive biochip without requiring bisulfite conversion. Sub-nanomolar concentrations (0.3 nM) of methylated DNA duplexes are detected using an MBD-GFP protein to transduce binding to either a fluorescent or colorimetric (visible hydrogel) signal formed by photopolymerization with short (<2 min) reaction times.

Thesis Supervisor: Hadley D. Sikes
Title: Assistant Professor of Chemical Engineering

Acknowledgements

This thesis was influenced by countless people to whom I owe many thanks. First, I would like to thank my thesis advisor, Professor Hadley Sikes, for her mentorship, encouragement, and giving me the opportunity to work on this project which combines two of my passions: protein engineering and improving cancer care. She has supported me in professional development and helped guide me throughout my time at MIT; for which, I am extremely grateful. I also want to thank the members of my thesis committee, Professors Paula Hammond and Bob Langer, for their guidance, inspiration, and direction while navigating through my doctoral research over the past four years.

I would like to thank many members of the Wittrup Lab, Tiffany Chen, Michael Jordi Mata-Fink, Santos, and Jim VanDeventer, for their instruction and very helpful conversations while yeast surface displaying and characterizing the MBD proteins. I am also very grateful to have had the opportunity to collaborate with Tanya Shatova of the Jensen Lab to develop and implement a microfluidic device with the MBD biochip.

I honestly could not have asked for better labmates. The Sikes Lab is a particularly collaborative bunch, and I truly believe this group is better than the sum of its parts. Our friendships forged over the lab benches of E19 will truly last as ΔS goes to infinity. Their encouragement and technical support were particularly welcome after failed PCR reactions, arrays that refused to polymerize, broken FPLCs, and nearly every other instrument in the lab just to name a few. I first want to recognize Dr. Jungkyu Lee who was my only loyal colleague after the rest of the lab left us to attend the David H. Koch School of Chemical Engineering Practice. I am particularly honored, privileged, and grateful to have worked alongside the three other members of the "Core Four." In alphabetical order, I thank Sohali Ali for his humor and introducing me to the game of cricket as we watched erratic streams of World Cup matches on his laptop as we "worked" on experiments; Kaja Kaastrup for her spirited political debates, debriefs on the latest (and most controversial) *On Point* episodes, starting/ending the birthday cake tradition, and of course, for her collaboration on the agarose slides that "magically" enabled very many useful biochip-based assays; and Joe Lim for his ninja-like mastery of all things PCR and OriginPlot in addition to being an all-around good friend, sports aficionado, and Chipotle burrito connoisseur.

I want to thank the undergraduate students who I had the privilege of mentoring: Angel Asante, Mary Boyd, and Miryam Saad. Thank you very much!

To my wife Emily: Words alone cannot express my gratitude for your support (including a semester of *very* late night TA office hours), for enduing "Grumpy Brandon" after experiments went badly, and for your unending encouragement. Graduate school can be quite difficult at times, but it was always better with you by my side.

To my parents and sister: Y'all have always encouraged and believed in me from the very beginning when I disassembled Mom's rocking chair…and Amanda's crib until I first said I want to be an engineer and go to MIT. I have many things to thank you for Mom and Dad, but in this context, I want to particularly recognize the hard work and sacrifices you made to get me to this very moment: writing my doctoral thesis at MIT! Thank you!

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Epigenetic Regulation by DNA Methylation

The word "epigenetics" is derived from the Latin root "epi-" and Greek "genesis" meaning "upon, over, or on top of" and "origin," respectively. Therefore, epigenetics involves the study of heritable, non-coding changes to the DNA molecule that comprises the hereditary information for life (*1*). There are several such epigenetic modifications that affect the structure of chromatin including DNA methylation, histone methylation, and histone acetylation (*2*). Each of these is worthy of many thesis research projects; however, this one will focus solely on DNA methylation and, more specifically, its detection for analytical purposes.

1.1.1 CpG Islands and DNA Methylation

A CpG dinucleotide arises in DNA when a cytosine base is sequentially followed by guanine (5' to 3') on the same DNA strand. Approximately 70% of the promoter regions for human genes have a CpG content that is higher than the normative level in the genome. These regions of high CpG dinucleotide content are known as CpG islands and are often co-incident with gene promoters (*3*). Cytosine bases in this context can exist in the normal state or with a methyl group appended at the 5 position of the pyrimidine ring (5-methylcytosine, 5mC).

In mammals, cytosine methylation is performed by the DNA methyltransferase (DNMT) class of enzymes by either one of two mechanisms: maintenance or *de novo*. Maintenance DNA methylation is performed by DNMT1 which restores symmetric methylation to a DNA strand left hemimethylated following DNA replication. The DNMT3 enzymes de novo methylate DNA at CpGs in which neither cytosine is methylated a priori thereby establishing genomic methylation patterns (*4-6*).

The majority, 80%, of globally distributed CpG dinucleoties exist methylated in non-embryonic cells (*7*). In contrast to the global genomic profile, current research has shown that dinucleotides in CpG islands, particularly when located within promoter regions, often occur unmethylated regardless of the transcription state of the gene (*8*). It

has been shown, however, that CpG methylation plays an important role in regulating gene expression during development and differentiation (*9, 10*). X-chromosome inactivation is a notable example of developmental long-term gene inactivation due to CpG island methylation (*11*). Repression is believed to be the result of DNA condensation and transcriptional inactivation (*12*).

1.1.2 Methyl Binding Domain Proteins

Hendrich & Bird identified a family of proteins containing methyl binding domains (MBD) capable of specifically binding methylated dsDNA: MBDs 1, 2, 4 and MeCP2 (*13*). The primary amino acid sequence alignment for the conserved MBD domain of these proteins in humans, mice, and frog (*X. laevis*) is shown below in Figure 1.1.

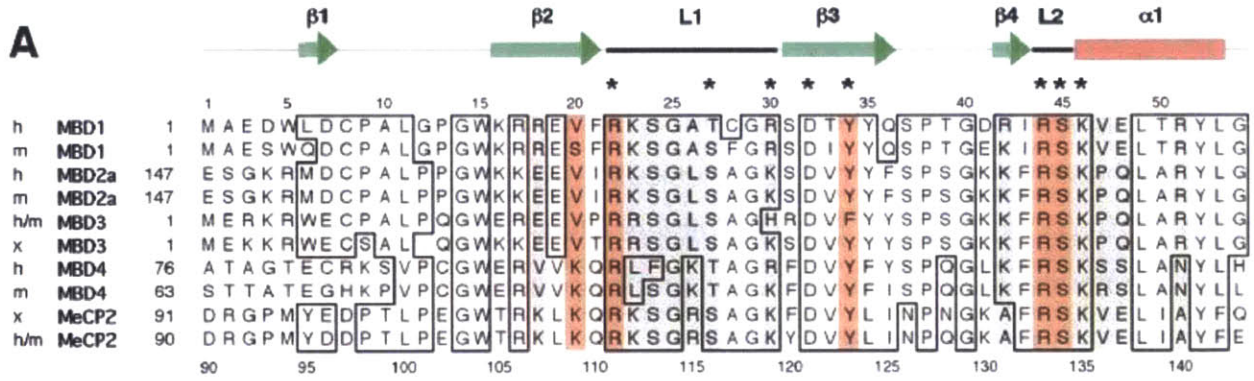


Figure 1.1: Primary amino acid sequence alignment of MBDs 1, 2, 3, 4, and MeCP2 from human (h), mouse (m) and *X. laevis* (x). Boxes designate conserved residues. MBD1 amino acids that interact with the DNA bases and backbone are shaded in red and gray, respectively (*13-15*). Reproduced from Ohki *et al.* (*14*).

MBD proteins have been studied extensively for their critical role in gene regulation in which they transduce DNA methylation into the appropriate chromatin state (*16*). The structures of human MBD1 (*14*), chicken MBD2 (*17*), and human/murine MeCP2 (*18*) have been elucidated using NMR (hMBD1 & cMBD2) and X-ray analysis (h/mMeCP2). MBD proteins generally have a beta sheet and an alpha helix in their secondary structrure, and it is believed that structured water molecules are critical to specifically recognizing methylated DNA (*18*). Binding affinities to DNA with various numbers and densities of methylated CpG dinucleotides have been performed for some MBD family

members (*19*). Murine MBD1 has been concatenated to improve its affinity (*20*) and expressed as an enhanced green fluorescent protein (eGFP) fusion (*21*); however, the MBD family proteins have largely not been engineered and systematically characterized for quantitative, analytical use.

1.2 DNA Methylation in Cancer

In healthy tissue, the majority of unmethylated CpG dinucleotides are found in promoter region CpG islands; however, early studies discovered substantial global hypomethylation in cancer cells is often accompanied by increased CpG methylation in other regions, particularly promoters (*22-25*). Recent work has implicated genomic hypomethylation in tumor formation (*26, 27*). Others have postulated two potential mechanisms for the role of DNA hypomethylation in cancer development: chromosomal instability and activation of normally silent regions of the genome (*22, 27, 28*).

Though the mechanisms differ, promoter hypermethylation results in nearly the same loss-of-function neoplastic effect as coding-region mutations on tumor-suppressor genes (*25*). Following the Knudson two-hit hypothesis for loss of homo/heterozygosity, (*29*) DNA hypermethylation can be responsible for producing the second hit in familial cancer or for both in nonfamilial malignancies (*30*). The retinoblastoma tumor-suppressor gene (*Rb*) was one of the first identified to have promoter region CpG islands that exist hypermethylated in cancer (*31, 32*). Since then, CpG island hypermethylation in the promoter regions for the *VHL* (E3 ubiquitin ligase tumor-suppressor in renal carcinoma), $p16^{INK4a}$ (cell cycle regulating multiple tumor suppressor 1 (MTS-1) in several cancers), *MLH1* (mismatch repair in colorectal cancer) and *BRCA1* (DNA damage repair in breast cancer) genes has been identified as a step in tumorigenesis (*33-36*). In the cases of the genes *MLH1* and *BRCA1* encoding repair enzymes, silencing allows mutations to propagate in the genome which illustrates the intimate relationship between genetic and epigenetic lesions (*28*). Further, Mehrotra et al. showed that genes in both local and distant metastasis frequently exhibit greater CpG methylation than the primary breast carcinoma; specifically, epigenetic inactivation of the *HIN-1* gene causes cells to lose control over cell growth leading to migration and

invasion (*37, 38*). It also appears that, because mutation of the gene is either very rare or absent, hypermethylation of the *CDH1* promoter is a necessary event for tumor invasion and metastasis (*39*).

1.3 Promoter Hypermethylation as a Cancer Biomarker

Hypermethylation analysis of tumor-suppressor CpG islands across a variety of cancers has revealed "hypermethylomes" characteristic of certain tumor types (*28, 40*), and medical research has begun to identify hypermethylated gene promoters that can be used as predictive, prognostic, and diagnostic biomarkers (*41*). These findings in combination with individualized epigenetic analysis could allow oncologists to better manage the care for patients with cancer.

Perhaps the best example of a clinically actionable mark is hypermethylation of the O^6-methylguanine-DNA methyltransferase (*MGMT*) (DNA damage repair in glioma) promoter which has been correlated with glioma tumor sensitivity to the DNA alkylating agents carmustine and temozolomide (*42*). Recent clinical studies have also shown that a subset of breast cancer patients with methylation-dependent silencing of *BRCA1* have tumors sensitive to cisplatin (*43*). In addition to predicting response to therapy, physicians can test for epigenetic silencing of the DNA mismatch repair gene MutL homolog 1 (*MLH1*) which has been correlated with microsatellite instability (MSI) and better prognosis for patients being treated with colon cancer (*35, 41*). As another example, hypermethylation of the genes *DAPK* and $p16^{INK4a}$ has been correlated to decreased survival in lung and colorectal cancer patients, respectively (*44, 45*). Finally, promoter hypermethylation analysis has been proposed and is being used to diagnose patients with cancer. Specifically, the US FDA approved the Cologuard® test in August 2014 which includes promoter hypermethylation analysis of N-myc downstream-regulated gene 4 (*NDRG4*) and the bone morphogenetic protein 3 (*BMP3*) (among other genetic targets) for colorectal cancer screening (*46*). Similar DNA methylation tests for the Septin9 (*SEPT9*) and short stature homeobox 2 (*SHOX2*) genes have been approved in Europe, but not in the US, for diagnosing colorectal and lung cancers, respectively. Hypermethylation at glutathione S-transferase pi 1 (*GSTP1*) has also

shown great promise (particularly when analyzed in combination with *APC* and *PTGS1*) as a biomarker for diagnosing prostate cancer (*47, 48*). More generally, profiling a group of approximately 10-100 genes for DNA hypermethylation with subsequent hierarchical cluster analysis has the potential to improve clinical prognosis predictions and disease classification (*49*). Hypermethylation clustering analysis of ten genes in neuroblastoma tumor samples has already been used to identify clinically relevant risk groups (*50*).

1.4 Detection of Promoter Hypermethylation

Because promoter methylation has been shown to have predictive, prognostic and diagnostic value, there has been great interest in developing methods for DNA methylation detection with increased sensitivity, specificity, and resolution to increase clinical value (*41*) and also for discovery purposes to generate reference methylome data (*51*).

1.4.1 Current Methods

The widely used methods for DNA methylation analysis can be coarsely divided into two categories: high-resolution (i.e. genome-wide) and site-specific. The ends of this spectrum represent a tradeoff between information content and cost (Figure 1.2).

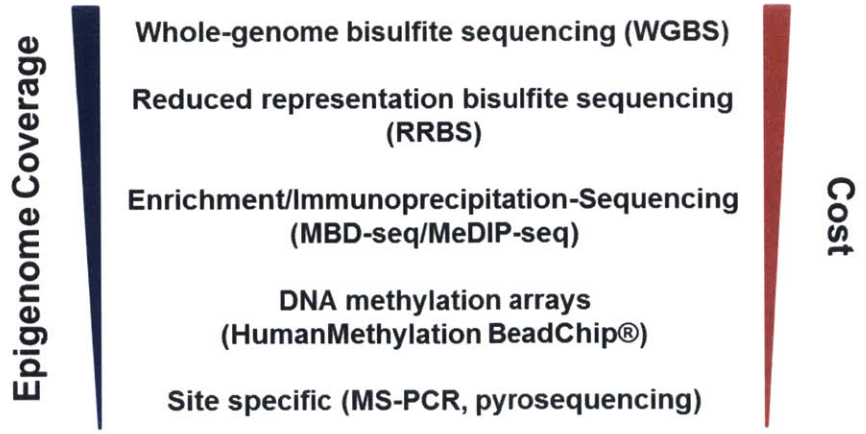


Figure 1.2: Widely used methods for DNA methylation offer a tradeoff between coverage and cost.

The high-resolution methods have been well studied and compared with each other (*52, 53*); these are the *de facto* standard for generating reference methylome data. While critically important for discovery purposes, these methods are generally not well suited for routine analysis of clinical samples where there may only be one to several areas of interest (*41*). I surveyed the twenty US hospitals identified as providing the best cancer care by U.S. News & World Report. At least two of these hospitals, at the time of this writing, do not currently offer DNA methylation testing in its pathology laboratory. Of respondents that do offer epigenetic screening (four institutions), all use some form of methylation-specific PCR (MSP) for routine analysis (*54*). Two also report using massively parallel sequencing based methods on a limited basis for clinical, non-research cases.

Essentially all of the methods currently used for specific detection of DNA methylation rely on sodium bisulfite conversion of unmethylated cytosine bases to uracil by deamination. Chemical conversion alone, however, can degrade more than 90% of the sample DNA (*55, 56*). Such losses incurred during DNA extraction and processing necessitate larger sample inputs on the order of hundreds of nanograms to micrograms of DNA (*57*) which can be present at lower amounts in some bodily fluids i.e. plasma (*58*) or from patients with very little resected/biopsied tissue available. Further, protocols must be assiduously optimized to minimize incomplete deamination of unmethylated cytosine bases and inappropriate conversion of methylated ones to thymine (*59*). Such errors lead to inaccurate results. Consequently, it should not be surprising that three of the four surveyed institutions currently testing for hypermethylation identified the bisulfite conversion step as an aspect of the method they would most like to change or eliminate.

1.4.2 Implications for Clinical and Translational Medicine

Only two gene promoters *MGMT* and *MLH1*, at maximum, are routinely analyzed for promoter hypermethylation at any one of the top twenty US cancer hospitals despite reports that more than sixty hypermethylated genes (at least twenty of which have been correlated with drug sensitivity) have predictive, prognostic, and/or diagnostic value for

cancer (*41*). For example, multiplex analysis of the promoters for the *GSTP1*, *APC*, and *PTGS1* genes has shown great potential for diagnosing prostate cancer (*48*); however, to date, their use has been limited to laboratory and investigational purposes only. One potential cause for this discrepancy is that the majority of these studies are conducted by analyzing primary tumor tissue (where sample amount is much less a concern) whereas many of the most compelling applications require analysis of bodily fluids such as plasma, serum, saliva, urine, and stool where sensitivity and/or specificity have, in general, been low (*60, 61*).

1.4.3 Alternatives to Widely Used Methods for DNA Methylation Analysis

Three general approaches have been devised to address the engineering challenges facing DNA methylation analysis (Figure 1.3). The first strategy for doing so eliminates the need for both bisulfite conversion and PCR amplification of target DNA. All of the methods adopting this strategy rely on a protein affinity agent, either a methyl binding domain (MBD) or antibody, for specific recognition of 5mC; however, each differs significantly in detection mechanism. Two of these involve capturing target DNA via direct hybridization to probe ssDNA on a biochip and detecting methylated CpG sites with MBD proteins. MBD binding is subsequently detected using either surface plasmon resonance (SPR) (*21*) or radical photopolymerization (*62*). SPR is not ideally suited for clinical analysis because it requires instrumentation not typical for most pathology labs. Radical polymerization has the advantage of exponential signal amplification, fast reaction times, and visible readout (*63*). However, the sample volume and required DNA concentrations necessitate more than 10^9 copies of each sequence to be interrogated which equate to gram quantities of tissue to be analyzed. Using a similar scheme, label-free, optical biosensors use MBDs or antibodies immobilized on an opto-fluidic ring resonator (*64*) or a silica optical microtoroid resonant cavity (*65*) to capture DNA and provide specific detection of 5mC and 5hmC, respectively. These methods are not yet capable of providing sequence-specific detection and need to be demonstrated using

clinically relevant sample sizes and types; although, microfluidic device integration is reported to be ongoing. While not fully understood biologically, the ability to discriminate between 5mC and 5hmC will likely further be valuable for basic scientific research. In addition, MBD binding to 5mC in single DNA molecules with fluorescence detection has been demonstrated in a high throughput nanofluidic device (*66*). This method offers simultaneous analysis of other chromatin modifications as well as selection and recovery of single DNA molecules (*67*). Either PCR or sequencing is still necessary, however, for site-specific epigenetic detection.

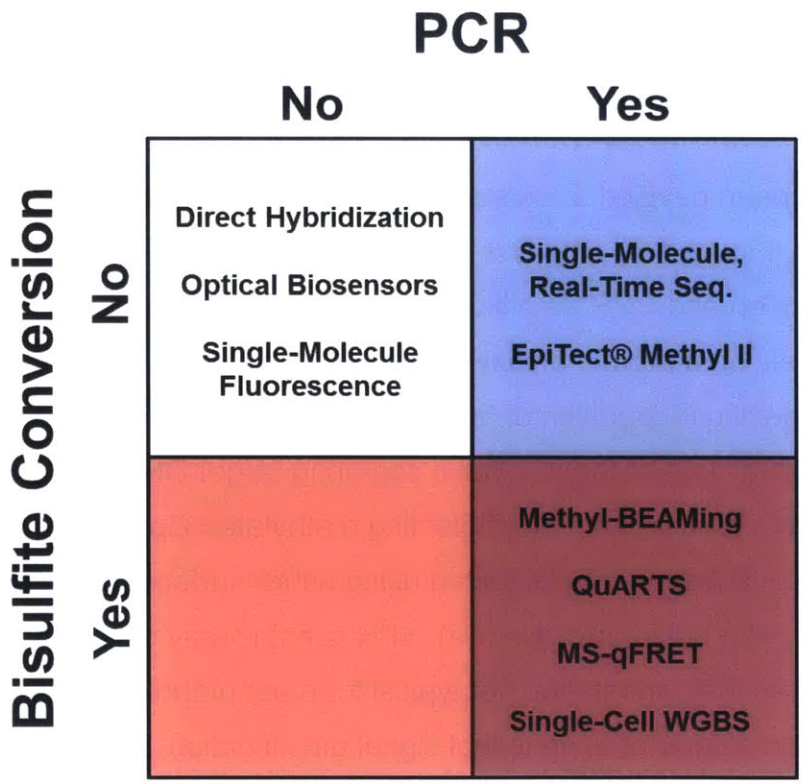


Figure 1.3: Matrix of approaches for DNA methylation analysis. Methods are shown on the basis of whether they require sodium bisulfite treatment of DNA and/or PCR amplification.

A second class of methods have been developed that leverage PCR to amplify unconverted DNA for methylation analysis. Single-molecule, real-time (SMRT) sequencing measures the change in DNA synthesis kinetics that arise when the translating polymerase encounters modified nucleotides in the template strand (*68*).

Primer design can be used to target the region of interest; however, SMRT is only capable of resolving the epigenetic mark 5mA and not 5mC or 5hmC to the single base-pair resolution. Alternatively, Qiagen has introduced the EpiTect® Methyl II PCR Assay which uses a combination of DNA restriction enzymes that cleave specifically in the presence or absence of DNA methylation and real-time PCR to provide quantitative methylation profiling for any of >37,000 promoter CpG islands. Using DNA from at least 10^4 cells (~100 ng), this method has been shown to correlate well with methods that utilize bisulfite conversion.

The third strategy is to mitigate and cope with the deleterious effects caused by bisulfite conversion and amplify the remaining signal to provide sensitive detection of DNA methylation. Many members of this class are capable of providing quantitative results with significantly less input DNA than current methods require. Methyl-BEAMing technology involves two sequential PCR reactions (*69*). The first is analogous to MSP, and the second involves single-molecule PCR on a microparticle in an o/w emulsion. Flow cytometry or sequencing can be used to quantify the number of methylated DNA fragments. While demonstrated using DNA from plasma and stool samples, the specialized bisulfite conversion protocol and lack of multiplex analysis make its clinical applicability uncertain. In a similar approach, MS-qFRET consists of a MSP reaction with forward and reverse primers functionalized with biotin and a fluorophore, respectively (*70*). The fluorescently labeled PCR products and a streptavidin coated quantum dot (QD) form a FRET pair when bound together and indicate DNA methylation with a change in the QD emission spectrum. The sensitivity of this approach has been demonstrated with sputum samples, and it integrates well with traditional equipment and off-the-shelf reagents. In another FRET-based approach, the quantitative allele-specific real-time target and signal amplification (QuARTS) method uses a gene-specific oligonucleotide probe to recognize a MSP amplicon which, through two sequential 5'-flap endonuclease reactions, liberates a fluorescent dye that can be detected in a qPCR like format (*71*). Not only does this approach provide significant signal amplification, it can accommodate multiplexed detection of three separate genes in a single PCR reaction for clinically relevant sample types i.e. stool. At the cutting edge of rare cell DNA methylation analysis, a method has been recently reported for

single-cell whole-genome bisulfite sequencing (72). DNA degradation from bisulfite treatment does reduce genome coverage; however, aggregating low-coverage data from several single-cell analyses has been shown to provide more complete information. While best suited for research applications at this point, this technology has the potential to greatly impact basic research and hypermethylome biomarker discovery.

1.5 Thesis Objectives

Given the potential for DNA methylation analysis to inform oncologists' management of their patient's care, the wealth of existing clinical data, and the ongoing large-scale mapping of reference human genomes by the Roadmap Epigenetics Consortium, the goal of this thesis is to develop clinically useful analytical tools to provide rapid, low-cost and sensitive detection of methylated DNA. In doing so, this thesis seeks to use protein engineering methods to produce methyl-CpG binding domain proteins with improved binding affinity and selectivity for methylated CpG dinucleotides as well as thermal stability. Further, this thesis integrates photopolymerization based signal amplification on a biochip to provide visual, sequence-specific readout of DNA hypermethylation.

The specific aims of this thesis are to:

- Characterize wild-type MBD family proteins and engineer a parent for improved thermal stability, specificity, and binding affinity to symmetrically methylated CpG dinucleotides in DNA using yeast surface display.
- Optimize expression of MBD protein constructs in bacterial cell cultures for rapid and low-cost production and purification of epigenotyping reagents.
- Synthesize and characterize the capability of dual-functional MBD containing conjugates to specifically detect dsDNA methylation patterns in a biochip format and initiate polymerization for visible positive readout.
- Develop a quantitative framework i.e. model of the performance of MBD protein reagents as a function of manipulated variables in epigenotyping assays.

1.6 Thesis Organization

This thesis is organized into six chapters. Chapter 1 provides an overview of DNA methylation as an epigenetic mark, its role in causing cancer, and its value as a biomarker. Chapter 1 also provides a review of the existing methods for DNA methylation analysis (both for research and clinical application), their advantages and drawbacks, and new strategies for DNA methylation analysis. Results of the work on each specific thesis aim are presented in the following chapters: Chapter 2 discuses engineering a high-affinity MBD protein for direct epigenotyping, Chapter 3 summarizes efforts to maximize the soluble expression yield of recombinant MBD proteins, Chapter 4 describes implementation of dual-functional MBD proteins to transduce DNA methylation to an observable readout on a biochip, and Chapter 5 describes a simple, quantitative framework for connecting MBD parameters and their performance in engineering applications. Finally, Chapter 6 summarizes the findings of this thesis and provides direction for future research and development.

Chapter 2

Characterization and Directed Evolution of a Methyl Binding Domain Protein

Abstract

Methyl binding domain (MBD) family proteins specifically bind double-stranded, methylated DNA which makes them useful for DNA methylation analysis. We displayed three of the core members MBD1, MBD2, and MBD4 on the surface of Saccharomyces cerevisiae cells. Using the yeast display platform, we determined the equilibrium dissociation constant of human MBD2 (hMBD2) to be 5.9±1.3 nM for binding to singly methylated DNA. The measured affinity for DNA with two methylated sites varied with the distance between the sites. We further used the yeast display platform to evolve the hMBD2 protein for improved binding affinity. Affecting five amino acid substitutions doubled the affinity of the wild-type protein to 3.1±1.0 nM. The most prevalent of these mutations, K161R, occurs away from the DNA binding site and improves the stability between the N and C-termini of the protein by forming a new hydrogen bond. The F208Y and L170R mutations added new non-covalent interactions with the bound DNA strand. We finally concatenated the high-affinity MBD variant and expressed it in Escherichia coli as a green fluorescent protein (GFP) fusion. Concatenating the protein from 1x to 3x improved binding 6-fold for an interfacial binding application.

2.1 Introduction

The structure of chromatin plays a significant role in gene expression and development for eukaryotic organisms (*73*). Methylation at the 5 position of the cytosine base, when followed by guanine (CpG) in the promoter region of a protein-coding gene, is an epigenetic modification that has been shown to be involved in DNA condensation and transcriptional inactivation (*12*). Aberrant DNA methylation patterns have been implicated in the development of human diseases such as cancer (*74*). Medical research has connected promoter methylation levels for certain genes to therapeutic response in patients. For example, glioma patients with a methylated promoter for the O^6-methylguanine-DNA methyltransferase (*MGMT*) gene exhibit particular sensitivity to alkylating agent chemotherapeutics (*42*), and breast cancer patients with methylation-dependent silencing of the breast cancer 1, early onset (*BRCA1*) gene have been shown to have tumors sensitive to cisplatin (*43*). Additionally, physicians can test for epigenetic silencing of the DNA mismatch repair gene MutL homolog 1 (*MLH1*) for its prognostic value for patients being treated with colon cancer (*35, 41*). Hypermethylation at glutathione S-transferase pi 1 (*GSTP1*) has also shown promise as a biomarker for diagnosing prostate cancer (*47*). Because promoter methylation has been shown to have predictive, prognostic and diagnostic value, there has been great interest in developing methods for DNA methylation detection with increased sensitivity, specificity, and resolution to increase clinical value (*41*) and also for discovery purposes to generate reference methylome data (*51*).

State of the art methods for DNA methylation detection (whole-genome bisulfite sequencing, reduced representation bisulfite sequencing, CpG specific arrays, and methylation-specific PCR) generally rely on sodium bisulfite conversion of unmethylated cytosine bases to uracil (*41*). Chemical conversion, however, can degrade more than 90% of the sample DNA (*55*), and protocols must be assiduously optimized to minimize incomplete deamination of unmethylated cytosine bases and inappropriate conversion of methylated ones to thymine (*59*). Such errors lead to inaccurate results. Alternatively, immunoprecipitation (IP) based methods such as MeDIP-seq and MBD-seq have been developed. These methods tend to require larger sample inputs (*75*) and are not

capable of providing single methyl CpG site resolution without bisulfite conversion (*76*). To avoid bisulfite conversion while still providing improved resolution, there have been several methods developed recently that use the very methyl binding domain (MBD) proteins involved in forming repressive complexes *in vivo* to transduce DNA methylation into a signal that can be measured directly (*21, 62, 66, 67, 77*) instead of simply providing sample enrichment as is the case with MBD-seq. These MBD proteins specifically recognize symmetrically methylated CpG dinucleotides in double stranded DNA (*13, 19, 20*), and therefore, have the potential to enable high resolution DNA methylation detection when paired with sequence specific probe DNA without requiring chemical conversion or sequencing of DNA. Current MBD-based methods require relatively large amounts of DNA (*21, 62, 77*) or are not sequence specific (*66, 67*). Clinical applications require that both these problems be addressed (*41*). A very high affinity MBD protein suitable for interfacial use and capable of recognizing a single methylated CpG site will thermodynamically provide a higher fractional coverage of these sites in DNA (*78*), which is particularly important when the total number of sites may be low. Such a reagent would support ongoing research to make methylation analysis on a single DNA molecule (*66, 67, 79, 80*) sequence specific.

Here, we report the display of methyl binding domains from murine MBD1 (mMBD1), human MBD2 (hMBD2), human MBD4 (hMBD4) and human/murine MeCP2 (h/mMeCP2) on the surface of *Saccharomyces cerevisiae* as a platform for systematically characterizing intrinsic binding properties and engineering variants with improved binding affinity to methylated DNA. We chose the highest affinity wild-type methyl binding domain protein hMBD2 as a parent for directed evolution via error prone PCR (epPCR) and flow cytometry screening. We isolated MBD2 variants exhibiting improved binding to methylated DNA and constructed a homology model of each variant using the published chicken MBD2 structure (*17*) to elucidate the molecular basis of the observed affinity enhancements. We further concatenated this MBD variant as a green fluorescent protein (GFP) fusion to create the highest affinity reagent reported for DNA methylation detection and demonstrated its utility in high-performance interfacial binding applications.

2.2 Materials and Methods

2.2.1 Displaying MBD proteins on the surface of *S. cerevisiae* yeast cells

The cDNA encoding the mMBD1 gene (amino acids 1-75) was PCR amplified from the pET-1xMBD (k19) construct (*20*) from Adrian Bird (University of Edinburgh). The forward 5'-GAC AGC TAG CAT GGC TGA GTC CTG G-3', and reverse 5'-GAC AGG ATC CAG CGT AGT CTG GGA C-3' primer pair was designed to append flanking 5' NheI and 3' BamHI restriction sites. The PCR reaction contained 1X Phusion HF reaction buffer (New England BioLabs), 10 nmol of each dNTP (New England BioLabs), 25 pmol of each primer (Integrated DNA Technologies), 10 ng pET-1xMBD construct, and 1.0 U Phusion DNA polymerase (New England BioLabs) in a final volume of 50 µL. The thermocycling profile was as follows: initial denaturation at 98°C for 30 sec followed by 30 cycles of denaturation at 98°C for 10 sec, annealing at 45°C for 30 sec, extension at 72°C for 30 sec, and a final extension at 72°C for 10 min. The PCR product and pCTCON-2 vector were double digested with NheI-HF and BamHI-HF restriction enzymes (New England BioLabs), gel purified, and ligated using T4 DNA ligase (New England BioLabs). The pCTCON-2/mMBD1 construct was transformed into electrocompetent NEB 5-alpha *E. coli* cells (New England BioLabs). The pCTCON-2/mMBD1 construct was also transformed into EBY100 *S. cerevisiae* yeast cells using the Frozen-EZ Yeast Transformation II Kit (Zymo Research) and plated onto SDCAA agar plates.

The cDNA encoding the hMBD2 gene (AAs 145-213) was PCR amplified from the pMal-c2X-MBD2 construct (*81*) from Indraneel Ghosh (University of Arizona). The forward 5'-TAC AGC TAG CGA AAG CGG CAA ACG-3', and reverse 5'-GAC AGG ATC CCA TTT TGC CGG TAC GA-3' primer pair was designed to append flanking 5' NheI and 3' BamHI restriction sites. The PCR reaction was carried out as described above. The thermocycling profile was as follows: initial denaturation at 98°C for 30 sec followed by 30 cycles of denaturation at 98°C for 10 sec, annealing at 60°C for 30 sec, extension at 72°C for 30 sec, and a final extension at 72°C for 10 min. All other steps were performed as described above.

The *S. cerevisiae* codon optimized (Gene Art—Life Technologies) cDNA encoding the hMBD4 (AAs 76-148) and h/mMeCP2 (AAs 78-162) gene including flanking 5' NheI and 3' BamHI restriction sites plus four nucleotide overhangs were ordered as gBlocks (Integrated DNA Technologies). These DNA fragments were double digested, ligated into pCTCON-2, and transformed separately into both NEB 5-alpha *E. coli* cells (New England BioLabs) and EBY100 *S. cerevisiae* yeast cells. All constructs were verified by sequencing.

2.2.2 Characterizing MBD binding to DNA oligonucleotides with varying methylation patterns

Quantitative equilibrium binding of DNA to yeast displayed MBD proteins was determined using the method described previously (*82*). EBY100 transformed with pCTCON-2/hMBD2 were grown in SDCAA media overnight at 30°C and 250 rpm. After reaching OD_{600} = 2-5, cultures were inoculated to OD_{600} = 1 in SGCAA and incubated at 20°C and 250 rpm for 40-48 h to induce surface display fusion expression. Induced EBY100 were resuspended to OD_{600} = 1 in PBSA (1xPBS, 0.1% w/v BSA). Five-hundred thousand EBY100 cells in PBSA were incubated with pre-hybridized DNA (synthesized by Integrated DNA Technologies) at concentrations ranging from 0.06-100 nM in volumes of PBSA ranging from 2225-200 µL to provide a 10-fold molar excess of DNA relative to the number of surface display fusions assuming 5×10^4 MBD/cell (*82*). The DNA oligonucleotides used for characterizing hMBD2 were derived from the *MGMT* gene as described previously (*21*) and functionalized with biotin on the 5' end of each target strand to facilitate fluorescence labelling (Figure 2.1).

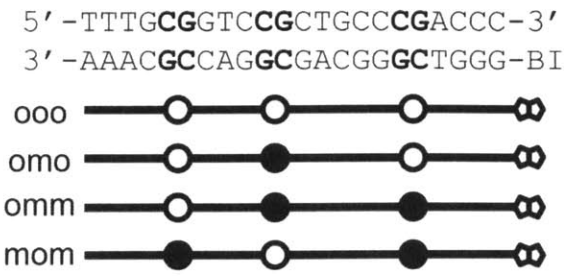


Figure 2.1: Synthetic DNA oligonucleotides derived from the *MGMT* gene. All oligos have the same sequence containing three CpG dinucleotides. The schematic shows the location and number of methylated CpGs for each test oligo. A 5' biotin was appended to one strand to facilitate detection using a streptavidin conjugated fluorophore.

Equilibrium binding was performed at room temperature for 45 min as described previously (*82*). The binding of methylated DNA to displayed MBD proteins was detected using streptavidin, Alexa Fluor® 647 (Life Technologies), and the fraction of EBY100 that expressed the surface display fusions was identified using the chicken anti-cMyc (Gallus Immunotech)/Alexa Fluor® 488 goat anti-chicken (Life Technologies) antibody pair. The dissociation constant (K_d) for each oligonucleotide was determined from an equilibrium binding titration curve fit obtained after plotting the mean fluorescence of the EBY100 cells displaying MBDs versus each DNA concentration (*82*). Each reported K_d value is the average of three biological replicates performed on separate days following the same protocol.

2.2.3 Human MBD2 library creation using error prone PCR

The GeneMorph II Random Mutagenesis Kit (Agilent) was used to perform epPCR on the hMBD2 gene. To affect 1-3 mutations per MBD2 gene (~5-15 mutations/kb), 250 ng of target DNA (7.75 µg plasmid construct) was used as the template for the epPCR reaction. The forward 5'-CGA CGA TTG AAG GTA GAT ACC CAT ACG ACG TTC CAG ACT ACG CTC TGC AG-3', and reverse 5'-CAG ATC TCG AGC TAT TAC AAG TCC TCT TCA GAA ATA AGC TTT TGT TC-3' primer pair (*82*) was used to produce a 367 bp product. The PCR reaction contained 1X Mutazyme II reaction buffer (Agilent), 40 nmol of each dNTP (New England BioLabs), 125 ng of each primer (Integrated DNA Technologies), 7.75 µg pCTCON-2/hMBD2 construct, and 2.5 U Mutazyme II DNA

polymerase (Agilent) in a final volume of 50 µL. The thermocycling profile was as follows: initial denaturation at 95°C for 2 min followed by 30 cycles of denaturation at 95°C for 30 sec, annealing at 58°C for 30 sec, extension at 72°C for 1 min, and a final extension at 72°C for 10 min. The epPCR product was gel purified and amplified using standard Taq based PCR to provide sufficient DNA material for library creation via transformation and homologous recombination in EBY100 yeast cells (*82*).

2.2.4 Library screening for MBD2 variants with improved binding affinity to methylated CpGs

The library was screened by incubating a number of EBY100 cells 10-fold greater than the calculated diversity (*82*). For the first library this corresponded to 2×10^9 cells for a diversity of 2×10^8. After the first round of fluorescence activated cell sorting (FACS), the number of cells screened was 10-fold greater than the number collected from the previous sort. Because the starting hMBD2 K_d was less than 10 nM, the library was enriched for high affinity MBD2 variants using a kinetic screen (*83*). The library was incubated with 100 nM biotinylated omo dsDNA while ensuring a 10-fold molar excess of DNA for 45 min at room temperature in order to saturate surface displayed MBDs with labeled DNA. The cells were then washed, resuspended in PBSA, and incubated with 100 nM unlabeled, competitor omo dsDNA at room temperature to distinguish clones by differences in the degree of labeling due to varying dissociation rate constants and, therefore, binding affinities; concurrently, the cMyc epitope tag of each surface display fusion was labeled with chicken anti-cMyc IgY diluted 1:250. The competition time was determined using the method described previously (*83*) and increased in successive rounds in the range of 90-120 min. The EBY100 population was washed and labeled using streptavidin, Alexa Fluor® 647 and Alexa Fluor® 488 goat anti-chicken secondary reagents (both diluted 1:100) on ice for 15 min. The library was washed and resuspended to a density of 10^7 cells/mL in sterile PBSF for sorting on a MoFlo XDP (Beckman Coulter). Diagonal sort gates were drawn to specify the fraction of the cells collected (Figure 2.2).

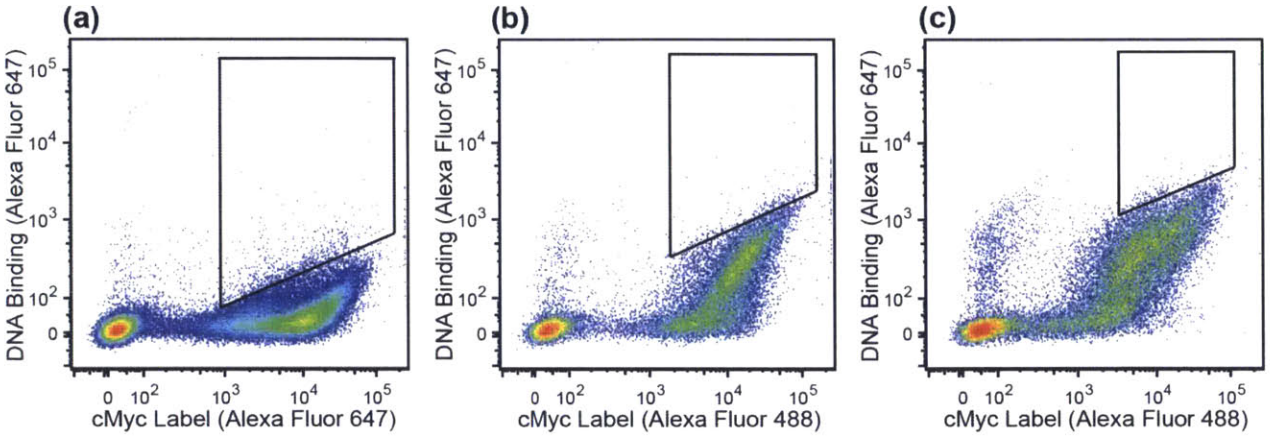


Figure 2.2: Flow cytometry screening of MBD yeast surface display libraries for variants with improved affinity. Dot plots show DNA binding and surface display fusion expression as well as polygonal gates used to isolate clones through directed evolution. (a) Yeast population after the first round of epPCR incubated with 4.5 nM omo DNA. (b) Yeast population during the first flow cytometry screen after the second round of epPCR incubated with 1 nM omo DNA. (c) Yeast population during the second flow cytometry screen after the second round of epPCR incubated with 1 nM omo DNA.

This value was decreased from 5%, to 1%, and to 0.1-0.2% over three consecutive rounds of flow cytometry following the method described previously (*83*). Yeast cells were collected in SDCAA media and subsequently propagated at 30°C and 250 rpm. A tenfold oversampling of the expanded cells was resuspended in SGCAA media for surface display fusion expression and sorting in the next round of screening. After the third round of FACS, the plasmids encoding the MBD2-derived variants were collected using the Zymoprep™ Yeast Plasmid Miniprep II (Zymo Research) and transformed into Mach 1 *E. coli* cells (Life Technologies). Individual clones were isolated and the MBD2 gene was sequenced using the forward primer 5'-CCC CTC AAC TAG CAA AGG CAG-3'.

After screening the first library, the plasmids collected from the final sort were subjected to a second round of mutagenesis by epPCR as described above to create another library with a calculated diversity of 1×10^8. This second library was screened using the same protocol above for the purpose of finding additional mutations giving rise to higher affinity MBD proteins.

2.2.5 Bacterial expression of MBD2 variant proteins

The cDNA for MBD2 variant 2/5 was codon optimized for expression in *E. coli* (Gene Art—Life Technologies) and used to create an MBD-GFP fusion analogous to that reported previously (*21*). The protein consists of an N-terminal His_6-tag followed by the nuclear localization sequence PKKKRKV, the MBD2 variant 2/5, a hemagglutinin (HA) tag, and a C-terminal enhanced green fluorescent protein (GFP) tag. A BsaI restriction site was included immediately preceding the MBD2 variant 2/5 to facilitate concatenation. The cDNA encoding the fusion was synthesized as a gBlock with flanking 5' EcoRI and 3' XhoI restriction sites plus four nucleotide overhangs, double digested, ligated into the pET-30b+ vector, and transformed into Mach 1 *E. coli* cells (Life Technologies). The miniprepped plasmid was subsequently transformed into BL21 (DE3) Tuner *E. coli* cells (Novagen) for expression.

To create the MBD2 variant 2/5 multimer, we designed a second gBlock consisting of the codon optimized cDNA for the MBD followed by the cDNA for a $(Gly_4\text{-}Ser)_2$ linker with flanking 5' and 3' BsaI restriction sites plus six nucleotide overhangs on each end. Both the pET-30b+/MBD2 variant 2/5 plasmid and second gBlock were digested with BsaI (New England Biolabs) and ligated using T4 DNA ligase (New England Biolabs) such that the digested gBlock was in large molar excess. The ligation product was transformed into Mach 1 *E. coli* cells and plated onto LB agar plates supplemented with kanamycin. Individual clones were screened for the number of incorporated MBD variant 2/5 monomer units on the basis of the size of the fragment obtained following double digestion with EcoRI and XhoI. The plasmid encoding the 3xMBD2 variant 2/5-GFP protein was transformed into BL21 (DE3) Tuner *E. coli* cells (Novagen) for expression. The 1x and 3xMBD2 variant 2/5 proteins were expressed (*84*) and purified under denaturing conditions with on-column refolding (*20*) using the protocols described previously.

2.2.6 MBD surface binding experiments and affinity determination

Clear glass slides coated with an agarose film were prepared (*85*) and printed (*62*) with pre-hybridized ooo probe/ooo target, omo probe/omo target, and omm probe/omm target oligonucleotides at 10 µM concentration in 3xSSC as described previously. A circular, 9 mm diameter isolator well was cut from Scotch 3M 665 tape and affixed to the biochip to define each test area. Each biochip was then rinsed under a stream of DI water and blown dry using compressed nitrogen gas. Biochips ready for testing were stored in the vacuum desiccator until needed.

NxMBD proteins were diluted in binding buffer (20 mM HEPES, pH 7.9, 3 mM $MgCl_2$, 10% v/v glycerol, 1 mM dithiothreitol, 100 mM KCl, 0.1% w/v BSA, 0.01% Tween-20, and 1 µM ssDNA) and pre-incubated for 10 min at room temperature. Each 40 µL NxMBD dilution was added to a separate test area and incubated for 40-45 min in a humid chamber at ambient temperature (approximately 20-22°C). Each slide was washed sequentially with 1xPBS/0.1% v/v Tween 20 (PBST), 1xPBS, and 18 MΩ DI water and blown dry using compressed nitrogen gas. The monoclonal mouse HA.11 clone 16B12 antibody (BioLegend) was diluted 1:100 in 1xPBS/0.1% w/v BSA (PBSA), added to each test area, and incubated for 10 min at 4°C in a humid chamber pre equilibrated to temperature. The slide was washed and dried as described previously. The secondary Alexa Fluor® 647 goat, anti-mouse antibody was diluted 1:100 in PBSA, added to each test area, and incubated for 10 min at 4°C in a humid chamber pre equilibrated to temperature. The slide was washed and dried as described previously before scanning with a GenePix 4000B fluorescent microarray scanner (Molecular Devices). Each fluorescence image was analyzed using ImageJ (NIH). The mean fluorescence intensity for each spot was determined by adjusting the threshold of the image to include the entire spot area and averaging the constituent pixel intensities. The values for all spots of the same DNA methylation pattern were averaged and plotted versus the NxMBD concentration in order to fit the data and determine the apparent equilibrium dissociation constant $K_{d,app}$.

2.3 Results and Discussion

2.3.1 Yeast surface display and characterization of methyl binding domain proteins

To establish a platform for characterizing and engineering methyl binding domain family proteins, we separately cloned the cDNA encoding the MBD domain from mMBD1 (AAs 1-75) (*20*), hMBD2 (AAs 145-213 (*13, 81*), hMBD4 (AAs 76-148) (*86*), and h/mMeCP2 (AAs 78-162) (*87*) into the pCTCON-2 yeast surface display vector. Each of these constructs is expressed as a fusion consisting of Aga2p (for yeast cell surface attachment), HA, MBD, and c-Myc (*82*). Display of each MBD was verified by fluorescently labelling the HA and c-Myc epitope tags with Alexa Fluor® 647 and 488, respectively, and flow cytometry analysis (Supplementary Fig. S3). All MBD proteins were successfully displayed on *S. cerevisiae* strain EBY100 except for h/mMeCP2 which exhibited a truncation likely due to either misfolding or proteolytic cleavage prior to being surface displayed (*88*).

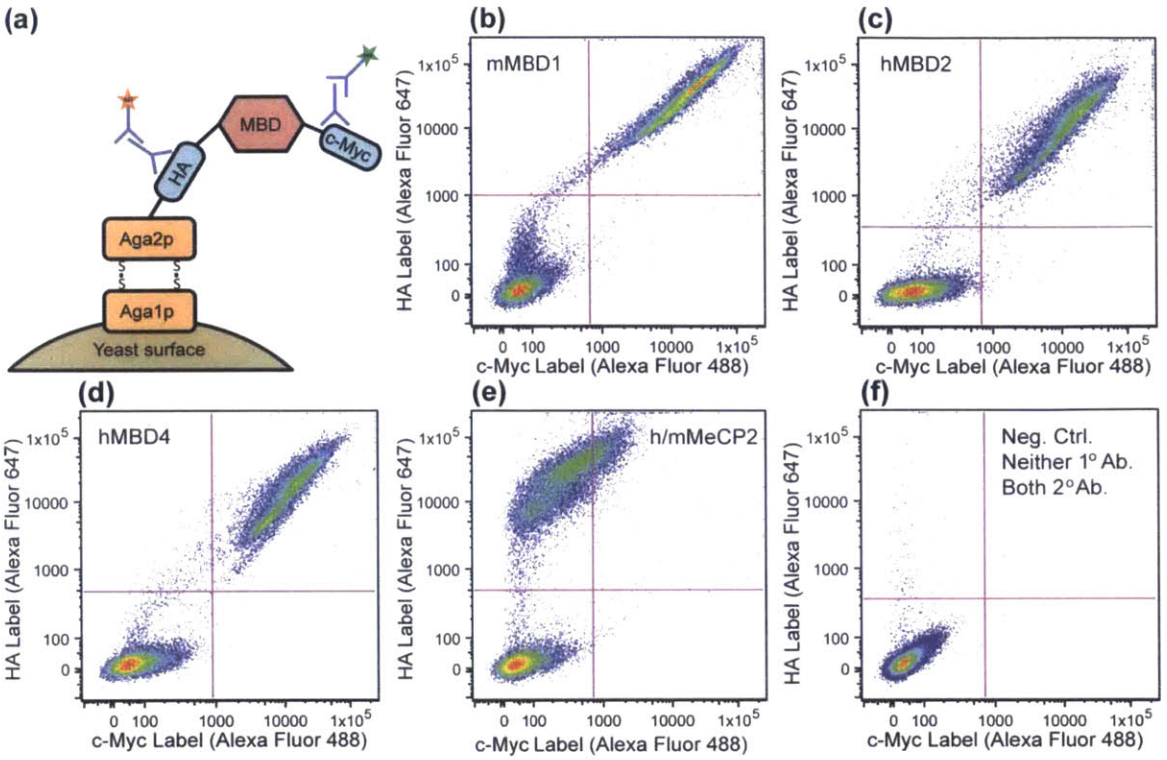


Figure 2.3: Yeast surface display of MBD family proteins. (a) Each MBD protein (red) is displayed as a fusion flanked by N and C-adjacent HA and c-Myc epitope tags (blue), respectively, on the surface of *S. cerevisiae*. Each fusion is anchored to the cell surface by disulfide bridges between the Aga1p and Aga2p proteins (orange). (b) Flow cytometry dot plot of yeast displaying mMBD1, hMBD2 (c), hMBD4 (d), and h/mMeCP2 (e) with their HA and c-Myc epitope tags labeled with Alexa Fluor® 647 and 488 conjugated antibodies, respectively, for detection of expression. The weak signal from h/mMeCP2 c-Myc labelling indicates a truncated surface display fusion. (f) Yeast displaying MBD proteins incubated without either primary antibody but both fluorescently labeled secondary antibodies shows that signal is specific to each epitope tag.

We initially screened the three fully displayed MBD family proteins across a range of methylated DNA concentrations to assess relative binding affinities (data not shown). Subsequently, equilibrium binding titration was used to quantitatively determine the affinity and selectivity of the methyl-CpG binding domain of hMBD2, the highest affinity MBD we displayed. In addition to an anti-c-Myc/Alexa Fluor® 488 antibody pair used to show surface display expression, yeast were equilibrated with biotinylated DNA at various concentrations followed by secondary labelling with streptavidin, Alexa Fluor® 647 (Figure 2.4a). The sequence of the DNA oligonucleotides used in this study was derived from the *MGMT* gene as described previously (*21*) and contains three CpG dinucleotides. Each oligonucleotide was synthesized having one of four different methylation patterns with no, one, or two methylated CpGs (Figure 2.1).

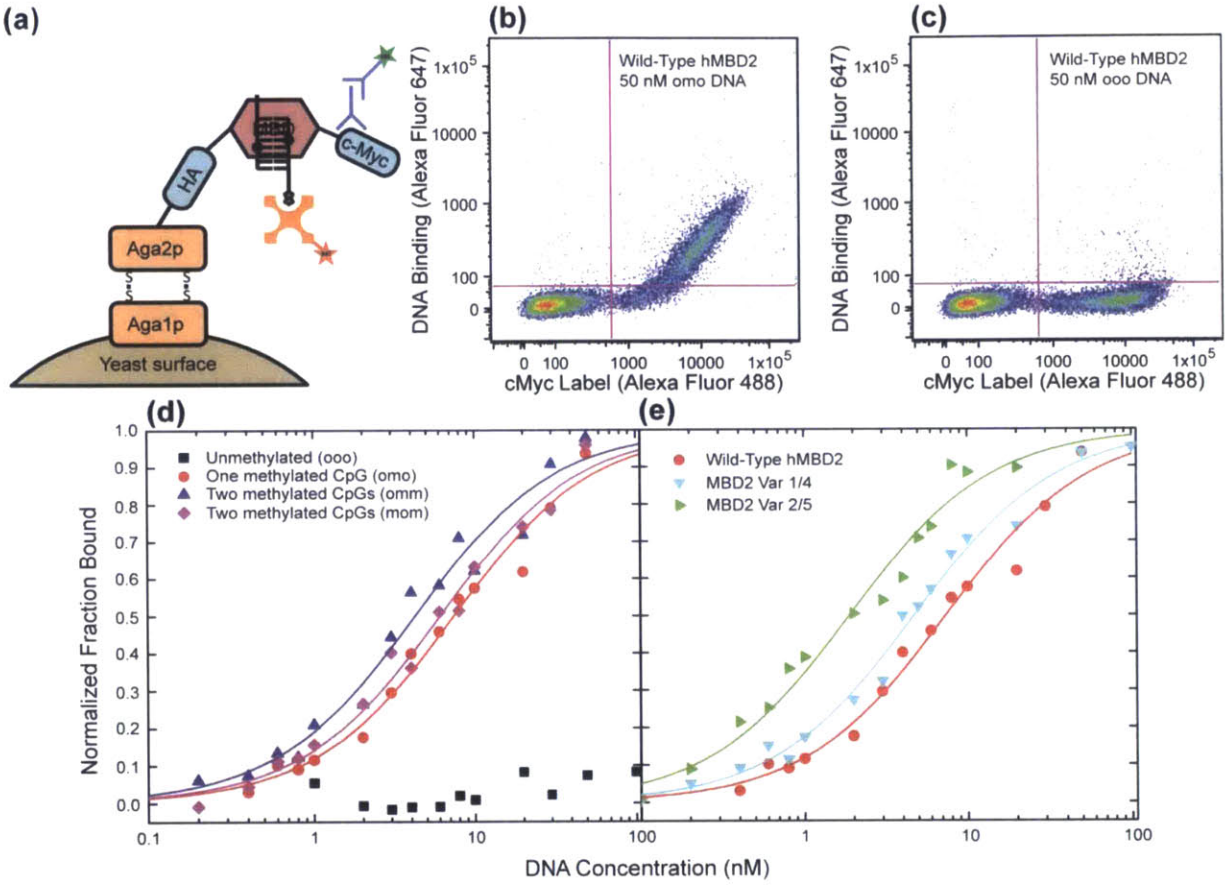


Figure 2.4: Detection and quantification of methylated DNA binding to yeast displayed MBD proteins. (a) Yeast displaying MBD proteins were incubated with biotinylated, methylated DNA and a primary anti-c-Myc antibody followed by labeling with streptavidin, Alexa Fluor® 647 and an Alexa Fluor® 488 secondary antibody, respectively. (b) Flow cytometry dot plot showing 50 nM omo DNA and (c) 50 nM ooo DNA binding to wild-type hMBD2. (d) Equilibrium binding titration curves for determining the affinity of wild-type hMBD2 binding to DNA with various DNA methylation patterns. The mean fluorescence of the displaying yeast population is normalized and plotted versus DNA concentration. Fitting the data yields the equilibrium dissociation constant (K_d) for each oligo. Each reported value (Table I) is the average of three such biological replicates (only one shown). (e) Titration curves for wild-type MBD2, variant 1/4, and variant 2/5 binding to omo DNA. Leftward shift of the binding curve indicates higher affinity binding.

The equilibrium dissociation constant for each oligo was determined by fitting the normalized mean fluorescence versus DNA concentration data for each of three biological replicates (Figure 2.4d & e). Each thermodynamic binding constant is reported as the average and standard deviation of the fit K_d values from these three independent titrations (Table 2.1). We would like to note that others have established fine affinity discrimination (*89*) and the equivalence of equilibrium dissociation constants

determined using yeast display titrations with solution binding measurements using equilibrium competition titration (*90*). The data for hMBD2 binding to unmethylated (ooo) DNA were unsuitable for fitting because saturation could not be achieved even at micromolar DNA concentrations. These data points have only been included for comparison with concentration-matched methylated oligos (Figure 2.4d). Human MBD2's selectivity and weak intrinsic non-specific binding to unmethylated DNA can be seen specifically at one concentration for yeast incubated with 50 nM singly methylated (omo) DNA (Figure 2.4b) and matched unmethylated (ooo) DNA (Figure 2.4c). Interestingly, hMBD2 binds doubly methylated DNA with a consecutive CpG methylation pattern (omm) with higher affinity (K_d = 4.2±1.0 nM) than DNA with an alternating meCpG arrangement (mom) (K_d = 6.5±0.6 nM) ($p<0.05$). The measured affinities for omo and mom DNA are statistically indistinguishable which implies the kinetics of dissociation occur at a similar rate when the methylated CpG dinucleotides are more distant. These results are consistent with previous observations that MBD2 "prefers more densely methylated DNA (*19*)."

Table 2.1: Equilibrium dissociation constants for wild-type hMBD2, variant 1/4, and variant 2/5 binding to DNA with one (omo) or two (omm or mom) methylated CpG sites.

MBD Clone	DNA Methylation Pattern	K_d by titration (nM)
WT hMBD2	omo	5.9±1.3
	omm	4.2±1.0
	mom	6.5±0.6
Var 1/4	omo	4.4±0.4
Var 2/5	omo	3.1±1.0

2.3.2 Affinity maturing hMBD2 using random mutagenesis and flow cytometry screening

A random mutant yeast display library of 10^8 hMBD2-derived clones was created and screened to isolate novel MBD proteins exhibiting increased binding affinity to DNA containing at least one methylated CpG dinucleotide. The library was screened by DNA dissociation kinetics such that clones with reduced off rates retained more biotinylated DNA, exhibited greater fluorescence when fluorescently labeled, and were separated using FACS (*83*). After the first round of epPCR, individual clones were isolated and the gene encoding each MBD variant was sequenced. We found six amino acid substitutions (Table 2.2) which combined to produce five unique MBD variants having one or two mutations each (Figure 2.5a).

Table 2.2: Mutations to hMBD2 and the frequency observed during rounds 1 and 2 of MBD directed evolution by error prone PCR and flow cytometry screening of the yeast surface display library.

	Mutation								
	M150T	**K161R**	**E163V**	**L170R**	**S175I**	**S175R**	**F187I**	**P191R**	**F208Y**
Round 1 Frequency	–	0.8	0.1	–	0.1	0.1	0.1	–	0.2
Round 2 Frequency	0.11	1	–	0.11	0.22	–	0.11	0.11	0.67

(a) Sequences of MBD variants from round 1:

WT: E S G K R M D C P A L P P G W K K E E V I R K S G L S A G K S D V Y Y F S P S G K K F R
S K P Q L A R Y L G N T V D L S S F D F R T G K M

#2: E S G K R M D C P A L P P G W K K E V V I R K S G L S A G K S D V Y Y F S P S G K K F R
S K P Q L A R Y L G N T V D L S S F D Y R T G K M

#3: E S G K R M D C P A L P P G W K R E E V I R K S G L S A G K S D V Y Y F S P S G K K F R
S K P Q L A R Y L G N T V D L S S F D F R T G K M

#4: E S G K R M D C P A L P P G W K R E E V I R K S G L S A G K S D V Y Y F S P S G K K F R
S K P Q L A R Y L G N T V D L S S F D Y R T G K M

#5: E S G K R M D C P A L P P G W K K E E V I R K S G L S A G K I D V Y Y F S P S G K K I R
S K P Q L A R Y L G N T V D L S S F D F R T G K M

10: E S G K R M D C P A L P P G W K R E E V I R K S G L S A G K R D V Y Y F S P S G K K F R
S K P Q L A R Y L G N T V D L S S F D F R T G K M

(b) Sequences of MBD variants from round 2:

WT: E S G K R M D C P A L P P G W K K E E V I R K S G L S A G K S D V Y Y F S P S G K K F R
S K P Q L A R Y L G N T V D L S S F D F R T G K M

#2: E S G K R M D C P A L P P G W K R E E V I R K S G L S A G K I D V Y Y F S P S G K K F R
S K P Q L A R Y L G N T V D L S S F D F R T G K M

#3: E S G K R T D C P A L P P G W K R E E V I R K S G L S A G K S D V Y Y F S P S G K K F R
S K P Q L A R Y L G N T V D L S S F D F R T G K M

#5: E S G K R M D C P A L P P G W K R E E V I R K S G R S A G K I D V Y Y F S P S G K K I R
S K P Q L A R Y L G N T V D L S S F D Y R T G K M

#9: E S G K R M D C P A L P P G W K R E E V I R K S G L S A G K S D V Y Y F S P S G K K F R
S K R Q L A R Y L G N T V D L S S F D Y R T G K M

Figure 2.5: Amino acid sequences of all unique MBD variants isolated from the first (a) and second (b) rounds of epPCR and screening.

One mutation K161R was found in 80% of the clones sequenced. We rapidly screened all five variants for binding to singly methylated (omo) DNA in parallel using reduced resolution equilibrium binding titrations and flow cytometry (Figure 2.6a).

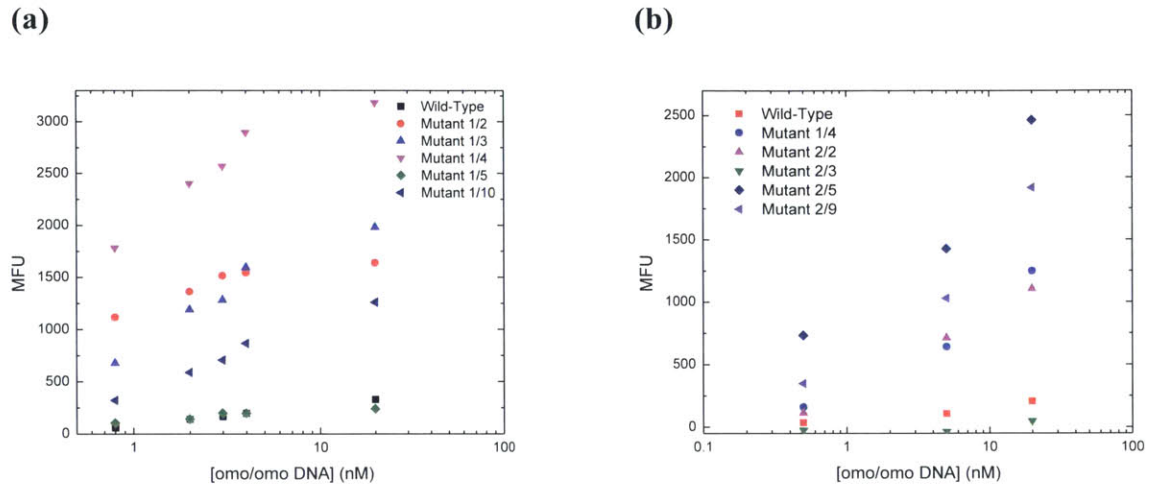


Figure 2.6: A reduced resolution, five-point equilibrium binding experiment was used to quickly screen the relative methylated DNA binding affinities of unique MBD variants relative to wild-type hMBD2 after the first (a) and second (b) rounds of epPCR. The variants from each round with the highest apparent binding affinity were selected for complete equilibrium binding titrations to omo DNA in triplicate to quantitatively determine K_d.

We aligned the sequence of the MBD variant 1/4 having the highest observed binding affinity with the wild-type primary structure (Figure 2.7) and performed complete equilibrium binding titrations to omo DNA in triplicate to quantitatively determine K_d (Figure 2.4e) (**Error! Reference source not found.**). Binding affinity improved approximately 25% from wild-type hMBD2 to 4.4±0.4 nM although not statistically different.

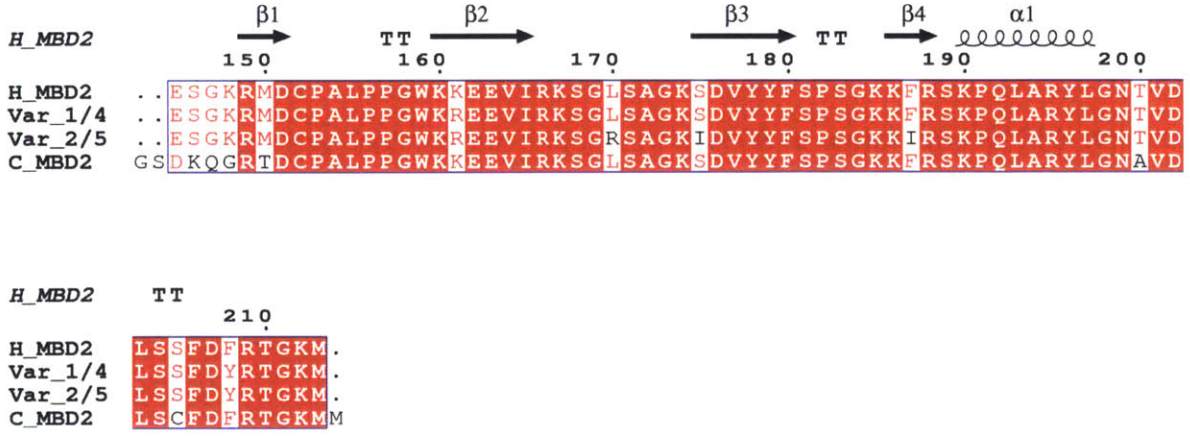


Figure 2.7: Sequence comparison of MBD2 proteins. The MBD variants 1/4 and 2/5 having two and five mutations, respectively, are shown below the wild-type hMBD2 sequence. The sequence of

chicken MBD2 is included for reference, and its secondary structure, determined from previous NMR analysis, is depicted above the sequence alignment.

In order to achieve further affinity enhancement, we implemented a second round of epPCR to mutagenize the cDNA encoding all MBD variants found to have improved K_d from the first round followed by flow cytometry screening as described previously. We observed three new amino acid substitutions following this round of evolution (Table II) as well as new combinations of mutations observed previously. The K161R mutation was present in every variant sequenced, and the F208Y was found in 67% of variants up from a 20% frequency in the first round. The four new MBD variants had two to five mutations each (Figure 2.5b). The highest affinity MBD variant was determined using the rapid flow cytometry screen described above (Figure 2.6b); MBD variant 2/5 contains five mutations (Figure 2.7) and has an affinity (K_d = 3.1±1.0 nM) approximately two-fold greater than wild-type hMBD2 (Figure 2.4e). Given the combination of mutations, this variant may have potentially arisen from recombination of variants 1/4 and 1/5 from the first round of evolution.

2.3.3 Structural modelling for MBD variants with improved binding affinity to meDNA

In order to determine the molecular basis of the observed affinity improvements, we used the SWISS-MODEL system (*91*) and the published chicken MBD2 NMR structure (2KY8 PDB) (*17*) to generate a homology model of the MBD2 variant 2/5. The kinetic library screening method is used to isolate variants with decreased off-rates (*83*). As such, forming new, non-covalent protein-DNA interactions slows the rate of MBD-DNA dissociation and results in improved binding affinity. In the case of hMBD2, mutation of phenylalanine to tyrosine at the 208 position adds a para substituted hydroxyl group to the aromatic side chain which donates a hydrogen bond to the DNA phosphate backbone (Figure 2.8a). Similarly, the L170R mutation restores an ionic interaction between the positively charged guanidinium group and the phosphate backbone of DNA native to MeCP2 but not present in wild-type MBD2 variants (*14, 17*).

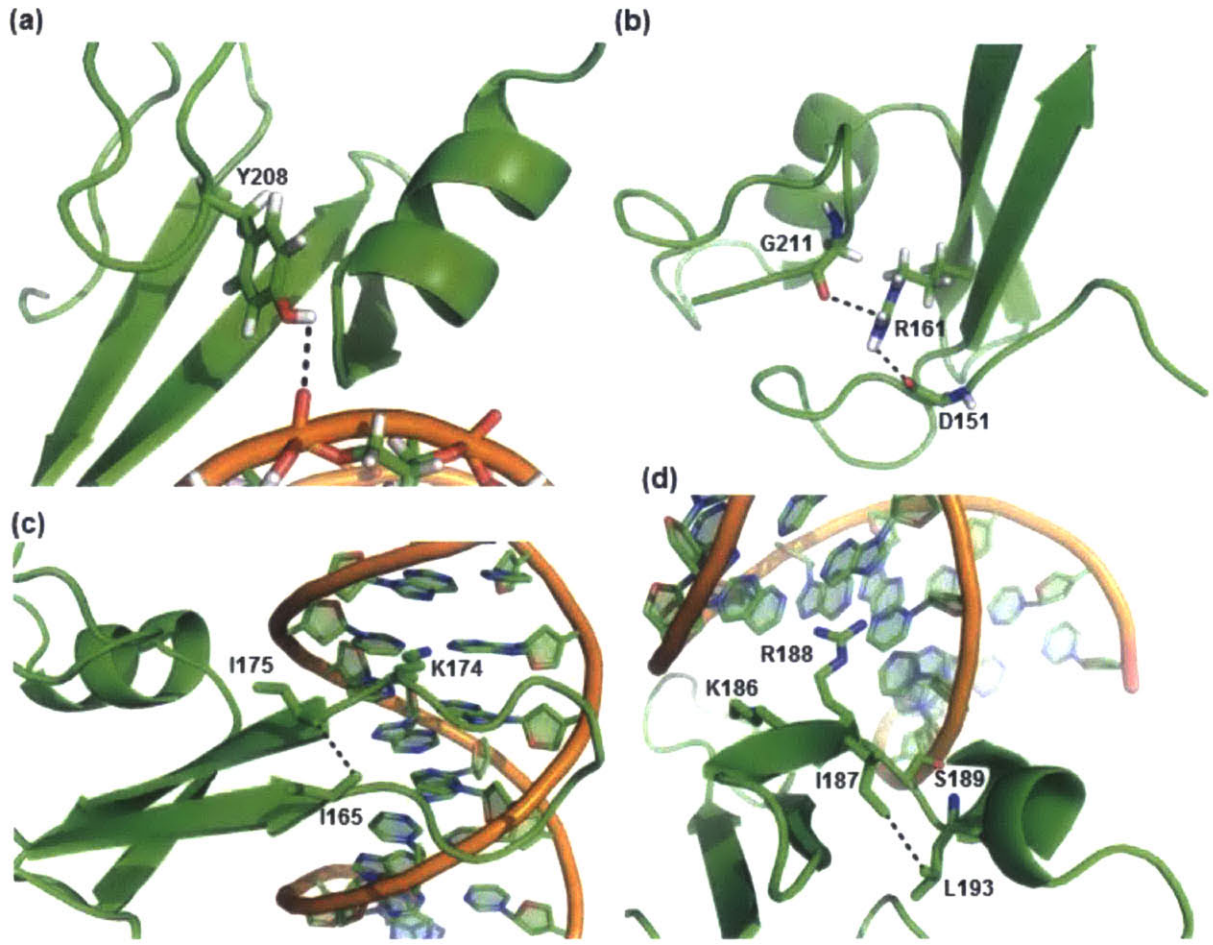


Figure 2.8: Structural analysis of amino acid substitutions in MBD variant 2/5. (a) The addition of the para substituted hydroxyl group of tyrosine relative to the wild-type phenylalanine forms a new hydrogen bond to the DNA phosphodiester backbone. (b) Mutating lysine 161 to arginine introduces a guanidinium group capable of forming an additional hydrogen bond to the main chain carbonyl of aspartic acid 151 in addition to that the wild-type lysine makes to the main chain carbonyl of glycine 211. (c). The side chains of isoleucines 165 and 175 form a hydrophobic interaction at the end of β2 and beginning of β3. (d) Isoleucine 187 and leucine 193 share a hydrophobic interaction between β4 and α1 in the vicinity of three residues lysine 186 (backbone), arginine 188 (bases), and serine 189 (backbone) known to interact with the bound DNA strand.

The frequency of which the K161R mutation is observed, if used as a surrogate for fitness, may indicate it is the most significant residue of those we found affecting MBD binding affinity. Despite being the highest affinity wild-type MBD reported (*19*), MBD2 is the only wild-type human or mouse MBD having a lysine at this position instead of an arginine. The hMBD2 K161 side chain forms a single hydrogen bond between its ε-amino group and the backbone of G211 in the wild-type protein (*17*). Mutating this

residue to arginine substitutes a resonance stabilized guanidium group for the ε-amino which allows for the formation of a second hydrogen bond to the backbone of D151 (Figure 2.8b). Together these two interactions allow R161 to stabilize the N- and C-terminal ends of the protein at the interface with the β sheet (see secondary structure in Figure 2.7). Missense mutation of the homologous residue R106 to tryptophan in MeCP2 has been implicated in the development of Rhett syndrome (*18*). Further, the R106W mutation has been shown to thermally destabilize the motif and reduce the binding affinity to methylated DNA by inducing changes in the MBD secondary structure (*92*).

The two mutations to isoleucine S175I and F187I appear to exist within a similar context in the MBD structure. Both are adjacent to residues known to form base-specific interactions: K174 with the guanine downstream of the CpG and R188 directly with the methylated CpG, respectively (*17*). D176 was also shown to form a CH…O hydrogen bond to the methyl group of 5mC in homologous h/mMeCP2 over a similar distance (~3.5 Å) (*18*). Further, I187 is one member of the four amino acid sequence KIRS in which all three other residues interact with the bound DNA strand. In both instances, the hydrophobic isoleucine side chains are oriented nearly opposite of those interacting with the DNA. The I175 side chain appears to engage in a hydrophobic interaction with I165 at the C-terminal end of the second β-strand (Figure 2.8c). Likewise, I187 forms a similar hydrophobic interaction with L193 residing in the α-helix (Figure 2.8d). The mechanism for the affinity enhancement from these mutations is unclear; however, it may be due to further stabilization of the local MBD structure or that the hydrophobic side chain positioning opposite the DNA-interacting side chains may allow the DNA binding residues greater freedom to form interactions with bound DNA.

2.3.4 Affinity enhancement by concatenation for interfacial binding applications

Starting with a wild-type mMBD1 (K_d=30 μM), others have reported a 60-fold improvement in MBD affinity (K_d=0.5 μM) for singly methylated DNA by concatenating four mMBD1s into a single peptide (*20*). Adopting this established method, we concatenated our highest affinity monomeric MBD variant in order to increase its

probability of forming MBD-meDNA interactions as well as enabling it to form multiple interactions with DNA strands having multiple sites of CpG methylation. We further expressed each MBD multimer as an enhanced green fluorescent protein (GFP) fusion to facilitate fluorescence detection, enhanced soluble expression in *E. coli*, and quantification of purified protein yield by 488 nm absorbance measurement (*21, 84*).

In order to further the development of high-performance, interfacial epigenotyping assays (*21, 62*), we evaluated our NxMBD variants on agarose coated slides (*85*) with immobilized dsDNA having no (ooo), one (omo), or two (omm) methylated CpG dinucleotides. We labeled the bound MBDs with an anti-HA/Alexa Fluor® 647 antibody pair and scanned them (Figure 2.9a). We determined the apparent, equilibrium dissociation constant ($K_{d,app}$) for each NxMBD by plotting the mean fluorescence from each group of spots versus the MBD concentration applied to the test site and fitting a monovalent, equilibrium binding model to the data. The 1x variant was found to bind singly methylated DNA with $K_{d,app}$=19.7±4.2 nM and doubly methylated DNA with $K_{d,app}$=18.0±2.8 nM (Figure 2.9b). Both these values are within each other's error and only show a small improvement in 1xMBD binding to doubly methylated DNA. The 3xMBD variant exhibits an approximately 6-fold improvement in binding to singly methylated DNA with $K_{d,app}$=2.90±0.42 nM and doubly methylated DNA with $K_{d,app}$=3.31±0.48 nM while exhibiting negligible binding to unmethylated DNA (Figure 2.9c).

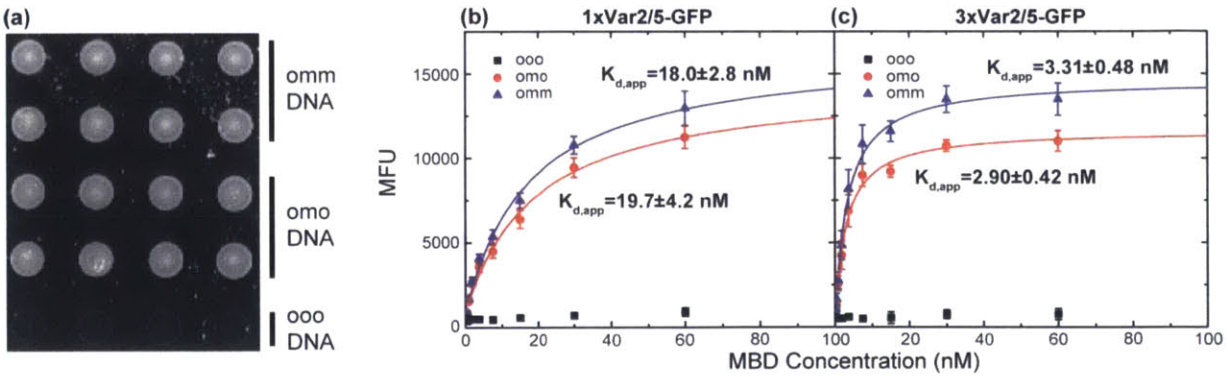


Figure 2.9: NxMBD2-Var2/5-GFP proteins bind to surface-immobilized DNA. (a) Fluorescent scan of 60 nM 1xMBD2-Var2/5-GFP binding to omm, omo, and ooo DNA on a biochip using an anti-HA/Alexa Fluor® 647 antibody pair for detection. (b) Titration curves were fitted to plots of the mean fluorescence from DNA spots versus the concentration of 1xMBD2-Var2/5-GFP (b) or

3xMBD2-Var2/5-GFP (c) applied to the array to determine the apparent dissociation constant $K_{d,app}$ of each reagent for interfacial binding.

Such binding affinity improvements while maintaining specificity allows us to preserve solution-like binding characteristics in a useful interfacial format where surface effects as well as MBD loss during wash steps can reduce the fractional MBD coverage. We can also estimate the fractional coverage of single methylated CpGs as a function of concentration for MBD proteins with varying K_d using a Langmuir adsorption model (Figure 2.10). The MBD proteins described here with single-digit nanomolar dissociation constants (10^{-9} M) can provide fractional coverages several fold higher than other MBDs having K_d values on the order of 100 nM (10^{-7} M) (*20, 21, 66*).

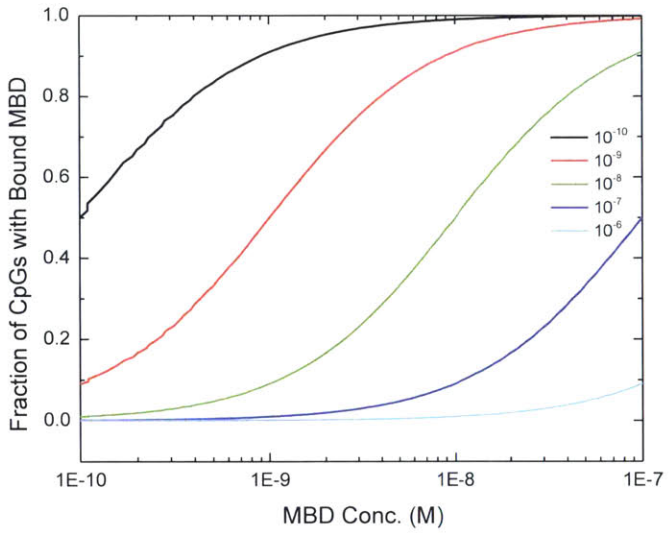


Figure 2.10: Modeling result for the fraction of CpG dinucleotides with bound MBD over various concentrations of MBDs with equilibrium dissociation constants ranging from 10^{-10} to 10^{-6} M (0.1-1000 nM). Higher binding affinities and higher MBD concentrations favor higher CpG fractional coverages.

2.4 Conclusions

We have demonstrated the yeast surface display and directed evolution of the hMBD2 protein. By introducing five amino acid substitutions, we improved the equilibrium dissociation constant approximately two-fold in solution binding experiments. Concatenation of the improved MBD variant from one to three repeat units improved binding affinity to methylated DNA at an interface six-fold while maintaining specificity for methylated CpG sites. As such, these reagents enable single-site methyl CpG resolution when paired with a sequence-specific probe and do not require chemical conversion of DNA. The eGFP and HA tags further allow these MBD proteins to transduce binding to methylated DNA to a directly observable signal which reduces assay complexity, reduces time, and eliminates the need for DNA sequencing.

We also demonstrated how such high-affinity MBD proteins can bind a significantly higher fraction of methylated CpG sites than existing MBD variants. Achieving high fractional coverages will be particularly beneficial for maximizing signal in applications where the total number of methylated CpGs is few e.g. small samples, single DNA molecules, and site-specific analyses. Eventually, these reagents could enable the development of technologies integrating DNA methylation analysis into the current pathology workflow which typically consists of thin tissue sections and fine needle aspirates.

Chapter 3

Expression and Purification Optimization for an MBD-GFP Fusion in *E. coli*

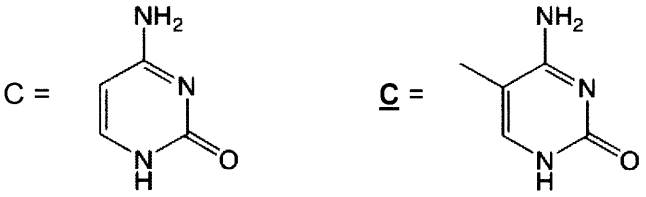
Abstract

DNA methylation is a major epigenetic modification in mammalian cells, and patterns involving methylation of cytosine bases, known as CpG methylation, have been implicated in the development of many types of cancer. Methyl binding domains (MBDs) excised from larger mammalian methyl-CpG-binding proteins specifically recognize methyl-cytosine bases of CpG dinucleotides in duplex DNA. Previous molecular diagnostic studies involving MBDs have employed Escherichia coli for protein expression with either low soluble yields or the use of time-consuming denaturation–renaturation purification procedures to improve yields. Efficient MBD-based diagnostics require expression and purification methods that maximize protein yield and minimize time and resource expenditure. This study is a systematic optimization analysis of MBD expression using both SDS–PAGE and microscopy and it provides a comparison of protein yield from published procedures to that from the conditions found to be optimal in these experiments. Protein binding activity and specificity were verified using a DNA electrophoretic mobility shift assay, and final protein yield was improved from the starting conditions by a factor of 65 with a simple, single-step purification

3.1 Introduction

Methylation of cytosine bases in DNA is an epigenetic modification that has far-reaching and, as of yet, incompletely understood implications in developmental biology (*93, 94*) and in pathology (*95, 96*). In mammals, a large body of work links particular methylation patterns in particular regions of the genome with several subtypes of cancer, including breast, colon, lung, glioma, leukemia, lymphoma, bladder, kidney, prostate, esophageal, stomach, liver, and ovarian cancers (*2, 97-100*). One of these patterns, termed CpG methylation, is described in Figure 3.1 and occurs in the promoter regions of tissue-specific subsets of genes in malignant cells (*40, 99*), whereas these dinucleotide duplexes in the gene promoters of normal cells are largely unmethylated (*101, 102*). Hypermethylation of CpG-rich regions of gene promoters, also known as CpG islands, has been demonstrated to effect the silencing of the downstream coding region; among other gene products, tumor suppressor proteins and DNA repair enzymes are expressed at lower levels as a result of this epigenetic modification (*99*).

(A)



(B)

5'-CG-3'	5'-**C**G-3' or	5'-CG-3'	5'-**C**G-3'
3'-GC-5'	3'-GC-5'	3'-G**C**-5'	3'-G**C**-5'
un-	hemi-		symmetric

Figure 3.1: (A) Cytosine, 5-methylcytosine, and (B) all possible CpG methylation patterns found in dsDNA. Underlining and bold text signify 5-methyl cytosine.

Table 3.1: DNA used in EMSA for MBD activity. Underlining and bold text indicate 5-methyl cytosine.

Methylation pattern	*DNA Sequence*
o_o_o	5'-NH_2-CGGCAAACGGCATCAAACGGCTTTGCGGTCCGCTGCCCGACCC-3'
o_o_o	3'-AAACGCCAGGCGACGGGCTGGG-5'
o_m_o	5'-NH_2-CGGCAAACGGCATCAAACGGCTTTGCGGTC**C**GCTGCCCGACCC-3'
o_m_o	3'-AAACGCCAGG**C**GACGGGCTGGG-5'

Many analytical approaches for detecting methyl-cytosine in DNA have been developed (*103-105*), and several rely on specific binding interactions with proteins. Antibodies raised against 5-methylcytosine (5-meC) are one class of protein that binds single-stranded methylated DNA. Immunoprecipitation using 5-meC antibodies is a widely used method of enriching a DNA sample for methylated sequences (*106-108*); however, this approach does not provide selectivity for CpG nucleotides or a particular CpG motif (hemi- vs. symmetric as defined in Figure 3.1). Methyl-CpG-binding proteins (*109*) are an alternative set of binding molecules that recognize only methyl-cystosine bases in CpG dinucleotides of DNA in its native duplex form (*110-112*). The methyl binding domains (MBDs) (*113*) of larger methyl-CpG-binding proteins (*109, 114, 115*) have been demonstrated to discriminate between unmethylated CpGs, hemi-methylated CpGs, and symmetrically methylated CpGs on the basis of the dissociation constants of each of the above molecular complexes (*19, 116, 117*).

The potential utility of MBDs in molecular diagnostic applications has been suggested in recent studies (*116, 118, 119*). In these proof-of-concept studies, MBD coding regions were excised from the larger mammalian (murine or human) gene, cloned into bacterial expression vectors, and expressed in *E. coli*. Yields of purified proteins were not reported, and some groups used lengthy denaturation-renaturation purification procedures to recover the large fraction of expressed protein in inclusion bodies. For diagnostic applications of MBDs, both in the assay development phase and in the end-use phase, expression and purification procedures that result in maximal protein yields with minimal investment of resources are highly desirable. The present study reports yields for published procedures (*116*) and a series of systematic optimization experiments coupled with a simple, single-step affinity purification that resulted in final yields improved by a factor of 65 over the starting conditions.

3.2 Materials and Methods

3.2.1 Bacterial strain and plasmids

A recombinant pET30(b+) plasmid containing the 1xMBD-GFP construct (*116*) was obtained from Steve Blair's lab (University of Utah, Salt Lake City). This fusion was created by inserting the GFP gene from the pUB-GFP plasmid (Addgene, Plasmid 11155) into the pET30(b+)-1xMBD plasmid (*120*) from Adrian Bird's lab (Edinburgh, U.K.). The sequence encoding the N-terminal His_6-tag and nuclear localization signal (NLS) preceding the cDNA of murine MBD1 (Entrez Gene 17190) amino acids 1-75 remained intact with downstream addition of the GFP gene. Amino acids 1-75 encode the methyl binding domain (MBD) of the larger MBD1 protein. *E. coli* BL21(DE3) cells were transformed with the recombinant 1xMBD-GFP plasmid via electroporation.

3.2.2 Optimization of protein expression conditions using SDS-PAGE and fluorescence microscopy

Protein expression conditions were systematically studied with the goal of maximizing the expression level of soluble protein. For these studies, a 14 mL culture tube (17 x 100 mm, VWR) containing 5 mL of TB medium (12 g tryptone, 24 g yeast extract from BD Biosciences and 4 mL glycerol from BDH dissolved in 900 mL of distilled water and autoclaved, then cooled and brought to 1 L with the addition of 100 ml of a sterile solution of 0.17M KH_2PO_4 and 0.72M K_2HPO_4) supplemented with 50 µM kanamycin was inoculated with 500 µL of overnight culture and incubated at 37°C with orbital shaking. When cultures reached an OD_{600} of 0.6, recombinant protein expression was induced at temperatures ranging from 20-37°C and IPTG concentrations from 0.05-1 mM. Due to slower rates of cellular processes at lower temperatures, longer expression times (in addition to the 3 hour time commonly reported in the literature for 37°C) were investigated for lower temperatures.

In addition to induction of protein expression using IPTG, auto-induction (*121*) conditions were studied. For these experiments, cultures were inoculated according to

the above protocol in Overnight Express™ Instant TB Medium (Novagen) supplemented with 50 uM kanamycin. Cultures were incubated at 37°C with orbital shaking until reaching an OD_{600} of 0.2-0.3. At this point, cultures were either moved to 18°C for 16-24 hours or maintained at 37°C for 8-24 hours. Cells were harvested in 1 mL samples via centrifugation at 12,000 g for 10 minutes. Both SDS-PAGE and fluorescence microscopy were used to compare the relative amounts of soluble and insoluble MBD-GFP protein produced with each set of conditions. For SDS-PAGE analysis, 1 mL cell pellets were lysed with 300uL Bugbuster HT Protein Extraction Reagent (EMD4Biosciences) and treated according to the manufacturer's instructions. The soluble portion was subsequently separated from insoluble cell debris via centrifugation at 12,000 g for 15 minutes. For every condition, approximately 0.25uL of cell pellet (i.e. insoluble fraction) and 10uL of supernatant (i.e. soluble fraction) were each diluted with 10uL gel loading buffer. These sample amounts were chosen because this loading provides a clear resolution of bands in PAGE analysis; intensities in pellet lanes can be meaningfully compared with intensities in other pellet lanes and intensities in lysate lanes may be meaningfully compared with intensities in other lysate lanes. After heat denaturation (95°C for 5 min, ice for 5 min), 20 µL of each sample and 15 µL of a MW standard (BioRad Precision Plus Protein™, Dual Color) were separated on a 4-15% SDS-PAGE gradient gel for 35 minutes at 150 V and visualized with Coomassie Brilliant Blue R250 stain. Culture volumes, lysis reagent volumes, and sample volumes for gel loading were each carefully controlled and held constant for all growth conditions.

In addition to SDS-PAGE analysis, light microscopy was used to visualize the degree of intracellular localization of MBD-GFP. Prior to microscope analysis, 1 mL cell pellets were resuspended and diluted 100-fold in Tris buffer (25 mM Tris-HCl, 150 mM NaCl, pH7.4). Intact cells were examined with both bright field and fluorescence microscopy using an Olympus IX81 microscope equipped with a 60X oil-immersion objective. Fluorescence images of MBD-GFP were acquired using a filter cube comprising Semrock 472/30 excitation, 495 dichroic and 520/35 emission filters. During fluorescence imaging, the lamp intensity setting (10%, Prior Lumen 200) and the image acquisition time (5ms) were each held constant for all samples. Binning was not used.

3.2.3 Protein purification

Larger-scale protein purification studies were carried out for growth conditions previously established in the literature and for the optimal conditions among those described above with the goal of quantitatively comparing yields for each set of conditions. Recombinant protein expression was induced using IPTG at an OD_{600} of 0.6 for two sets of growth conditions: 37°C for 3 hours with 1 mM IPTG in accordance with previous work (*116*) and 20°C for 16 hours with 0.05 mM IPTG in accordance with the results the optimization experiments described above. MBD-GFP expression in auto-induction medium was accomplished by allowing cultures to reach an OD_{600} of 0.2 at 37°C and then moving the cultures to 18°C for 24 hours. Cultures expressed for 3 hours at 37 °C were prepared in a 1L flask containing 250 mL of TB medium supplemented with 50uM kanamycin and inoculated with 5 mL of overnight culture. Because of the high level of expression at 20 °C for 16 hours, purification cultures at 20 °C were prepared in a 250mL flask containing 50 mL of TB medium supplemented with 50uM kanamycin and inoculated with 2 mL overnight culture to prevent column saturation during purification. Auto-induction cultures were prepared in a 250 mL flask containing 50 mL of Overnight Express™ Instant TB Medium supplemented with 50 uM kanamycin and inoculated with 2 mL overnight culture. Cells were harvested by centrifugation at 4,000g at 4°C and subsequently lysed with 4 mL of Bugbuster HT solution following the manufacturer's instructions. The soluble fraction was separated from insoluble cell debris by centrifugation at 12,000g for 15 minutes. The clarified lysate was filtered using a Pall Acrodisc PF Syringe Filter 0.8/0.2 and loaded onto a 1 mL HisTrap™ Fast Flow column (GE Healthcare, USA) for purification on an ÄKTA purifier 10™ FPLC System (GE Healthcare, USA). The column was equilibrated and washed with binding buffer (300 mM NaCl, 100 mM NaH_2PO_4, 10 mM Tris, 10 mM BME, and 10 mM imidazole, pH 7.4), and MBD-GFP protein was eluted using a linear gradient from 0 to 100% elution buffer (300 mM NaCl, 100 mM NaH_2PO4, 10 mM Tris, 10 mM BME, and 250 mM imidazole, pH 7.4) over twenty column volumes (CV) followed by a 5 CV isocratic elution at 100% elution buffer. The fractions containing only 1xMBD-GFP were pooled, concentrated, and buffer-exchanged into storage buffer

(PBS/10% glycerol with 1mM dithiothreitol) using a Millipore 30,000 MWCO spin filter. Protein samples from all stages of purification were separated using a 4-15% SDS-PAGE gradient gel and visualized with Coomassie Brilliant Blue R250 stain. Protein concentration was quantified using the BCA Assay with bovine serum albumin standards (Pierce), and purified 1xMBD-GFP protein was stored at -80°C in single-use aliquots for later use in methylation profiling assays.

3.2.4 Protein activity assay

The CpG-rich target oligonucleotide sequences from the O6-methylguanine DNA methyl- transferase (MGMT) gene and promoter region shown in Table 1 were adapted from those used in (*116*). The abbreviation "m" indicates a methylated CpG while "o" indicates a non-methylated CpG dinucleotide. A primary amine with a six-carbon linker was attached to the 5' end of two oligos to facilitate later attachment to a surface. Further modifications include the addition of a 21 nt spacer sequence to distance the MGMT promoter sequence from the surface in heterogeneous phase assays. All oligonucleotides shown in Table 1 were purchased from Integrated DNA Technologies (Coralville, IA) and reconstituted in 1x TE buffer for long-term storage. Complementary oligonucleotides were diluted to 4 μM in 3x SSC buffer and hybridized using the following protocol: 95°C for 5 min; ramp temperature -1°C/min for 70 minutes; hold at 4°C, with the goal of obtaining 22 bp duplexes with 21 nt tails terminated with a primary amine from the oligos listed in Table 1.

To verify MBD binding activity, an electrophoretic mobility shift assay (EMSA) was employed to detect a bandshift upon MBD binding to methylated dsDNA oligonucleotides (*122*). For each of three DNA methylation motifs—a nonmethylated duplex (ooo/ooo), a hemimethylated duplex (ooo/omo) and symmetrically methylated duplex (omo/omo)—0.5 nmol (22 μg) of purified 1xMBD-GFP protein was pre-incubated in binding buffer (10 mM HEPES, 3 mM $MgCl_2$, 10% glycerol, 1 mM dithiothreitol and 100 mM KCl, pH 7.9) at room temperature for 10 minutes before adding 0.036 nmol pre-hybridized dsDNA. The final concentrations of 1xMBD-GFP and pre-hybridized dsDNA in the 50 μL reaction were 10 and 0.72 μM, respectively. After incubation for 30 minutes

at room temperature, 9 µL of each sample, including three containing only pre-hybridized DNA diluted to 0.72 µM with DEPC treated water, was mixed with 1 µL of Novex TBE Hi-Density Sample Buffer (5x) (Invitrogen, Carlsbad, CA). Samples were loaded onto a 6% polyacrylamide/0.5xTBE DNA Retardation Gel (Invitrogen) with 0.1 µg of a New England BioLabs (Ipswich, MA) Quick-Load®Low Molecular Weight DNA Ladder and run at constant 100 V for 55 minutes. The gel was stained with SYBR Green I in 0.5x Novex TBE Buffer (Invitrogen) for 40 minutes and imaged under 302 nm illumination.

3.3 Results and Discussion

Figure 3.2 presents evidence of overexpression of the MBD from murine MBD1 fused to a GFP tag (MBD-GFP, MW=44kD) by *E. coli*. A systematic study of the effects of temperature and IPTG induction strength on protein expression levels in the insoluble and soluble fractions following cell lysis and centrifugation revealed that temperature is important for maximizing soluble yield and that induction strength in the range 0.05-1 mM is not. Lanes 2 and 3 of Figure 3.2**Error! Reference source not found.**A correspond to the insoluble and soluble fractions, respectively, for previously published expression conditions (*116*). In accordance with previous observations for cultures maintained at 37°C, the vast majority of the protein product is found in the insoluble fraction. Decreasing the concentration of IPTG used to induce protein expression from 1 mM, the published condition, to 0.05 mM did not have a substantial effect on the soluble yield. Figure 3.2**Error! Reference source not found.**B and Figure 3.2C demonstrate a dramatic temperature effect; as temperature is lowered to 30°C (Figure 3.2B) and further to 20°C (Figure 3.2C), the soluble yield increases. At these temperatures, as at 37°C, an IPTG induction level of 0.05 mM was sufficient for overexpression of MBD-GFP, but increasing the IPTG level to 1 mM did not have an impact on soluble yield.

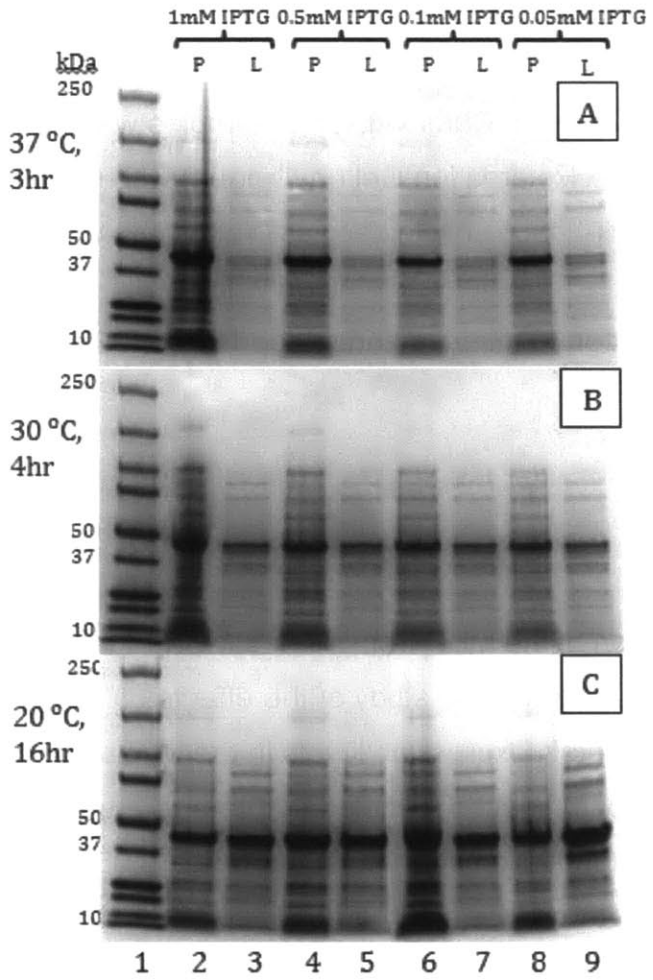


Figure 3.2: Expression study of 1xMBD-GFP at 37 ^{0}C (A), 30 ^{0}C (B), and 20 ^{0}C (C) with SDS-PAGE analysis. Test cultures were expressed at induction strengths of 1mM, 0.5mM, 0.1mM, and 0.05mM IPTG for each temperature. Cultures grown at 30 and 20 ^{0}C were expressed longer than 3 hours to accommodate slower growth at lower temperatures. Soluble fractions were separated from insoluble cell debris via centrifugation after lysis, and samples from cell pellet (P) and crude lysate (L) were treated with gel loading buffer and separated side-by-side on a 4-15% SDS-PAGE gradient gel.

SDS-PAGE analysis of protein expression using auto-induction medium is presented in Figure 3.3. A comparison of soluble protein levels for auto-induced cultures at 37°C with IPTG-induced cultures at 37°C shows that the soluble level of the target protein and all other soluble cellular proteins is higher for auto-induced cultures (i.e. soluble protein yield is increased due to a higher culture density). The trend of increasing soluble expression level with decreased culture temperature was also

evident; decreased culture temperature increased the soluble yield of MBD-GFP produced in auto-induction medium.

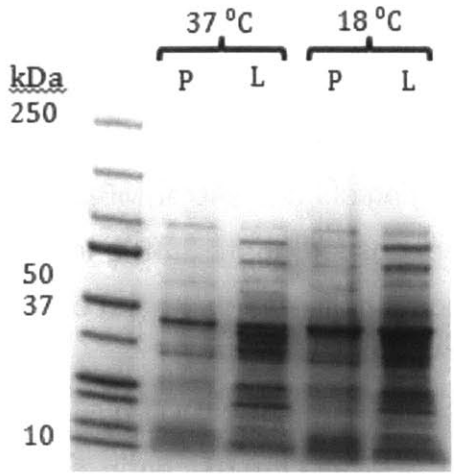


Figure 3.3: SDS-PAGE analysis of auto-induction expression of MBD-GFP at 37 °C and 18 °C. Test cultures (5mL) were grown to an OD_{600} of 0.2 at 37 °C with orbital shaking and either moved to 18 °C or kept at 37 °C for 24 hours. Cells were harvested and treated according to the same protocol as the IPTG induced expression samples.

In addition to analysis by SDS-PAGE, expression level and the intracellular localization of MBD-GFP proteins in aggregates was observed using fluorescence microscopy. Figure 3.4 presents paired bright field and fluorescence images of intact *E. coli* cells and line intensity scans across representative cells for cultures maintained at 37°C (Figure 3.4A-C) and at 20°C (Figure 3.4D-F). Comparisons of panel A with panel B and of panel D with panel E show that all of the cells that are visible in bright field imaging mode express MBD-GFP. Cells are not uniformly oriented with respect to the surface in these images; some cells are standing on end (perpendicular to the surface) and others are oriented flat and parallel with the surface. Cells with the latter orientation were chosen for intensity as a function of distance analyses. Line intensity scans, shown in Figure 3.4C and Figure 3.4F, demonstrate that the MDB-GFP expression level is much higher for cultures maintained at 20°C. Fluorescence from samples grown using the two sets of conditions was imaged in a single session with a constant lamp intensity setting and image exposure time. Signal to noise ratios, defined as the average signal within a cell minus the average background with this difference divided by the

standard deviation of the background, are on the order of 10 for protein expression at 37°C and on the order of 100 for protein expression at 20°C. In addition, the recombinant protein is more uniformly distributed throughout the cell in 20°C cultures. Fluorescently imaged cells from cultures maintained at 37°C have a punctate appearance and the line intensity scan of a representative cell oriented parallel to the surface exhibits more pronounced peaks in comparison with that for a similarly oriented cell representative of growth at 20°C. For accurate visualization of intracellular MBD-GFP localization, it was necessary to use different y-axis ranges for Figure 3.4B and Figure 3.4E. On a 0-4000 intensity scale, Figure 3.4B appears uniformly dark, and on a 0-400 intensity scale, the fluorescent signal in Figure 3.4E appears saturated. To summarize our findings for expression of recombinant MBD from murine MBD1 using the indicated bacterial expression vector and strain, SDS-PAGE and fluorescence analysis indicated 20°C as the optimal temperature of those investigated for maximizing soluble MBD-GFP expression.

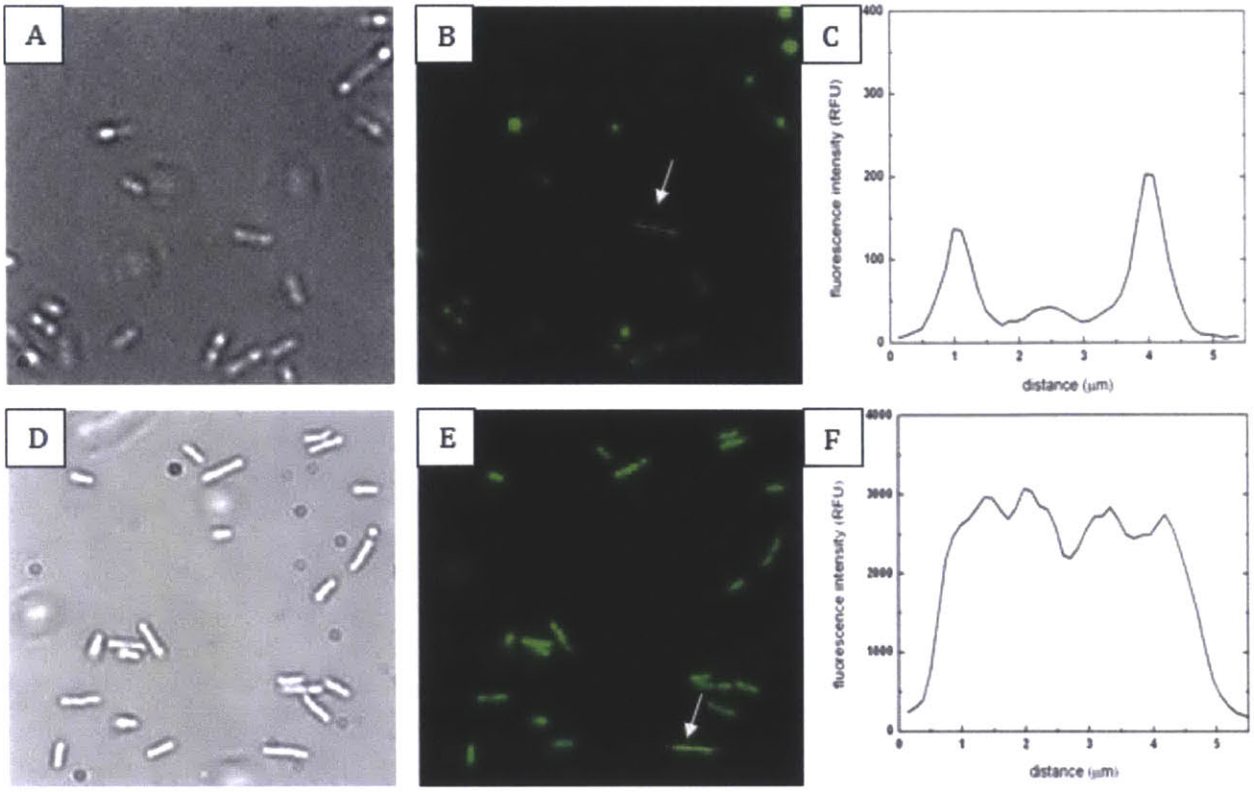


Figure 3.4: Microscopy comparison of MBD-GFP expression at 37^{0}C for 3 hours with 1mM IPTG (literature conditions) to expression at 20^{0}C for 16 hours with 0.05mM IPTG (optimal conditions, this study). Bright field images of expression at 37^{0}C (A) and 20^{0}C (D) were compared to fluorescence images of the same cells (B and E, respectively). Fluorescence images were acquired consecutively with an identical lamp intensity and exposure time. (B) is displayed on a scale of 0 to 4000 and (E) is displayed on a scale of 0-400. On a scale of 0-4000, (B) appears uniformly dark. Line-scans of cells expressing MBD-GFP at 37 ^{0}C (C) and at 20^{0}C (F) show relative fluorescence across the length of the cells. Note the change in y-axis scale in (C) and (F). Line-scans shown are representative of the samples.

Previous studies of the effect of temperature on the soluble recombinant yield of other mammalian proteins have been performed. In many (*123-125*) but not all (*126*) cases, decreasing the expression temperature increased soluble yield. In some cases, decreasing the temperature increased the soluble yield for short (~3 hour) expression times but not for longer (18-24 hour) times (*126*). These results have been interpreted in terms of a number of different molecular mechanisms. The initial increase in soluble yield may be understood in terms of the slower protein synthesis rate at lower growth temperatures and a better match with the capabilities of cellular protein folding machinery. For some proteins, decreased soluble yield at low temperatures with longer expression times has been explained via concentration-dependent aggregation

phenomena among correctly folded proteins due to hydrophobic interactions. MBD-GFP remained soluble as its concentration increased within the cell for 16-hour expression times at 20°C. Longer expression times (up to 20 hours) did not appreciably change the soluble yield.

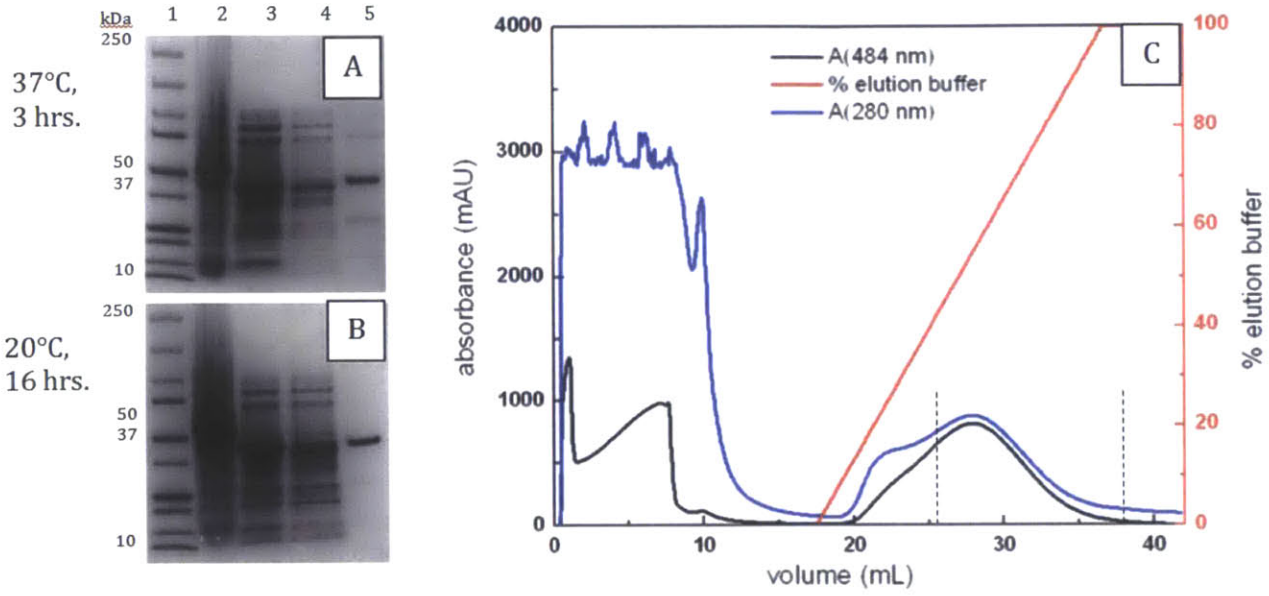


Figure 3.5: Comparison of IMAC purification of MBD-GFP produced at the starting (A) and optimal (B, C) conditions. Lane 1 contains MW standards, Lane 2 contains samples of the insoluble fraction, Lane 3 contains clarified lysate, Lane 4 contains a sample of the flow-through during column loading, and Lane 5 contains 1.4 μg of the purified product (according BCA assay results). For the 37°C condition (A), both the lysate and flow-through were undiluted. For the 20°C condition (B), the lysate was diluted 3X and the flow-through was undiluted. (C) Chromatogram for the purification represented in (B) with pooled fractions indicated using dashed lines.

In order to obtain purified proteins and quantify yield at both the starting and the optimized conditions, expression cultures were scaled up from 5 mL to 250 mL (37°C) and 50 mL (20°C) and the N-terminal His_6-tag in the MBD-GFP construct was utilized in a single-step IMAC purification (i.e. without denaturation/renaturation). Figure 3.5 shows purification gels for 37°C (Figure 3.5A), for 20°C (Figure 3.5B) and a chromatogram (Figure 3.5C) for the purification represented in Figure 3.5B. Purified protein samples were quantified by BCA assay, and lane 5 of each gel was loaded with an equal amount of protein according to the BCA assay results. Preparations using the optimized, 20°C expression conditions were of higher purity. For previously published expression

conditions (Figure 3.5A) and the linear gradient IMAC purification described herein, BCA assays suggest a yield of 0.9 ±0.1 mg/L of MBD-GFP compared with 58.5 ±0.7 mg/L for the optimized conditions. The highest purified yield from the soluble fraction obtained using auto-induction media was 8.3 ±0.2 mg/L (see Figure 3.6). The 75 amino acids comprising the MBD domain account for 19.3% of the molecular weight of the entire construct (8.5 kDa out of 44.2 kDa). Thus, the yield of the MBD domain alone increased from 174 μg/L for previously reported literature conditions to 1.6 mg/L for auto-induction medium and even further to 11.3 mg/L for induction using IPTG at the optimal temperature among those investigated. We deemed this increased yield suitable for methlyation profiling assays since a single batch provided sufficient protein for initial verification of binding activity and hundreds to thousands of subsequent methylation profiling assays following initial protein activity quality control assays.

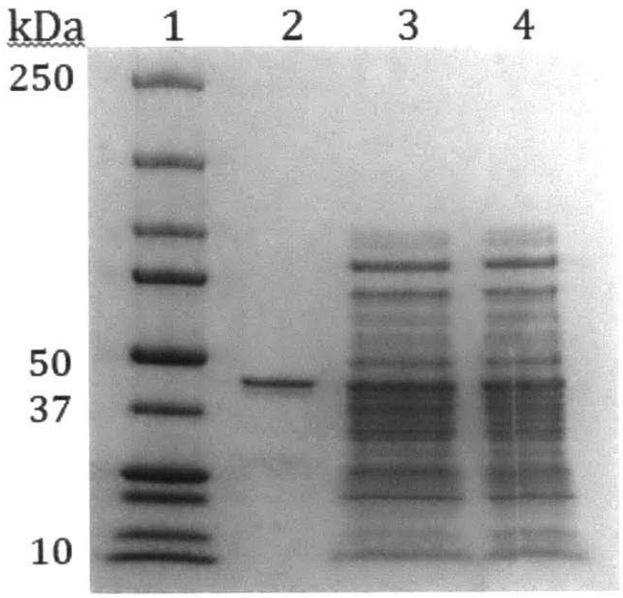


Figure 3.6: IMAC Purification of MBD-GFP after Auto-induction for 24 hours at 18 ^{0}C. Lane 1 contains MW standards, Lane 2 contains pooled MBD-GFP fractions, Lane 3 contains IMAC flow-through, Lane 4 contains clarified lysate diluted 5X.

Figure 3.7 presents the results of an EMSA of methylated-DNA binding activity. The standards in lane 1 of this gel are DNA fragments, and the gel is stained for DNA. DNA bound by a protein runs more slowly than unbound DNA and thus, protein-DNA

complexes are localized near the top of the gel. Lane 4 of this EMSA gel demonstrates that MBD-GFP expressed and purified according to the above conditions preferentially binds a single symmetrically methylated CpG site (omo/omo, sequences in Table 3.1) as indicated by the band that ran similarly to the 300 bp standard and the lack of a bright band between 25 and 50 bp. At the concentrations specified in the Methods section, MBD-GFP does not bind an ooo/ooo duplex (lane 2) or a hemimethylated ooo/omo duplex (lane 3) to the same extent that it binds an omo/omo duplex as indicated by the DNA band between 25 and 50 bp and the lack of a band near the 300 bp standard. We note that all of the SYBR-stained oligo bands exhibit decreased intensity in the presence of MBD-GFP, and that the hemimethylated duplex between 25 and 50 bp is intermediate in intensity when compared with the much dimmer omo/omo duplex between 25 and 50 bp (most of which has been bound by MBD-GFP and runs similarly to the 300 bp standard) and the slightly brighter ooo/ooo duplex that was observed in between the 25 and 50 bp standards.

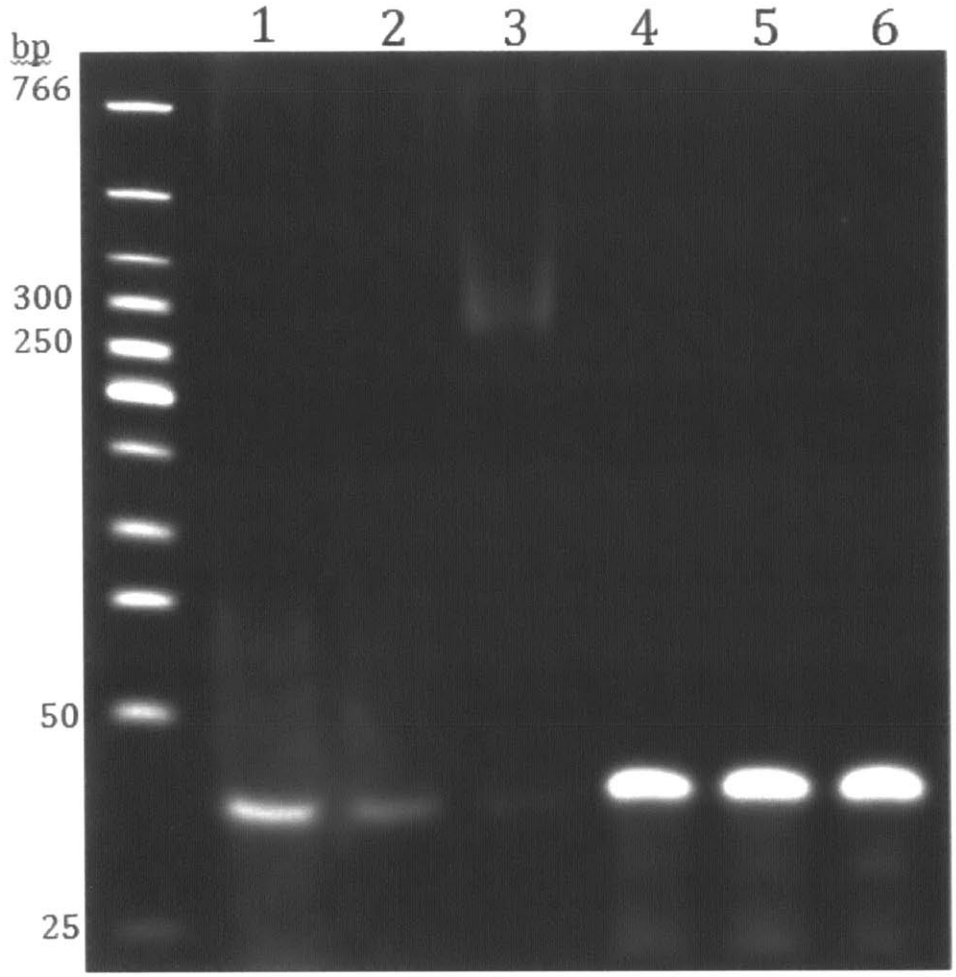


Figure 3.7: Electrophoretic mobility shift assay (EMSA) using purified MBD-GFP and methylated DNA oligonucleotides. MBD-GFP was incubated with pre-hybridized dsDNA of known methylation patterns. Test samples included MBD-GFP incubated with ooo/ooo (1), ooo/omo(2), and omo/omo (3) hybridized duplexes. Pre-hybridized dsDNA oligos ooo/ooo (4), ooo/omo (5), and omo/omo (6), were run as controls.

The binding affinities of MBD-GFP for unmethylated, hemimethylated, and symmetrically methylated CpG sites in DNA have been previously reported (*116*). The K_d values are 72 ±5 μM, 12 ±5 μM, and 0.244 ±0.002 μM for the particular unmethylated, hemimethylated, and symmetrically methylated sequences, respectively, that were used in the present study (Table 3.1). The presence of the GFP tag did not interfere with DNA binding as evidenced by surface plasmon resonance (SPR) measurements. These K_d values provide guidance regarding suitable practical applications of the MBD-GFP construct. For example, in order to use this protein in a

pull-down assay, the concentration of methylated CpG sites in the sample would need to be above ~0.2 μM and the concentration of unmethylated CpG sites should be below ~70 μM in order to obtain enrichment of methylated DNA. This requirement may be met with certain clinical specimens (DNA isolated from biopsy samples, 5 μg or more) and not with others (much lower abundance circulating, cell-free DNA). MBD constructs with superior binding affinities for methylated DNA have been reported that are more suitable for highly sensitive analyses though each has tradeoffs in terms of production. Multimerization of the same MBD gene we have worked with here yielded a 50-fold improvement in the dissociation constant as described in (*120*) though a decrease in the expression level was observed. The MBD domain from human MBD2 has been expressed in a bivalent, antibody-like construct using insect culture techniques, and this recombinant protein was successfully used to pull down candidate genes from as little as 160 pg-10 ng of restricted genomic DNA(*111*). One of our current research directions and a topic of forthcoming reports is the use of evolutionary protein engineering techniques to produce higher affinity MBD-GFP proteins that may be produced efficiently in *E. coli*. In these experiments, the GFP tag generates a convenient signal that may be used to normalize binding data for expression level.

3.4 Conclusions

We have presented an optimization study of the heterologous expression of a mammalian methylated DNA-binding protein domain in *E. coli* and provided conditions for the high-level soluble expression and single-step purification of this protein. SDS–PAGE analysis and microscopy techniques were used to qualitatively and quantitatively compare the soluble yield for expression conditions that vary in temperature, IPTG induction level and time. When compared to the standard conditions presented in Ref. [26] and elsewhere, optimal expression conditions set forth in this paper in combination with a single-step IMAC purification resulted in a 65-fold increase in purified 1 MBD-GFP yield (58.5 ± 0.7 mg/ L). This advance will aid efforts to utilize MBD proteins in methylation profiling assays with minimal investment of resources.

Chapter 4

Hybridization-Based Epigenotyping Using an MBD-GFP Protein

Abstract

Hypermethylation of CpG islands in gene promoter regions has been shown to be a predictive biomarker for certain diseases. Most current methods for methylation profiling are not well-suited for clinical analysis. Here, we report the development of an inexpensive device and an epigenotyping assay with a format conducive to multiplexed analysis.

4.1 Introduction

Accurately diagnosing and characterizing cancer is critical to improving patient outcomes. Medical research coupled with analytical tests has started to allow physicians to identify cohorts of patients and deliver therapy developed for their particular needs. For example, markers such as BRCA1 and HER2 are routinely used for screening, prognosis, or treatment selection (*127*). In addition to somatic mutations, research over the past decade has shown that epigenetic disruptions in gene expression, particularly tumor suppressors, can lead to neoplastic disease (*22*). Current indications show it is likely the FDA will soon evaluate DNA methylation profiles, specifically in the context of the genes PITX2 (*128*) (breast cancer) and SEPT9 (*129*) (colon cancer), for use as cancer biomarkers because such diagnostics have the potential to provide additional information to that available from the current methods used in the clinic.

In mammals, enzymatic methylation of the cytosine base occurs only in the context of the dinucleotide sequence CpG, that is, when cytosine is sequentially followed by guanosine on the same DNA strand (*22*). The human genome has short stretches of DNA with a high density of CpG dinucleotides termed "CpG islands." Many CpG islands are coincident with gene promoters (*130*). These CpGs are usually unmethylated (*131*). However, in colorectal cancer, human mutL homolog 1 (*hMLH1*) promoter hypermethylation has been shown to contribute to genetic hypermutability as a consequence of inactivated DNA mismatch repair function (*35*). Others have correlated promoter hypermethylation to patients' responsiveness to select chemotherapeutics. For example, a subset of glioblastoma patients with silenced *MGMT* is particularly responsive to treatment with alkylating agents because there is no active enzyme to remove alkyl adducts from DNA (*42*). CpG island hypermethylation has further been shown to be unique to the tissue of origin (*132*) which adds to the promise of DNA hypermethylation profiles as cancer biomarkers.

Methylation-specific PCR (*54*) is the *de facto* standard for methylation profiling in diagnostic settings where the genes of interest are known. Methylation-specific PCR, however, requires sodium bisulfite conversion of unmethylated cytosine bases to urasil

in the target DNA. Chemical conversion is costly, and the reaction alone requires 12-16 h. Methods must be assiduously optimized to minimize incomplete deamination of unmethylated cytosine bases and inappropriate conversion of methylated ones to thymine; both errors can lead to inaccurate results (*59*). The degenerate genomic sequence following conversion can also constrain the primer sequence used for downstream PCR amplification (*133*). Further, readout is not direct. The converted DNA must be amplified using a separate PCR reaction for each CpG interrogated and analyzed by gel electrophoresis making it unsuitable for multiplexing. Luo et al. proposed an alternative in which an MBD protein facilitates detection (*77*). However, this format necessitates not only the bisulfite conversion and PCR steps required for methylation-specific PCR but also enzymatic remethylation of the DNA. Recently, Yu et al. reported using surface plasmon resonance (SPR) to detect MBD binding to methylated DNA (*21*). While requiring fewer steps, this method is also less amenable to multiplexed detection and the reliance on ProteOn™ GLC Sensor Chips (BioRad) significantly increases the cost per test.

Here, we report the development of a simple method for detecting methylated DNA fragments from the *MGMT* gene promoter. A defining feature is that target oligonucleotides from the test sample hybridize directly to capture probes printed in 300 μm diameter spots on an inexpensive biochip without requiring bisulfite conversion. We detect methylated DNA duplexes using an MBD protein engineered to facilitate detection using either fluorescence or photopolymerization-based signal amplification with short reaction times (*63*).

4.2 Materials and Methods

4.2.1 MBD Protein Cloning, Expression, Purification and Characterization

The biotin acceptor peptide sequence MAGGLNDIFEAQKIEWHE was appended to the C-terminal end of the murine MBD1-green fluorescent protein (MBD-GFP) construct (*21*) from Steve Blair's Lab (University of Utah, Salt Lake City) for *in vivo* biotinylation by the biotin ligase enzyme BirA. This was accomplished using PCR-driven overlap

extension. In short, we PCR amplified the MBD-GFP gene from the pET-30b vector starting 4 bp upstream of the EcoRI restriction site and ending immediately before the 3' TAA stop codon following the GFP cDNA. Fwd Primer: 5'-ATCCGAATTCGATGCCAAAAAAGA-3' (24 nt, GC = 37.5%, T_m = 55.6°C) Rev Primer: 5'-CTTGTACAGCTCGTCCATGC-3' (20 nt, GC = 55.0%, T_m = 56.0°C) Concurrently, we synthesized the cDNA for the biotin acceptor sequence from primers designed using the DNAWorks software from the NIH (*134*). The construct began at the 5' end with a 30 bp sequence complimentary to the end of the GFP cDNA (ATCACTCCCGGCATGGACGAGCTGTACAAG) followed by the cDNA for the acceptor sequence, the ochre stop codon, a one nucleotide spacer "A," the terminal XhoI restriction site and a 4 nt 3' overhang to facilitate restriction enzyme binding. The MBD1-GFP PCR product and biotin acceptor construct were joined using PCR-driven overlap extension (*135*). Fwd Primer: 5'-ATCCGAATTCGATGCCAAAAAAGA-3' (24 nt, GC = 37.5%, T_m = 55.6°C) Rev Primer: 5'-GGTGCTCGAGTTTATTCATGCCATTCAATTTTCTGCG-3' (37 nt, GC = 43.2%, T_m = 63.5°C). The gene structure is shown in Figure 4.1. The pET-30b vector and PCR product were double digested with EcoRI-HF and XhoI (New England BioLabs), gel purified, and ligated using T4 ligase (New England BioLabs). The pET-30b/MBD1-GFP-B construct was co-transformed with pACYC184/BirA (Avidity) into chemically competent *E. coli* BL21 (DE3) cells (New England BioLabs).

Figure 4.1: Representation of the MBD protein structure. Amino acids 1-75 of the murine MBD1 protein are fused to a downstream enhanced green fluorescent protein (EGFP) tag. While not used directly for fluorescent detection, the EGFP tag significantly increased the soluble yield when expressed in *E. coli*. An N-terminal His_6 tag is used for immobilized metal affinity chromatographic (IMAC) purification of the protein from bacterial cells used to express it, and a C-terminal biotin acceptor peptide sequence (MAGGLNDIFEAQKIEWHE) provides the recognition site for the BirA biotin ligase enzyme to covalently attach a biotin to the underlined lysine residue. 410 AA. Theoretical pI/MW: 5.98 / 46.3 kDa.

E. coli BL21 (DE3) cells containing both plasmids were inoculated from a glycerol cell stock into a 14 mL culture tube (VWR Scientific) containing 5 mL of Difco LB Broth (BD Biosciences) supplemented with 10 µg/mL chloramphenicol (Calbiochem) and 50 µg/mL kanamycin (Calbiochem) and grown overnight at 37°C and 250 rpm orbital shaking. These cells were then subcultured in a 1 L flask containing 250 mL of TB medium (12 g tryptone (BD Biosciences), 24 g yeast extract (BD Biosciences), and 4 mL glycerol (BDH Chemicals) dissolved in 900 mL of 18 MΩ deionized (DI) water, autoclaved, cooled, and brought to 1 L with the addition of 100 mL of a sterile solution of 0.17 M KH_2PO_4 (Macron Chemicals) and 0.72 M K_2HPO_4 (BDH Chemicals)) supplemented with 50 µg/mL kanamycin and incubated at 37°C and 250 rpm orbital shaking. When the expression culture reached an OD_{600} of 0.6, MBD expression was induced by adding IPTG (Omega Bio-Tek) and D-Biotin (Sigma-Aldrich) to final concentrations of 1 mM and 50 µM, respectively, and decreasing the incubator temperature to 20°C for 16 h.(*84*) Cells were harvested in five 50 mL conical tubes (BD Falcon) by centrifuging at 4,000g and 4°C for 10 min.

One tube was subsequently lysed with BugBuster HT Protein Extraction Reagent (EMD Millipore) following the manufacturer's instructions. The soluble fraction was separated from insoluble cell debris by centrifugation at 12,000g for 15 min. The clarified lysate was filtered using an Acrodisc PF Syringe Filter 0.8/0.2 (Pall) and loaded onto a 1 mL HisTrap™ Fast Flow column (GE Healthcare) for purification on an ÄKTA purifier 10™ FPLC System (GE Healthcare). The column was equilibrated and washed with binding buffer (*20*) (300 mM NaCl (Mallinckrodt Chemicals), 100 mM $Na_xH_xPO_4$ (Mallinckrodt Chemicals), 10 mM Tris-HCl (BDH Chemicals), 10 mM 2-mercaptoethanol (J.T. Baker), and 10 mM imidazole (Alfa Aesar), pH 8), and MBD protein was eluted using a linear gradient from 0% to 100% elution buffer(*20*) (300 mM NaCl, 100 mM $Na_xH_xPO_4$, 10 mM Tris-HCl, 10 mM 2-mercaptoethanol, and 250 mM imidazole, pH 8) over 20 column volumes (CV) followed by a 5 CV isocratic elution at 100% elution buffer. The pH of the binding and elution buffers was raised from 7.4 to 8 to eliminate a co-eluting impurity present in our prior expression and purification work with MBD-GFP fusion proteins (*84*). The fractions containing only MBD protein were pooled, concentrated, and buffer exchanged into storage buffer (1xPBS (AMRESCO), 10%

glycerol and 1 mM dithiothreitol (MP Biomedicals)) using a 30,000 MWCO spin filter (Millipore). MBD concentration was assayed using the bicinchoninc acid (BCA) protein assay kit – reducing agent compatible (Thermo Scientific) with BSA standards. Approximately 1.5 µg of MBD protein was loaded onto a 4–15% SDS–PAGE gradient gel (Bio-Rad) and visualized with Coomassie Brilliant Blue R250 stain (AMRESCO); see Figure 4.2a. MBD binding activity to methylated DNA was verified by electromobility shift assay (EMSA) using the method described previously;(*84*) see Figure 4.2b. Purified protein was stored at -80°C in 20 µL single-use aliquots for later use in methylation profiling assays.

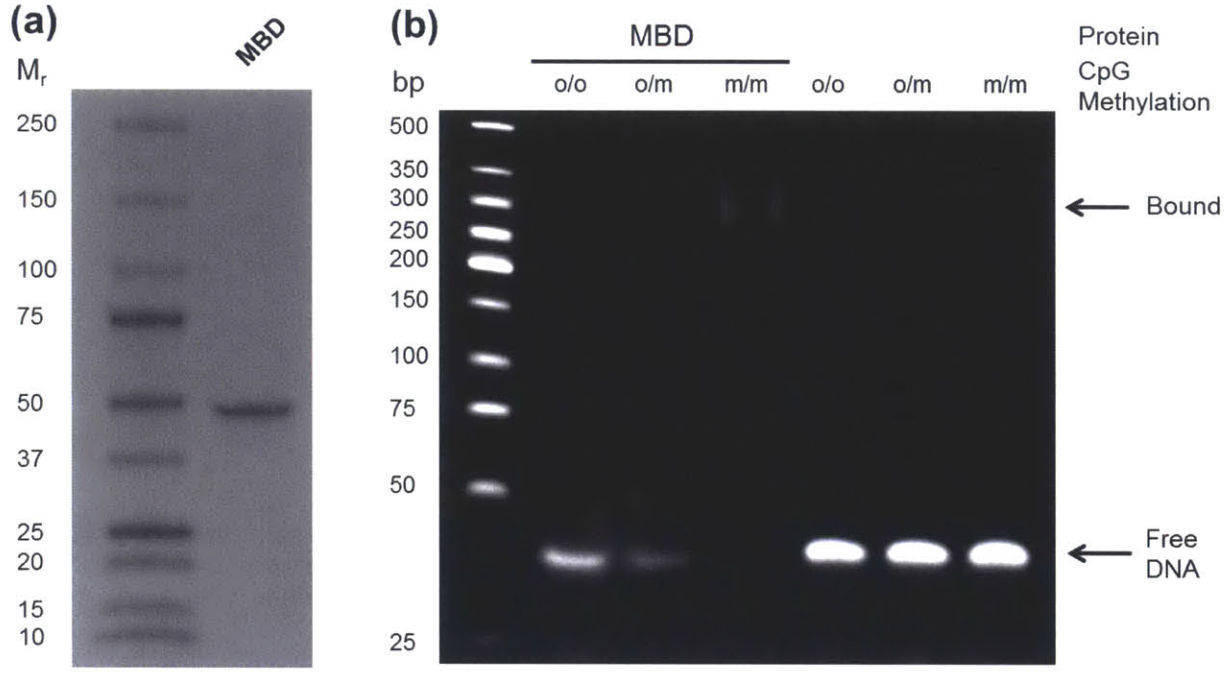


Figure 4.2: Characterization of the purified MBD protein. (a) Coomassie blue stained 4–15% SDS–PAGE gradient gel shows purified MBD protein (MW = 46.3 kDa). (b) Electromobility shift assay (EMSA) shows the capacity of purified MBD to bind methylated DNA. MBD protein was incubated with pre-hybridized dsDNA oligonucleotides with non- (o/o), hemi- (o/m) and singly, symmetrically-methylated (m/m) CpGs; concentration matched DNA samples were prepared without protein as controls. DNA with bound MBD protein migrates through the gel at a retarded rate thus appearing to have larger size when compared to the DNA ladder standard.

Streptavidin mutein matrix (Roche) was used for the final polishing affinity chromatography step to remove any MBD protein not biotinylated *in vivo*. Roche's batch purification protocol and 50 µL of resin were used to purify 125 µg of monobiotinylated

MBD after already being purified using IMAC as described above. Three elution fractions were collected and pooled. An UltraCruz™ Micro G-25 (Santa Cruz Biotechnology) gel filtration spin column was used to remove free biotin remaining from eluting the biotinylated MBD protein. Pure MBD was buffer exchanged into storage buffer as described above. The final polishing purification step was not routinely performed after it was found not to increase the signal from MBD binding to methylated DNA in the biochip assay; see Figure 4.3.

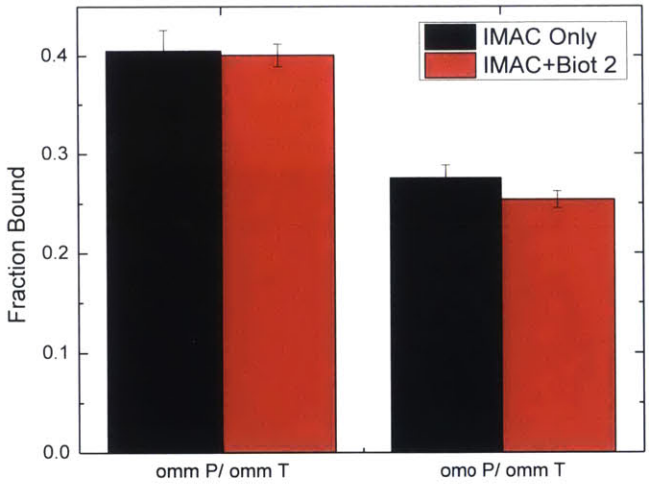


Figure 4.3: Comparison of the signal from doubly (omm) and singly (omo) methylated DNA using MBD purified using IMAC only or IMAC plus a streptavidin mutein affinity chromatography polishing step. The polishing step was not found to improve the fraction bound over the IMAC only case and was thus not routinely used.

4.2.2 Preparing Biochip Test Surfaces

Standard clear glass microscope slides (VWR Scientific) coated with an aldehyde functionalized agarose film were prepared using 2 mL of a 0.2% w/v SeaKem® LE agarose (Lonza) solution and the method previously described (*85*). The concentration of agarose in the film casting solution was decreased to 0.2% w/v from 1% reported by Afanassiev *et al.* to reduce nonspecific uptake of the eosin Y stain used to visualize the photopolymerized hydrogels formed usain PBA detection. The ooo, omo, and omm

probe oligonucleotides previously designed to capture target DNA sequences from the *MGMT* promoter (*21*) were ordered from Integrated DNA Technologies with a 5' hydrazide functionalization (instead of biotin used by Yu *et al.*) to facilitate coupling to the reactive aldehyde groups on the biochip surface. Additionally, two, 50 nt single-stranded oligonucleotides (5'-CATCACACAACATCACACAACATCACGTATATAAAACGGAACGTCGAAGG-3') were designed as controls; both had 5' hydrazide functionalization and one had a biotin covalently attached to the 3' end. Each probe was diluted to 10 μM in 3xSSC buffer (G Biosciences) and spotted at 50% relative humidity according to the layout illustrated in Figure 4.5a using a GeSiM Nano-Plotter 2.1 equipped with a Nano Tip. Each spot was formed from ten separate 300 pL depositions. After spotting, each biochip was transferred to a vacuum desiccator for overnight incubation. A circular, 9 mm diameter isolator well was cut from Scotch 3M 665 tape and affixed to the biochip to define each test area. Each biochip was then rinsed under a stream of DI water and blown dry using compressed nitrogen gas. Biochips ready for testing were stored in the vacuum desiccator until needed.

4.2.3 Microfluidic Device Fabrication

Microfluidic devices were fabricated using photolithography techniques.(*136*) A 4" silicon wafer was coated with a 60 μm thick layer of SU-8 2050 (MicroChem) and patterned using a 25,400 dpi transparency mask (CAD/Art Services, Inc.). Next, a thin layer of polydimethylsiloxane (PDMS) (Sylgard 184, Dow Corning) was cast onto the SU-8 mold and cured for 15 min at 80°C until solid. Then, a supporting glass piece was cut from a standard microscope slide (VWR Scientific) to be 20 mm x 12 mm, and was positioned over the microfluidic channel. The rest of the PDMS was poured onto the mold and cured for 2 h at 80°C. The PDMS was then peeled off and each device was cut out separately. The inlet and outlet holes were punched using a 1.5 mm ID x 1.91 mm OD Harris Uni-Core Puncher (Ted Pella, Inc.). The microfluidic channel, shown in Figure 4.10a, features 2 mm wide inlet and outlet regions that expand to a 9 mm diameter circular test area. The channel was designed to be 27 mm long, in order to

allow for a standard glass slide (25 mm wide) to be used as an additional support above the PDMS device; see Figure 4.10b. The device was then positioned above the test area on the biochip slide, and clamped on using standard binder clips (Staples).

4.2.4 Detecting Methylated DNA with MBD, Fluorescence and PBA

To block each test area against nonspecific protein binding, each isolator well was filled with 40 µL of 1% w/v BSA in 1xPBS and incubated for 15 min in a humid chamber at ambient temperature. Excess BSA was then washed away with 1xPBS, and the entire biochip was submersed in 1xPBS at 4°C for 10 min. Following a rinse with DI water, dilutions of omm target (*21*) oligonucleotides were prepared in hybridization buffer (6xSSC, 5xDenhardt's (BioExpress)), and 40 µL of each were pipetted onto separate test areas. Biochips were then incubated in a humid chamber (defined here as a pipette tip box partially filled with DI water) for 4 h at ambient temperature to allow for DNA capture. Each biochip was rinsed with 1xPBS followed by DI water to remove unhybridized target DNA. Next, 40 µL of MBD protein diluted in binding buffer (20 mM HEPES (EMD Millipore), pH 7.9, 3 mM $MgCl_2$ (BDH Chemicals), 10% v/v glycerol, 1 mM dithiothreitol, 100 mM KCl (BDH Chemicals), 0.1% w/v BSA, and 0.01% v/v Tween-20 (Sigma-Aldrich)) was pipetted onto each test area, and the biochip was incubated for 30 min in a humid chamber at 4°C. After MBD incubation, each test site was rinsed with PBST (1xPBS, 0.1% Tween-20) followed by 1xPBS, and the entire biochip was submersed in 1xPBS at 4°C for 25 min. Following a rinse with DI water, 40 µL of 0.1 µM streptavidin-Cy3 or streptavidin-eosin in 1xPBS, 5xDenhardt's, and 0.5% w/v BSA was pipetted onto each test area, and the biochip was incubated for 5 min in a humid chamber covered in foil at ambient temperature. Unbound labeling reagent was removed by washing sequentially with PBST, 1xPBS, and DI water followed by blow drying using compressed nitrogen gas. The biochip was either scanned in the case of streptavidin-Cy3 fluorescent detection or polymerized using an aqueous PEG diacrylate monomer and methods previously described (*78*).

Alternatively, the reusable microfluidic device was attached to a rinsed and dried biochip using binder clips. FEP 1/16" OD x 0.020" ID tubing (Upchurch Scientific) was

inserted into each port on the device and mated to 0.89 mm ID Tygon-LFL tubing (Saint-Gobain) which was then run through a Masterflex C/L 77120-52 peristaltic pump (Cole Parmer). All of the tubing connections were self-sealing, and no glue was necessary. A 1% w/v BSA in 1xPBS solution was pumped through the microfluidic device at 100 µL/min for 15 min followed by 1xPBS for 10 min. A solution of omm target ssDNA in 6xSSC/5xDenhardt's was recirculated at 100 µL/min for 4 h. A 1xPBS solution was pumped through the microfluidic device at 100 µL/min for 5 min followed by DI water for another 5 min. MBD protein diluted to 40 µg/mL in binding buffer was recirculated at 100 µL/min for 30 min. PBST, 1xPBS, and DI water were sequentially pumped through the microfluidic device for 5 min each at 100 µL/min. Fluorescent or photoinitiator labeling and detection was performed as described above.

4.2.5 Biochip Analysis

Biochips labeled with streptavidin-Cy3 were scanned using a Molecular Devices GenePix 4000B fluorescent microarray scanner. Each fluorescence image was analyzed using ImageJ (NIH). The mean fluorescence intensity for each spot was determined by adjusting the threshold of the image to include the entire spot area and averaging the constituent pixel intensities. Fraction bound values were computed by dividing the average signal from replicate spots by the average signal from the biotinylated ssDNA control spots. The background corrected intensity (BCI) was computed by subtracting the mean background intensity (signal from unmethylated ssDNA) from the mean fluorescence intensity value from replicate spots. Cy3 surface densities were then calculated from BCI values using a calibration curve generated by scanning and analyzing a Full Moon Biosystems calibration array. Dividing the Cy3 surface densities by the known average number of Cy3 dye molecules per streptavidin produced the surface density of binding events.

Each polymerized test area was imaged using the digital camera in the ampliPHOX® Reader (InDevR) imaging bay.

4.2.6 Quantitative MBD Hybridization Analysis

The MBD diffusion coefficient (D) was estimated to be 1.1×10^{-10} $m^2(s)^{-1}$ using the Stokes-Einstein equation. All calculations used 2.88×10^6 $(M\cdot s)^{-1}$ and 0.4 $(s)^{-1}$ MBD association (k_a) and dissociation (k_d) rates, respectively (*21*). The depletion zone thickness was approximated as $\delta=L(Pe_s)^{-1/3}$, where Pe_s is defined as the shear Peclet number ($Pe_s=6(L/H)^2Pe$) (*137, 138*) L is the length of the binding area, which was approximated as 900 μm (equivalent to three spot diameters), H is the channel height, Pe is the Peclet number (Pe=UH/D), and U is the fluid velocity. The Damköhler number ($Da=k_aC_sL_d/D$) was calculated based on the surface density of DNA duplexes (C_s) in each spot and the depletion zone thickness (L_d).

4.3 Results and Discussion

Our epigenotyping assay follows the scheme shown in Figure 4.4. First, single-stranded, 5'-hydrazide modified DNA capture probes are spotted onto an aldehyde functional biochip according to the layout illustrated in Figure 4.5a. Any remaining unreacted aldehyde groups are quenched with BSA. DNA target oligonucleotides in a test sample then hybridize to the corresponding sequence-specific capture probes on the biochip. Next, biotinylated MBD proteins bind to DNA duplexes with symmetrically methylated CpGs. Streptavidin functionalized with either eosin (photoinitiator) or Cy3 dye then transduces MBD binding into a colorimetric or fluorescent readout as depicted in Figure 4.5b. For Cy3 labeling, the biochip is directly scanned to produce the type of image shown in Figure 4.5c. For colorimetric detection, an aqueous polyethylene glycol diacrylate monomer is polymerized to form hydrogels specifically at the sites with surface-bound eosin when irradiated with green (522 nm) light for approximately two minutes (*63*). The hydrogels become visible to the unaided eye after staining with a red dye for two minutes as shown in Figure 4.5d (*63*).

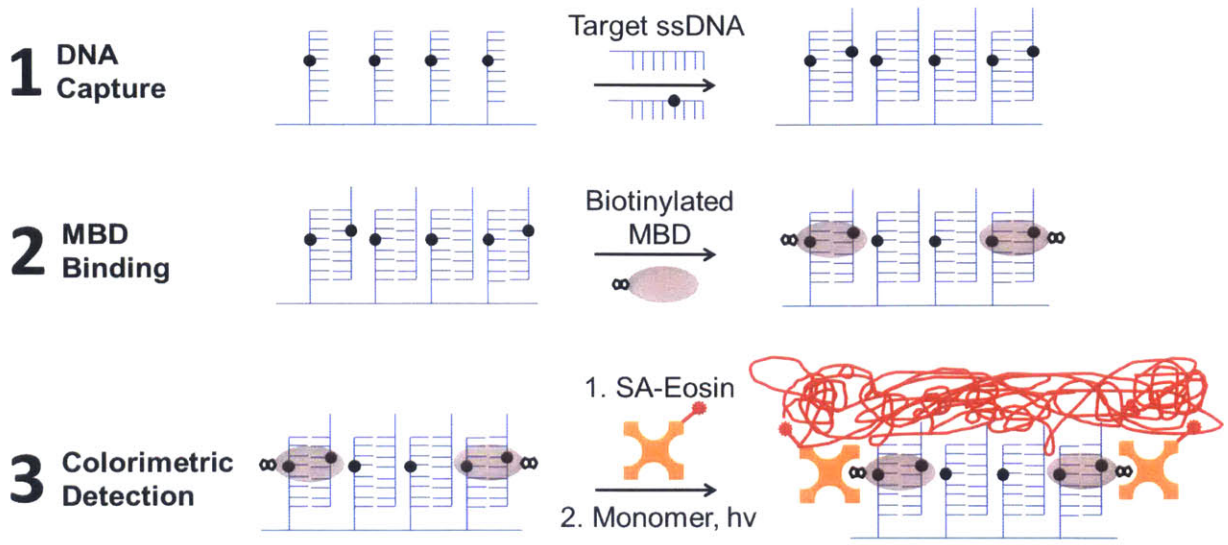


Figure 4.4: Methylated and hydrazide functionalized probe ssDNA oligos are spotted onto activated agarose coated slides with subsequent blocking using 1% w/v BSA. Target ssDNA oligos for analysis are hybridized to the sequence-specific probes on the biochip. A biotinylated MBD protein is then added to bind methylated CpG dinucleotides followed by streptavidin-Cy3 for fluorescence or streptavidin-eosin for polymerization readout.

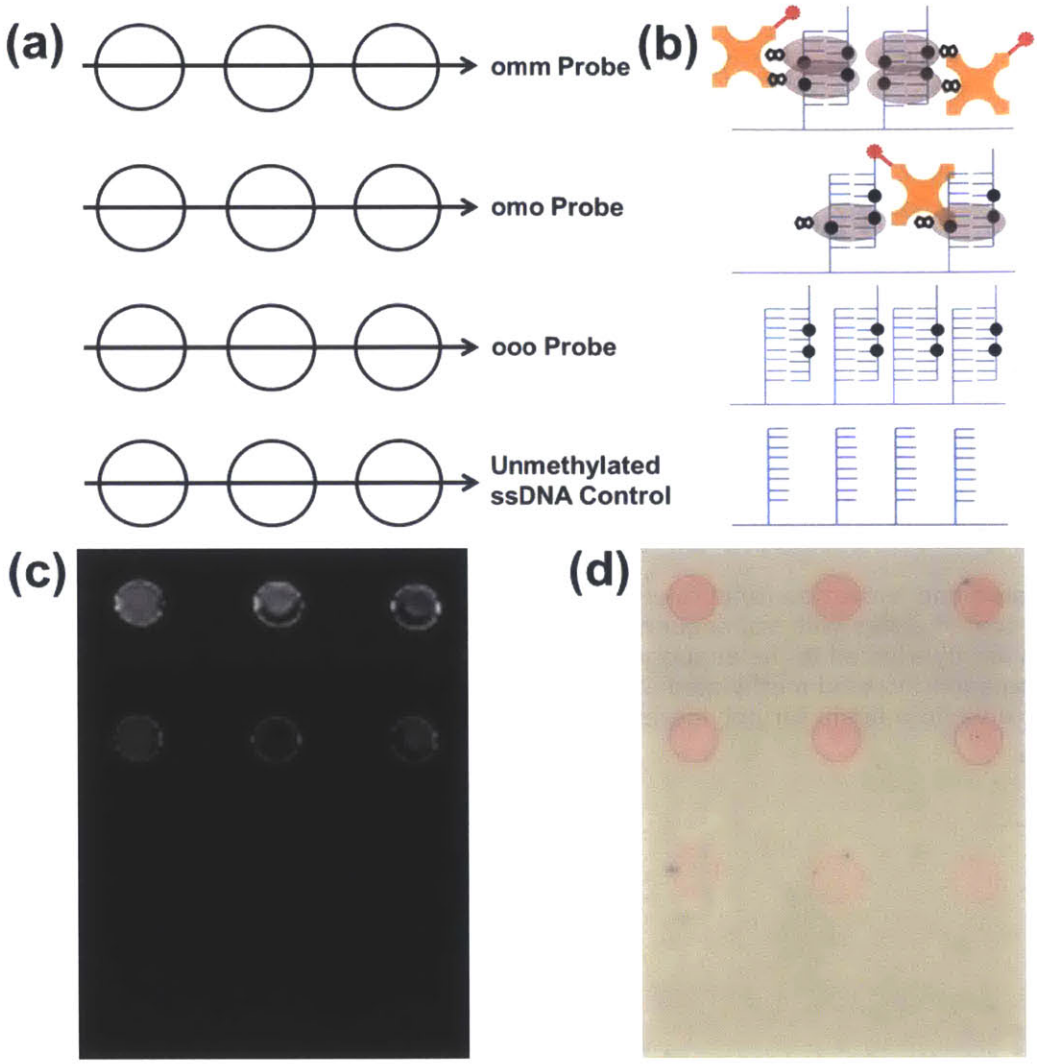


Figure 4.5: (a) Biochips were spotted with capture probes having two (omm), one (omo), or no (ooo) methylated cytosine nucleotides in order to epigenotype the target oligos. Unmethylated ssDNA served as the negative control. (b) Schematic representation of the area within each group of spots following hybridization with 100 nM doubly methylated (omm) target ssDNA and detection using the scheme shown in Figure 4.4. (c & d) Fluorescence, or colorimetric, readout of MBD binding to methylated DNA.

Several design choices were key to obtaining the results presented in Figure 4.5. The epigenotyping assay we describe here uses a standard glass microscope slide coated with an agarose film (*85*). The agarose film coating, with its 3D surface structure, shields biomolecules from the charged glass substrate and provides a solution-like environment conducive for nucleic acid and protein binding events (*78, 139*). Experimentally, the agarose film was essential to our assay because self-assembled

monolayer (SAM) coated glass surfaces, when evaluated, suffered from high nonspecific binding and signals not distinguishable from background; see Figure 4.6.

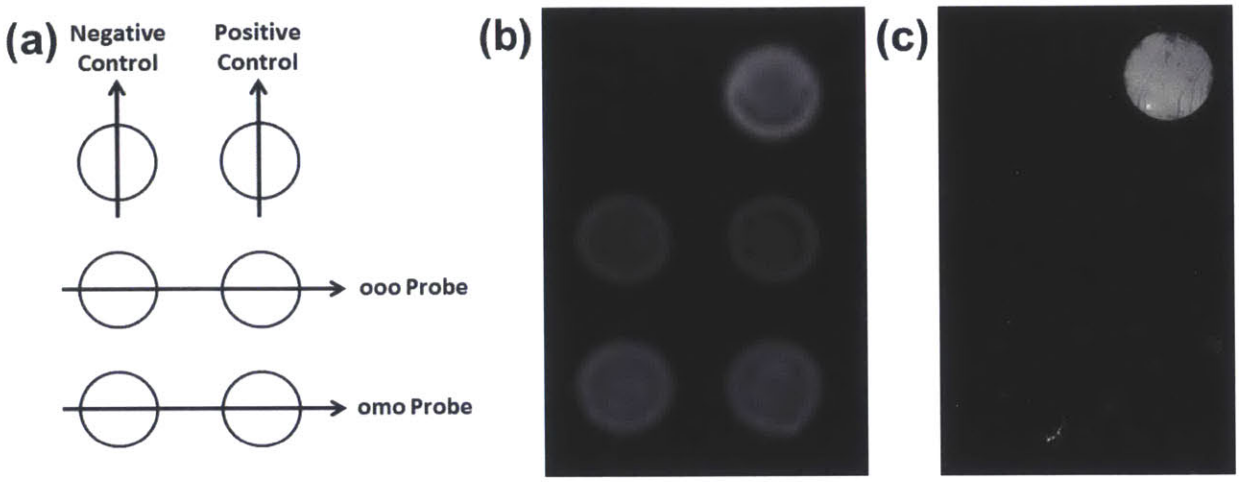


Figure 4.6: Biochip epigenotyping assay performance on a SCHOTT Nexterion Slide AL. (a) Biochip spotting layout for probe ssDNA. (b) Hybridization of biotinylated target ssDNA and streptavidin-Cy3 labeling shows that duplex DNA is being formed at each of the ooo and omo probe spots. (c) Hybridization of omo target ssDNA followed by 40 µg/mL MBD detection and streptavidin-Cy3 labeling. Fluorescence scanning shows high background signal from static MBD binding and omo probe/omo target signal not distinguishable from background on the self-assembled monolayer surface.

Agarose film coated surfaces also benefit from being significantly less expensive than SAMs. The capture probe oligos were modified with a hydrazide which has a significantly higher reactivity toward aldehyde functional surfaces than primary amines leading to detectable DNA hybridization at sub-nanomolar concentrations as a result of the higher probe density (*140, 141*); see Figure 4.7.

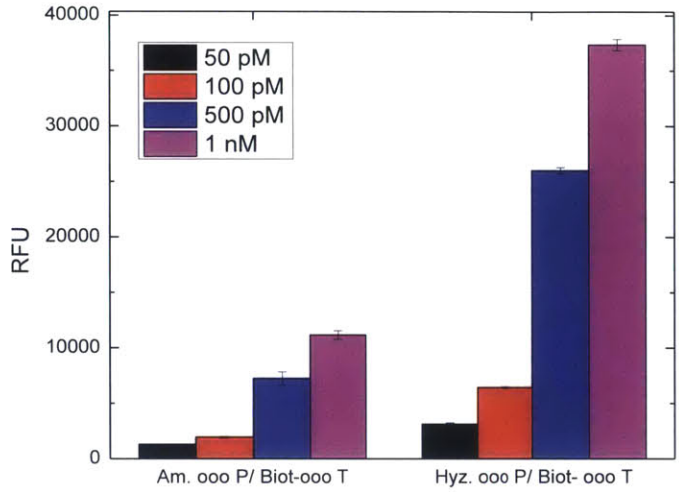


Figure 4.7: Comparison of capture performance of probe ssDNA oligos with either a primary amine (Am.) or hydrazide (Hyz.) modification to facilitate coupling to the aldehyde functional, agarose biochip. Sub-nanomolar concentrations of biotinylated target ssDNA were hybridized to each array and detected using streptavidin-Cy3 and fluorescence scanning. Spots printed using hydrazide functionalized probe produced a signal approximately threefold greater than those printed with the primary amine functionalized probe when incubated with target ssDNA at the same concentration.

Each DNA probe was spotted at 10 μM to saturate the aldehyde functional surface with capture probe oligonucleotides. Further, we sought to print the smallest spots (300 μm diameter) that allowed for unaided, colorimetric readout; such feature size miniaturization has also been shown to improve surface capture performance (*142*). The average surface density of the biotinylated control oligo was measured to be 900±40 oligos/μm^2. When the un- (ooo), singly- (omo), and doubly-methylated (omm) capture probes were hybridized with 100 nM complimentary, biotinylated target oligos, the average surface density of the duplex decreased slightly to 850±50 duplex/μm^2; see Figure 4.8 for further detail. All further binding signals were reported as a fraction of the capture probe density, assumed equal to the measured density of biotinylated ssDNA (900±40 oligos/μm^2).

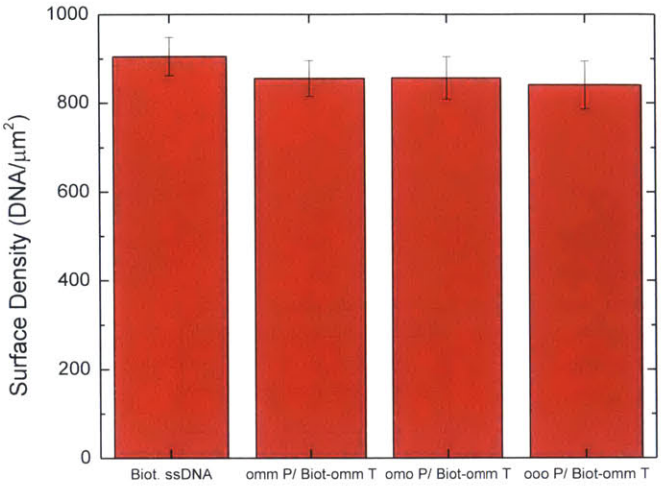
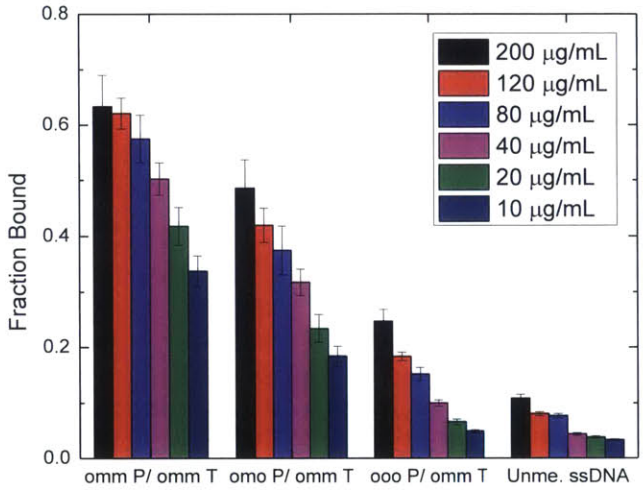


Figure 4.8: The surface density of the biotinylated ssDNA control oligonucleotide and capture probes hybridized with 100 nM biotinylated target ssDNA were determined using streptavidin-Cy3 and fluorescence quantification. The 900±40 DNA/μm^2 surface density of the biotinylated ssDNA control was taken as the maximum for the hydrazide functionalized probes used in this study. Hybridization to each of the three (ooo, omo and omm) capture probes was equivalent within error; with 100 nM target, 94% of capture probes had bound target.

The MBD protein is expressed as a GFP fusion. In addition to being a fluorescent reporter, the GFP also significantly improves the soluble yield of the fusion when induced using the protocol we optimized previously (*84*). We further added a 15 amino acid C-terminal AviTag™ (Avidity) with an internal lysine residue which the co-expressed BirA biotin ligase enzyme biotinylates. Such *in* vivo biotinylation greatly simplifies MBD production and eliminates the risk of denaturation and loss of binding activity following chemical biotinylation which requires organic solvents. The MBD concentration used in the assay was tuned to maximize signal from duplex DNA with both singly, and doubly, symmetrically methylated CpGs and minimize signal from hemi-methylated (defined as signal from unmethylated (ooo) probe spots hybridized with methylated (omo or omm) target oligos) and background signal (defined as signal from unmethylated ssDNA). Using a MBD concentration of 40 μg/mL optimized signal intensity, specificity, and the amount of MBD protein used per assay (1.6 μg) as shown in Figure 4.9; thus, it was used for all further experiments. Maintaining a low background signal is critical for achieving low fluorescence signal-to-noise ratios (SNR). However,

the ratio is less critical for the colorimetric assay because no hydrogels will be formed as long as the background signal is below the photopolymerization amplification threshold (*78*).



MBD Conc. (μg/mL)	ommP : oooP
200	2.6
120	3.4
80	3.8
40	5.0
20	6.3
10	6.9

Figure 4.9: Comparison of fraction bound as a function of varying MBD concentration on arrays hybridized with 100 nM omm target ssDNA. The fraction of DNA probes with MBD bound decreases as a function of MBD concentration. While the highest MBD concentrations give larger fractional values, they have lower omm P:ooo P ratios indicating sub-optimal assay specificity. Using 40 μg/mL MBD decreases the fraction bound by as much as 13% but increases the omm P:ooo P ratio twofold relative to 200 μg/mL and has a non-specific binding level to unmethylated ssDNA comparable to the lowest MBD concentrations.

Nucleic acid hybridization is typically performed under inefficient, diffusion-limited conditions in conventional microarray experiments (*143*). Numerous microfluidic devices that use mixing have been reported to improve DNA hybridization (*143-145*). We fabricated a reusable microfluidic device from PDMS, shown in Figure 4.10a, to implement a recirculating mixing method similar to that described by Lee et al. for the purpose of improving DNA hybridization as well as standardize the work flow and provide an assay format suitable for automation. Consistent with previous reports, hybridizing target DNA using the microfluidic device produced MBD binding signals approximately one-third higher than those achieved using static DNA incubation over the times studied; see Figure 4.10c (*144*). However, experimental results shown in

Figure 4.10d indicated no similar improvement in the fraction bound using the flow system for MBD binding.

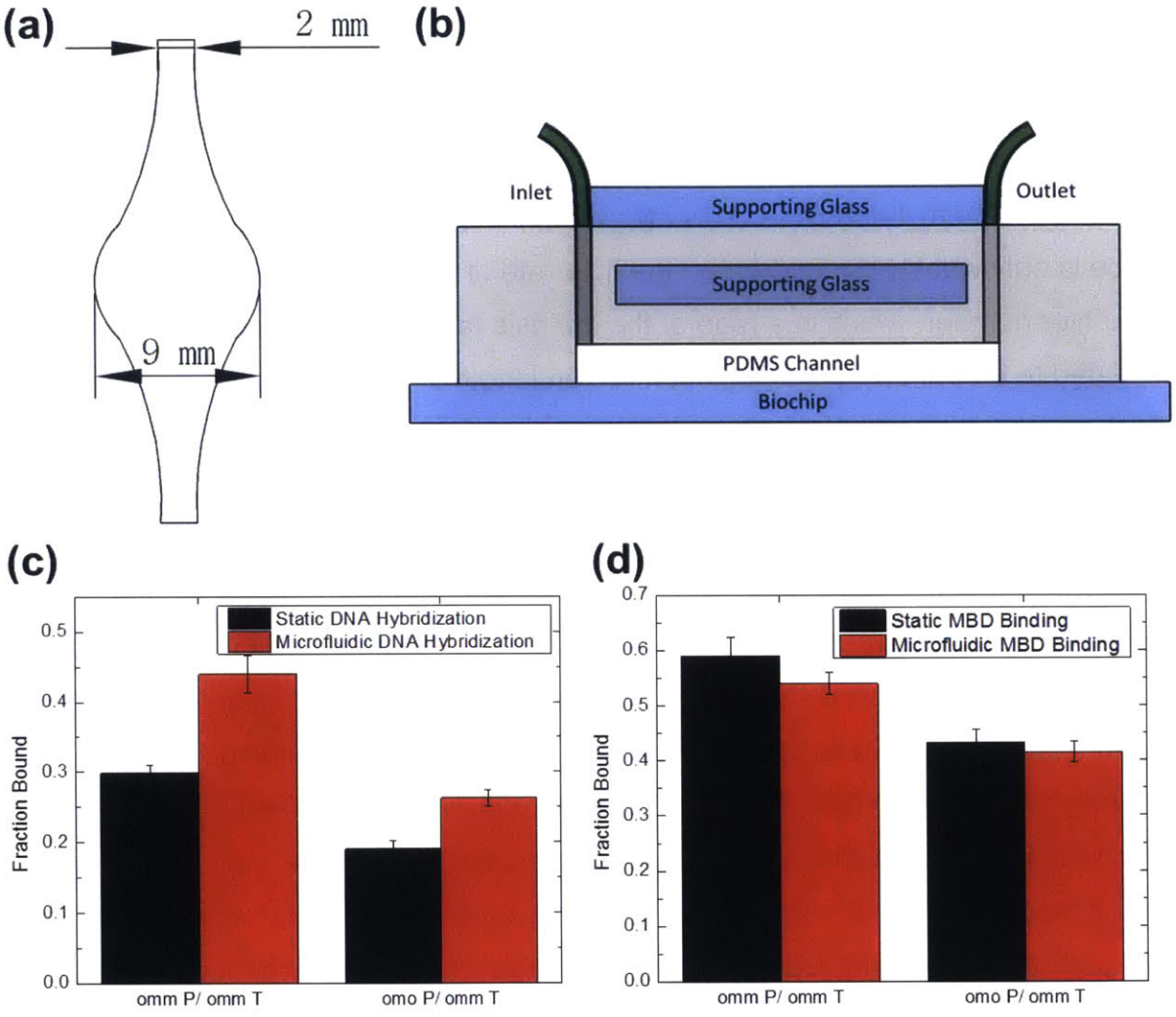


Figure 4.10: A microfluidic device was mated to the biochip. (a) Diagram of the microfluidic channel etched in PDMS. (b) Supporting glass was added to the unetched side of the microfluidic device, and it was clamped to the biochip. Inlet and outlet tubing was connected to each end of the channel, fed into a microcentrifuge tube reservoir, and passed through a peristaltic pump for recirculation. (c) Comparison of the signal from static MBD binding to doubly (omm) and singly (omo) methylated probes hybridized with either 100, 10 or 1 nM doubly (omm) methylated target DNA under static or microfluidic (uFl) conditions. Microfluidic hybridization increases the signal from MBD binding by approximately one-third for target concentrations 10 nM and greater. (d) Comparison of the signal from MBD binding under static or 100 µL/min recycling conditions to doubly and singly methylated duplexes. The microfluidic device does not improve signal due to a weak dependence of MBD binding on flow rate and a partial hybridization reaction limit.

In order to more completely understand whether convection, diffusion and/or reaction rate were limiting the MBD binding step for the purpose of improving assay sensitivity, we calculated both the Peclet and the Damköhler numbers for the flow system (*137, 138*). The Peclet number is a measure of the importance of convection relative to diffusion. It was found to be 1700 which implies the presence of a thin (approximately 7 μm) depletion region of MBD above the binding surface. Any MBD above the depletion region is swept down the channel without having a chance to encounter the methylated DNA. This also means the mass transport toward the binding surface is only weakly dependent on the flow rate of the system. Furthermore, the Damköhler number, which is a ratio of the intrinsic rate of reaction to diffusion, was calculated to be 0.3 using the depletion region length-scale. Because this value is less but not *much* less than unity, the system is only partially reaction-limited. The predicted weak dependence of MBD binding on flow rate and the partial reaction limit both seem to explain the equivalent performance of both systems for the MBD binding step.

The signal from MBD binding versus the concentration of target ssDNA with two methylated CpG sites is shown in Figure 4.11 for capture probes with two, one or no complimentary, methylated CpG sites. We determined the limit of quantification for MBD binding to spots with two or one symmetrically methylated CpG sites using our epigenotyping assay to be 0.3 nM and 1 nM, respectively, as defined by the target concentration that produces a signal greater than that from the un-methylated probe plus ten standard deviations (*146*). Sample size is the current drawback of the system as the reported sensitivity requires a larger number of cells than what can be obtained from a needle biopsy. In a clinical sample, the two *MGMT* alleles may be heterozygously methylationed at a CpG dinucleotide in the interrogated promoter sequence. All cells in the population would, however, share this methylation pattern due to the clonal expansion of tumor cells and heritability of DNA methylation (*7*).Therefore, the LOD curve would reside between the unmethylated (ooo) and singly methylated (omo) curves in that case.

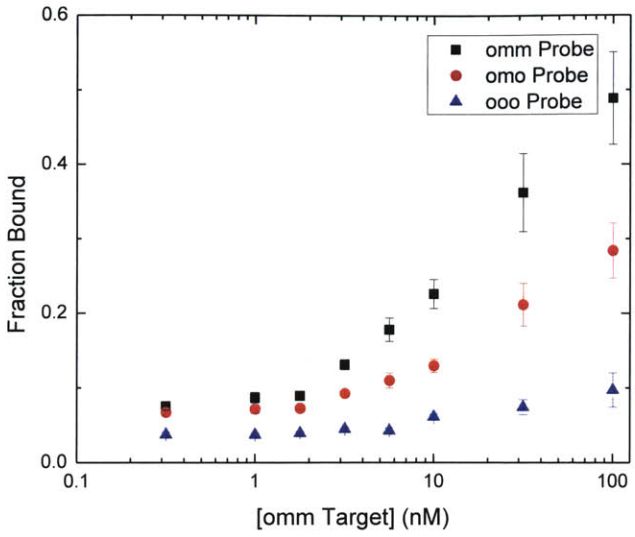


Figure 4.11: The limit of quantification (defined as the minimum concentration producing a signal greater than that from the unmethylated (ooo) probe mean plus 10 SDs) for doubly (omm) and singly (omo) symmetrically methylated DNA duplexes was determined to be 0.3 nM and 1 nM, respectively. Each sample was on-chip hybridized with the stated concentration with omm target ssDNA and detected with 40 µg/mL MBD followed by streptavidin-Cy3 labeling.

4.4 Conclusions

We have demonstrated a novel method for directly hybridizing and detecting methylated DNA fragments on a biochip. This simple method requires less time to perform than the existing methods which rely on bisulfite conversion of unmethylated cytosine bases, PCR, and/or next-generation DNA sequencing. Eliminating sodium bisulfite treatment further reduces the risk of DNA degradation and inaccuracies resulting from incomplete conversion. Our method is also able to achieve a low cost per test by using inexpensive test surfaces and a recombinant MBD protein that can be produced efficiently in *E. coli*. Fluorescence detection requires a microarray scanner for readout; however, the colorimetric assay requires only a green light source and digital camera for imaging the visible, stained hydrogels.

New biochips can easily be customized to interrogate a specific gene or panel of genes by simply designing unique capture probes specific to the sequences of interest. Further, microfluidic integration provides continuous delivery of test solutions,

standardizes the assay, and provides a platform well-suited for automation. Quantitative experimental and theoretical analysis also enables future work to develop technologies using engineered MBD proteins for single-cell epigenetic analyses.

Chapter 5

Quantitative Evaluation of MBD Proteins for Methylated DNA Enrichment Reactions

Abstract

Methods to enrich fractionated genomic samples for methylated DNA (meDNA) have become useful tools for whole-genome DNA methylation mapping at a lower cost than whole-genome bisulfite conversion. Either a methyl-CpG binding domain (MBD) protein or an antibody is used to selectively bind meDNA on the surface of beads. Here, we describe a quantitative framework for analyzing key meDNA capture parameters for the purpose of better understanding and improving enrichment steps prior to sequencing. Our analysis shows that higher affinity capture reagents such as MBD2s are expected to capture nearly 10-fold more DNA than lower affinity MBDs such as MeCP2 at equivalent MBD loadings. Unmethylated CpGs only increase the likelihood that a fragment will be pulled-down only in the absence of a methylated CpG; further, additional methylated CpGs after the first two do not significantly increase the likelihood that a fragment will be pulled-down. Finally, the majority of DNA fragments captured by higher affinity MBD reagents tend to have fewer methylated CpGs in contrast to lower affinity MBDs where avidity binding effects are significantly more important for retaining captured, meDNA fragments. Therefore, the MBD protein and elution conditions used should be chosen based on the desired meDNA fragment profile.

5.1 Introduction

DNA methylation is an epigenetic mark that is known to regulate gene expression in eukaryotic cells. Methylation occurs predominately at the 5-position of the cytosine pyrimidine ring when followed by guanine on the same DNA strand in a CpG dinucleotide sequence (*9*). CpG islands, regions of high CpG content, are often coincident with gene promoters (*3*). Hypermethylation of these promoter-coincident CpG islands can silence the downstream gene (*147, 148*). Such aberrant DNA methylation can be both a causal event in the development of disease (*25, 149*) and be used as biomarkers for diagnosing and predicting patients' response to treatment (*41*). As such, discovering such epigenetic and DNA methylation signatures in the human genome has become of increased interest (*51*).

Several methods exist for determining the content of the entire human methylome. Many of these methods rely on selectively converting unmethylated cytosine bases to uracil by treatment with sodium bisulfite and detecting sequence alterations using sequencing or array methods (*52, 53*). Methods to enrich genomic DNA samples for methylated sequences combined with massively parallel DNA sequencing have also proven useful for genome-wide methylation analysis without requiring sodium bisulfite conversion of DNA (*52, 150*). One of two approaches, methyl-CpG binding domain protein-sequencing (MBD-seq) or methylated DNA immunoprecipitation-sequencing (MeDIP-seq), is generally used. Numerous kits have been developed that use MBD proteins to capture methylated DNA from fractionated genomic samples (De Meyer). These kits work by immobilizing a member of the MBD family of proteins on a stationary phase to capture methylated DNA that can be subsequently eluted and analyzed. The fragments of genomic DNA have varying numbers of CpGs which may or may not be methylated (*3*). Previous work has compared these kits for use in epigenotyping DNA from cell lines (*151*); however, no reports to date have described a quantitative framework for optimizing DNA capture on the basis of parameters for the pull-down reaction. Here, we present a model based on the thermodynamics of MBD-DNA binding and material balance to determine how key variables such as MBD affinity, the number of beads used for the pull-down reaction, and DNA fragment type affect the results of

the enrichment step. The purpose is to provide a better understanding this key step and identify parameters for optimizing pull-down of the desired methylated DNA fragment types.

5.2 Methods

5.2.1 Inputs

The input values were chosen to model the performance of the MethylMiner® kit. The capture reaction contains 0.2 µg of human genomic DNA as monodisperse 200 bp fragments in a 200 µL reaction to emulate the conditions used by De Meyer *et al.* (*151*). All MethylMiner® kit-specific parameters followed the manufacturer's instructions. The number of fragments of each type "*i*" was estimated from the data published by Saxonov *et al.* (*3*). The authors identified two regions, 1kb upstream regions and exons, as having higher observed CpG fractions of 0.042 and 0.028, respectively, than the rest of the genome which has an observed CpG fraction of approximately 0.01. We multiplied the average fragment size (200 bp) by each of these fractions to approximate the average number of CpGs in the fragments from each region and also by 0.7 as a correction for the number of methylated CpGs relative to total CpGs. The resulting number of methylated CpGs in fragments from each region was set as the average for three separate normal distributions. These distributions were multiplied by the number of fragments from each region (determined by multiplying the weighted fraction of the region length to the genome length and the number of 200 bp DNA fragments) to approximate the number of fragments containing a given number of methylated CpGs. Fragments with the same number of methylated CpGs were aggregated and treated as a DNA fragment of type "*i*" in the model.

5.2.2 Numerical Analysis

The thermodynamic and mole balance equations were solved simultaneously as a system in MATLAB.

5.3 Results and Discussion

Binding of methylated DNA (D) in solution to MBD (M) proteins immobilized on the surface of beads follows the reaction scheme

$$[M]+[D_i]\underset{k_{-i}}{\overset{k_i}{\rightleftharpoons}}[MD_i] \tag{1}$$

where [M] is the bead surface concentration of MBD proteins (pmol/cm^2), [D_i] is the bulk concentration of DNA fragments of type "*i*" (pmol/mL), [MD_i] is the bead surface concentration of MBD proteins with bound DNA of type "*i*" (pmol/cm^2), k_i is the association rate constant, and k_{-i} is the dissociation rate constant. The rate of formation of the MBD-DNA complex can be represented as a second order formation rate and first order dissociation.

$$\frac{d[MD_i]}{dt}=k_i[M][D_i]-k_{-i}[MD_i] \tag{2}$$

At equilibrium, the MBD-DNA complex concentration becomes invariant with time; thus, setting this quantity equal to zero produces

$$\frac{[MD_i]}{[M][D_i]}=\frac{k_i}{k_{-i}}=K_i \tag{3}$$

with rearrangement where K_i is the thermodynamic equilibrium constant for DNA fragments of type "*i*.". The methylated DNA capture reaction is also governed by mole balances for both the MBD and DNA species. The MBD protein site balance

$$[M]_0-[M]-\sum_i[MD_i]=0 \tag{4}$$

relates the occupation of MBD binding sites with DNA where $[M]_0$ is the initial bead surface concentration i.e. density of MBD proteins, $[M]$ is the surface concentration of

MBDs without bound DNA, and $[MD_i]$ is the surface concentration of MBDs with bound DNA fragments of type "*i*." Multiplying each of the terms in equation 4 by the total bead surface area in the reaction *A* (a function of both bead diameter and total number) establishes conservation of MBD binding sites. Similarly, the DNA mole balance

$$V\left([D_i]_0 - [D_i]\right) - A[MD_i] = 0 \tag{5}$$

relates the conservation of each DNA type as the sum of fragments in bulk solution of volume V and surface-bound MBD-DNA complexes over all beads with total surface area A. In the limit of *i*=1 for a reaction consisting of a single fragment type, this system has an analytical solution:

$$\frac{[D_i]}{[D_i]_0} = \frac{-\left(\frac{1}{K_i[D_i]_0} + \frac{A}{V}\frac{[M]_0}{[D_i]_0} - 1\right) + \sqrt{\left(\frac{1}{K_i[D_i]_0} + \frac{A}{V}\frac{[M]_0}{[D_i]_0} - 1\right)^2 + \frac{4}{K_i[D_i]_0}}}{2} \tag{6}$$

Plotting one minus this ratio relates the fraction of DNA fragments pulled down from bulk solution. The total number of beads used for enrichment and the affinity of the MBD proteins used for capture are variables affecting the pull-down of methylated DNA fragments. As such, we varied these two parameters systematically and computed the pull-down fraction using equation 6 (Figure 5.1). In general, higher affinity MBDs and higher bead concentrations yield greater DNA pull-down. The MethylMiner® kit uses human MBD2 ($K_d \approx 6\times10^{-9}$ M) proteins and 6×10^6 beads per microgram of DNA; our model predicts 80-90% recovery of methylated DNA fragments from those inputs. Other enrichment kits use the MBDs of MBD1 and MeCP2 with binding affinities on the order of 10^{-7} M. One would, therefore, expect these kits to be less sensitive than MethylMiner® based on the model prediction. We previously reported engineering a human MBD2-derived variant with a dissociation constant of 3×10^{-9} M. This affinity improvement yields approximately a 5% improvement in pull-down over more than an order of magnitude of bead loadings (5×10^5-10^7/reaction); these benefits become more significant with decreasing DNA input.

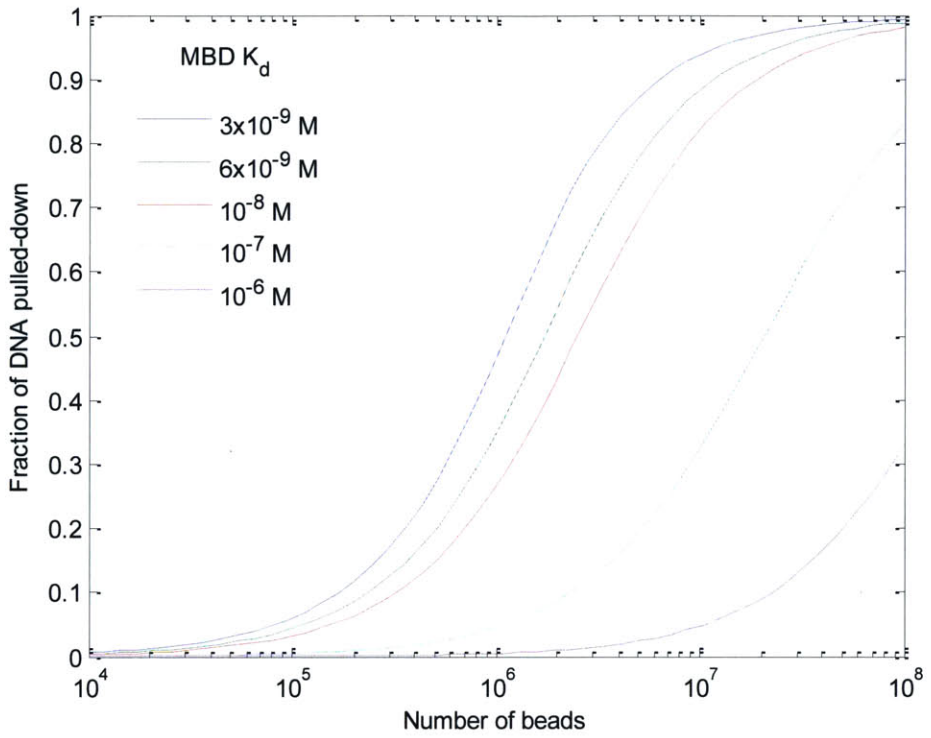


Figure 5.1: The DNA pull-down fraction is a function of both MBD-meDNA binding affinity and the number of beads used for the capture reaction.

Equilibrium constants have been shown to vary by orders of magnitude across the various members of the MBD family (*19*) and have a functional dependence on the absolute number of methylated CpG sites (*21*). Because the true value of K_i is not experimentally known for every DNA fragment with any possible combination of methylated and unmethylated CpG dinucleotide sites, an effective equilibrium constant is calculated

$$\bar{K}_i = \sum_{j=1}^{n} K_j \tag{7}$$

as the sum of the binding constants for each of the "*j*" CpG dinucleotides (methylated or unmethylated) contained in the DNA fragment of type "*i*" (*152*). Using this method, we compared effective equilibrium constants computed using our model to experimentally measured values for DNA binding to human MBD2 (Figure 5.2).

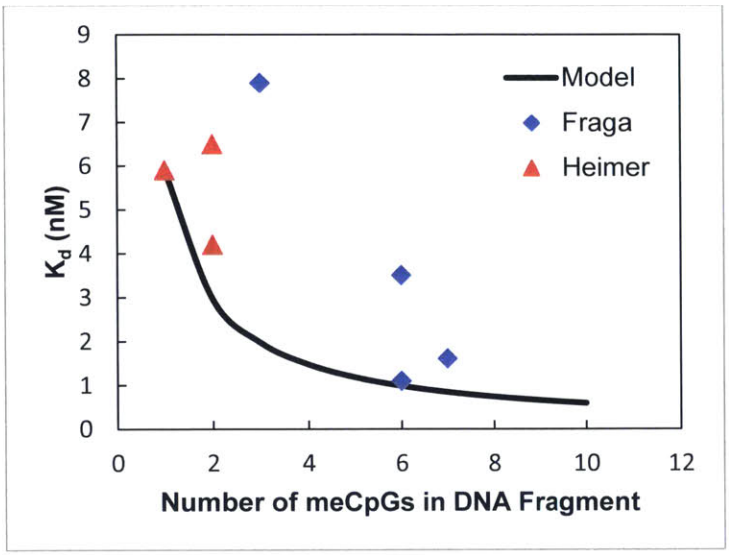


Figure 5.2: Human MBD2 equilibrium dissociation constant as a function of the number of methylated CpGs for experimental (Fraga *et al.*, Heimer & Sikes) and model (this work) systems.

While these predicted values correlate well for some numbers of methylated CpGs, it is particularly important to note that significant differences arise when comparing fragments having the same number of methylated CpGs. This observation has two implications: 1. MBD binding affinity appears to be a function of variables in addition to the absolute number of methylated CpGs (i.e. meCpG spacing or context), and 2. Further work is necessary to develop a more rigorous method for accurately estimating MBD binding affinities to DNA fragments in the absence of experimental data. Replacing the individual K_i value in equation 6 with an effective equilibrium constant calculated using equation 7 allowed us to model the effect the number of methylated and unmethylated CpG dinucleotides has on DNA binding to bead-immobilized MBD proteins in isolation from other pull-down variables. We calculated an effective binding constant for each of 147 DNA fragments having every possible combination of 0-6 methylated and 0-20 unmethylated CpG dinucleotides and computed the pull-down fraction for each fragment type (Figure 5.3).

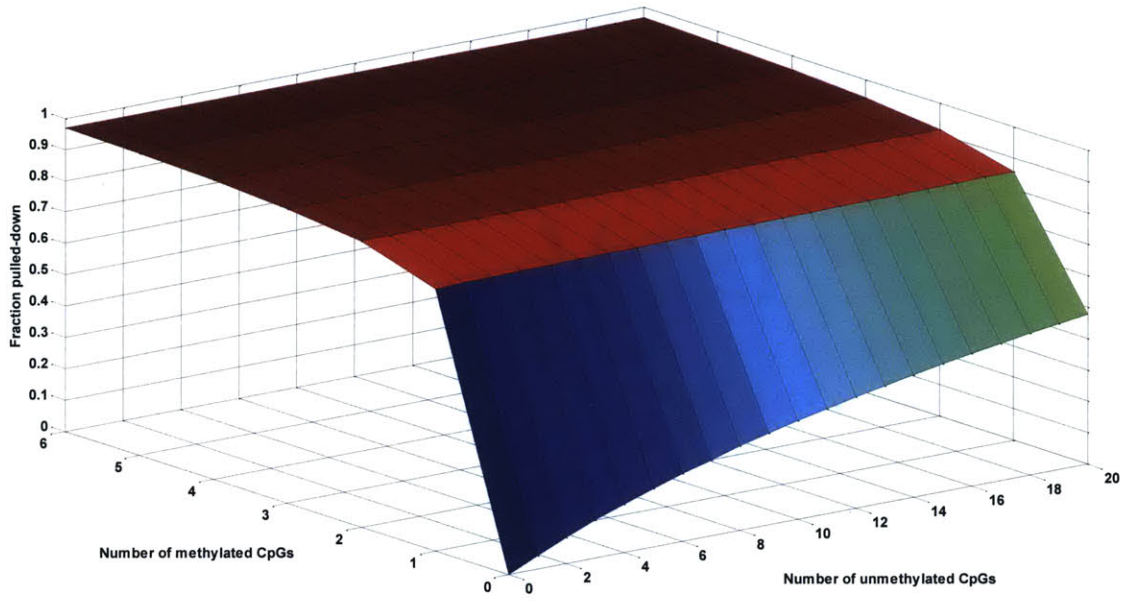


Figure 5.3: The pull-down fraction varies with the number of methylated and unmethylated CpGs in the input DNA fragments. In the absence of a methylated CpG, the pull-down fraction varies significantly with the number of unmethylated CpGs; however, after the first methylated CpG, additional methylated or unmethylated CpGs have little to no effect on the pull down fraction.

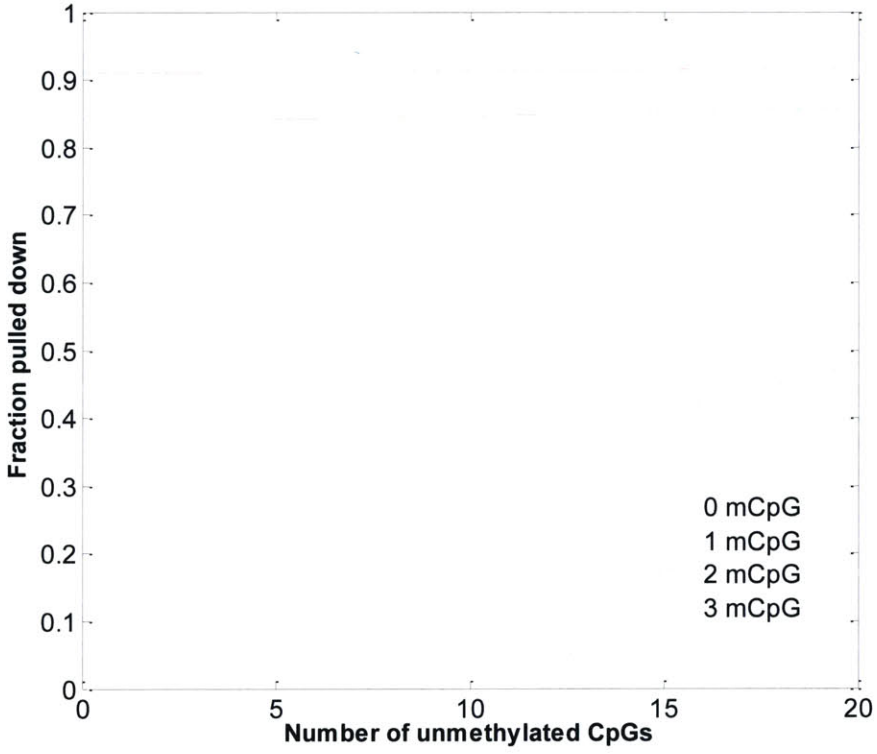


Figure 5.4: DNA fragment pull down fractions as a function of increasing unmethylated CpG content for fragments having no, one, two, or three methylated CpGs.

These results indicate that only in the absence of a methylated CpG will the pull-down fraction be a strong function of the number of unmethylated CpGs (Figure 5.4). Further, the greatest increase in the pull-down fraction is achieved by adding a single methylated CpG to an umethylated fragment, and after reaching two methylated CpGs per fragment, there appears to be negligible change in the pull-down fraction. This clearly has implications for MBD-seq based approaches which could potentially pull-down fragments with dense, unmethylated CpGs such as CpG islands for actively transcribed genes; doing so could produce inaccurate methylation signatures. As such, performing a low-salt concentration wash immediately after capture to liberate these weakly bound fragments has the potential to ensure the accuracy of data obtained from sequencing pulled-down DNA.

In order to model pull-down with the heterogeneous mixture of fragments resulting from sonicating genomic DNA, we treated each fragment type as a different species based on the number of methylated CpGs such that the complex surface concentration for each species "*i*" follows the expression

$$[MD_i] = \bar{K}_i [M][D_i] \tag{8}$$

which governs the thermodynamics of DNA binding to bead-immobilized MBD proteins. The total MBD occupancy is the sum of the concentrations of each MBD-DNA complex for all fragment types. When equation 8 is solved simultaneously for each fragment type "*i*" along with the MBD and DNA mole balance equations 4 and 5, respectively, we can compute the mole fraction of each fragment type relative to all pulled-down fragments. These mole fractions can be plotted versus the number of methylated CpGs contained in each DNA fragment to produce a profile of the DNA fragments pulled down (Figure 5.5). These results show that the majority of the fragments pulled-down have 1-3 methylated CpGs and qualitatively match experimental data collected using the MethylMiner® kit reported by De Meyer *et al.* (*151*). The total DNA recovery (defined as the percent of input DNA pulled down) computed using the model is approximately 65% or 4- to 5-fold greater than the experimental results reported by De Meyer *et al.* This discrepancy may be due to overestimation of either: 1. Fragment binding because the model does not provide for multivalent surface interactions of multi-methylated fragments which could otherwise potentially occlude binding of lesser methylated

fragments or 2. The number of methylated fragments in the input DNA due to the relatively crude approximation used here.

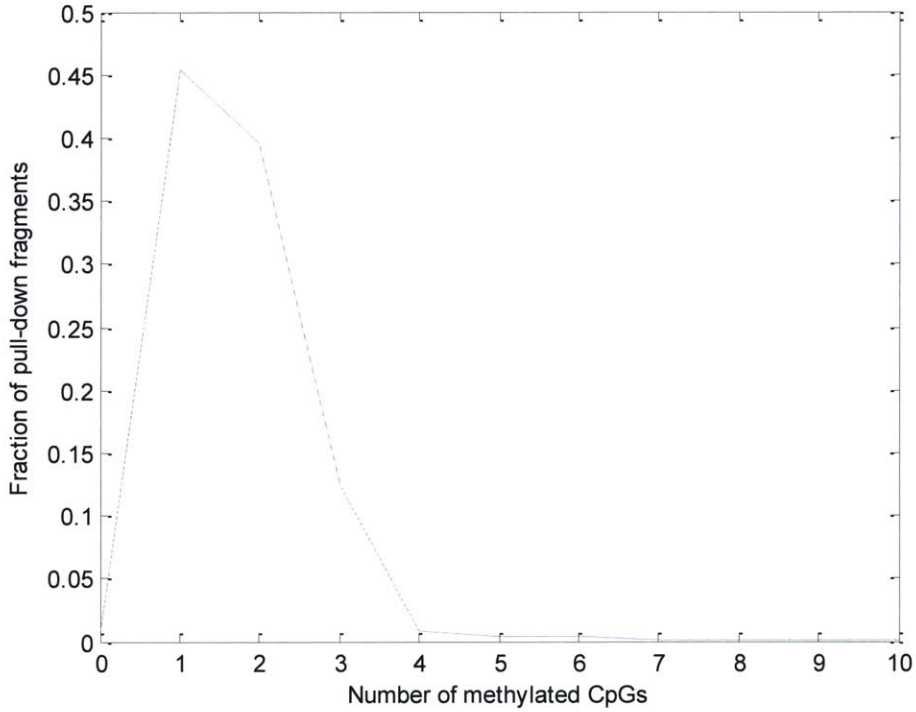


Figure 5.5: Model pull-down results. The fraction of captured fragments is shown as a function of methylated CpG content.

The recovery was also computed for each fragment type and shows a monotonic and asymptotic approach to unity with increasing methylated CpG content (Figure 5.6). Taken together, Figure 5.5 and Figure 5.6 show that the low prevalence of highly methylated fragments pulled-down is not because they suffer from low recovery but rather are present at a much lower concentration relative to other fragments in the overall pull-down reaction. These data are consistent for competitive-type binding reactions where the more densely methylated DNA fragments with higher effective binding constants "out compete" those with lower affinities and are, therefore, recovered at higher rates (Figure 5.6).

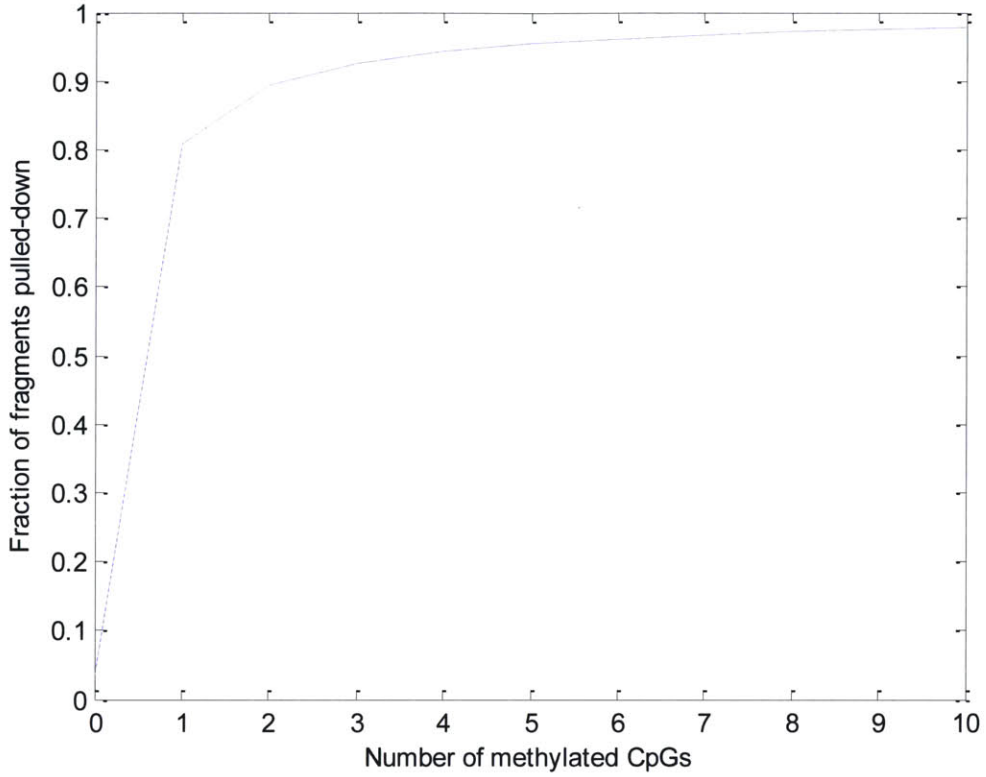


Figure 5.6: The fraction of fragments pulled-down from the bulk increases monotonically with the number of methylated CpGs contained in the DNA fragment.

5.4 Conclusions

Here, we have described a model useful for quantitatively estimating the effect of key parameters on the type and quantity of DNA captured in MBD-seq pull-down reactions. MBD2 and MBD2-derived mutants are expected to capture nearly 10-fold more DNA than lower affinity MBDs at the same bead loading. Of fragments pulled down, the presence of unmethylated CpGs increases the likelihood that a fragment will be pulled-down only in the absence of a methylated CpG which suggests that a low concentration salt wash would significantly reduce the number of false methylation reads. Conversely, greater than 90% of fragments with more than two methylated CpGs are pulled-down under the conditions studied.

Our results predict that high affinity MBD reagents capture nearly all methylated fragments from heterogeneous genomic samples. The vast majority of these fragments, however, have fewer than three methylated CpGs indicating that MBD2-derived kits are best suited for analyzing regions of less dense methylation unless a series of elutions with increasing salt concentration are used. This contrasts with the MeCP2-based MethylCap® kit which exhibits a pull-down profile biased toward more highly methylated DNA (*151*). Our model results indicate these lower-affinity MBDs ($K_d \approx 10^{-7}$ nM) (*19*) are quite poor for capturing fragments with few methylated CpGs; however, binding avidity effects may be what allow highly methylated fragments to be retained despite low-affinity individual MBD-meCpG interactions. While generally consistent with experimental data, model accuracy could be improved with further work in three areas: 1. Developing a more rigorous method for calculating the effective binding affinity constant, 2. Simulating multivalent DNA-MBD interactions, and 3. Computing input DNA fragment composition using full methylome sequencing data for a well-studied cell line.

Chapter 6

Conclusions and Future Directions

6.1 Conclusions

Promoter CpG island hypermethylation has been shown to be not only a causal event in the development of cancer but also a biomarker useful for informing clinical decisions; however, essentially all currently used methods rely on sodium bisulfite treatment of DNA to provide specific detection of methylated cytosine. Such treatment degrades >90% of the input DNA and can lead to inaccurate results if unmethylated cytosine bases are not converted to uracil in the chemical reaction. Further, all but one of the top cancer hospitals in the US reported the bisulfite conversion as the aspect of their DNA methylation analysis they would most like to have changed and/or eliminated. As such, I sought an alternative method that requires neither bisulfite conversion of DNA nor PCR.

The goal of this thesis was to develop clinically useful analytical tools to provide rapid, low-cost, and sensitive detection of methylated DNA. In doing so, I engineered the human MBD2 protein for improved binding affinity and selectivity for methylated CpG dinucleotides using yeast surface display. I further optimized the soluble expression and purification of MBD-green fluorescent protein fusions in *E. coli*. Finally, I demonstrated these proteins as useful reagents for sequence-specific detection of cytosine methylation in DNA from the *MGMT* gene on a biochip with either fluorescent or colorimetric (visible hydrogel) readout.

6.1.1 Characterization and Directed Evolution of a Methyl Binding Domain Protein

Previous work identified a family of proteins that share a conserved MBD domain that specifically recognizes methylated DNA (*13*). In Chapter 2, I described the successful yeast surface display of all the core members MBD1, MBD2, and MBD4 except for MeCP2 which exhibited a truncation. The human MBD2 was selected for mutagenesis and affinity matured for binding to methylated DNA. By introducing five amino acid substitutions, I improved the equilibrium dissociation constant to singly methylated DNA approximately two-fold to 3.1±1.0 nM with negligible change to binding in solution binding experiments. The most prevalent of these mutations, K161R, occurs away from the DNA binding site and improves the stability between the N and C-termini of the

protein by forming a new hydrogen bond. The F208Y and L170R mutations add new non-covalent interactions with the bound DNA strand. And, two mutations to isoleucine at the 175 and 187 positions form new hydrophobic interactions stabilizing the β-sheet and β-sheet/α-helix interface, respectively, in the vicinity of the DNA binding regions.

I expressed the best variant as a GFP fusion and evaluated it in an interfacial binding application where the measured binding affinity decreased approximately 6-fold. To restore the solution binding affinity, I concatenated the improved MBD variant from one to three repeat units. I further showed how these high-affinity agents can bind a significantly higher fraction of methylated CpG sites than existing MBD variants. Achieving high fractional coverages will be particularly beneficial for maximizing signal in applications where the total number of methylated CpGs is few e.g. small samples, single DNA molecules, and site-specific analyses. Eventually, these reagents could enable the development of technologies integrating DNA methylation analysis into the current pathology workflow which typically consists of thin tissue sections and fine needle aspirates.

6.1.2 Expression and Purification Optimization for an MBD-GFP Fusion in E. coli

Protein reagents can significantly increase the cost of producing biosensors and clinical diagnostics. Previous reports expressing MBD and MBD-GFP fusion proteins in BL21 (DE3) *E. coli* reported low yields and required denaturing purification conditions (*21, 120*). As such, I sought to optimize the heterologous expression of a recombinant MGD-GFP fusion protein in *E. coli* and provide conditions for the high-level soluble expression and single-step purification of this protein.

SDS–PAGE analysis and fluorescent microscopy were used to qualitatively and quantitatively compare the soluble yield for expression conditions across various induction temperatures, strengths (i.e. IPTG concentrations), and times. The induction temperature and time exhibited the strongest correlation with both soluble and overall expression of the MBD-GFP constructs. The optimal expression conditions identified through this study were 20°C for 16 h with 0.05 mM IPTG compared to previous reports of 37°C for 3 h and 1 mM IPTG resulting in a 65-fold increase in purified MBD-GFP yield

(58.5 ± 0.7 mg/ L). The increase in soluble yield was the primary factor enabling the single-step purification because denaturation was not required to extract protein aggregated in inclusion bodies. A benzonase nuclease was included in the lysis buffer to digest DNA and prevent MBD binding to *E. coli* host cell DNA. Together these two findings aided my efforts to utilize MBD proteins in methylation profiling assays with minimal investment of resources.

6.1.3 Hybridization Based Epigenotyping Using an MBD-GFP Protein

The foundational goal of this thesis was to develop a novel technology for detecting methylated DNA without requiring either bisulfite conversion or PCR. To do so, I designed a simple method for directly hybridizing and detecting methylated DNA fragments on a biochip. It requires less time to perform than the existing methods which rely on bisulfite conversion of unmethylated cytosine bases, PCR, and/or next-generation DNA sequencing. It has a low cost per test which I was able to achieve by using inexpensive, agarose coated glass test surfaces and a recombinant MBD protein that can be produced efficiently in *E. coli*. Fluorescence detection requires a microarray scanner for readout; however, the colorimetric assay requires only a green light source and digital camera for imaging the visible, stained hydrogels.

I successfully detected 22 bp singly and doubly methylated oligonucleotides derived from the *MGMT* gene at concentrations as low as 1 nM and 0.3 nM, respectively. In an attempt to improve mass transfer of DNA to the surface-immobilized probes, I implemented a recirculating microfluidic device which improved DNA hybridization by approximately one-third; however, MBD binding did not improve similarly. Quantitative analysis revealed MBD binding was only partially reaction limited and a weak function of flow rate through the microfluidic device. With the reported limit of detection and required sample volume for use in the system, this method requires the DNA from more than 10^9 cells derived from approximately 4 g of cells. As such, sensitivity is the major limitation of this method and would need to be improved for it to be useful in clinical applications.

Advantages of this method include the flexibility to interrogate a specific gene or panel of genes by simply designing unique capture probes specific to the sequences of interest. Further, microfluidic integration provides continuous delivery of test solutions, standardizes the assay, and provides a platform well-suited for automation.

6.2 Future Directions

This thesis has made significant contributions in two areas: 1. yeast surface display and engineering of DNA binding proteins and 2. bisulfite conversion and PCR free DNA methylation detection. Through the protein engineering, I have developed a high-affinity MBD protein and characterized the structural implications of the mutations driving the improvement improving our understanding of methylated DNA detection. Further, I have described the advantages, disadvantages, and possible uses for these proteins in DNA methylation analysis. This section describes future research directions that build upon the advances of MBD engineering and DNA methylation detection.

6.2.1 Engineered Methyl Binding Domain Proteins

The best MBD2-derived variant I have described in this thesis has the highest reported affinity for a protein that specifically binds methylated, double-stranded DNA. From my work, it seems unlikely that further, significant binding affinity enhancement can be achieved through additional rounds of positive selection on MBD2-derived mutants. It is possible, however, that by further exploring the MBD fitness landscape via mutagenesis of another parent or the adding rounds of negative selection could produce a higher affinity MBD protein than what I have reported here.

In addition to ways I have already described above, this engineered MBD has several, potentially interesting uses. Many researchers perform an enrichment step to concentrate methylated DNA before sequencing. In one variation of this approach, MBD-seq, MBD proteins immobilized on beads selectively bind methylated DNA which can be subsequently eluted for analysis. Some have already approached us wishing to

license our MBD variants for use in these types of applications. Similarly, it is also conceivable one could use the engineered MBD proteins displayed on EBY100 *S. cerevisiae* yeast cells for the very same purpose. Such application would constitute a self-replenishing pull-down reagent in which the yeast cell replaces relatively costly, functionalized beads. A third potential application for a population of these cells could be to bind fractionated, genomic DNA hybridized with promoter-specific, fluorescently labeled probe oligonucleotides. As labeled DNA fragments bind to yeast surface displayed MBD proteins, each cell will exhibit a fluorescent signal easily detectible using conventional flow cytometry. Using multiple probes and colors will facilitate multiplexed DNA methylation analysis.

6.2.2 Clinically Applicable DNA Methylation Analysis

Chapter 4 identified the sensitivity of the direct-hybridization, biochip based DNA methylation analysis method to be a limiting factor for its utility in the clinic. In order to provide sensitive, promoter-specific methylation analysis, a histology-based approach that draws upon the successful engineering of MBD proteins could be used to test for DNA methylation in formalin fixed paraffin embedded (FFPE) thin sections ubiquitous in pathology labs. Fluorescence in situ hybridization (FISH) has been well established as a method for visualizing specific DNA features within individual cells, and FISH assays are already used by pathologists for clinical diagnostics. More specifically, algorithms for designing multiple, singly-labeled FISH probes have been developed to provide sequence-specific detection in individual nucleic acid molecules (*153*). Results can be obtained within a day, and FISH has been used to distinguish between well differentiated liposarcoma and benign tumors and to test for gene rearrangement that is characteristic of certain types of prostate cancer using only a core needle biopsy, a sample that is too small to give reliable results using traditional imaging methods (*154, 155*). In collaboration with Brooke Tam and my thesis advisor Prof. Hadley Sikes, we have proposed a method to epigenotype single-cells that uses a traditional FISH assay with fluorescently labeled oligonucleotide probes to identify gene-specific CpG islands. When combined with MBD protein-based DNA methylation analysis, co-localization of

the fluorescent signals from FISH probe hybridization and MBD binding, or lack thereof, will indicate the methylation status of the gene of interest.

Chapter 1, Section 1.4.3 reviewed several methods that have been recently proposed as clinically viable alternatives to the bisulfite conversion-based methods. Two of these methods were label-free, optical biosensors. Based on data the authors presented comparing MBD proteins to antibodies, the engineered MBD proteins I described in Chapter 2 would likely improve the sensitivity of these biosensors. Further these engineered MBD proteins (81-fold higher than the MBD1) could be integrated into the single-DNA molecule assay described by Cipriany *et al.* (*66, 67*). Sequence specificity could be achieved if fluorescently labeled (and specifically methylated) probes were introduced and hybridized before MBD binding. Co-localization of probe and MBD fluorescence would enable direct sorting using the multiplexed nanofluidic system the authors describe.

References

1. A. Bird, Perceptions of epigenetics. *Nature* **447**, 396-398 (2007).
2. P. A. Jones, S. B. Baylin, The Epigenomics of Cancer. *Cell* **128**, 683-692 (2007).
3. S. Saxonov, P. Berg, D. L. Brutlag, A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 1412-1417 (2006).
4. J. K. Kim, M. Samaranayake, S. Pradhan, Epigenetic mechanisms in mammals. *Cell. Mol. Life Sci.* **66**, 596-612 (2009).
5. M. G. Goll, T. H. Bestor, EUKARYOTIC CYTOSINE METHYLTRANSFERASES. *Annual Review of Biochemistry* **74**, 481-514 (2005).
6. X. Cheng, R. M. Blumenthal, Mammalian DNA Methyltransferases: A Structural Perspective. *Structure* **16**, 341-350 (2008).
7. R. Jaenisch, A. Bird, Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* **33**, 245-254 (2003).
8. A. Bird, DNA methylation patterns and epigenetic memory. *Genes & Development* **16**, 6-21 (2002).
9. P. W. Laird, R. Jaenisch, THE ROLE OF DNA METHYLATION IN CANCER GENETICS AND EPIGENETICS. *Annual Review of Genetics* **30**, 441-464 (1996).
10. E. Li, T. H. Bestor, R. Jaenisch, Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915-926 (1992).
11. T. Mohandas, R. Sparkes, L. Shapiro, Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. *Science* **211**, 393-396 (1981).
12. A. P. Wolffe, M. A. Matzke, Epigenetics: Regulation Through Repression. *Science* **286**, 481-486 (1999).
13. B. Hendrich, A. Bird, Identification and Characterization of a Family of Mammalian Methyl-CpG Binding Proteins. *Mol. Cell. Biol.* **18**, 6538-6547 (1998).

14. I. Ohki, N. Shimotake, N. Fujita, J.-G. Jee, T. Ikegami, M. Nakao, M. Shirakawa, Solution Structure of the Methyl-CpG Binding Domain of Human MBD1 in Complex with Methylated DNA. *Cell* **105**, 487-497 (2001).
15. P. A. Wade, A. Gegonne, P. L. Jones, E. Ballestar, F. Aubry, A. P. Wolffe, Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. *Nat Genet* **23**, 62-66 (1999).
16. M. Fatemi, P. A. Wade, MBD family proteins: reading the epigenetic code. *Journal of Cell Science* **119**, 3033-3037 (2006).
17. J. N. Scarsdale, H. D. Webb, G. D. Ginder, D. C. Williams, Solution structure and dynamic analysis of chicken MBD2 methyl binding domain bound to a target-methylated DNA sequence. *Nucleic Acids Research*, (2011).
18. K. L. Ho, I. W. McNae, L. Schmiedeberg, R. J. Klose, A. P. Bird, M. D. Walkinshaw, MeCP2 Binding to DNA Depends upon Hydration at Methyl-CpG. *Molecular cell* **29**, 525-531 (2008).
19. M. F. Fraga, E. Ballestar, G. Montoya, P. Taysavang, P. A. Wade, M. Esteller, The affinity of different MBD proteins for a specific methylated locus depends on their intrinsic binding properties. *Nucleic Acids Research* **31**, 1765-1774 (2003).
20. H. F. Jørgensen, K. Adie, P. Chaubert, A. P. Bird, Engineering a high-affinity methyl-CpG-binding protein. *Nucleic Acids Research* **34**, e96 (2006).
21. Y. Yu, S. Blair, D. Gillespie, R. Jensen, D. Myszka, A. H. Badran, I. Ghosh, A. Chagovetz, Direct DNA Methylation Profiling Using Methyl Binding Domain Proteins. *Analytical Chemistry* **82**, 5012-5019 (2010).
22. J. G. Herman, S. B. Baylin, Gene Silencing in Cancer in Association with Promoter Hypermethylation. *New England Journal of Medicine* **349**, 2042-2054 (2003).
23. A. P. Feinberg, B. Vogelstein, Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* **301**, 89-92 (1983).
24. A. P. Feinberg, C. W. Gehrke, K. C. Kuo, M. Ehrlich, Reduced Genomic 5-Methylcytosine Content in Human Colonic Neoplasia. *Cancer Research* **48**, 1159-1161 (1988).
25. P. A. Jones, S. B. Baylin, The fundamental role of epigenetic events in cancer. *Nat Rev Genet* **3**, 415-428 (2002).
26. M. F. Fraga, M. Herranz, J. Espada, E. Ballestar, M. F. Paz, S. Ropero, E. Erkek, O. Bozdogan, H. Peinado, A. Niveleau, J.-H. Mao, A. Balmain, A. Cano, M. Esteller, A Mouse Skin Multistage Carcinogenesis Model Reflects the Aberrant DNA Methylation Patterns of Human Tumors. *Cancer Research* **64**, 5527-5534 (2004).
27. F. Gaudet, J. G. Hodgson, A. Eden, L. Jackson-Grusby, J. Dausman, J. W. Gray, H. Leonhardt, R. Jaenisch, Induction of Tumors in Mice by Genomic Hypomethylation. *Science* **300**, 489-492 (2003).
28. M. Esteller, Epigenetics in Cancer. *New England Journal of Medicine* **358**, 1148-1159 (2008).
29. A. G. Knudson, Two genetic hits (more or less) to cancer. *Nat Rev Cancer* **1**, 157-162 (2001).
30. M. Esteller, M. F. Fraga, M. Guo, J. Garcia-Foncillas, I. Hedenfalk, A. K. Godwin, J. Trojan, C. Vaurs-Barrière, Y.-J. Bignon, S. Ramus, J. Benitez, T. Caldes, Y. Akiyama, Y. Yuasa, V. Launonen, M. J. Canal, R. Rodriguez, G. Capella, M. A. Peinado, A. Borg, L. A. Aaltonen, B. A. Ponder, S. B. Baylin, J. G. Herman, DNA methylation patterns in hereditary human cancers mimic sporadic tumorigenesis. *Human Molecular Genetics* **10**, 3001-3007 (2001).
31. V. Greger, E. Passarge, W. Höpping, E. Messmer, B. Horsthemke, Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma. *Human Genetics* **83**, 155-158 (1989).

32. T. Sakai, J. Toguchida, N. Ohtani, D. Yandell, J. Rapaport, T. P. Dryja, Allele-specific hypermethylation of the retinoblastoma tumor-suppressor gene. *Am J Hum Genet* **48**, 880-888 (1991).
33. J. G. Herman, F. Latif, Y. Weng, M. I. Lerman, B. Zbar, S. Liu, D. Samid, D. S. Duan, J. R. Gnarra, W. M. Linehan, Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *Proceedings of the National Academy of Sciences* **91**, 9700-9704 (1994).
34. A. Merlo, J. G. Herman, L. Mao, D. J. Lee, E. Gabrielson, P. C. Burger, S. B. Baylin, D. Sidransky, 5[prime] CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers. *Nat Med* **1**, 686-692 (1995).
35. J. G. Herman, A. Umar, K. Polyak, J. R. Graff, N. Ahuja, J.-P. J. Issa, S. Markowitz, J. K. V. Willson, S. R. Hamilton, K. W. Kinzler, M. F. Kane, R. D. Kolodner, B. Vogelstein, T. A. Kunkel, S. B. Baylin, Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proceedings of the National Academy of Sciences* **95**, 6870-6875 (1998).
36. M. Esteller, J. M. Silva, G. Dominguez, F. Bonilla, X. Matias-Guiu, E. Lerma, E. Bussaglia, J. Prat, I. C. Harkes, E. A. Repasky, E. Gabrielson, M. Schutte, S. B. Baylin, J. G. Herman, Promoter Hypermethylation and BRCA1 Inactivation in Sporadic Breast and Ovarian Tumors. *Journal of the National Cancer Institute* **92**, 564-569 (2000).
37. J. Mehrotra, M. Vali, M. McVeigh, S. L. Kominsky, M. J. Fackler, J. Lahti-Domenici, K. Polyak, N. Sacchi, E. Garrett-Mayer, P. Argani, S. Sukumar, Very High Frequency of Hypermethylated Genes in Breast Cancer Metastasis to the Bone, Brain, and Lung. *Clinical Cancer Research* **10**, 3104-3109 (2004).
38. I. Krop, M. T. Parker, N. Bloushtain-Qimron, D. Porter, R. Gelman, H. Sasaki, M. Maurer, M. B. Terry, R. Parsons, K. Polyak, HIN-1, an Inhibitor of Cell Growth, Invasion, and AKT Activation. *Cancer Research* **65**, 9659-9669 (2005).
39. Y. Kudo, S. Kitajima, I. Ogawa, M. Hiraoka, S. Sargolzaei, M. R. Keikhaee, S. Sato, M. Miyauchi, T. Takata, Invasion and Metastasis of Oral Cancer Cells Require Methylation of E-Cadherin and/or Degradation of Membranous β-Catenin. *Clinical Cancer Research* **10**, 5455-5463 (2004).
40. M. Esteller, P. G. Corn, S. B. Baylin, J. G. Herman, A Gene Hypermethylation Profile of Human Cancer. *Cancer Research* **61**, 3225-3229 (2001).
41. H. Heyn, M. Esteller, DNA methylation profiling in the clinic: applications and challenges. *Nat Rev Genet* **13**, 679-692 (2012).
42. M. E. Hegi, A.-C. Diserens, T. Gorlia, M.-F. Hamou, N. de Tribolet, M. Weller, J. M. Kros, J. A. Hainfellner, W. Mason, L. Mariani, J. E. C. Bromberg, P. Hau, R. O. Mirimanoff, J. G. Cairncross, R. C. Janzer, R. Stupp, MGMT Gene Silencing and Benefit from Temozolomide in Glioblastoma. *New England Journal of Medicine* **352**, 997-1003 (2005).
43. D. P. Silver, A. L. Richardson, A. C. Eklund, Z. C. Wang, Z. Szallasi, Q. Li, N. Juul, C.-O. Leong, D. Calogrias, A. Buraimoh, A. Fatima, R. S. Gelman, P. D. Ryan, N. M. Tung, A. De Nicolo, S. Ganesan, A. Miron, C. Colin, D. C. Sgroi, L. W. Ellisen, E. P. Winer, J. E. Garber, Efficacy of Neoadjuvant Cisplatin in Triple-Negative Breast Cancer. *Journal of Clinical Oncology* **28**, 1145-1153 (2010).
44. X. Tang, F. R. Khuri, J. J. Lee, B. L. Kemp, D. Liu, W. K. Hong, L. Mao, Hypermethylation of the Death-Associated Protein (DAP) Kinase Promoter and Aggressiveness in Stage I Non-Small-Cell Lung Cancer. *Journal of the National Cancer Institute* **92**, 1511-1516 (2000).
45. M. Esteller, S. González, R. A. Risques, E. Marcuello, R. Mangues, J. R. Germà, J. G. Herman, G. Capellà, M. A. Peinado, K-ras and p16 Aberrations Confer Poor Prognosis in Human Colorectal Cancer. *Journal of Clinical Oncology* **19**, 299-304 (2001).
46. D. A. Ahlquist, H. Zou, M. Domanico, D. W. Mahoney, T. C. Yab, W. R. Taylor, M. L. Butz, S. N. Thibodeau, L. Rabeneck, L. F. Paszat, K. W. Kinzler, B. Vogelstein, N. C. Bjerregaard, S. Laurberg,

H. T. Sørensen, B. M. Berger, G. P. Lidgard, Next-Generation Stool DNA Test Accurately Detects Colorectal Cancer and Large Adenomas. *Gastroenterology* **142**, 248-256 (2012).
47. L. Van Neste, J. G. Herman, G. Otto, J. W. Bigley, J. I. Epstein, W. Van Criekinge, The Epigenetic promise for prostate cancer diagnosis. *The Prostate* **72**, 1248-1261 (2012).
48. P. J. Bastian, J. Ellinger, A. Wellmann, N. Wernert, L. C. Heukamp, S. C. Müller, A. von Ruecker, Diagnostic and Prognostic Information in Prostate Cancer with the Help of a Small Set of Hypermethylated Gene Loci. *Clinical Cancer Research* **11**, 4097-4106 (2005).
49. K. D. Siegmund, P. W. Laird, I. A. Laird-Offringa, A comparison of cluster analysis methods using DNA methylation data. *Bioinformatics* **20**, 1896-1904 (2004).
50. M. Alaminos, V. Davalos, N.-K. V. Cheung, W. L. Gerald, M. Esteller, Clustering of Gene Hypermethylation Associated With Clinical Risk Groups in Neuroblastoma. *Journal of the National Cancer Institute* **96**, 1208-1219 (2004).
51. C. Roadmap Epigenomics, A. Kundaje, W. Meuleman, J. Ernst, M. Bilenky, A. Yen, A. Heravi-Moussavi, P. Kheradpour, Z. Zhang, J. Wang, M. J. Ziller, V. Amin, J. W. Whitaker, M. D. Schultz, L. D. Ward, A. Sarkar, G. Quon, R. S. Sandstrom, M. L. Eaton, Y.-C. Wu, A. R. Pfenning, X. Wang, M. Claussnitzer, Y. Liu, C. Coarfa, R. A. Harris, N. Shoresh, C. B. Epstein, E. Gjoneska, D. Leung, W. Xie, R. D. Hawkins, R. Lister, C. Hong, P. Gascard, A. J. Mungall, R. Moore, E. Chuah, A. Tam, T. K. Canfield, R. S. Hansen, R. Kaul, P. J. Sabo, M. S. Bansal, A. Carles, J. R. Dixon, K.-H. Farh, S. Feizi, R. Karlic, A.-R. Kim, A. Kulkarni, D. Li, R. Lowdon, G. Elliott, T. R. Mercer, S. J. Neph, V. Onuchic, P. Polak, N. Rajagopal, P. Ray, R. C. Sallari, K. T. Siebenthall, N. A. Sinnott-Armstrong, M. Stevens, R. E. Thurman, J. Wu, B. Zhang, X. Zhou, A. E. Beaudet, L. A. Boyer, P. L. De Jager, P. J. Farnham, S. J. Fisher, D. Haussler, S. J. M. Jones, W. Li, M. A. Marra, M. T. McManus, S. Sunyaev, J. A. Thomson, T. D. Tlsty, L.-H. Tsai, W. Wang, R. A. Waterland, M. Q. Zhang, L. H. Chadwick, B. E. Bernstein, J. F. Costello, J. R. Ecker, M. Hirst, A. Meissner, A. Milosavljevic, B. Ren, J. A. Stamatoyannopoulos, T. Wang, M. Kellis, Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330 (2015).
52. C. Bock, E. M. Tomazou, A. B. Brinkman, F. Muller, F. Simmer, H. Gu, N. Jager, A. Gnirke, H. G. Stunnenberg, A. Meissner, Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nat Biotech* **28**, 1106-1114 (2010).
53. R. A. Harris, T. Wang, C. Coarfa, R. P. Nagarajan, C. Hong, S. L. Downey, B. E. Johnson, S. D. Fouse, A. Delaney, Y. Zhao, A. Olshen, T. Ballinger, X. Zhou, K. J. Forsberg, J. Gu, L. Echipare, H. O'Geen, R. Lister, M. Pelizzola, Y. Xi, C. B. Epstein, B. E. Bernstein, R. D. Hawkins, B. Ren, W.-Y. Chung, H. Gu, C. Bock, A. Gnirke, M. Q. Zhang, D. Haussler, J. R. Ecker, W. Li, P. J. Farnham, R. A. Waterland, A. Meissner, M. A. Marra, M. Hirst, A. Milosavljevic, J. F. Costello, Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nat Biotech* **28**, 1097-1105 (2010).
54. J. G. Herman, J. R. Graff, S. Myöhänen, B. D. Nelkin, S. B. Baylin, Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proceedings of the National Academy of Sciences* **93**, 9821-9826 (1996).
55. C. Grunau, S. J. Clark, A. Rosenthal, Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. *Nucleic Acids Research* **29**, e65 (2001).
56. A. Okamoto, Chemical approach toward efficient DNA methylation analysis. *Organic & Biomolecular Chemistry* **7**, 21-26 (2009).
57. R. Lister, M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q.-M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A. H. Millar, J. A. Thomson, B. Ren, J. R. Ecker, Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315-322 (2009).

58. F. Diehl, M. Li, D. Dressman, Y. He, D. Shen, S. Szabo, L. A. Diaz, S. N. Goodman, K. A. David, H. Juhl, K. W. Kinzler, B. Vogelstein, Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 16368-16373 (2005).
59. D. P. Genereux, W. C. Johnson, A. F. Burden, R. Stöger, C. D. Laird, Errors in the bisulfite conversion of DNA: modulating inappropriate- and failed-conversion frequencies. *Nucleic Acids Research* **36**, e150 (2008).
60. M. L. Gonzalgo, C. P. Pavlovich, S. M. Lee, W. G. Nelson, Prostate Cancer Detection by GSTP1 Methylation Analysis of Postbiopsy Urine Specimens. *Clinical Cancer Research* **9**, 2673-2677 (2003).
61. P. Cairns, M. Esteller, J. G. Herman, M. Schoenberg, C. Jeronimo, M. Sanchez-Cespedes, N.-H. Chow, M. Grasso, L. Wu, W. B. Westra, D. Sidransky, Molecular Detection of Prostate Cancer in Urine by GSTP1 Hypermethylation. *Clinical Cancer Research* **7**, 2727-2730 (2001).
62. B. W. Heimer, T. A. Shatova, J. K. Lee, K. Kaastrup, H. D. Sikes, Evaluating the sensitivity of hybridization-based epigenotyping using a methyl binding domain protein. *Analyst* **139**, 3695-3701 (2014).
63. K. Kaastrup, H. D. Sikes, Polymerization-based signal amplification under ambient conditions with thirty-five second reaction times. *Lab on a Chip* **12**, 4055-4058 (2012).
64. J. D. Suter, D. J. Howard, H. Shi, C. W. Caldwell, X. Fan, Label-free DNA methylation analysis using opto-fluidic ring resonators. *Biosensors and Bioelectronics* **26**, 1016-1020 (2010).
65. R. M. Hawk, A. M. Armani, Label free detection of 5'hydroxymethylcytosine within CpG islands using optical sensors. *Biosensors and Bioelectronics* **65**, 198-203 (2015).
66. B. R. Cipriany, R. Zhao, P. J. Murphy, S. L. Levy, C. P. Tan, H. G. Craighead, P. D. Soloway, Single Molecule Epigenetic Analysis in a Nanofluidic Channel. *Analytical Chemistry* **82**, 2480-2487 (2010).
67. B. R. Cipriany, P. J. Murphy, J. A. Hagarman, A. Cerf, D. Latulippe, S. L. Levy, J. J. Benítez, C. P. Tan, J. Topolancik, P. D. Soloway, H. G. Craighead, Real-time analysis and selection of methylated DNA by fluorescence-activated single molecule sorting in a nanofluidic channel. *Proceedings of the National Academy of Sciences* **109**, 8477-8482 (2012).
68. B. A. Flusberg, D. R. Webster, J. H. Lee, K. J. Travers, E. C. Olivares, T. A. Clark, J. Korlach, S. W. Turner, Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Meth* **7**, 461-465 (2010).
69. M. Li, W.-d. Chen, N. Papadopoulos, S. N. Goodman, N. C. Bjerregaard, S. Laurberg, B. Levin, H. Juhl, N. Arber, H. Moinova, K. Durkee, K. Schmidt, Y. He, F. Diehl, V. E. Velculescu, S. Zhou, L. A. Diaz Jr, K. W. Kinzler, S. D. Markowitz, B. Vogelstein, Sensitive digital quantification of DNA methylation in clinical samples. *Nat Biotech* **27**, 858-863 (2009).
70. V. J. Bailey, H. Easwaran, Y. Zhang, E. Griffiths, S. A. Belinsky, J. G. Herman, S. B. Baylin, H. E. Carraway, T.-H. Wang, MS-qFRET: A quantum dot-based method for analysis of DNA methylation. *Genome Research* **19**, 1455-1461 (2009).
71. H. Zou, H. Allawi, X. Cao, M. Domanico, J. Harrington, W. R. Taylor, T. Yab, D. A. Ahlquist, G. Lidgard, Quantification of Methylated Markers with a Multiplex Methylation-Specific Technology. *Clinical Chemistry* **58**, 375-383 (2012).
72. M. Farlik, Nathan C. Sheffield, A. Nuzzo, P. Datlinger, A. Schönegger, J. Klughammer, C. Bock, Single-Cell DNA Methylome Sequencing and Bioinformatic Inference of Epigenomic Cell-State Dynamics. *Cell Reports* **10**, 1386-1397 (2015).
73. T. Hashimshony, J. Zhang, I. Keshet, M. Bustin, H. Cedar, The role of DNA methylation in setting up chromatin structure during development. *Nat Genet* **34**, 187-192 (2003).

74. A. P. Feinberg, Phenotypic plasticity and the epigenetics of human disease. *Nature* **447**, 433-440 (2007).
75. P. W. Laird, Principles and challenges of genome-wide DNA methylation analysis. *Nat Rev Genet* **11**, 191-203 (2010).
76. K. R. Pomraning, K. M. Smith, M. Freitag, Genome-wide high throughput analysis of DNA methylation in eukaryotes. *Methods* **47**, 142-150 (2009).
77. J. Luo, W. Zheng, Y. Wang, Z. Wu, Y. Bai, Z. Lu, Detection method for methylation density on microarray using methyl-CpG-binding domain protein. *Analytical Biochemistry* **387**, 143-149 (2009).
78. K. Kaastrup, L. Chan, H. D. Sikes, Impact of Dissociation Constant on the Detection Sensitivity of Polymerization-Based Signal Amplification Reactions. *Analytical Chemistry* **85**, 8055-8060 (2013).
79. D. Wang, S. Bodovitz, Single cell analysis: the new frontier in 'omics'. *Trends in Biotechnology* **28**, 281-290 (2010).
80. E. Shapiro, T. Biezuner, S. Linnarsson, Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat Rev Genet* **14**, 618-630 (2013).
81. J. R. Porter, C. I. Stains, D. J. Segal, I. Ghosh, Split β-Lactamase Sensor for the Sequence-Specific Detection of DNA Methylation. *Analytical Chemistry* **79**, 6702-6708 (2007).
82. G. Chao, W. L. Lau, B. J. Hackel, S. L. Sazinsky, S. M. Lippow, K. D. Wittrup, Isolating and engineering human antibodies using yeast surface display. *Nat. Protocols* **1**, 755-768 (2006).
83. E. T. Boder, K. D. Wittrup, Optimal Screening of Surface-Displayed Polypeptide Libraries. *Biotechnology Progress* **14**, 55-62 (1998).
84. M. E. Boyd, B. W. Heimer, H. D. Sikes, Functional heterologous expression and purification of a mammalian methyl-CpG binding domain in suitable yield for DNA methylation profiling assays. *Protein Expression and Purification* **82**, 332-338 (2012).
85. V. Afanassiev, V. Hanemann, S. Wölfl, Preparation of DNA and protein micro arrays on glass slides coated with an agarose film. *Nucleic Acids Research* **28**, e66 (2000).
86. B. Hendrich, U. Hardeland, H.-H. Ng, J. Jiricny, A. Bird, The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. *Nature* **401**, 301-304 (1999).
87. X. Nan, R. R. Meehan, A. Bird, Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic Acids Research* **21**, 4886-4892 (1993).
88. E. V. Shusta, M. C. Kieke, E. Parke, D. M. Kranz, K. D. Wittrup, Yeast polypeptide fusion surface display levels predict thermal stability and soluble secretion efficiency. *Journal of Molecular Biology* **292**, 949-956 (1999).
89. J. J. Van Antwerp, K. D. Wittrup, Fine Affinity Discrimination by Yeast Surface Display and Flow Cytometry. *Biotechnology Progress* **16**, 31-37 (2000).
90. D. Lipovšek, S. M. Lippow, B. J. Hackel, M. W. Gregson, P. Cheng, A. Kapila, K. D. Wittrup, Evolution of an Interloop Disulfide Bond in High-Affinity Antibody Mimics Based on Fibronectin Type III Domain and Selected by Yeast Surface Display: Molecular Convergence with Single-Domain Camelid and Shark Antibodies. *Journal of Molecular Biology* **368**, 1024-1041 (2007).
91. M. Biasini, S. Bienert, A. Waterhouse, K. Arnold, G. Studer, T. Schmidt, F. Kiefer, T. G. Cassarino, M. Bertoni, L. Bordoli, T. Schwede, SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*, (2014).
92. R. P. Ghosh, R. A. Horowitz-Scherer, T. Nikitina, L. M. Gierasch, C. L. Woodcock, Rett Syndrome-causing Mutations in Human MeCP2 Result in Diverse Structural Changes That Impact Folding and DNA Interactions. *Journal of Biological Chemistry* **283**, 20523-20534 (2008).
93. M. M. Suzuki, A. Bird, DNA methylation landscapes: provocative insights from epigenomics. *Nature Reviews Genetics* **9**, 465-476 (2008).

94. A. Meissner, T. S. Mikkelsen, H. C. Gu, M. Wernig, J. Hanna, A. Sivachenko, X. L. Zhang, B. E. Bernstein, C. Nusbaum, D. B. Jaffe, A. Gnirke, R. Jaenisch, E. S. Lander, Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**, 766-U791 (2008).
95. K. D. Robertson, DNA methylation and human disease. *Nature Reviews Genetics* **6**, 597-610 (2005).
96. A. K. Maunakea, I. Chepelev, K. J. Zhao, Epigenome Mapping in Normal and Disease States. *Circulation Research* **107**, 327-339 (2010).
97. P. A. Jones, S. B. Baylin, The fundamental role of epigenetic events in cancer. *Nature Reviews Genetics* **3**, 415-428 (2002).
98. P. W. Laird, Cancer epigenetics. *Human Molecular Genetics* **14**, R65-R76 (2005).
99. M. Esteller, Molecular origins of cancer: Epigenetics in cancer. *New England Journal of Medicine* **358**, 1148-1159 (2008).
100. M. E. Maradeo, P. Cairns, Translational application of epigenetic alterations: Ovarian cancer as a model. *Febs Letters* **585**, 2112-2120 (2011).
101. J. G. Herman, S. B. Baylin, Mechanisms of disease: Gene silencing in cancer in association with promoter hypermethylation. *New England Journal of Medicine* **349**, 2042-2054 (2003).
102. M. Weber, I. Hellmann, M. B. Stadler, L. Ramos, S. Paabo, M. Rebhan, D. Schubeler, Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature Genetics* **39**, 457-466 (2007).
103. M. Esteller, Cancer epigenomics: DNA methylomes and histone-modification maps. *Nature Reviews Genetics* **8**, 286-298 (2007).
104. P. W. Laird, Principles and challenges of genome-wide DNA methylation analysis. *Nature Reviews Genetics* **11**, 191-203 (2010).
105. C. Bock, E. M. Tomazou, A. B. Brinkman, F. Muller, F. Simmer, H. C. Gu, N. Jager, A. Gnirke, H. G. Stunnenberg, A. Meissner, Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nature Biotechnology* **28**, 1106-U1196 (2010).
106. T. A. Down, V. K. Rakyan, D. J. Turner, P. Flicek, H. Li, E. Kulesha, S. Graf, N. Johnson, J. Herrero, E. M. Tomazou, N. P. Thorne, L. Backdahl, M. Herberth, K. L. Howe, D. K. Jackson, M. M. Miretti, J. C. Marioni, E. Birney, T. J. P. Hubbard, R. Durbin, S. Tavare, S. Beck, A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nature Biotechnology* **26**, 779-785 (2008).
107. M. Weber, J. J. Davies, D. Wittig, E. J. Oakeley, M. Haase, W. L. Lam, D. Schubeler, Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nature Genetics* **37**, 853-862 (2005).
108. L. G. Acevedo, M. Bieda, R. Green, P. J. Farnham, Analysis of the mechanisms mediating tumor-specific changes in gene expression in human liver tumors. *Cancer Research* **68**, 2641-2651 (2008).
109. B. Hendrich, A. Bird, Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Molecular and Cellular Biology* **18**, 6538-6547 (1998).
110. S. H. Cross, J. A. Charlton, X. S. Nan, A. P. Bird, PURIFICATION OF CPG ISLANDS USING A METHYLATED DNA-BINDING COLUMN. *Nature Genetics* **6**, 236-244 (1994).
111. C. Gebhard, L. Schwarzfischer, T. H. Pham, R. Andreesen, A. Mackensen, M. Rehli, Rapid and sensitive detection of CpG-methylation using methyl-binding (MB)-PCR. *Nucleic Acids Research* **34**, (2006).
112. T. Rauch, H. W. Li, X. W. Wu, G. P. Pfeifer, MIRA-assisted microarray analysis, a new technology for the determination of DNA methylation patterns, identifies frequent methylation of homeodomain-containing genes in lung cancer cells. *Cancer Research* **66**, 7939-7947 (2006).

113. X. S. Nan, R. R. Meehan, A. Bird, DISSECTION OF THE METHYL-CPG BINDING DOMAIN FROM THE CHROMOSOMAL PROTEIN MECP2. *Nucleic Acids Research* **21**, 4886-4892 (1993).
114. H. F. Jorgensen, A. Bird, MeCP2 and other methyl-CpG binding proteins. *Mental Retardation and Developmental Disabilities Research Reviews* **8**, 87-93 (2002).
115. I. Ohki, N. Shimotake, N. Fujita, M. Nakao, M. Shirakawa, Solution structure of the methyl-CpG-binding domain of the methylation-dependent transcriptional repressor MBD1. *Embo Journal* **18**, 6653-6661 (1999).
116. Y. N. Yu, S. Blair, D. Gillespie, R. Jensen, D. Myszka, A. H. Badran, I. Ghosh, A. Chagovetz, Direct DNA Methylation Profiling Using Methyl Binding Domain Proteins. *Analytical Chemistry* **82**, 5012-5019 (2010).
117. K. Inomata, I. Ohki, H. Tochio, K. Fujiwara, H. Hiroaki, M. Shirakawat, Kinetic and thermodynamic evidence for flipping of a methyl-CpG binding domain on methylated DNA. *Biochemistry* **47**, 3266-3271 (2008).
118. C. I. Stains, J. L. Furman, D. J. Segal, I. Ghosh, Site-specific detection of DNA methylation utilizing mCpG-SEER. *Journal of the American Chemical Society* **128**, 9761-9765 (2006).
119. J. F. Luo, W. L. Zheng, Y. Wang, Z. X. Wu, Y. F. Bai, Z. H. Lu, Detection method for methylation density on microarray using methyl-CpG-binding domain protein. *Analytical Biochemistry* **387**, 143-149 (2009).
120. H. F. Jorgensen, K. Adie, P. Chaubert, A. P. Bird, Engineering a high-affinity methyl-CpG-binding protein. *Nucleic Acids Research* **34**, (2006).
121. F. W. Studier, Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification* **41**, 207-234 (2005).
122. J. A. Ceglarek, A. Revzin, STUDIES OF DNA-PROTEIN INTERACTIONS BY GEL-ELECTROPHORESIS. *Electrophoresis* **10**, 360-365 (1989).
123. C. H. Schein, M. H. M. Noteborn, FORMATION OF SOLUBLE RECOMBINANT PROTEINS IN ESCHERICHIA-COLI IS FAVORED BY LOWER GROWTH TEMPERATURE. *Bio-Technology* **6**, 291-294 (1988).
124. T. C. Lin, J. Rush, E. K. Spicer, W. H. Konigsberg, Cloning and expression of a T4 DNA-polymerase. *Proceedings of the National Academy of Sciences of the United States of America* **84**, 7000-7004 (1987).
125. H. Pan, D. Clary, P. D. Sadowski, IDENTIFICATION OF THE DNA-BINDING DOMAIN OF THE FLP RECOMBINASE. *Journal of Biological Chemistry* **266**, 11347-11354 (1991).
126. V. Derbyshire, M. Astatke, C. M. Joyce, REENGINEERING THE POLYMERASE DOMAIN OF KLENOW FRAGMENT AND EVALUATION OF OVERPRODUCTION AND PURIFICATION STRATEGIES. *Nucleic Acids Research* **21**, 5439-5448 (1993).
127. J. A. Ludwig, J. N. Weinstein, Biomarkers in Cancer Staging, Prognosis and Treatment Selection. *Nat Rev Cancer* **5**, 845-856 (2005).
128. O. Hartmann, F. Spyratos, N. Harbeck, D. Dietrich, A. Fassbender, M. Schmitt, S. Eppenberger-Castori, V. Vuaroqueaux, F. Lerebours, K. Welzel, S. Maier, A. Plum, S. Niemann, J. A. Foekens, R. Lesche, J. W. M. Martens, DNA Methylation Markers Predict Outcome in Node-Positive, Estrogen Receptor-Positive Breast Cancer with Adjuvant Anthracycline-Based Chemotherapy. *Clinical Cancer Research* **15**, 315-323 (2009).
129. T. deVos, R. Tetzner, F. Model, G. Weiss, M. Schuster, J. Distler, K. V. Steiger, R. Grützmann, C. Pilarsky, J. K. Habermann, P. R. Fleshner, B. M. Oubre, R. Day, A. Z. Sledziewski, C. Lofton-Day, Circulating Methylated SEPT9 DNA in Plasma Is a Biomarker for Colorectal Cancer. *Clinical Chemistry* **55**, 1337-1346 (2009).

130. D. Macleod, R. R. Ali, A. Bird, An Alternative Promoter in the Mouse Major Histocompatibility Complex Class II I-Aβ Gene: Implications for the Origin of CpG Islands. *Molecular and Cellular Biology* **18**, 4433-4443 (1998).
131. A. P. Bird, CpG-rich islands and the function of DNA methylation. *Nature* **321**, 209-213 (1986).
132. J. F. Costello, Fruhwald, Michael C., Smiraglia, Dominic J., Rush, Laura J., Robertson, Gavin P., Gao, Xin, Wright, Fred A., Feramisco, Jamison D., Peltomaki, Paivi, James C., Schuller, David E., Yu, Li, Bloomfield, Clara D., Caligiuri, Michael A., Yates, Allan, Nishikawa, Ryo, Su Huang, H.-J., Petrelli, Nicholas J., Zhang, Xueli, O'Dorisio, M. S., Held, William A., Cavenee, Webster K., Plass, Christoph, Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nat Genet* **24**, 132-138 (2000).
133. S. J. Clark, A. Statham, C. Stirzaker, P. L. Molloy, M. Frommer, DNA methylation: Bisulphite modification and analysis. *Nat. Protocols* **1**, 2353-2364 (2006).
134. D. M. Hoover, J. Lubkowski, DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Research* **30**, e43 (2002).
135. K. L. Heckman, L. R. Pease, Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat. Protocols* **2**, 924-932 (2007).
136. D. C. Duffy, J. C. McDonald, O. J. A. Schueller, G. M. Whitesides, Rapid Prototyping of Microfluidic Systems in Poly(dimethylsiloxane). *Analytical Chemistry* **70**, 4974-4984 (1998).
137. T. M. Squires, R. J. Messinger, S. R. Manalis, Making it stick: convection, reaction and diffusion in surface-based biosensors. *Nat Biotech* **26**, 417-426 (2008).
138. T. Gervais, K. F. Jensen, Mass transport and surface reactions in microfluidic systems. *Chemical Engineering Science* **61**, 1102-1121 (2006).
139. M. Dufva, Fabrication of high quality microarrays. *Biomolecular Engineering* **22**, 173-184 (2005).
140. S. Raddatz, J. Mueller-Ibeler, J. Kluge, L. Wäß, G. Burdinski, J. R. Havens, T. J. Onofrey, D. Wang, M. Schweitzer, Hydrazide oligonucleotides: new chemical modification for chip array attachment and conjugation. *Nucleic Acids Research* **30**, 4793-4802 (2002).
141. L. M. Johnson, R. R. Hansen, M. Urban, R. D. Kuchta, C. N. Bowman, Photoinitiator Nucleotide for Quantifying Nucleic Acid Hybridization. *Biomacromolecules* **11**, 1133-1138 (2010).
142. D. S. Dandy, P. Wu, D. W. Grainger, Array feature size influences nucleic acid surface capture in DNA microarrays. *Proceedings of the National Academy of Sciences* **104**, 8223-8228 (2007).
143. J. Liu, B. A. Williams, R. M. Gwirtz, B. J. Wold, S. Quake, Enhanced Signals and Fast Nucleic Acid Hybridization By Microfluidic Chaotic Mixing. *Angewandte Chemie International Edition* **45**, 3618-3623 (2006).
144. H. H. Lee, J. Smoot, Z. McMurray, D. A. Stahl, P. Yager, Recirculating flow accelerates DNA microarray hybridization in a microfluidic device. *Lab on a Chip* **6**, 1163-1170 (2006).
145. R. H. Liu, R. Lenigk, R. L. Druyor-Sanchez, J. Yang, P. Grodzinski, Hybridization Enhancement Using Cavitation Microstreaming. *Analytical Chemistry* **75**, 1911-1917 (2003).
146. G. L. Long, J. D. Winefordner, Limit of Detection A Closer Look at the IUPAC Definition. *Analytical Chemistry* **55**, 712A-724A (1983).
147. G. S. Watts, R. O. Pieper, J. F. Costello, Y. M. Peng, W. S. Dalton, B. W. Futscher, Methylation of discrete regions of the O6-methylguanine DNA methyltransferase (MGMT) CpG island is associated with heterochromatinization of the MGMT transcription start site and silencing of the gene. *Molecular and Cellular Biology* **17**, 5612-5619 (1997).
148. M. Gonzalez-Zulueta, C. M. Bender, A. S. Yang, T. Nguyen, R. W. Beart, J. M. Van Tornout, P. A. Jones, Methylation of the 5' CpG Island of the p16/CDKN2 Tumor Suppressor Gene in Normal and Transformed Human Tissues Correlates with Gene Silencing. *Cancer Research* **55**, 4531-4535 (1995).

149. M. Esteller, J. G. Herman, Cancer as an epigenetic disease: DNA methylation and chromatin alterations in human tumours. *The Journal of Pathology* **196**, 1-7 (2002).
150. K. A. Aberg, J. L. McClay, S. Nerella, L. Y. Xie, S. L. Clark, A. D. Hudson, J. Bukszár, D. Adkins, S. S. Consortium, C. M. Hultman, P. F. Sullivan, P. K. E. Magnusson, E. J. C. G. van den Oord, MBD-seq as a cost-effective approach for methylome-wide association studies: demonstration in 1500 case–control samples. *Epigenomics* **4**, 605-621 (2012).
151. T. De Meyer, E. Mampaey, M. Vlemmix, S. Denil, G. Trooskens, J.-P. Renard, S. De Keulenaer, P. Dehan, G. Menschaert, W. Van Criekinge, Quality Evaluation of Methyl Binding Domain Based Kits for Enrichment DNA-Methylation Sequencing. *PLoS ONE* **8**, e59068 (2013).
152. A. Sethi, B. Goldstein, S. Gnanakaran, Quantifying Intramolecular Binding in Multivalent Interactions: A Structure-Based Synergistic Study on Grb2-Sos1 Complex. *PLoS Comput Biol* **7**, e1002192 (2011).
153. A. Raj, P. van den Bogaard, S. A. Rifkin, A. van Oudenaarden, S. Tyagi, Imaging individual mRNA molecules using multiple singly labeled probes. *Nat Meth* **5**, 877-879 (2008).
154. J. Weaver, P. Rao, J. R. Goldblum, M. J. Joyce, S. L. Turner, A. J. F. Lazar, D. Lopez-Terada, R. R. Tubbs, B. P. Rubin, Can MDM2 analytical tests performed on core needle biopsy be relied upon to diagnose well-differentiated liposarcoma[quest]. *Mod Pathol* **23**, 1301-1306 (2010).
155. J.-M. Mosquera, R. Mehra, M. M. Regan, S. Perner, E. M. Genega, G. Bueti, R. B. Shah, S. Gaston, S. A. Tomlins, J. T. Wei, M. C. Kearney, L. A. Johnson, J. M. Tang, A. M. Chinnaiyan, M. A. Rubin, M. G. Sanda, Prevalence of TMPRSS2-ERG Fusion Prostate Cancer among Men Undergoing Prostate Biopsy in the United States. *Clinical Cancer Research* **15**, 4706-4711 (2009).