Published in final edited form as:

Mol Syst Des Eng. 2016 October 1; 1(3): 273–277. doi:10.1039/C6ME00073H.

Engineering affinity agents for the detection of hemi-methylated CpG sites in DNA

B. E. Tarn, K.-J. Sung, and H. D. Sikes

Department of Chemical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA, 02139, USA

Abstract

Wild-type methyl-CpG-binding domain (MBD) proteins specifically bind symmetrically methylated DNA sequences, and assays have been developed that use these proteins for profiling DNA methylation. Here, we use directed evolution in the yeast surface display format to identify a new protein variant that binds hemi-methylated CpG dinucleotides.

Graphical abstract

A methyl-binding domain protein was engineered to bind to hemi-methylated DNA and tested in a biochip-based methylation detection assay.

Introduction

DNA methylation patterns, and predominantly the methylation of cytosine bases in CpG dinucleotides, have a strong correlation with numerous cancer subtypes. Currently, CpG methylation of the MGMT gene promoter is used to guide treatment decisions for patients with glioblastoma,¹ and MLH1 promoter methylation is used to determine whether patients with certain types of colorectal cancer have a sporadic or familial form of the disease.^{2–4} The utility of promoter methylation as a biomarker for cancer diagnostics is demonstrated by the commercially available tests Cologuard® and Epi ProColon® for colorectal cancer screening.^{5–7} While there are only a few gene promoters that are routinely evaluated for

methylation at this time,^{8,9} studies have shown tumor-associated promoter methylation of more than 45 genes in various types of cancers.¹⁰

Current clinical methylation detection assays involve bisulfite conversion of unmethylated cytosines to uracil while leaving the methylated cytosines unchanged, producing modified DNA sequences that can be distinguished by sequencing or PCR.^{9,11} Such methods have numerous drawbacks. For example, large quantities of DNA are required because 90% can be degraded during the conversion step, and without careful optimization and control of the reaction conditions, false positives and negatives are common.^{12,13} One of the methods proposed to eliminate the need for bisulfite treatment is to use an affinity agent, such as a methyl-CpG binding domain protein (MBD), to detect methylation of cytosine bases. Such affinity agents are commonly used in assays such as MethylCap-Seq, where genomic DNA is enriched for fragments containing methylated cytosines and then sequenced to identify the methylated regions.¹⁴ In addition, several assays based on DNA microarrays have been developed to profile methylation patterns of sequences of interest.

Early methylation-specific microarrays involved either bisulfite conversion followed by hybridization on an array of DNA probes specific to the sequences of the converted unmethylated or methylated states¹⁵ or enrichment of methylated sequences in genomic DNA via proteins that bind methylated CpGs followed by hybridization with DNA arrays to compare the prevalence of a specific sequence in the enriched and unenriched samples.¹⁶ More recently, researchers have used DNA probes to capture a specific DNA sequence followed by MBD binding as a simple method to determine the sequence-specific methylation status of the gene or region of interest. Yu *et al.*¹⁷ hybridized target sequences to probes on a chip and detected MBD binding with surface plasmon resonance. This method, however, required a 1 μ M concentration of target DNA.¹⁷ Heimer *et al.*¹⁸ performed the hybridization and MBD binding steps on an agarose-coated surface and reduced the limit of detection to 1 nM while detecting bound MBD directly with fluorescence labeling. This method eliminates the need for bisulfite conversion, PCR, sequencing, and enrichment steps, providing many advantages over conventional methylation detection assays.

Wild-type MBD proteins bind specifically to symmetrically methylated DNA, in which the cytosines on both DNA strands are methylated. Therefore, the capture probes must contain methylated CpGs in order for these assays to work. Such probes cannot be produced by standard enzymatic methods, such as PCR; instead, they require chemical synthesis. The incorporation of modified phosphoramidites into synthetic oligos yields probes that are much more expensive than those that only require unmodified bases. Developing a protein to recognize hemi-methylated DNA, where the cytosine bases are only methylated on one of the two DNA strands, would allow for the detection of a methylated sequence from a patient's sample bound to an unmethylated capture probe. This offers a key advantage in assay development. In other applications, where researchers are developing DNA microarrays to detect specific genetic mutations¹⁹ or viral RNA²⁰, several probes for each target sequence must often be tested experimentally, even after using specialized software to design the probes. This is also an important problem when designing probes for CpG islands, where high GC content and repeat sequences add additional challenges to probe design. If probes for the methylation assay need not be methylated, a larger number of

potential probes could be screened more easily, making the process of advancing microarray assays to the clinics much more feasible and efficient.

Here, we report the development of an MBD variant that binds to hemi-methylated DNA with high affinity. A library created from human MBD2, previously shown to have the highest affinity for methylated DNA of the MBD family,²¹ was used as a starting point. Variants with improved affinity for hemi-methylated DNA were isolated and analyzed in a yeast surface display construct.²² The top performing variant was produced as a GFP fusion protein²³ and tested in an array-based binding assay on a microscope slide to demonstrate its functionality in such detection assays.

Results and discussion

To characterize the binding affinity, the MBD proteins were displayed on the surface of *S. cerevisiae* using the pCTCON-2 vector.²² The binding affinity of wild-type human MBD2 for a DNA oligo with a single methyl group on one strand was evaluated using equilibrium binding titrations with flow cytometry. The sequence and methylation patterns of the test DNA used for characterization are shown in Figure 1a. These sequences are based on a region of the MGMT gene, as described previously.^{17,21} In the equilibrium binding titrations, expression of the MBD protein was verified by labeling the cMyc tag of the fusion protein with Alexa Fluor® 488. The cells expressing the protein were incubated with the biotinylated DNA at a range of concentrations and labeled with streptavidin, Alexa Fluor® 647. As shown in Figure 1b, the wild type MBD2 protein binds to symmetrically methylated DNA with high affinity but shows almost no binding to the hemi-methylated DNA sample, even at concentrations as high as 100 nM. Because binding was not observed, a dissociation constant could not be determined using this method. MBD2 was the focus of this study because it was previously demonstrated to have the greatest affinity for methylated DNA of the members of the MBD family.²¹ However, the results demonstrating that MBD2 binds to symmetrically methylated DNA with much higher affinity than hemi-methylated DNA agree with previously reported data for MBD1 and MeCP2, shown in Figure 1c, that show affinity differences of an order of magnitude or more between the two methylation states.

Beginning with the error-prone PCR library generated by Heimer et al.²¹, variants of the protein human MBD2 were displayed on the surface of yeast cells, and those with improved affinity for hemi-methylated DNA were selected using an equilibrium binding assay. The selection process is depicted in Figure 2. In the early rounds of selection, cells expressing the MBD library were incubated with hemi-methylated DNA attached to magnetic beads, allowing the cells that bind to the DNA to be separated from the larger library. In later rounds, the cells isolated from the magnetic bead selections were incubated with biotinylated, hemi-methylated DNA that was then labeled with a streptavidin-conjugated fluorophore. In this assay, cells expressing proteins with the highest affinities had the largest number of fluorophore-labeled DNA molecules attached, giving the brightest signal during flow cytometry. These cells were isolated using fluorescence-activated cell sorting (FACS). The amino acid sequences of the proteins isolated after the selection procedure are shown in the ESI (Table S1). All of the variants isolated had the K161R mutation and 70% had the F208Y mutation, two mutations described previously that allow for the formation of an

additional hydrogen bond to stabilize the protein structure and to bind to the DNA backbone, respectively.²¹ The F187I mutation, which is adjacent to the arginine residue that interacts with the methylated cytosine base,²⁵ was also found in 50% of the isolated proteins.

Six unique protein variants were compared (Figure S3), and the top-performing protein, variant h4, was characterized by equilibrium binding titration. Figure 3a shows the improvement in the binding affinity of the engineered protein for hemi-methylated DNA over that of the wild-type MBD2 protein. Completion of equilibrium binding titrations in triplicate gives a dissociation constant of 5.6 ± 1.4 nM, a value nearly identical to the wild type protein's dissociation constant with symmetrically methylated DNA. Figure 3b shows that the new protein binds to hemi-methylated DNA and symmetrically methylated DNA with similar affinity while retaining good specificity for these constructs over unmethylated DNA. Three of the four mutations in variant h4, K161R, F187I, and F208Y, have been described previously. The fourth mutation, T200S, is a small change from a threonine to the slightly smaller serine. Depicted in Figure 4, which is based on the NMR structure for chicken MBD2,²⁵ this mutation is located far from the DNA binding site. This residue is not conserved across the MBD family: it is found as alanine in MBD1, threonine in human MBD2, asparagine in MBD4, and valine in MeCP2. However, none of the wild type MBD proteins nor any of the proteins isolated from the library except for variant h4 have the S200 residue. Nevertheless, this mutation appears to play an important role in binding to hemi-methylated DNA.

To determine whether the new protein can function to distinguish between hemi-methylated and unmethylated DNA in the interfacial binding assays previously developed,^{17,18} binding experiments were performed with soluble MBD2 variant h4 and DNA arrays printed on agarose-coated glass slides. The MBD2 variant h4 was cloned into the pET30b bacterial expression vector and expressed as a fusion protein with eGFP and a biotin acceptor sequence. The slides were printed with hemi-methylated DNA as well as unmethylated DNA. Biotinylated MBD bound to the DNA was labeled with streptavidin, Alexa Fluor® 647 and detected by fluorescence imaging. In the resulting image, found in Figure 5a, MBD bound to the hemi-methylated DNA is easily visible while the spots printed with unmethylated DNA show little binding and are very difficult to identify by eye, a visual distinction that is confirmed by the quantitative results shown in Figure 5b.

Conclusions

The chemical conversion-based methods currently used for clinical methylation analyses have many disadvantages, including DNA degradation during sample treatment, and affinity agent-based methylation assays are a promising alternative. There are also limitations that must be overcome before these assays will be useful for clinical applications, however. Protein-DNA binding assays are highly dependent on the concentration of probe/target duplexes on the surface of the slide. Therefore, one key limitation is that the assay is not currently sensitive enough to detect the pM DNA concentrations that could be obtained from many patient samples. Further work to improve both the probe density and the efficiency of DNA hybridization could lead to higher concentrations of probe/target duplexes, effectively addressing this problem. Another disadvantage of the assay as described in this paper is that

equipment for fluorescence detection was also required for assay readout; however, an alternative detection mechanism that allows detection by eye has already been developed for this system.¹⁸

Compared to other protein-based assays, a method that doesn't require methylated capture probes can be developed more quickly and easily into an affordable test suitable for clinical use. Here, an MBD protein was developed to bind hemi-methylated DNA by introducing four mutations into human MBD2 through directed evolution with yeast surface display. The resulting protein variant binds to singly, hemi-methylated DNA with a dissociation constant of 5.6 ± 1.4 nM while the wild-type protein has an affinity too low to be quantified using the yeast surface display method. Specificity over unmethylated DNA was retained with the engineered protein, and the protein was proven to distinguish between hemi-methylated and unmethylated DNA in an interfacial binding assay performed with DNA affixed to a slide. These results demonstrate that variant h4 with unmethylated DNA probes can be used in place of the wild-type MBD proteins with methylated probes used in previously developed epigenotyping assays.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by an NSF graduate research fellowship (KJS), the Massachusetts Institute of Technology Center for Environmental Health Sciences (through support from the National Institutes of Health-National Institute of Environmental Health Sciences) [P30-ES002109], and partially by the National Institutes of Health-National Cancer Institute [P30CCA14051 to the Massachusetts Institute of Technology Flow Cytometry Core Facility]. We thank Dr. Brandon Heimer for sharing his expertise and providing advice related to this project.

Notes and references

1. Hegi ME, Diserens AC, Godard S, Dietrich PY, Regli L, Ostermann S, Otten P, Van Melle G, De Tribolet N, Stupp R. Clin Cancer Res. 2004; 10:1871–1874. [PubMed: 15041700]
2. Cunningham JM, Christensen ER, Tester DJ, Kim CY, Roche PC, Burgart LJ, Thibodeau SN. Cancer Res. 1998; 58:3455–60. [PubMed: 9699680]
3. Herman JG, Umar A, Polyak K, Graff JR, Ahuja N, Issa JPI, Markowitz S, Willson JKV, Hamilton SR, Kinzler KW, Kane MF, Kolodner RD, Vogelstein B, Kunkel TA, Baylin SB. Proc Natl Acad Sci. 1998; 95:6870–6875. [PubMed: 9618505]
4. Veigl ML, Kasturi L, Olechnowicz J, Ma A, Lutterbaugh JD, Periyasamy S, Li GM, Drummond J, Modrich PL, Sedwick WD, Markowitz SD. Proc Natl Acad Sci. 1998; 95:8698–8702. [PubMed: 9671741]
5. Imperiale TF, Ransohoff DF, Itzkowitz SH, Levin TR, Lavin P, Lidgard GP, Ahlquist DA, Berger BM. N Engl J Med. 2014; 370:1287–97. [PubMed: 24645800]
6. Potter NT, Hurban P, White MN, Whitlock KD, Lofton-Day CE, Tetzner R, Koenig T, Quigley NB, Weiss G. Clin Chem. 2014; 000:1–9.
7. Pratt VM. Clin Chem. 2014; 60:1141–2. [PubMed: 25028508]
8. Noehammer C, Pulverer W, Hassler MR, Hofner M, Wielscher M, Vierlinger K, Liloglou T, McCarthy D, Jensen TJ, Nygren A, Gohlke H, Trooskens G, Braspenning M, Van Criekinge W, Egger G, Weinhaeusel A. Epigenomics. 2014; 6:603–22. [PubMed: 25531255]
9. Heimer BW, Tam BE, Minkovsky A, Sikes HD. Wiley Interdiscip Rev Nanomed Nanobiotechnol. 2016; doi: 10.1002/wnan.1407

10. Heyn H, Esteller M. *Nat Rev Genet.* 2012; 13:679–692. [PubMed: 22945394]
11. Herman JG, Graff JR, Myöhänen S, Nelkin BD, Baylin SB. *Proc Natl Acad Sci U S A.* 1996; 93:9821–9826. [PubMed: 8790415]
12. Grunau C, Clark SJ, Rosenthal A. *Nucleic Acids Res.* 2001; 29:E65–5. [PubMed: 11433041]
13. Okamoto A. *Org Biomol Chem.* 2009; 7:21–26. [PubMed: 19081937]
14. Brinkman AB, Simmer F, Ma K, Kaan A, Zhu J, Stunnenberg HG. *Methods.* 2010; 52:232–6. [PubMed: 20542119]
15. Gitan RS, Shi H, Chen CM, Yan PS, Huang THM. *Genome Res.* 2002; 12:158–64. [PubMed: 11779841]
16. Gebhard C, Schwarzfischer L, Pham TH, Schilling E, Klug M, Andreessen R, Rehli M. *Cancer Res.* 2006; 66:6118–28. [PubMed: 16778185]
17. Yu Y, Blair S, Gillespie D, Jensen R, Myszkka D, Badran AH, Ghosh I, Chagovetz A. *Anal Chem.* 2010; 82:5012–5019. [PubMed: 20507169]
18. Heimer BW, Shatova TA, Lee JK, Kaastrup K, Sikes HD. *Analyst.* 2014; 139:3695–701. [PubMed: 24824477]
19. Waldmüller S, Freund P, Mauch S, Toder R, Vosberg HP. *Hum Mutat.* 2002; 19:560–9. [PubMed: 11968089]
20. Gall A, Hoffmann B, Harder T, Grund C, Höper D, Beer M. *J Clin Microbiol.* 2009; 47:327–34. [PubMed: 19052173]
21. Heimer BW, Tam BE, Sikes HD. *Protein Eng Des Sel.* 2015; 28:543–51. [PubMed: 26384511]
22. Chao G, Lau WL, Hackel BJ, Sazinsky SL, Lippow SM, Wittrup KD. *Nat Protoc.* 2006; 1:755–68. [PubMed: 17406305]
23. Boyd ME, Heimer BW, Sikes HD. *Protein Expr Purif.* 2012; 82:332–8. [PubMed: 22326799]
24. Valinluck V, Liu P, Kang JI, Burdzy A, Sowers LC. *Nucleic Acids Res.* 2005; 33:3057–64. [PubMed: 15917437]
25. Scarsdale JN, Webb HD, Ginder GD, Williams DC. *Nucleic Acids Res.* 2011; 39:6741–52. [PubMed: 21531701]
26. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Gallo Cassarino T, Bertoni M, Bordoli L, Schwede T. *Nucleic Acids Res.* 2014; 42:W252–8. [PubMed: 24782522]

Design, System, Application

Directed evolution is a valuable technique for developing proteins with desirable catalytic or binding properties. Used in conjunction with a yeast surface display platform, a large number of proteins with random mutations can be quickly screened to select those with the desired properties. Here, a protein that binds to hemi-methylated DNA with high affinity and excellent specificity with respect to unmethylated DNA is desired. Current affinity agent-based methylation detection assays require symmetrically methylated DNA sequences for a positive readout. With a protein that binds hemi-methylated DNA, we can alter these assays to, instead, use a simple, unmethylated DNA probe to capture a DNA sequence of interest and profile its methylation status. This allows for the development of simpler, less expensive assays that can more easily be developed into clinical methylation tests to aid in the detection and treatment of several types of cancers.

Fig. 1.

a) The DNA sequence of the probe/target pairs and methylation states used for protein assessment are shown with the methylated cytosine bases bolded and underlined. The sequence is from the *MGMT* gene and contains three CpG dinucleotides, one of which is methylated in the hemi-methylated and symmetrically methylated states used in this paper.

b) An equilibrium binding titration of wild-type human MBD2. The fraction of MBD bound to DNA was determined based on the normalized mean fluorescence. Wild-type MBD2 binds symmetrically methylated DNA with high specificity and shows no detectable binding to hemi-methylated DNA at the concentrations of interest.

c) Reported dissociation constants for MBD1 and MeCP2 also show affinity differences of one or more orders of magnitude between symmetrically methylated and hemi-methylated DNA.^{17,21,24}

Fig. 2.

MBD variants were expressed on the surface of *S. cerevisiae*. Cells expressing variants with improved affinity for hemi-methylated DNA were selected first with magnetic beads coated in hemi-methylated DNA and then with fluorescence activated cell sorting.

Fig. 3.

a) An equilibrium binding titration shows the binding affinity for the engineered variant H4 in comparison with the wild-type hMBD2 protein. b) The engineered hMBD2 protein H4 is characterized by equilibrium binding titrations with hemi-methylated, symmetrically methylated, and unmethylated DNA.

Fig. 4. The structure of the engineered protein, variant H4, is shown in association with DNA with the T200S mutation highlighted. This structure was constructed using the SWISS-MODEL system²⁶ based on the crystal structure of chicken MBD2.²⁵

Fig. 5.

a) An image of the DNA array after labeling of MBD2 Variant h4 shows that the hemi-methylated and unmethylated spots are easily distinguishable. b) Quantitative analysis shows a 7.8-fold higher signal from binding to hemi-methylated DNA as compared to unmethylated DNA in the arrays.