Electronic Supplementary Material (ESI) for Molecular Systems Design & Engineering.

Supplementary Information

Directed Evolution of a Methyl-CpG-Binding Protein to Recognize Hemi-Methylated DNA

Brooke E. Tam, Ki-Joo Sung, and Hadley D. Sikes*

Department of Chemical Engineering, Massachusetts Institute of Technology,
Building E19-502C, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA

* To whom correspondence should be addressed.

telephone: 617-253-5224

fax: 617-253-2272

e-mail: sikes@mit.edu

Figures and Tables

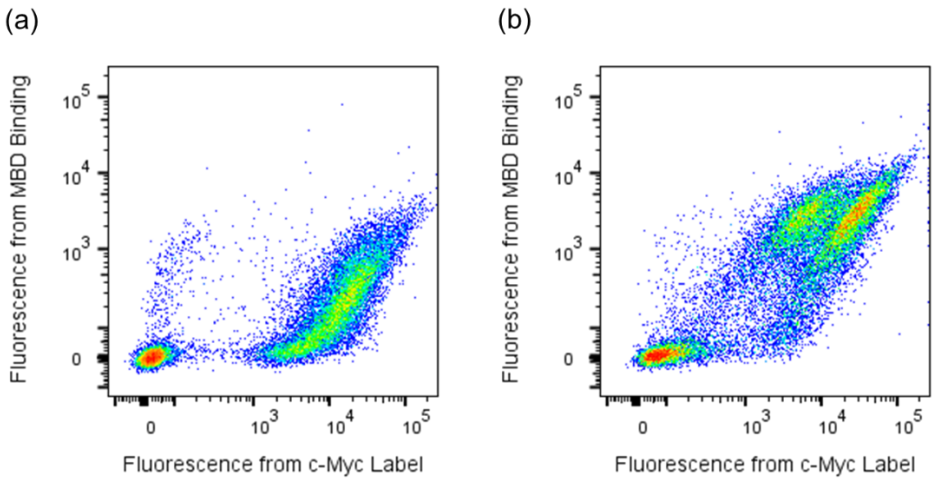

Figure S1: Analysis of yeast-displayed protein variants encoded by the error-prone PCR library using flow cytometry. Following magnetic bead sorting, flow cytometry was used during the enrichment process to select variants with improved binding to hemi-methylated DNA. Protein expression is shown on the x-axis and DNA bound by the protein on the y-axis. (a) The MBD library after two rounds of selection with magnetic beads. (b) The MBD library after two rounds of selection with magnetic beads and two rounds of selection with flow cytometry.

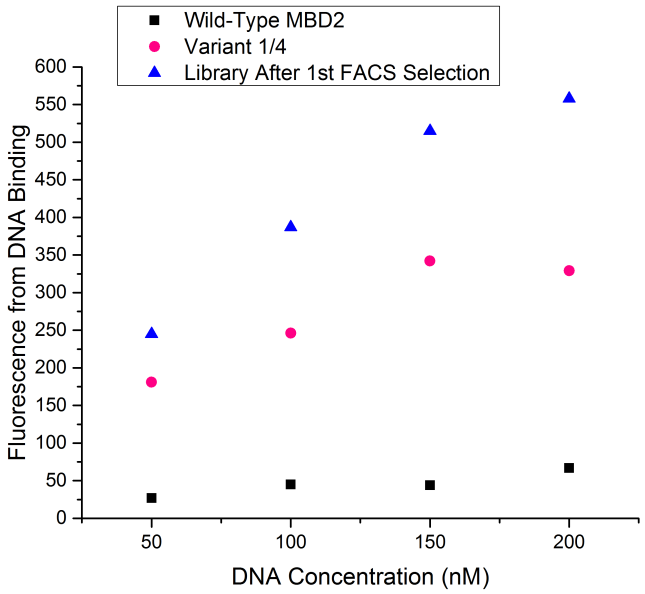

Figure S2: The signal from hemi-methylated DNA binding to yeast displayed protein variants encoded by the error-prone PCR library after two magnetic bead selections and one selection with flow cytometry (blue triangles) was compared with binding to the yeast displayed wild-type MBD2 protein (black squares) and Variant 1/4 isolated by Heimer *et al.*[1] (red circles). This survey of four DNA concentrations motivated the isolation of individual clones from the library and the more extensive titrations presented in the main text.

Table S1: Sequences of MBD variants isolated from the error-prone PCR library that was enriched for binding to hemi-methylated DNA (mutations in red)

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
h1	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 **I** R S K P Q L A R Y L G N T V D L S S F D F R T **C** K M
h2	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
h3	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
h4	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 **I** R S K P Q L A R Y L G N **S** V D L S S F D **Y** R T G K M
h5	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 **Y** 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
h6	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
h7	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 **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
h8	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 **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

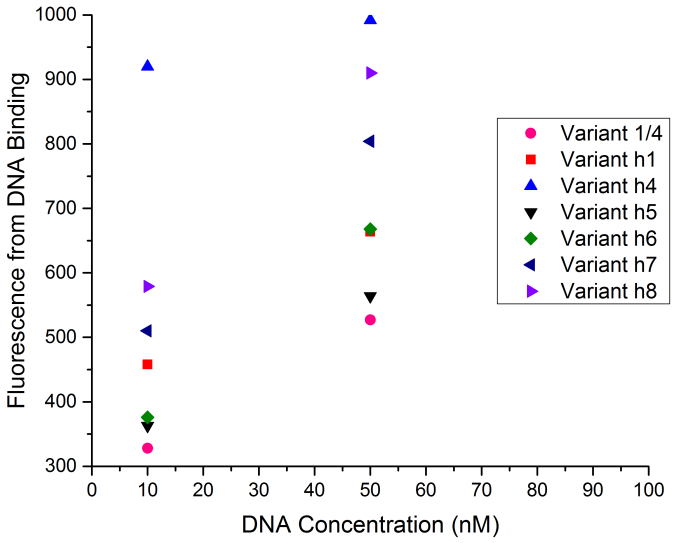

Figure S3: The binding of hemi-methlyated DNA by different MBD variants isolated from the error-prone PCR library was compared in the yeast display format using two DNA concentrations to identify those variants with the best performance. Further characterization of the top performer is presented in the main text.

Experimental Methods

Characterization of Binding Affinities using Yeast Surface Display

MBD proteins were displayed on the surface of EBY100 *S. cerevisiae* yeast cells as described previously.[1] The cells containing the pCTCON-2 vector with the MBD insert were grown overnight in SDCAA medium at 30°C and 250 rpm. To induce protein expression, after the SDCAA cultures reached an OD_{600} between 2 and 5, the cells were resuspended in SGCAA medium to an OD_{600} of 1 and incubated at 20°C and 250 rpm for 36-48 hours. The cells were then resuspended in PBS with 0.1% BSA and an equilibrium binding titration was performed by incubating the cells expressing the MBD protein with biotinylated DNA oligomers at a range of concentrations between 0.05 and 100 nM for 45 min at room temperature. Total reaction volumes were chosen to ensure 10-fold excess of DNA in each sample, calculated based on the protein expression level identified by Chao *et al*.[2] Expressed protein and bound DNA were labeled with chicken anti-cMyc/Alexa Fluor® 488 goat anti- chicken and streptavidin, Alexa Fluor® 647, respectively. The extent of binding was evaluated using flow cytometry, and dissociation constants were calculated using the method described by Chao *et al*.[2]

Screening the MBD Library for Improved Affinity for Hemi-Methylated DNA

The MBD library from the second round of error-prone PCR performed by Heimer *et al.*[1] was expressed on the surface of *S. cerevisiae* as described above. To enrich for protein variants that bind to hemi-methylated DNA, biotinylated DNA with a single methylated cytosine on one strand was incubated with Dynabeads® Biotin Binder. For each selection, the Dynabeads® were washed with 1 mL of 1xPBS containing 0.1% BSA and incubated with 55 nM biotinylated, hemi-methylated DNA for 30 min at 4°C. A total of $4x10^9$ cells expressing the MBD library were then incubated with the DNA coated beads for 2 hours at 4°C to capture those expressing proteins with good binding characteristics. After the incubation, the beads with cells attached were separated from unbound cells with a magnet and resuspended in SDCAA medium, pH 4.5, supplemented with pen-strep (1:100 dilution). The captured cells were grown overnight at 30°C and 250 rpm. The bead selection was repeated with $2x10^8$ cells from the enriched library. After the second selection with magnetic beads, the cells were again grown up and protein expression was induced. Two additional selections for hemi-methylated DNA were performed using fluorescence-activated cell sorting (FACS). For the first FACS selection, binding reactions were prepared as described for characterization by flow cytometry and a gate was drawn during sorting to capture the top 1% of cells. This top 1% was defined using a diagonal sort window, as described by Chao *et al*.[2] In the second FACS selection, the top 0.37% of the cells were isolated. The plasmids encoding the selected proteins were extracted using the Zymoprep™ Yeast Plasmid Miniprep II kit, transformed into Mach 1 E. coli, and grown on LB plates containing 100 µg/ml ampicillin. Ten single colonies were selected, and for each of these colonies, the MBD insert was sequenced. After sequencing, plasmids containing unique clones were transformed into EBY100 *S. cerevisiae* and expressed on the surface. To compare the clones, binding reactions were performed as described above with two DNA concentrations, 10 nM and 50 nM. After a comparison of binding affinities among the isolated clones, titrations were performed to determine the dissociation constant of the top performing variant.

Soluble Protein Expression

The sequence encoding the top performing variant, h4, was PCR amplified from the pCTCON-2 vector using Phusion HF polymerase with the forward primer 5'-GCCTGAATTCTGAAAGCGGCAAACG-3', which includes an EcoRI restriction site, and the reverse primer 5'- CATTTTGCCGGTACGATAATCAAAGCTGCTC-3'. In this reaction, the DNA was denatured at 95°C for 6 min, then 30 cycles were performed with 30 sec each of denaturation at 95°C, annealing at 56°C, and extension at 72°C. A 10 min final extension was performed at 72°C. Splicing by overlap extension was used to append an eGFP tag and a biotin accepter sequence to MBD2 variant h4. First, a 3-primer PCR reaction was used to add a linker sequence to the MBD variant. This reaction used the forward primer 5'-GCCTGAATTCTGAAAGCGGCAAACG-3', the long reverse primer 5'-CGTAGTCTGGCACGTCGTATGGGTACATTTTGCCGGTACGATAATCAAAGCTG-3' for adding the linker group, and the short reverse primer 5'-CGTAGTCTGGCACGTCGTATGGG-3' for amplifying the product containing the linker group with the same PCR conditions as the first reaction. The eGFP tag and biotin accepter sequence were amplified from another plasmid using the forward primer 5'-TACCCATACGACGTGCCA-3'and the reverse primer 5'-TGGTGCTCGAGTTTATTCATGC-3', which added an XhoI restriction site. The eGFP reaction proceeded as described above except the annealing temperature was reduced to 52°C and the extension time increased to 1 min. For the splicing by overlap extension reaction, the forward primer 5'-GCCTGAATTCTGAAAGCGGCAAACG-3' and reverse primer 5'-TGGTGCTCGAGTTTATTCATGC-3 were used to amplify the full MBD-GFP fusion protein using touchdown PCR. An annealing temperature of 61°C was used for the first cycle, and this temperature was decreased by 1°C for each of the next eight cycles. The final annealing temperature of 53°C was then used for an additional 30 cycles. The resulting PCR product was cloned into the pET30b vector using the EcoRI and XhoI restriction sites. After sequencing to verify that the intended product was obtained, the pET30b vector containing the insert was transformed into DE3 Tuner *E. coli* and grown in LB broth supplemented with Kanamycin. To express the fusion protein, the cells were grown in TB medium to an OD_{600} of 0.6 and then protein expression was induced by the addition of 0.05 mM IPTG. The cells were incubated at 20°C for 16 hours, pelleted, and lysed using BugBuster® HT protein extraction reagent according to the manufacturer's protocol for soluble protein.

Biochip Experiments

Glass slides were coated with 0.2% SeaKem® LE agarose (Lonza) and arrays of pre-hybridized DNA were printed, as described previously.[3] Each slide contained two rows of hemi-methylated and one row of unmethylated target DNA. The slides were left to dry in a vacuum desiccator overnight. Wells were cut from Scotch 3M tape and placed around the arrays on the slide. The wells were rinsed with 18 MΩ DI water and dried under compressed air. Blocking was performed by incubating the wells with 40 µl of 1% BSA at room temperature for 15 min. After the blocking reaction, the wells were rinsed with PBS and 18 MΩ DI water and dried with compressed air before 40 ul of the clarified cell lysate containing MBD2 variant h4, diluted in binding buffer[4] (20 mM HEPES, pH 7.9, 3 mM MgCl2, 10% (v/v) glycerol, 1 mM dithiothreitol, 100 mM KCl, 0.1% (w/v) BSA, 0.01% Tween-20), was added. The DNA arrays and protein solution were incubated at room temperature for 45 minutes, after which the wells were washed consecutively with 1xPBS/0.1% Tween 20, 1xPBS, and 18 MΩ DI water and dried with compressed air. Bound protein was labeled with streptavidin, Alexa Fluor® 647 diluted 1:100 in 1xPBS/0.1% BSA (PBSA) for 10 min at 4°C and the wells were washed and dried again, as described above. All incubation steps were performed in a humid chamber that had been equilibrated to the desired incubation temperature. Fluorescence was detected using a GenePix 4000B scanner (Molecular Devices) with 635 nm excitation. Quantitative results were obtained by calculating the mean fluorescence and background fluorescence

for each spot within the DNA array using the GenePix 6.1 software. For each methylation pattern, the fluorescence intensity was averaged over all of the spots within the well.

Works Cited

1 B. W. Heimer, B. E. Tam and H. D. Sikes, *Protein Eng. Des. Sel.*, 2015, **28**, 543–51.

2 G. Chao, W. L. Lau, B. J. Hackel, S. L. Sazinsky, S. M. Lippow and K. D. Wittrup, *Nat. Protoc.*, 2006, **1**, 755–68.

3 B. W. Heimer, T. A. Shatova, J. K. Lee, K. Kaastrup and H. D. Sikes, *Analyst*, 2014, **139**, 3695–701.

4 Y. Yu, S. Blair, D. Gillespie, R. Jensen, D. Myszka, A. H. Badran, I. Ghosh and A. Chagovetz, *Anal. Chem.*, 2010, **82**, 5012–5019.