Site Specific Detection of DNA Methylation Utilizing mCpG-SEER

Cliff I. Stains, Jennifer L. Furman, David J. Segal, and Indraneel Ghosh

Contribution from the Department of Chemistry, University of Arizona, Tucson, Arizona 85721 and the Department of Medical Pharmacology and Toxicology, University of California, Davis, California 95616

ghosh@email.arizona.edu

Supplementary Materials

Effect of Target Site Spacing on SEER-GFP

General Materials and Methods: All spectra were taken on a Photon Technology International spectrofluorometer with excitation and emission wavelengths of 468 nm and 505 nm respectively. Slit widths were set to 5 nm for excitation and 10 nm for emission. All refolding experiments were conducted using 3.5KD MWCO Slide-A-Lyzer Dialysis Cassettes (Pierce). All DNA constructs used in refolding are shown in Figure S1 and were obtained HPLC purified from IDT. Oligos were annealed in 1x BamHI Buffer (NEB) using the following procedure: heating to 95 ^{o}C for 7 min, cooling to 56 ^{o}C at a rate of 1 ^{o}C/min, equilibrating at 56 ^{o}C for 5 min, and finally cooling to 25 ^{o}C at a rate of 1 ^{o}C/min using a Techne Genius thermocycler. All refolding experiments were conducted at 4 ^{o}C in uncovered chambers. Duplicate experiments were compared by the use of a standard, 5(6)-carboxyflourescein (FAM), obtained from Sigma prepared at 20 nM in Buffer A (10 mM Tris-HCl @ pH=7.5, 100 mM NaCl, 1 mM DTT, and 100 μM $ZnCl_2$). FAM emission spectra were acquired by excitation at 490 nm. SEER-GFP data from duplicate experiments were made relative to FAM emission at 512 nm for comparison.

Sensitivity of SEER-GFP to Target Site Spacing: Spectra were acquired from samples which contained 5 μM NGFP-Zif268, 5 μM CGFP-PBSII, and 2.5 μM of each target DNA. Spectra were taken four days post-refolding and were normalized to the final DNA concentration after dialysis (using the absorbance at 260 nm) and then to the 20 nM FAM emission. Refolding experiments were repeated, separately, and the trends are plotted below (Figure S2).

Cloning

General Materials and Methods: All enzymes were obtained from NEB, dNTP's were purchasd from Fermentas. A pUC57 plasmid containing an optimized *E. coli* gene encoding for human MBD2[1,2] (residues 147 – 215) was purchased from GenScript.

MBD2 Cloning: Sequences were confirmed by dideoxyoligonucleotide sequencing at the University of Arizona DNA Sequencing Facility (Figure S3). A map of this plasmid is shown below (Figure S4).

Figure S1. DNA constructs used in the refolding experiments to test the effect of spacing of SEER-GFP. Blue and Red indicate the respective Zif268 and PBSII sites. Numbers indicate the distance between binding sites in base pairs.

Spacing

0:
GCGTAGCGTGGGCGGTGTGGAAACACCG

3:
GCGTAGCGTGGGCGTAAGTGTGGAAACACCG

6:
GCGTAGCGTGGGCGTTAGTCGTGTGGAAACACCG

10:
GCGTAGCGTGGGCGTAGGACGATAGTGTGGAAACACCG

13:
GCGTAGCGTGGGCGTTAGTCACTAGAGGTGTGGAAACACCG

16:
GCGTAGCGTGGGCGTTAGTCACTAGAGGACGTGTGGAAACACCG

20:
GCGTAGCGTGGGCGTTAGTCACTAGAGGACGATAGTGTGGAAACACCG

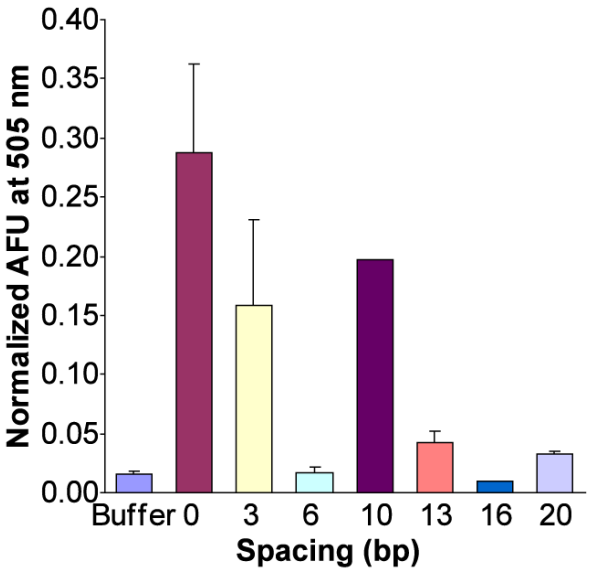

Figure S2. The effect of target site spacing on SEER-GFP flourescence.

NGFP-Zif268 Cloning: An NGFP-Zif268 construct was cloned previously by our lab and was used in these experiments, its sequence is given below (Figure S3).[3]

Purification of CGFP-MBD2

Purification of Soluble CGFP-MBD2 by IMAC: Cells were re-suspended in Buffer A and lysed using standard sonication protocols and clarified for 30 minutes at 18,000 rcf. This lysate was passed over Ni- NTA agarose beads (Qiagen) and eluted with Buffer A containing increasing concentrations of imidazole (2, 10, 20, 50, and 500 mM sequentially). CGFP-MBD2 eluted in the 50 – 500 mM imidazole fractions. Fractions containing CGFP-MBD2 were found to have high concentrations of DNA (as determined by the A_{260}/A_{280}), therefore CGFP-MBD2 was further purified under denaturing conditions. CGFP-MBD2 obtained above was diluted into an equivalent volume of Buffer A containing 8 M Urea (4 M Urea final). This sample was re-exposed to Ni-NTA agarose beads, and the protein was eluted with Buffer A containing 4 M Urea and increasing concentrations of imidazole (2, 10, 20, 50, and 500 mM sequentially). Fractions containing CGFP-MBD2 were pooled, concentrated, and dialyzed into Buffer A containing 4 M Urea. Concentrations were obtained using protein absorbance measurements at 280 nm ($\varepsilon = 14440\ M^{-1}\ cm^{-1}$).

Purification of Insoluble CGFP-MBD2 by IMAC: Inclusion bodies obtained from the initial lysis above were solubilized by re-suspension in Buffer A containing 8 M Urea. After incubating on ice for one hour, this solution was diluted to 4 M Urea by addition of an equivalent volume of Buffer A. The solution was clarified by centrifugation at 18,000 rcf for 30 minutes. The solubilized protein was exposed to Ni-NTA resin and CGFP-MBD2 was eluted with Buffer A containing 4 M Urea and increasing concentrations of imidazole (2, 10, 20, 50, and 500 mM sequentially). Concentrations were obtained using protein absorbance measurements at 280 nm ($\varepsilon = 14440\ M^{-1}\ cm^{-1}$).

Characterization of the mCpG-SEER Proteins

SDS-PAGE: Equivalent amounts of purified NGFP-Zif268 (32.7 kD) and CGFP-MBD2 (19.7 kD) were loaded on a 15 % SDS-PAGE gel (Figure S5).

MALDI: Samples of the refolded mCpG-SEER proteins from above were sent for MALDI-MS analysis. MALDI mass spectra were acquired on a Bruker Reflex-III MALDI/TOF, the masses obtained are shown below.

NGFP-Zif268 MH^+ calculated is 32686; found: 32648
CGFP-MBD2 MH^+ calculated is 19590; found: 19572

Figure S3. A.) The DNA and amino acid sequence of CGFP-MBD2. CGFP is shown in green, MBD2 is shown in red, and the linker is shown in pink. B.) The DNA and amino acid sequence of NGFP-Zif268. NGFP is shown in green, Zif268 is shown in blue, and the linker shown in pink.

A. CGFP-MBD2:

```
      1 ATGGGCAGCA GCCATCACCA TCATCACCAC
        M  G  S  S   H  H  H   H  H  H

     31 AGCCAGGATC CGAATTCGGA AAGCGGCAAA CGCATGGATT GCCCGGCGCT
        S  Q  D  P   N  S  E   S  G  K   R  M  D  C   P  A  L

     81 GCCGCCGGGT TGGAAAAAAG AAGAAGTGAT TCGTAAAAGC GGCCTGAGCG
         P  P  G   W  K  K  E   E  V  I   R  K  S   G  L  S  A

    131 CGGGCAAAAG CGATGTGTAT TATTTTAGCC CGAGCGGCAA AAAATTTCGT
          G  K  S   D  V  Y   Y  F  S  P   S  G  K   K  F  R

    181 AGCAAACCGC AGCTGGCGCG TTATCTGGGC AACACCGTGG ATCTGAGCAG
        S  K  P  Q   L  A  R   Y  L  G   N  T  V  D   L  S  S

    231 CTTTGATTTT CGTACCGGCA AAATGACCGG TGGGGGTGGC GGTTCAGGCG
         F  D  F   R  T  G  K   M  T  G   G  G  G   G  S  G  G

    281 GTGGGGGTTC TGGTGGGGGT GGTACCAAGA ATGGAATCAA AGTGAACTTC
          G  G  S   G  G  G   G  T  K  N   G  I  K   V  N  F

    331 AAGACCCGCC ACAACATTGA AGATGGAAGC GTTCAACTAG CAGACCATTA
        K  T  R  H   N  I  E   D  G  S   V  Q  L  A   D  H  Y

    381 TCAACAAAAT ACTCCAATTG GCGATGGCCC TGTCCTTTTA CCAGACAACC
         Q  Q  N   T  P  I  G   D  G  P   V  L  L   P  D  N  H

    431 ATTACCTGTC CACACAATCT GCCCTTTCGA AAGATCCCAA CGAAAAGAGA
          Y  L  S   T  Q  S   A  L  S  K   D  P  N   E  K  R

    481 GACCACATGG TCCTTCTTGA GTTTGTAACA GCTGCTGGGA TTACACATGG
        D  H  M  V   L  L  E   F  V  T   A  A  G  I   T  H  G

    531 CATGGATGAA CTGTACAACT GA
         M  D  E   L  Y  N  *
```

B. NGFP-Zif268:

```
    1 ATGGCAGATC TGATGGCTAG CAAAGGAGAA GAACTCTTCA
      M  A  D  L  M  A  S  K  G  E  E  L  F  T

   41 CTGGAGTTGT CCCAATTCTT GTTGAATTAG ATGGTGATGT TAACGGCCAC
      G  V  V  P  I  L  V  E  L  D  G  D  V  N  G  H

   91 AAGTTCTCTG TCAGTGGAGA GGGTGAAGGT GATGCAACAT ACGGAAAACT
      K  F  S  V  S  G  E  G  E  G  D  A  T  Y  G  K  L

  141 TACCCTGAAG TTCATCTGCA CTACTGGCAA ACTGCCTGTT CCATGGCCAA
      T  L  K  F  I  C  T  T  G  K  L  P  V  P  W  P  T

  191 CACTAGTCAC TACTCTGTGC TATGGTGTTC AATGCTTTTC AAGATACCCG
      L  V  T  T  L  C  Y  G  V  Q  C  F  S  R  Y  P

  241 GATCATATGA AACGGCATGA CTTTTTCAAG AGTGCTATGC CCGAAGGTTA
      D  H  M  K  R  H  D  F  F  K  S  A  M  P  E  G  Y

  291 TGTACAGGAA AGGACCATCT TCTTCAAAGA TGACGGCAAC TACAAGACAC
      V  Q  E  R  T  I  F  F  K  D  D  G  N  Y  K  T  R

  341 GTGCTGAAGT CAAGTTTGAA GGTGATACCC TTGTTAATAG AATCGAGTTA
      A  E  V  K  F  E  G  D  T  L  V  N  R  I  E  L

  391 AAAGGTATTG ACTTCAAGGA AGATGGCAAC ATTCTGGGAC ACAAATTGGA
      K  G  I  D  F  K  E  D  G  N  I  L  G  H  K  L  E

  441 ATACAACTAT AACTCACACA ACGTTCCCAT CATGGCAGAC AAACAAGGTC
      Y  N  Y  N  S  H  N  V  P  I  M  A  D  K  Q  G  L

  491 TGCAGGGCGG TTCAGGCGGT GGGGGTTCTG GCGGGGGTGG GTACCCCGGG
      Q  G  G  S  G  G  G  G  S  G  G  G  G  Y  P  G

  541 GAACGCCCTT ACGCTTGCCC AGTGGAGTCC TGTGATCGCC GCTTCTCCCG
      E  R  P  Y  A  C  P  V  E  S  C  D  R  R  F  S  R

  591 CTCCGACGAG CTCACCCGCC ACATCCGCAT CCACACAGGC CAGAAGCCCT
      S  D  E  L  T  R  H  I  R  I  H  T  G  Q  K  P  F

  641 TCCAGTGCCG CATCTGCATG CGCAACTTCA GCCGCAGCGA CCACCTCACC
      Q  C  R  I  C  M  R  N  F  S  R  S  D  H  L  T

  691 ACCCACATCC GCACCCACAC AGGCGAAAAG CCCTTTGCCT GCGACATCTG
      T  H  I  R  T  H  T  G  E  K  P  F  A  C  D  I  C

  741 TGGAAGAAAG TTTGCCAGGA GCGATGAACG CAAGAGGCAT ACCAAGATCC
      G  R  K  F  A  R  S  D  E  R  K  R  H  T  K  I  H

  791 ACTTGCGGCA GAAGGACCTC GAGTCTGGTA AAGAAACCGC TGCTGCGAAA
      L  R  Q  K  D  L  E  S  G  K  E  T  A  A  A  K

  841 TTTGAACGCC AGCACATGGA CTCGTCTACT AGCGCAGCTT AA
      F  E  R  Q  H  M  D  S  S  T  S  A  A  *
```

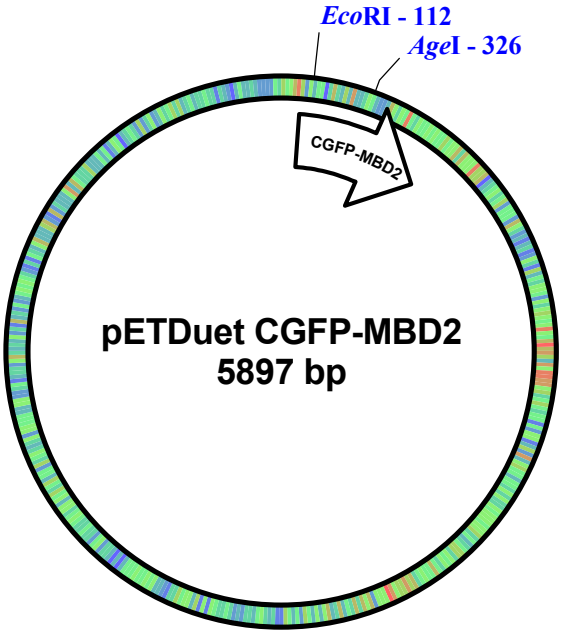

Figure S4. A schematic of the pETDuet CGFP-MBD2 plasmid showing the position of the CGFP-MBD2 gene and restriction enzymes used.

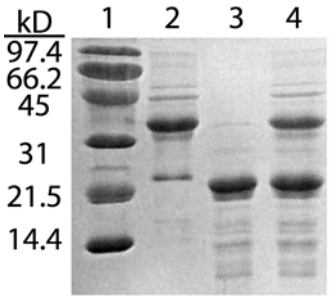

Figure S5. MW standards (lane 1); NGFP-Zif268 (lane 2); CGFP-MBD2 (lane 3); and equal amounts of each protein (lane 4).

Effect of Target Site Spacing on mCpG-SEER

Sensitivity of mCpG-SEER to Target Site Spacing: Spectra were acquired from samples which contained 5 μM NGFP-Zif268, 5 μM CGFP-MBD2, and 2.5 μM of each target DNA (Table S1). Spectra were taken seventeen days post-refolding, fluorescence values were normalized to the 2 bp spacing after background subtraction. Refolding experiments were repeated, separately, and the trends are plotted below (Figure S6).

Specificity of mCpG-SEER at 2bp Spacing

Using 100 nM Target DNA: Samples were prepared and refolded as described for the 1 μM DNA targets, using 100 nM of each DNA target. Fluorescence spectra were acquired 7 hours post-refolding (Figure S7).

Table S1. Oligos used to test mCpG-SEER dependence on target site spacing.

Spacing		
0	*1*	*2*
GCGTA$_m$CGCGCCCACGCCACCG	GCGTA$_m$CGTCGCCCACGCCACCG	GCGTA$_m$CGTACGCCCACGCCACCG
CGCATGC$_m$GCGGGTGCGGTGGC	CGCATGC$_m$AGCGGGTGCGGTGGC	CGCATGC$_m$ATGCGGGTGCGGTGGC
3	*4*	*5*
GCGTA$_m$CGTAGCGCCCACGCCACCG	GCGTA$_m$CGTAGGCGCCCACGCCACCG	GCGTA$_m$CGTAGGACGCCCACGCCACCG
CGCATGC$_m$ATCGCGGGTGCGGTGGC	CGCATGC$_m$ATCCGCGGGTGCGGTGGC	CGCATGC$_m$ATCCTGCGGGTGCGGTGGC
6	*7*	*8*
GCGTA$_m$CGTAGGACCGCCCACGCCACCG	GCGTA$_m$CGTAGGACGCGCCCACGCCACCG	GCGTA$_m$CGTAGGACGACGCCCACGCCACCG
CGCATGC$_m$ATCCTGGCGGGTGCGGTGGC	CGCATGC$_m$ATCCTGCGCGGGTGCGGTGGC	CGCATGC$_m$ATCCTGCTGCGGGTGCGGTGGC
9	*10*	*11*
GCGTA$_m$CGTAGGACGATCGCCCACGCCACCG	GCGTA$_m$CGTAGGACGATACGCCCACGCCACCG	GCGTA$_m$CGTAGGACGATAACGCCCACGCCACCG
CGCATGC$_m$ATCCTGCTAGCGGGTGCGGTGGC	CGCATGC$_m$ATCCTGCTATGCGGGTGCGGTGGC	CGCATGC$_m$ATCCTGCTATTGCGGGTGCGGTGGC
12	*13*	
GCGTA$_m$CGTAGGACGATAACCGCCCACGCCACCG	GCGTA$_m$CGTAGGACGATAACCCGCCCACGCCACCG	
CGCATGC$_m$ATCCTGCTATTGGCGGGTGCGGTGGC	CGCATGC$_m$ATCCTGCTATTGGGCGGGTGCGGTGGC	

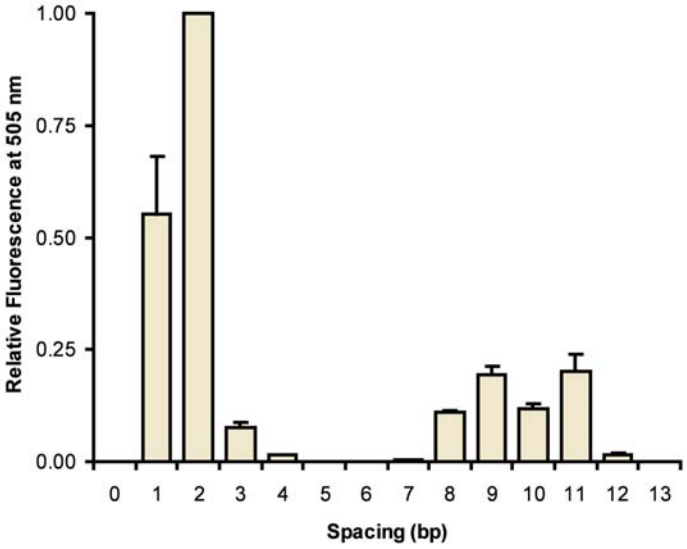

Figure S6. The reassembly of the mCpG-SEER system as a function of target site spacing.

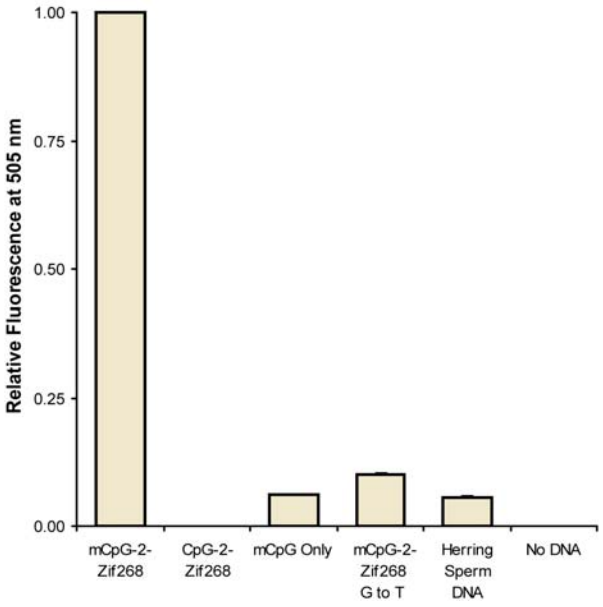

Figure S7. Specificity of the mCpG-SEER system using 100 nM, 2bp spacing targets.

References

(1) Hendrich, B.; Abbott, C.; McQueen, H.; Chambers, D.; Cross, S.; Bird, A. *Mammalian Genome* **1999**, *10*, 906-912.
(2) Hendrich, B.; Bird, A. *Mol.Cell. Biol.* **1998**, *18*, 6538-6547.
(3) Stains, C. I.; Porter, J. R.; Ooi, A. T.; Segal, D. J.; Ghosh, I. *J. Am. Chem. Soc.* **2005**, *127*, 10782-10783.

Complete Reference 3 Citation

Costello, J. F.; Fruhwald, M. C.; Smiraglia, D. J.; Rush, L. J.; Robertson, G. P.; Gao, X.; Wright, F. A.; Feramisco, J. D.; Peltomaki, P.; Lang, J. C.; Schuller, D. E.; Yu, L.; Bloomfield, C. D.; Caligiuri, M. A.; Yates, A.; Nishikawa, R.; Huang, H. J. S.; Petrelli, N. J.; Zhang, X. L.; O'Dorisio, M. S.; Held, W. A.; Cavenee, W. K.; Plass, C. *Nat. Genet.* **2000**, *24*, 132-138.