J|A|C|S
ARTICLES
Published on Web 07/12/2006

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

Received February 6, 2006; Revised Manuscript Received May 21, 2006; E-mail: ghosh@email.arizona.edu

Abstract: Currently there are no direct methods for the sequence-specific detection of DNA-methylation at CpG dinucleotides, which provide a possible diagnostic marker for cancer. Toward this goal, we present a methodology termed mCpG-**SE**quence **E**nabled **R**eassembly (mCpG-SEER) of proteins utilizing a split green fluorescent protein (GFP) tethered to specific DNA recognition elements. Our system, mCpG-SEER, employs a zinc-finger attached to one-half of GFP to target a specific sequence of dsDNA, while a methyl-CpG binding domain protein attached to the complementary half of GFP targets an adjacent methylated CpG dinucleotide site. We demonstrate that the presence of both DNA sites is necessary for the reassembly and concomitant fluorescence of the reassembled GFP. We further show that the GFP-dependent fluorescence reaches a maximum when the methyl-CpG and zinc-finger sites are separated by two base pairs and the fluorescence signal is linear to 5 pmol of methylated target DNA. Finally, the specificity of this reporter system, mCpG-SEER, was found to be >40-fold between a methylated versus a nonmethylated CpG target site.

Introduction

Underlying the complex cellular regulatory pathways are intricate patterns of covalent modifications of DNA comprising the epigenome.[1] These modifications are integral to normal cellular processes such as transcriptional regulation.[2] DNA methylation in higher eukaryotes occurs at cytosines in CpG dinucleotides by the transfer of a methyl group from S-adenosylmethionine to the C5-position of cytosine catalyzed by DNA methyltranferases. In normal cells methylation is predominantly found at repetitive genomic regions, such as transposable elements and satellite DNA. Strikingly, CpG islands associated with the promoter regions of genes involved in cancer are generally unmethylated in normal cells but aberrantly methylated in tumor cells.[3,4] Thus, the detection of aberrant patterns of methylation is an emerging area of early detection for human carcinomas. To recognize specific sites of DNA methylation we would need a system that can recognize specific sequences of double-stranded (dsDNA) as well as methylated DNA.

The site-specific recognition of dsDNA can be achieved by using either sequence-specific DNA-binding proteins or designed polyamides. Of special note is the zinc-finger class of DNA-binding proteins that consist of tandem repeats of a $\beta\beta\alpha$ domain, stabilized by coordination to zinc. The α-helix of this domain can specifically recognize a 3 bp tract of dsDNA in the major groove.[5] Moreover, protein design and selection approaches have helped identify an almost complete lexicon of zinc-fingers for targeting 3 bp tracts.[6−10] These seminal advances allow for the design of sequence-specific proteins to recognize virtually any dsDNA target of interest.[11] Another elegant approach for dsDNA detection has also been recently demonstrated using sequence-specific polyamides.[12,13] However, neither zinc-fingers nor polyamides alone can currently detect subtle chemical modifications such as DNA methylation in a sequence-specific manner.

We present a new approach for the direct detection of specific sites of cytosine methylation building upon our recently developed methodology for the detection of double-stranded DNA (dsDNA) called **SE**quence **E**nabled **R**eassembly (SEER). Direct detection of specific sequences of double-stranded DNA (dsDNA) utilizing the SEER approach was recently demonstrated using modular zinc-finger DNA-binding domains attached to split reporter proteins, such as GFP[14] and β-lacta-

[†] University of Arizona.
[‡] University of California, Davis.

(1) Hotchkiss, R. D. *J. Biol. Chem.* **1948**, *175*, 315−332.
(2) Jaenisch, R.; Bird, A. *Nat. Genet.* **2003**, *33*, 245−254.
(3) Costello, J. F.; et al. *Nat. Genet.* **2000**, *24*, 132−138.
(4) Esteller, M.; Corn, P. G.; Baylin, S. B.; Herman, J. G. *Cancer Res.* **2001**, *61*, 3225−3229.
(5) Wolfe, S. A.; Nekludova, L.; Pabo, C. O. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 183-212.
(6) Choo, Y.; Klug, A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 11163−11167.
(7) Dreier, B.; Beerli, R. R.; Segal, D. J.; Flippin, J. D.; Barbas, C. F., III. *J. Biol. Chem.* **2001**, *276*, 29466−29478.
(8) Dreier, B.; Fuller, R. P.; Segal, D. J.; Lund, C. V.; Blancafort, P.; Huber, A.; Koksch, B.; Barbas, C. F., III. *J. Biol. Chem.* **2005**, *280*, 35588−35597.
(9) Elrod-Erickson, M.; Pabo, C. O. *J. Biol. Chem.* **1999**, *274*, 19281−19285.
(10) Segal, D. J.; Dreier, B.; Beerli, R. R.; Barbas, C. F., III. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 2758−2763.
(11) Segal, D. J. *Methods* **2002**, *26*, 76−83.
(12) Fechter, E. J.; Olenyuk, B.; Dervan, P. B. *J. Am. Chem. Soc.* **2005**, *127*, 16685−16691.
(13) Warren, C. L.; Kratochvil, N. C.; Hauschild, K. E.; Foister, S.; Brezinski, M. L.; Dervan, P. B.; Phillips, G. N., Jr.; Ansari, A. Z. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 867−872.

mase.[15] In our current design we chose GFP as our split-reporter protein due to its undetectable background fluorescence.[16,17] We envisioned that the direct detection of mCpG sites could be achieved by redesigning the SEER system to recognize CpG methylation. In this new design we envisioned that a sequence specific DNA-binding protein could recognize a dsDNA site adjacent to the mCpG site, while a second DNA-binding protein, selective for mCpG sites, could then be used to detect site-specific methylation. We hypothesized that the two DNA binding proteins attached to two halves of GFP would localize on adjacent sites on a target dsDNA resulting in the reassembly of the parent GFP fold. The DNA-localized split-GFP would then autocatalyze fluorophore formation, thus producing an observable signal that directly reports on the methylation status of a specific methylated CpG site.

To recognize methylated DNA we have focused on the naturally occurring methyl-CpG binding domain (MBD) family of proteins that consists of MBD1, MBD2, MBD3, MBD4 and MeCP2.[18,19] NMR structures of MBD1 (with and without DNA)[20,21] and of MeCP2 (without DNA),[22] have shown that these MBDs have similar structures and make specific contacts in the major groove of DNA. We envisioned that by using a protein from this family we could detect methylation in CpG islands by redesigning our previously described SEER methodology.

As a first test of our proposed detection system for mCpG sites, we chose the well-characterized MBD2 protein from humans. MDB2 has a binding affinity of 2.7 nM for mCpG sites, while its binding affinity for nonmethylated-CpG sites is at least 70-fold less.[23] We hypothesized that this difference in binding affinities would allow us to selectively target mCpG sites versus nonmethylated-CpG sites. Since numerous sites on a genome are methylated, we also needed to introduce sequence specificity, which can be readily achieved by either natural or designed zinc-fingers. As proof of concept, we employed the well-studied Zif268 zinc-finger (K_d = 6 nM)[24] to recognize a site adjacent to the mCpG site. Molecular modeling suggested that attachment of the zinc-finger specificity domain to the N-terminal GFP fragment and the MBD to the C-terminal GFP fragment would result in a viable protein−DNA complex positioned for reassembly of the GFP halves (Figure 1). Each protein fragment was designed with a 15-amino acid linker separating the split-protein reporter from the DNA-binding domain to ensure conformational flexibility as 30 amino acids can theoretically span a distance of approximately 100 Å. We considered possible steric clashes due to the juxtaposition of the MBD and zinc-finger domains (Figure 1, inset) on DNA and chose to focus initial experiments in the context of a dsDNA target containing a 10 bp spacing (34 Å) between the mCpG and zinc-finger site. The 10 bp spacing had been shown to be optimal in the context of our previous SEER systems. We hypothesized that this separation between binding sites would both minimize steric clashes and position the mCpG-SEER proteins on the same surface of dsDNA.

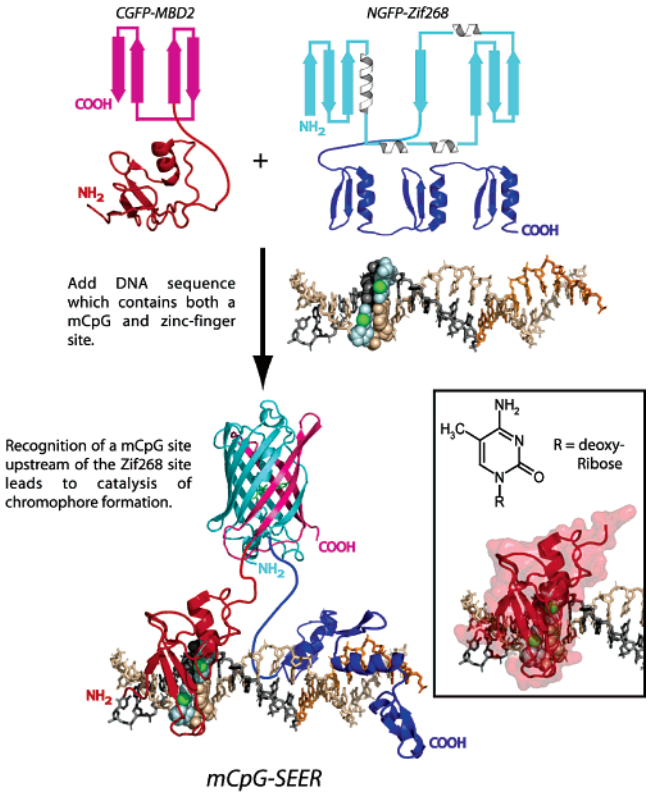

Figure 1. mCpG-SEER strategy. CGFP−MBD2 (pink and red) and NGFP−Zif268 (cyan and blue) are shown in ribbon form. The mCpG island is shown as spheres, cytosines are colored cyan with methyl groups highlighted in green, the zinc-finger binding site is highlighted in orange. (Inset) Structure of 5-methylcytosine as well as the surface of an MBD bound to DNA.

Experimental Section

General Materials. All materials were obtained from Sigma-Aldrich unless otherwise noted. 2XYT media was purchased from Becton Dickinson, and $ZnCl_2$ was obtained from EM Sciences. Urea and DTT were obtained from Research Products International.

Cloning, Expression, and Purification of mCpG-SEER Proteins. A plasmid containing an *Escherichia coli* optimized gene for human MBD2 (residues 147−215) was obtained from GenScript. The following primers were used to clone MBD2 into an existing cassette (containing the C-terminal portion of GFP, residues 158−238, separated by a 15-amino acid linker): 5′-GCGTATGAATTCGGAAAGCG-GCAAACGC-3′ and 5′-CGGTTAACCGGTCATTTTGCCGGTACG-3′. The pETDuet plasmid containing the CGFP−MBD2 gene was then transformed into electrocompetent BL21(DE3) cells (Novagen). An overnight culture of these cells was used to inoculate a 1-L culture of 2XYT with Amp (Research Products International) at an OD_{600} of 0.05. Protein expression was induced at an OD_{600} of 1.32 using 1 mM IPTG (Research Products International) for 3 h. Cells were pelleted and frozen overnight at −20 °C. Cells were then lysed using sonication and clarified by centrifugation. SDS-PAGE analysis showed that CGFP−MBD2 was expressed in both the soluble and insoluble fractions.

(14) Stains, C. I.; Porter, J. R.; Ooi, A. T.; Segal, D. J.; Ghosh, I. *J. Am. Chem. Soc.* **2005**, *127*, 10782−10783.
(15) Ooi, A. T.; Stains, C. I.; Ghosh, I.; Segal, D. J. *Biochemistry* **2006**, *45*, 3620−3625.
(16) Ghosh, I.; Hamilton, A. D.; Regan, L. *J. Am. Chem. Soc.* **2000**, *122*, 5658−5659.
(17) Magliery, T. J.; Wilson, C. G. M.; Pan, W. L.; Mishler, D.; Ghosh, I.; Hamilton, A. D.; Regan, L. *J. Am. Chem.* Soc. **2005**, *127*, 146−157.
(18) Hendrich, B.; Abbott, C.; McQueen, H.; Chambers, D.; Cross, S.; Bird, A. *Mamm. Genome* **1999**, *10*, 906−912.
(19) Hendrich, B.; Bird, A. *Mol. Cell. Biol.* **1998**, *18*, 6538−6547.
(20) Ohki, I.; Shimotake, N.; Fujita, N.; Jee, J. G.; Ikegami, T.; Nakao, M.; Shirakawa, M. *Cell* **2001**, *105*, 487−497.
(21) Ohki, I.; Shimotake, N.; Fujita, N.; Nakao, M.; Shirakawa, M. *EMBO J.* **1999**, *18*, 6653−6661.
(22) Wakefield, R. I. D.; Smith, B. O.; Nan, X. S.; Free, A.; Soteriou, A.; Uhrin, D.; Bird, A. P.; Barlow, P. N. *J. Mol. Biol.* **1999**, *291*, 1055−1065.
(23) Fraga, M. F.; Ballestar, E.; Montoya, G.; Taysavang, P.; Wade, P. A.; Esteller, M. *Nucleic Acids Res.* **2003**, *31*, 1765−1774.
(24) Pavletich, N. P.; Pabo, C. O. *Science* **1991**, *252*, 809−817.

Table 1. Specificity Targets for mCpG-SEER

2 bp Targets	10 bp Targets
$_m$CpG-2-Zif268	$_m$CpG-10-Zif268
GCGTA$_m$CGTACGCCCACGCCACCG	GCGTA$_m$CGTAGGACGATACGCCCACGCCACCG
CGCATGC$_m$ATGCGGGTGCGGTGGC	CGCATGC$_m$ATCCTGCTATGCGGGTGCGGTGGC
CpG-2-Zif268	CpG-10-Zif268
GCGTACGTACGCCCACGCCACCG	GCGTACGTAGGACGATACGCCCACGCCACCG
CGCATGCATGCGGGTGCGGTGGC	CGCATGCATCCTGCTATGCGGGTGCGGTGGC
$_m$CpG Only	$_m$CpG Only
GCGTA$_m$CGTAGCACATAGGCACCG	GCGTA$_m$CGTAGGACGATAGCACATAGGCACCG
CGCATGC$_m$ATCGTGTATCCGTGGC	CGCATGC$_m$ATCCTGCTATCGTGTATCCGTGGC
$_m$CpG-2-Zif268 G to T	$_m$CpG-10-Zif268 G to T
GCGTA$_m$CGTACGCCCACGCCACCG	GCGTA$_m$CGTAGGACGATACGCCCACGCCACCG
CGCATGC$_m$ATGCGTGTGCGGTGGC	CGCATGC$_m$ATCCTGCTATGCGTGTGCGGTGGC

[a] Red, blue, and pink indicate the respective MBD2, Zif268, and mutation sites.

CGFP−MBD2 from either fraction was purified separately under denaturing conditions using Ni-NTA resin (Qiagen) using Buffer A (10 mM Tris-HCl at pH = 7.5, 100 mM NaCl, 1 mM DTT, and 100 μM $ZnCl_2$) plus 4 M Urea. The protein was characterized by SDS-PAGE and mass spectrometry (Supporting Information), concentrations were determined using absorbance measurements at 280 nm (ϵ = 14 440 M^{-1} cm^{-1}).

The NGFP−Zif268 fusion protein (containing residues 1−157 of GFP) was expressed and purified as described previously.[14] NGFP−Zif268 was characterized by SDS-PAGE and mass spectrometry (Supporting Information), concentrations were determined using absorbance measurements at 280 nm (ϵ = 17 210 M^{-1} cm^{-1}).

Detailed information including the sequences of NGFP−Zif268 and CGFP−MBD2 as well as detailed purification procedures can be found in the Supporting Information.

Refolding Expirements. HPLC-purified target DNA sequences were obtained from IDT, and herring sperm DNA was purchased from Invitrogen. Oligos were annealed in 1× BamHI Buffer (NEB) using the following procedure: heating to 95 °C for 7 min, cooling to 56 °C at a rate of 1 °C/min, equilibrating at 56 °C for 5 min, and finally cooling to 25 °C at a rate of 1 °C/min using a Techne Genius thermocycler.

All refolding experiments were conducted using 3.5 kD MWCO Slide-A-Lyzer Dialysis Cassettes (Pierce). Sequences for all DNA targets used for testing the specificity and dependence on target-site spacing of mCpG-SEER are given in Table 1 and the Supporting Information (Table S1), respectively.

For experiments used to determine mCpG-SEER specificity and dependence on target site spacing, NGFP−Zif268 (5 μM), CGFP−MBD2 (20 μM), and target DNA (2.5 μM for 10 bp targets; 1 μM for 2 bp targets) were mixed in Buffer A containing 4 M Urea. These samples were refolded by gradually dialyzing into Buffer A containing 2, 1, and 0.5 M urea and then twice into Buffer A with no urea (final concentration of urea = 350 nM) over a period of 2 days. All refolding experiments were conducted at 4 °C in uncovered chambers; after refolding, each sample was placed at 4 °C. Fluorescence spectra were acquired 2 days post-refolding for the 10 bp targets and 7 h post-refolding for the 2 bp targets.

For experiments used to determine the limit of detection of dsDNA for mCpG-SEER, NGFP−Zif268 (5 μM), CGFP−MBD2 (20 μM), and decreasing concentrations of target DNA (500, 250, 100, and 50 nM) containing a 2 bp spacing between the mCpG and zinc-finger sites were mixed in Buffer A containing 4 M urea. These samples were refolded by gradually dialyzing into Buffer A containing 2, 1, 0.5 M urea, and twice into Buffer A with no urea (final concentration of urea = 350 nM) over a period of 2 days. All refolding experiments were conducted at 4 °C in uncovered chambers; after refolding, each sample was placed at 4 °C. Fluorescence spectra were acquired 2 days post-refolding.

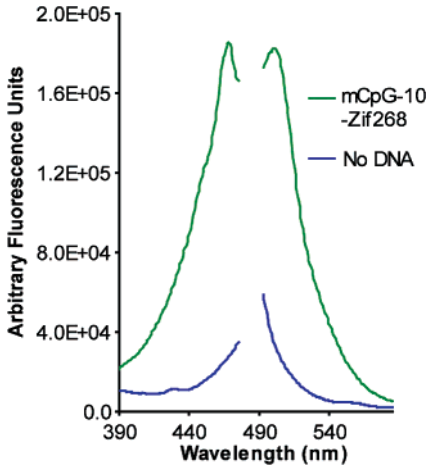

Figure 2. Fluorescence excitation (monitored at 505 nm) and emission (excited at 468 nm) spectra of NGFP−Zif268 (5 μM) + CGFP−MBD2 (20 μM) in the presence (green) or absence (blue) of 2.5 μM target DNA (mCpG-10-Zif268) in buffer A.

Fluorescence Measurements. All spectra were acquired on a Photon Technology International spectrofluorometer with excitation and emission wavelengths of 468 and 505 nm, respectively. Slit widths were set to 5 nm for excitation and 10 nm for emission. Relative intensities were determined by subtracting background emission values at 505 nm for refolded mCpG-SEER proteins at the same concentration with no DNA; these samples were then normalized relative to the sample with the highest fluorescence intensity.

Results and Discussion

Initial mCpG Detection. Initial modeling studies suggested that a target oligo containing a 10 bp separation between the mCpG and zinc-finger sites would allow the DNA-binding proteins to be positioned approximately on the same face of dsDNA while avoiding noticeable steric clashes upon binding. To test this idea 5 μM of NGFP−Zif268 and 20 μM of CGFP−MBD2, were refolded in the presence or absence of a methylated (mCpG-10-Zif268) dsDNA target which contained a 10 bp spacing between the mCpG and Zif268 sites. Fluorescence spectra were acquired 2 days post refolding by exciting at 468 nm and monitoring emission at 505 nm. Signal due to GFP reassembly was observed only for samples containing both halves of GFP in the presence of the methylated target oligonucleotide (Figure 2), whereas no fluorescence was observed without DNA. These experiments clearly validated our design and indicated that the designed mCpG-SEER system could detect mCpG islands and that the split-protein reporter did not spontaneously reassemble.

mCpG-SEER Specificity. To further address the specificity of the mCpG-SEER system, we independently varied each target site and assayed the ability of the GFP fragments to refold and produce a fluorescence signal. Control DNA sequences containing a 10 bp spacer between the MBD2 and Zif268 sites were designed as follows: (1) no methylation (CpG-10-Zif268), (2) no Zif268 site (mCpG only), and (3) a single bp mutation in the Zif268 site (mCpG-10-Zif268, G to T). Mixtures of the two proteins, NGFP−Zif268 (5 μM) and CGFP−MBD2 (20 μM), were allowed to refold separately in the presence of 2.5 μM of each target dsDNA sequence as well as an equivalent amount of herring sperm DNA (Figure 3). Importantly, a very low level

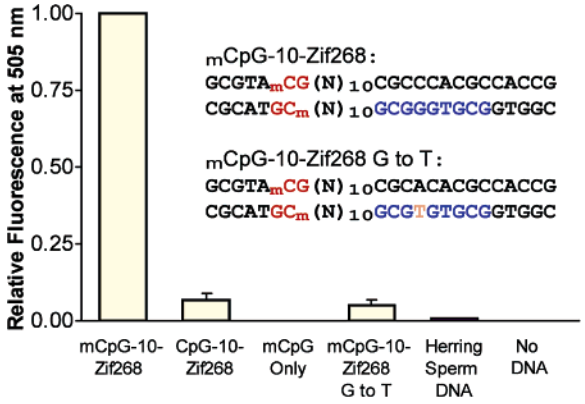

Figure 3. Fluorescence emission at 505 nm of NGFP−Zif268 (5 μM) + CGFP−MBD2 (20 μM) in the presence of indicated double-stranded DNA controls (2.5 μM each).

of fluorescence was observed for the CpG-10-Zif268 site, which was 15-fold lower than the cognate methylated target (Figure 3).

These data clearly demonstrate the ability of mCpG-SEER to discern between methylated- versus nonmethylated-CpG sites via the use of the methyl-CpG binding domain. Additionally, the requirement for the adjacent zinc-finger binding site was verified by the lack of fluorescence in the presence of the target that contained an arbitrary DNA sequence adjacent to a methylated-CpG site. Furthermore, the exquisite sequence specificity of the zinc-finger was validated by almost a complete loss of fluorescence by a single-base pair mutation (G to T) in the zinc-finger target site. Last the absence of an observable signal for the sample containing nonspecific herring sperm DNA reconfirms the requirement for both a mCpG site and a zinc-finger site to be in close proximity to induce reassembly of GFP.

mCpG-SEER Dependence on Target Site Spacing. To further address the requirements for target site spacing between the CpG and zinc-finger sites we devised a set of experiments to probe the GFP reassembly-dependent fluorescence signal intensity as a function of target site distance. Fourteen dsDNA targets were tested with spacings of zero through thirteen bp's between the mCpG and zinc-finger binding sites. Mixtures of the two proteins, NGFP−Zif268 (5 μM) and CGFP−MBD2 (20 μM), were allowed to refold separately in the presence of 2.5 μM of each target dsDNA sequence (Figure 4). An analogous experiment using 5 μM of each mCpG-SEER protein was also performed and produced similar results (Supporting Information, Figure S6).

The results of these experiments yield several interesting insights into the requirements for maximum signal generation for the mCpG-SEER system. No reassembly is observed when DNA-binding domains are juxtaposed next to one another (0 bp spacing), presumably caused by the inaccessibility of both target sites simultaneously by the DNA-binding proteins (molecular modeling indicates that MBD2 occupies a surface area greater than just the mCpG site on dsDNA, Figure 1 inset). Second, a sharp increase in fluorescence is observed for target site spacings at 1 and 2 bp spacing. These results suggest that at 2 bp the two GFP halves are best oriented for reassembly probably due to the relatively close proximity (the local concentration of each half determined by modeling is approximately 5 mM). Third, the helical nature of dsDNA is visualized in the repeating maxima and minima of GFP

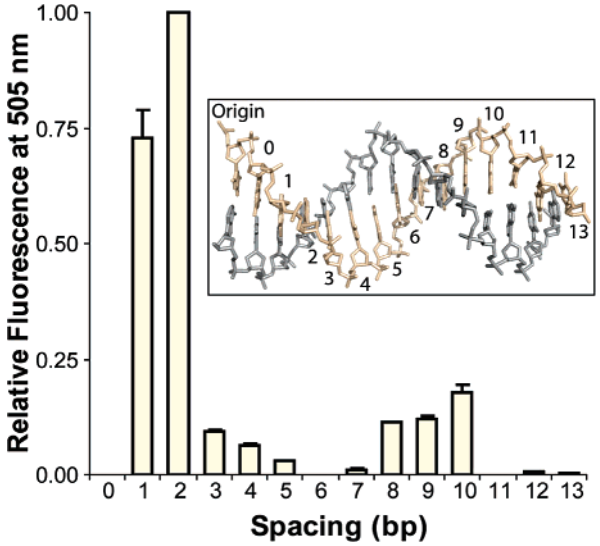

Figure 4. mCpG-SEER fluorescence emission spectra at 505 nm. NGFP−Zif268 (5 μM) + CGFP−MBD2 (20 μM) in the presence of target dsDNA (2.5 μM each) as a function of increasing spacing between target sites.

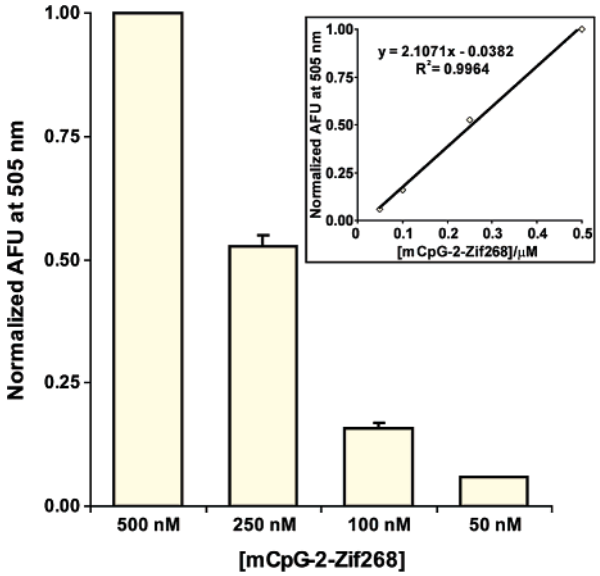

Figure 5. Relative fluorescence emission at 505 nm of NGFP−Zif268 (5 μM) + CGFP−MBD2 (20 μM) in the presence of the decreasing concentrations of mCpG-2-Zif268 target DNA. (Inset) Linear fit to the data.

reassembly as observed by the fluorescence intensity changes as a function of DNA length and pitch. We attribute this phenomenon to the increasing distance as well as positioning of each DNA-binding protein of the surface of dsDNA. As the distance between target sites is increased, the orientations of the binding sites also change along the helical axis. This yields discrete maxima and minima and is further supported by results from similar experiments using our SEER-GFP system (Supporting Information, Figure S2) and SEER-Lac system.[15]

mCpG-SEER Limit of Detection. The limit of detection of mCpG-SEER was determined for an oligo containing a 2 bp separation between the mCpG site and the zinc-finger site (mCpG-2-Zif268). Mixtures of the two mCpG-SEER proteins, NGFP−Zif268 (5 μM) and CGFP−MBD2 (20 μM), were allowed to refold in the presence of decreasing amounts of mCpG-2-SEER target DNA (Figure 5). These experiments show that 50 nM dsDNA (5 pmol) is readily visible over background

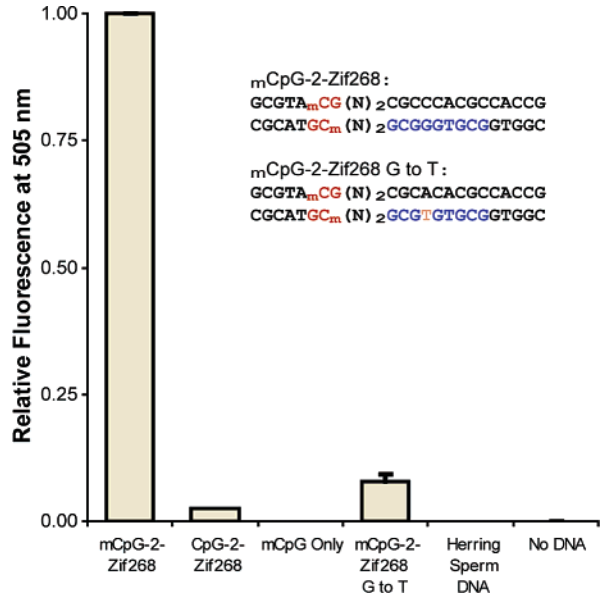

Figure 6. Fluorescence emission at 505 nm of NGFP−Zif268 (5 μM) + CGFP−MBD2 (20 μM) in the presence of indicated double-stranded DNA controls (1 μM each).

and that the signal scales linearly with dsDNA concentration (Figure 5, inset).

Specificity of mCpG-SEER at 2bp Spacing. For a final set of experiments we interrogated the sensitivity for mCpG islands of mCpG-SEER at the optimal 2 bp spacing between target sites. Mixtures of the two mCpG-SEER proteins, NGFP−Zif268 (5 μM) and CGFP−MBD2 (20 μM), were allowed to refold in the presence of DNA controls (Table 1) at 1 μM and an equivalent amount of herring sperm DNA. Results of these experiments are given in Figure 6.

These experiments reconfirmed the requirement for an adjacent zinc-finger site to achieve efficient reassembly, and the choice of the 2 bp spacing resulted in significant optimization of signal from the target site. The specificity for our system can be observed in the >40-fold difference in signal due to GFP reassembly between the methylated and nonmethylated target dsDNA site. It is possible that this high degree of specificity at high protein concentrations may be due to a decreased affinity of MBD2 for unmethylated CpG sites when fused to unfolded split proteins such as CGFP.[23] A second mCpG-SEER experiment was also carried out using 100 nM target DNA (2 bp spacing) and yielded similar results (Supporting Information, Figure S7), with no observable signal from the nonmethylated target site.

Conclusion

Previous work in protein design has demonstrated the ability to assemble multifunctional protein complexes using specific domains[25−29] and the ability to select proteins, through the use of phage display, which bind unnatural DNA nucleobases.[30] We build upon these concepts by employing a naturally occurring DNA-binding protein which recognizes the methylation of cytosine residues in CpG sites to assemble a ternary protein complex for the site-specific detection of CpG methylation. This detection system builds upon our previously described methodology for detecting DNA in its native state utilizing split-reporter proteins termed SEER.

The specificity of mCpG-SEER for methylation will allow for a sensitive alternative method for the direct detection of mCpG sites. Future experiments in this area will be focused on optimizing conditions to generate efficient reassembly of the current split-GFP or other GFP variants[31,32] as well as analyzing the ability of mGpG-SEER to detect mCpG in a genomic context where numerous mCpG sites will be present.

The sensitivity of mCpG-SEER to the helical turn of dsDNA yields insights into the application of this system for detection of mCpG sites on specific promoters. Unexpectedly, we found that, while the mCpG-SEER signal is dependent on the helical nature of dsDNA, a 2 bp separation between the mCpG island and zinc-finger site yielded maximal GFP reassembly. Systems which employ zinc-fingers, which are specific for promoter sequences 2 bp's downstream of a suspected mCpG site, would therefore be the most sensitive. These principles will be used in future experiments to determine the methylation status of specific promoter sites of genes associated with human cancers such as BRCA1,[33] p15,[34] and GSTP1.[35] Thus, our approach toward the direct recognition of methylated DNA may be utilized in the early diagnosis of human carcinomas based on site-specific patterns of methylation.

We have clearly demonstrated the successful detection of site-specific CpG island methylation in dsDNA with our rationally designed mCpG-SEER system. This strategy could be expanded to other split-proteins, such as β-lactamase[15,36] or luciferase,[37] which could further amplify signal by substrate turnover and possibly increase sensitivity. We believe that our approach is modular, and given the available lexicon of zinc-fingers,[7,8,10] almost any sequence of interest adjacent to a mCpG site could be targeted. Thus, mCpG-SEER represents a new and potentially useful method for the direct detection of CpG methylation, which may find applications in delineating the epigenome and in cancer research.

Acknowledgment. C.I.S. and J.L.F. were supported by an NIH training grant. We thank members of the Ghosh lab and Bernard Futscher for helpful comments. Models were rendered using PyMol; DeLano, W. L. (www.pymol.org).

Supporting Information Available: Details of cloning, sequencing, purification, characterization, and the complete ref 3 citation are available free of charge via the Internet at http://pubs.acs.org.

JA060681J

(25) Devit, M.; Cullen, P. J.; Branson, M.; Sprague, G. F., Jr.; Fields, S. *Genome Res.* **2005**, *15*, 560−5.
(26) Kim, C. Y.; Alekseyev, V. Y.; Chen, A. Y.; Tang, Y.; Cane, D. E.; Khosla, C. *Biochemistry* **2004**, *43*, 13892−8.
(27) Nevado, J.; Gaudreau, L.; Adam, M.; Ptashne, M. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 2674−7.
(28) Park, S. H.; Zarrinpar, A.; Lim, W. A. *Science* **2003**, *299*, 1061−4.
(29) Tan, W.; Zhu, K.; Segal, D. J.; Barbas, C. F., III; Chow, S. A. *J. Virol.* **2004**, *78*, 1301−13.
(30) Simon, M. D.; Shokat, K. M. *J. Am. Chem. Soc.* **2004**, *126*, 8078−9.
(31) Shaner, N. C.; Steinbach, P. A.; Tsien, R. Y. *Nat. Methods* **2005**, *2*, 905−909.
(32) Hu, C. D.; Kerppola, T. K. *Nat. Biotechnol.* **2003**, *21*, 539−545.
(33) Rice, J. C.; Ozcelik, H.; Maxeiner, P.; Andrulis, I.; Futscher, B. W. *Carcinogenesis* **2000**, *21*, 1761−1765.
(34) Herman, J. G.; Jen, J.; Merlo, A.; Baylin, S. B. *Cancer Res.* **1996**, *56*, 722−727.
(35) Song, J. Z.; Stirzaker, C.; Harrison, J.; Melki, J. R.; Clark, S. J. *Oncogene* **2002**, *21*, 1048−1061.
(36) Galarneau, A.; Primeau, M.; Trudeau, L. E.; Michnick, S. W. *Nat. Biotechnol.* **2002**, *20*, 619−622.
(37) Paulmurugan, R.; Gambhir, S. S. *Anal. Chem.* **2003**, *75*, 1584−1589.