

Potential role of DNA methylation as a facilitator of target search processes for transcription factors through interplay with methyl-CpG-binding proteins

Catherine A. Kemme, Rolando Marquez, Ross H. Luu and Junji Iwahara*

Department of Biochemistry and Molecular Biology, Sealy Center for Structural Biology and Molecular Biophysics, University of Texas Medical Branch, Galveston, TX 77555-1068, USA

Received February 01, 2017; Revised April 19, 2017; Editorial Decision April 21, 2017; Accepted May 06, 2017

ABSTRACT

Eukaryotic genomes contain numerous non-functional high-affinity sequences for transcription factors. These sequences potentially serve as natural decoys that sequester transcription factors. We have previously shown that the presence of sequences similar to the target sequence could substantially impede association of the transcription factor Egr-1 with its targets. In this study, using a stopped-flow fluorescence method, we examined the kinetic impact of DNA methylation of decoys on the search process of the Egr-1 zinc-finger protein. We analyzed its association with an unmethylated target site on fluorescence-labeled DNA in the presence of competitor DNA duplexes, including Egr-1 decoys. DNA methylation of decoys alone did not affect target search kinetics. In the presence of the MeCP2 methyl-CpG-binding domain (MBD), however, DNA methylation of decoys substantially (~10–30-fold) accelerated the target search process of the Egr-1 zinc-finger protein. This acceleration did not occur when the target was also methylated. These results suggest that when decoys are methylated, MBD proteins can block them and thereby allow Egr-1 to avoid sequestration in non-functional locations. This effect may occur *in vivo* for DNA methylation outside CpG islands (CGIs) and could facilitate localization of some transcription factors within regulatory CGIs, where DNA methylation is rare.

INTRODUCTION

A pre-requisite for transcriptional activation of genes is the association of transcriptional factors with cis-regulatory elements such as promoters and enhancers in the genome (1). To regulate genes, transcription factors must first locate functionally important sites within cis-regulatory elements.

Sequence specificity in DNA-binding of transcription factors is important for their target association. Eukaryotic transcription factors recognize relatively short (typically <10 bp) sequences through direct base readout or indirect shape readout (2) and bind specifically to target sequences with high affinity. In many cases, however, sequence specificity alone is clearly insufficient for eukaryotic transcription factors to uniquely locate functional target sites because an overwhelming number of non-functional high-affinity sites are also present in the genome (3–5). These sites can serve as natural decoys that sequester transcription factors.

This problem is exemplified by the zinc-finger transcription factor Egr-1 (also known as Zif268 or NGFI-A), which recognizes 9-bp target sequences as a monomer (6). In mammals, Egr-1 is induced by particular stimuli or stress to cells and plays important roles in the nervous and cardiovascular systems (7–10). Because Egr-1 regulates ~10² genes (11,12) and each cis-regulatory element may contain only up to several Egr-1 sites, the total number of functional targets for Egr-1 is estimated to be ~10²–10³ sites. Upon induction, the number of Egr-1 molecules can reach ~10⁴ per nucleus (13). However, the genome contains a far greater number of non-functional sites that are identical or similar to the target sequence and exhibit high affinities for Egr-1. The sequence specificity and binding free energy as a function of DNA sequence have been well studied for Egr-1 (14–16). It was shown that the genome contains millions of high-affinity sequences for Egr-1, and even though ~90% of them are buried in nucleosomes, ~10⁶ sites should remain accessible (15). Due to the huge number of decoys, it seems difficult for Egr-1 to adequately occupy each functional target because sequestration of Egr-1 molecules may occur in off-target locations.

Despite these circumstances, how can Egr-1 reach functionally important targets? This is probably relevant to Egr-1's co-localization with CpG islands (CGIs) *in vivo*, which has been shown in some cell types in several genome-wide ChIP-on-chip and ChIP-seq studies (17–19). CGIs are regions of DNA with a high density of CpG dinucleotides

*To whom correspondence should be addressed. Tel: +1 409 747 1403; Fax: +1 409 772 6334; Email: j.iwahara@utmb.edu

(CpGs) present within 200–3,000 bp. The human genome contains ~25,000 CGIs, most of which are sites of transcription initiation (20). In fact, the majority (~70%) of human gene promoters are associated with CGIs, although CGIs represent only 0.8% of the human genome (21). Perhaps surprisingly, CGIs are typically unmethylated for active genes, although CpGs are the substrates of major DNA methyltransferases and ~85% of CpGs in the human genome are methylated (22). Because the Egr-1 consensus sequence, GCGTGGGCG, contains two CpGs, it is reasonable to consider that functionally important Egr-1 sites within CGIs are unmethylated, whereas Egr-1 decoys outside CGIs are methylated.

Given this information, one may think that DNA methylation may diminish affinity of decoys, allowing Egr-1 to locate unmethylated targets within CGIs. CpG methylation moderately changes structural properties of DNA and thereby affects some transcription factors (23). However, our recent biochemical study showed that the affinity of Egr-1 for its recognition sequence is unaffected by CpG methylation (24). The crystal structures of Egr-1 complexes with unmethylated and methylated target DNA (6,24–26) also show that Egr-1 is well suited to recognize both unmethylated and methylated targets. Thus, DNA methylation alone cannot drive Egr-1 to unmethylated recognition sequences within CGIs.

Based on these considerations, we formulated a hypothetical model to explain how Egr-1 locates unmethylated targets within CGIs. This model is schematically depicted in Figure 1 and would work in the following manner. Because CGIs represent only 0.8% of the genome, the vast majority of natural decoys for Egr-1 should be located outside CGIs and therefore methylated at CpGs. Due to this methylation, methyl-CpG-binding domain (MBD) proteins occupy these high-affinity non-functional sites, allowing Egr-1 to avoid being trapped there. In this manner, MBD proteins would indirectly guide Egr-1 to unmethylated targets within CGIs. Supporting this model, the genome-wide analysis of bisulfite sequencing and ChIP-seq data show that DNA regions ± 300 bp from Egr-1 ChIP-seq peak centers are unmethylated (27). Our model differs from previous models for competition between MBD proteins and other transcription factors (20,28). Our competitive model focuses on decoys outside CGIs, whereas the previous models focus on competition for functional sites within CGIs. Due to the large number of decoys in the genome, our model requires a high expression level of MBD proteins. The methyl-CpG-binding protein 2 (MeCP2) satisfies this requirement in some cell types; for instance, neuronal cells are estimated to contain as many as 10^7 MeCP2 molecules per nucleus (29). In cell types where MeCP2 expression is lower, other MBD proteins, such as Mbd1 and Mbd2, are often more abundant, according to the protein abundance database PaxDb (<http://pax-db.org>). Thus, it seems that the overall expression level of MBD proteins is considerably high in many cell types, satisfying the condition required for our model.

In the current paper, we present our biochemical study as the first step to assess this model. We used a simplified system involving the DNA-binding domains of Egr-1 and MeCP2 together with synthetic DNA duplexes. The

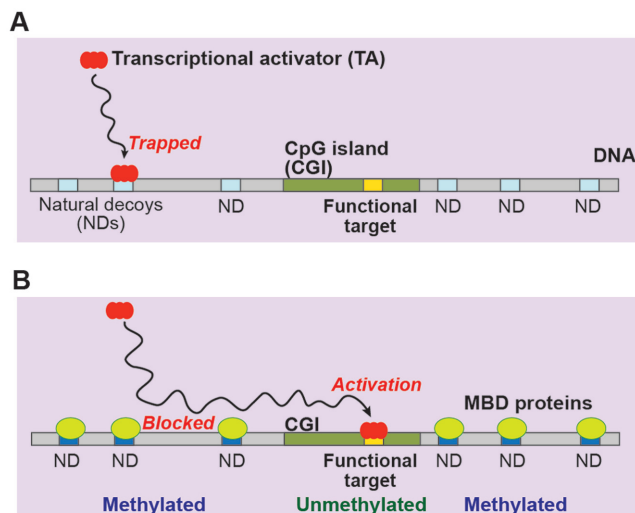


Figure 1. DNA methylation as a potential mechanism that facilitates the association of transcriptional activators with functional targets in CpG islands (CGIs) through competitive interplay with methyl-CpG-binding domain (MBD) proteins. The genome contains numerous non-functional high-affinity sites that are identical or similar to the target sequences of transcription activators. When these sites are accessible, they can serve as natural decoys that sequester transcriptional activators in non-functional locations (A). In vertebrate genomes, CpGs in CGI promoters of active genes are unmethylated, whereas ~85% of CpGs outside CGIs are methylated. For activators that recognize sequences containing CpG (e.g. GCGTGGGCG for Egr-1; see also Figure 6), it is likely that their functional targets in CGIs are unmethylated, whereas decoys outside CGIs are methylated. DNA methylation outside CGIs may enhance the association of transcriptional activators with unmethylated targets in CGIs, because highly expressed MBD proteins block sequestration of the transcription activators in methylated decoys (B).

Egr-1 DNA-binding domain comprised three zinc fingers is well suited for biochemical and biophysical research on the target search process (13,30–35). Taking advantage of this system, we have performed kinetic studies on the Egr-1 zinc-finger protein in the presence and absence of the MeCP2 MBD, when it is challenged with methylated CpG and unmethylated CpG sites within Egr-1 decoys. Our data demonstrate that DNA methylation of decoys accelerates Egr-1–target association through competitive interplay with the MeCP2 MBD, supporting our hypothetical model.

MATERIALS AND METHODS

Preparation of Egr-1 zinc-finger protein

The DNA-binding domain of the human Egr-1 protein (residues 335–423) comprised three zinc fingers was expressed in *Escherichia coli* and purified using affinity, size-exclusion and cation-exchange columns, as previously described (32). This construct was used in our previous nuclear magnetic resonance (NMR) (32–38), X-ray (24) and fluorescence (13,15,30,31) studies. For simplicity's sake, this construct is referred to as the Egr-1 zinc-finger protein hereafter. The Egr-1 zinc-finger protein was quantified using the Pierce bicinchoninic acid (BCA) protein assay kit (Thermo Fisher).

Preparation of MeCP2 MBD

A synthetic gene encoding the MBD of the human MeCP2 (residues 77–167) was sub-cloned into the XmaI/XhoI sites of the pET-49b vector (Novagen). *Escherichia coli* strain BL21(DE3) transformed with this plasmid was cultured at 37°C in 4 l of M9 media containing kanamycin (30 µg/ml). At OD₆₀₀ ≈ 0.8, protein expression was induced by 0.6 mM isopropyl β-D-thiogalactopyranoside, and the culture was continued at 37°C for additional 2 h. The *E. coli* cells were harvested and disrupted by sonication in a buffer containing 50 mM Tris•HCl (pH 7.5), 1 mM ethylenediaminetetraacetic acid (EDTA), 500 mM NaCl, 2 mM dithiothreitol, 5% glycerol, 1% Triton X-100 (Sigma-Aldrich) and a Roche protease inhibitor cocktail (one tablet per 50 ml). After centrifugation at 30,000 × *g* and 4°C for 20 min, the supernatant of the lysate was loaded on a glutathione S-transferase (GST)-Prep FF 16/10 column (GE Healthcare) equilibrated with 50 mM Tris•HCl (pH 7.5), 400 mM NaCl and 1% Triton X-100. The GST-MeCP2 MBD fusion protein was eluted using 50 mM Tris•HCl (pH 7.5), 400 mM NaCl and 10 mM glutathione. The fusion protein was cleaved with 100 units of HRV-3C protease (GenWay Biotech). After confirming the cleavage by polyacrylamide gel electrophoresis, the reaction mixture was concentrated to ~10 ml with an Amicon Ultra-15 device, and loaded onto a Sephacryl S100 size-exclusion column (GE Healthcare) equilibrated with 50 mM Tris•HCl (pH 7.5), 1,000 mM NaCl, 2 mM EDTA for separation. Fractions containing the MeCP2 MBD protein were pooled, and the buffer was exchanged to 50 mM Tris•HCl (pH 7.0), 200 mM NaCl and 5% glycerol. The protein solution was loaded onto a Resource-S cation-exchange column (GE Healthcare) equilibrated with 50 mM Tris•HCl (pH 7.0), 200 mM NaCl and 5% glycerol and then eluted with a gradient of 200–500 mM NaCl. The MeCP2 MBD was quantified by UV absorbance at 280 nm based upon an extinction coefficient of 11,460 M⁻¹ cm⁻¹ (<http://web.expasy.org/protparam/>).

Preparation of DNA

Chemically synthesized DNA strands were purchased from Integrated DNA Technologies, Inc. Each strand was purified with a Mono-Q anion-exchange column installed on a ÄKTA Purifier system (GE Healthcare). The sequences of the DNA duplexes used in the current study are shown in Figure 2A. The sequence of the decoy DNA duplexes (L_S and mL_S) was chosen based on our previous study (13). This 28-bp sequence contains a quasi-specific site matching 6 bp out of the 9 bp of the Egr-1 recognition sequence and exhibits ~400-fold higher affinity than completely non-specific 28-bp DNA L (13). The duplexes L_S and mL_S are identical in sequence, but differ in that the CpG sequence in the quasi-specific site is methylated in mL_S. All 5-methylcytosine for CpG methylation were introduced when DNA was chemically synthesized. A fluorescein amidite (FAM) labeled 143-bp DNA duplex containing an Egr-1 target sequence was generated by polymerase chain reaction and purified as described previously (30). This FAM-labeled 143-bp probe DNA was used in our previous stopped-flow kinetics studies (13,30,31,34). All other DNA duplexes were

prepared through annealing of complementary strands and removal of excess single-stranded DNA as described previously (30).

Stopped-flow fluorescence experiments

The target search kinetics of Egr-1 were measured at 20°C using an Applied Photophysics SX20-LED stopped-flow spectrofluorometer. In these experiments, the following two solutions were rapidly mixed in a 1:1 volume (~0.5 ml) ratio by the stopped-flow device: a solution of the Egr-1 zinc-finger protein, and a DNA solution of FAM-labeled probe DNA (143-bp) and competitor DNA, with the addition of the MeCP2 MBD in some experiments. Both solutions were in a buffer containing 10 mM Tris•HCl (pH 7.5), 0.2 µM ZnCl₂ and 150 mM KCl. Immediately after the flow for mixing had been stopped, the time course data of fluorescence intensity were collected for a period of 4–50 s with time intervals ranging from 0.02 to 0.05 s. A light-emitting diode with maximum intensity at 470 nm was used for excitation of the FAM fluorophore. The emission light that passed through a long-pass filter with a cutoff at 515 nm was recorded. This configuration with no monochromator involved increases sensitivity in fluorescence detection. However, compared to observations with monochromators, it also increases non-fluorescent background and reduces the percentage change in the FAM fluorescence intensity upon the target association. The total concentration of the competitor 28 bp duplexes was kept constant at 2 µM, though the concentration of the quasi-specific duplexes as Egr-1 decoys was either 0 or 150 nM. The concentration of the probe DNA (D_{tot}) was 2.5 nM, whereas the concentrations of the protein (P_{tot}) and competitor (C_{tot}) were varied. To create a pseudo-first-order condition to simplify the kinetic analysis (39), all binding reactions were conducted under conditions of $D_{\text{tot}} \ll P_{\text{tot}} \ll C_{\text{tot}}$. The apparent pseudo-first-order kinetic rate constant (k_{app}) for target association was determined from the time course of fluorescence intensity, $I(t)$, by non-linear least-squares fitting with $I(t) = (I_0 - I_{\infty})\exp(-k_{\text{app}}t) + I_{\infty}$, where I_0 and I_{∞} represent the intensities at time zero and infinite time, respectively. Rate constants k_{app} were measured at various concentrations of the Egr-1 zinc-finger protein. For each kinetic rate constant, the measurement was replicated 8–10 times. MATLAB software (MathWorks) was used for non-linear least-squares fitting.

Competition assays for methylated and non-methylated DNA duplexes

Affinities of the methylated and unmethylated 28-bp DNA duplexes (the sequences shown in Figure 2A) for the Egr-1 zinc-finger protein and the MeCP2 MBD were measured using fluorescence-based competitive binding assays with an ISS PC1 spectrofluorometer. To measure the affinities of the Egr-1 zinc-finger protein, the 143-bp FAM-labeled DNA probe was used with one of the 28-bp DNA duplexes as competitor DNA. For measuring the affinities of the MeCP2 MBD, a FAM-labeled 19-bp DNA duplex, FAM-CTGGAACGGAATTCTTCTA in which the underlined CpG is methylated, was used as a probe. FAM fluorescence anisotropy was measured as a function of the concentra-

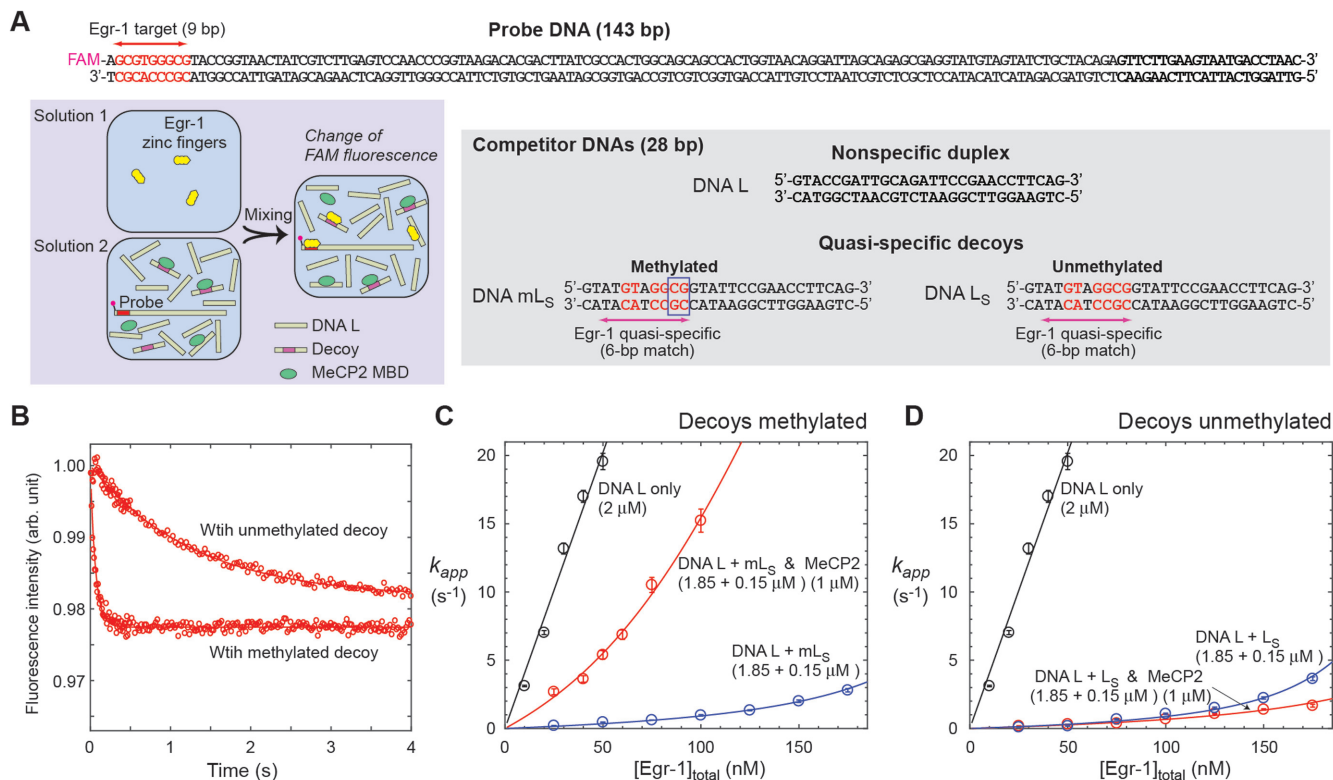


Figure 2. Stopped-flow kinetic assays for investigating how DNA methylation of decoys affects the target search kinetics for the Egr-1 zinc-finger protein in the presence and absence of the MeCP2 MBD. (A) Macromolecular components mixed in the stopped-flow experiment. The FAM-labeled probe DNA (143 bp), non-specific DNA L (28 bp) and quasi-specific decoy DNA L_S were used in our previous study (13). The decoy DNA m_LS is identical to DNA L_S in sequence, but is methylated at the quasi-specific site. (B) The time-course data of FAM fluorescence intensity recorded for the binding reactions using 2.5 nM probe DNA, 100 nM Egr-1 zinc-finger protein, 1850 nM non-specific DNA L, 150 nM unmethylated (DNA L_S) or methylated (DNA m_LS) decoy and 1,000 nM MeCP2 MBD. (C and D) Apparent pseudo-first-order kinetic rate constants for the target association of the Egr-1 zinc-finger protein in the presence of competitor DNAs. The Egr-1 quasi-specific sites in the decoys were methylated in panel C and unmethylated in panel D. The concentration of the probe DNA was 2.5 nM. The buffer was 10 mM Tris•HCl (pH 7.5), 150 mM KCl and 0.2 μM ZnCl₂ for all experiments. The presence of the MeCP2 MBD substantially accelerated the target search kinetics of the Egr-1 zinc-finger protein only when the decoys were methylated.

tion of the competitor DNA. Excitation and emission wavelengths were 490 and 521 nm, respectively. The concentrations of the fluorescent probe and the protein were 2.5 and 50 nM in the assays for the Egr-1 zinc-finger protein and 10 and 100 nM in those for the MeCP2 MBD, respectively. FAM fluorescence anisotropy was also measured in the absence of protein. Each anisotropy measurement was conducted for a 2 ml solution in a buffer containing 10 mM Tris•HCl (pH 7.5), 0.2 μM ZnCl₂ and 150 mM KCl. For each competitor DNA, the dissociation constants K_d were determined from the anisotropy data by non-linear least-squares fitting using MATLAB software, as described (40).

RESULTS

In our previous study, using a stopped-flow fluorescence method, we demonstrated that the presence of quasi-specific sequences, which are similar to the target sequence, impedes target DNA association of the Egr-1 zinc-finger protein (13). In the current study, using a similar system together with the MeCP2 MBD and methylated DNA, we examine the above-mentioned hypothetical model on the association of Egr-1 with its unmethylated targets in the presence of decoys (Figure 1). The components we used in these

stopped-flow experiments are shown in Figure 2A. To measure the Egr-1-target association kinetics, we collected time-course data for FAM fluorescence intensity upon mixing a solution of the Egr-1 zinc-finger protein with a solution of FAM-labeled probe DNA, non-specific DNA, decoy DNA and in some experiments, the MeCP2 MBD. Typical examples of the time-course fluorescence data are shown in Figure 2B. As observed in our previous studies (13,30,31,34), the association of the Egr-1 zinc-finger protein with the target on the probe caused a decrease in fluorescence intensity. As shown in the Supplementary Data (Supplementary Figure S1), the MeCP2 MBD does not interact with the Egr-1 zinc-finger protein directly. However, indirect interplay between these proteins through competition for methylated decoys enhances the apparent activity of the Egr-1 zinc-finger protein for unmethylated targets, as described below.

Unmethylated and methylated quasi-specific sites impede target search of Egr-1

We first examined whether methylation of the quasi-specific decoy DNA affects the target search by the Egr-1 zinc-finger protein in the absence of any other proteins. Figure 2C and D shows the results of kinetic measurements us-

ing methylated and unmethylated quasi-specific DNA duplexes, respectively. The data points shown in blue are the results from the experiments with 2.5 nM probe DNA, 1850 nM non-specific 28-bp DNA, 150 nM 28-bp decoy DNA containing a quasi-specific sequence and various concentrations of the Egr-1 zinc-finger protein. Note that the concentration of the decoy was set to be 60-fold greater than that of the target. To clarify the impact of the decoys, these figures also show the results from the experiments with 2000 nM non-specific 28-bp DNA and no decoy (shown in black). For both methylated and unmethylated quasi-specific sites, the presence of decoy DNA was found to substantially impede the target association of the Egr-1 zinc-finger protein. This effect occurred to a similar degree regardless of the decoy DNA's methylation state.

These results suggest that DNA methylation itself does not diminish the ability of a quasi-specific site to trap the Egr-1 zinc-finger protein. We verified this by measuring the relative affinities of the unmethylated and methylated quasi-specific DNA duplexes using competitive binding assays (Figure 3A). In this assay, anisotropy of FAM fluorescence from the 143-bp probe DNA was measured to determine the equilibrium population of the protein-bound target at various concentrations of the competitor DNA. Fluorescence anisotropy is related to the fluorophore's effective ro-

tational correlation time, which can change upon molecular association (41). The FAM fluorescence anisotropy was measured to be 0.077 for the free probe and 0.161 for the protein-bound probe. The anisotropy measurements were conducted at various concentrations of the 28-bp competitor DNA duplexes with (mL_S) and without (L_S) the CpG methylation of the quasi-specific site (magenta and blue data points, respectively, in Figure 3A). The anisotropy data showed that at relatively low concentrations of the competitor, the target on the fluorescent probe was predominantly in the protein-bound state. However, an increase in competitor concentration caused a decrease in the population of the protein-bound target due to transfer of the protein to the competitor. The concentration dependence data were very similar for methylated and unmethylated competitors, indicating that CpG methylation does not diminish the Egr-1 zinc-finger protein's affinity for quasi-specific sites.

MeCP2 accelerates target search of Egr-1 when quasi-specific DNA is methylated

To test our hypothetical model (Figure 1), we examined whether the MeCP2 MBD would influence the association of the Egr-1 zinc-finger protein with unmethylated target when decoys were methylated. For this purpose, we conducted the same stopped-flow kinetic assays in the presence of the MeCP2 MBD in solution with the 143-bp DNA probe and methylated or unmethylated competitor DNA. Data points shown in red in Figure 2C and D show apparent pseudo-first-order rate constants measured for Egr-1-target association in the presence of 1 μ M MeCP2 MBD.

When the MeCP2 MBD was present and the decoy was methylated in the system, the Egr-1 zinc-finger protein exhibited substantially faster association with the unmethylated target on the probe DNA (Figure 2C). At 1 μ M MeCP2 MBD, there was a 12-fold increase in the rate constant for target association of the Egr-1 zinc-finger protein, compared to that in the absence of the MeCP2 MBD. As shown in Figure 4A, we found that this acceleration was dependent on the concentration of the MeCP2 MBD but became saturated at concentrations far higher than the decoy concentration. These results strongly suggest that the MeCP2 MBD accelerates the association of the Egr-1 zinc-finger protein with unmethylated target by blocking the methylated decoys as predicted by our hypothetical model (Figure 1).

In contrast, the acceleration of Egr-1's target association by the MeCP2 MBD was not observed when the decoy DNA was unmethylated (Figure 2D). In fact, the presence of MeCP2 MBD slightly slowed down the target association of the Egr-1 zinc-finger protein in the experiments with the unmethylated decoy. The slower target search of the Egr-1 zinc-finger protein in this situation may be due to additional binding of the MeCP2 MBD to the probe DNA. To illustrate the impact of decoy DNA methylation on the Egr-1 zinc-finger protein, Figure 4B shows fold increases in target association rate constants upon the decoy methylation. In the presence of 1 μ M MeCP2 MBD, the decoy methylation resulted in 14- to 28-fold acceleration of the target association of the Egr-1 zinc-finger protein.

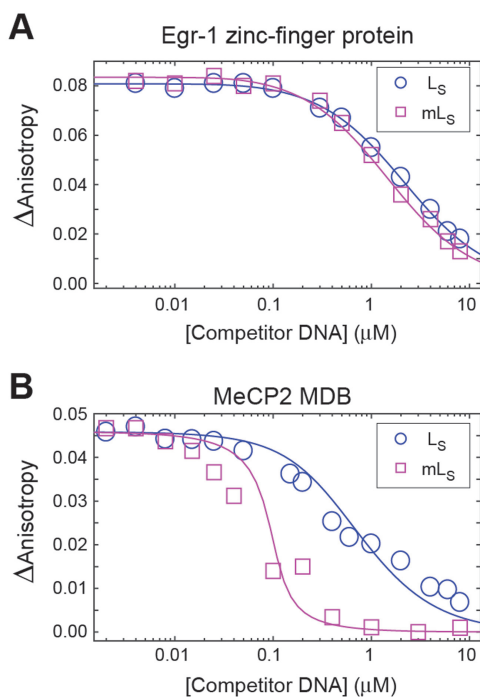


Figure 3. Equilibrium competitive binding assays to investigate the impact of CpG methylation on the affinity of the decoy DNA for the Egr-1 zinc-finger protein (A) and the MeCP2 MBD (B). In panel A, FAM fluorescence anisotropy was measured for solutions containing 2.5 nM FAM-labeled 143-bp DNA containing the Egr-1 target (see Figure 2A), 50 nM Egr-1 zinc-finger protein and various concentrations of DNA L_S or mL_S . In panel B, fluorescence anisotropy for an FAM-labeled 19-bp DNA containing a methyl-CpG was measured for solutions containing 10 nM probe, 100 nM MeCP2 MBD and various concentrations of DNA L_S or mL_S . The vertical axes show the difference between fluorescence anisotropy values for each sample and free probe DNA solutions.

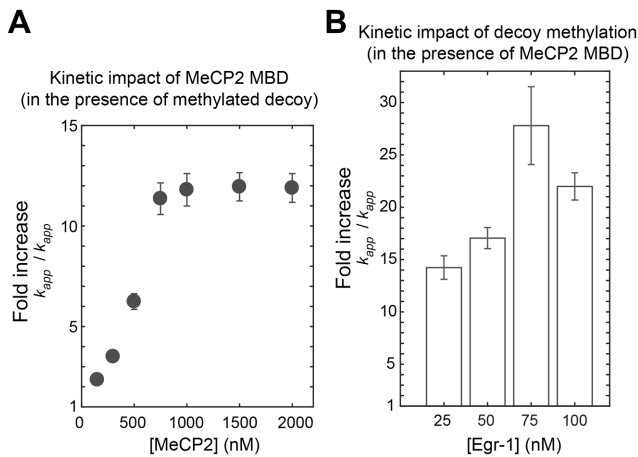


Figure 4. Acceleration of target association of the Egr-1 zinc-finger protein by the MeCP2 MBD and decoy methylation. (A) Dependence on the concentration of the MeCP2 MBD when the Egr-1 decoys were methylated. (B) Impact of the decoy DNA methylation when 1 μ M MeCP2 MBD was present. Fold increases in the apparent pseudo-first-order rate constant k_{app} for the target association are plotted in each panel.

Although unmethylated and methylated decoys could almost equally impede the target search of the Egr-1 zinc-finger protein, the MeCP2 MBD markedly accelerated Egr-1 target association only when the decoys were methylated. This is likely due to specific binding of the MeCP2 MBD to methyl-CpGs at Egr-1 quasi-specific sites. We verified this by conducting equilibrium competitive binding assays and comparing the relative affinities of the MeCP2 MBD for these 28-bp quasi-specific DNA duplexes (Figure 3B). The affinity of the MeCP2 MBD for the methylated DNA (the dissociation constant $K_d = 2.3$ nM) was 54-fold higher than that for the unmethylated DNA ($K_d = 124$ nM), which is consistent with a previous study (42).

No acceleration by the MeCP2 MBD when the target is also methylated

If blocking methylated decoys by the MeCP2 MBD directly causes acceleration of Egr-1 target association as shown in Figure 2, it is expected that such acceleration will not occur when the target sequence on the probe DNA is also methylated. This is because the MeCP2 MBD should block the methylated target sequence on the probe DNA as well. We examined whether this is indeed the case by conducting the same experiment using another probe DNA (33 bp) in which the target sequence is methylated (Figure 5). Although CpG methylation of the same 143-bp DNA would be preferable, we used this DNA because site-specific incorporation of methyl-CpG was difficult to achieve with the FAM-labeled 143-bp probe DNA. Using the methylated 33-bp probe DNA together with the methylated decoy DNA mL_S , we measured kinetic rate constants for target association of the Egr-1 zinc-finger protein in the presence and absence of the MeCP2 MBD. As expected, the MeCP2 MBD did not accelerate Egr-1 target association in this case (compare red and blue data points in Figure 5) but rather slowed it. In contrast, when the same 33-bp probe was unmethylated, the MeCP2 MBD significantly accelerated Egr-

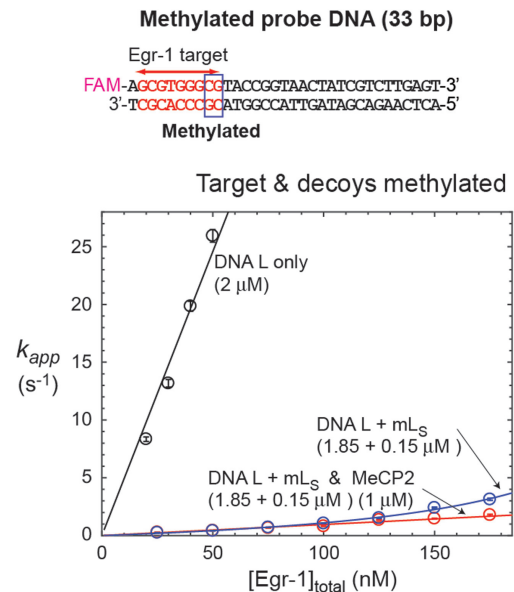


Figure 5. The MeCP2 MBD does not accelerate target association of the Egr-1 zinc-finger protein when the target is also methylated. The graph shows the apparent kinetic rate constants measured for the target association of the Egr-1 zinc-finger protein using the same conditions as those in Figure 2C except that the Egr-1 target in the 33-bp probe DNA was methylated.

1 target association (Supplementary Figure S2), as seen for the 143-bp probe DNA containing an unmethylated target. These results suggest that when Egr-1 targets are methylated, the MeCP2 MBD binds to them and thereby blocks Egr-1's association with the methylated targets.

DISCUSSION

In vertebrates, DNA methylation is an important epigenetic mechanism that is associated with various biological processes such as development, genomic imprinting, and X-chromosome inactivation (43). Abnormalities in DNA methylation status are associated with cancer and many other diseases (22,44). While the regulatory mechanisms of DNA methylation and demethylation remain under active investigation (27,45), it has been well established that DNA methylation within CGIs is directly associated with gene silencing (21,43,46). Based on our biochemical data, we argue that DNA methylation outside CGIs may indirectly enhance activities of transcriptional activators at CGIs through competitive interplay with MBD proteins.

DNA methylation as potential facilitator of transcription factor-CGI interactions

Our current biochemical data show that DNA methylation of decoys can greatly facilitate association of the Egr-1 zinc-finger protein with its unmethylated target through competitive interplay with the MeCP2 MBD. In this manner, MBD proteins may indirectly guide Egr-1 to unmethylated targets within CGIs by blocking methylated decoys outside CGIs. We speculate that this model may be applicable for some other transcriptional activators that bind to CGI promoters. In fact, there are many transcriptional activators

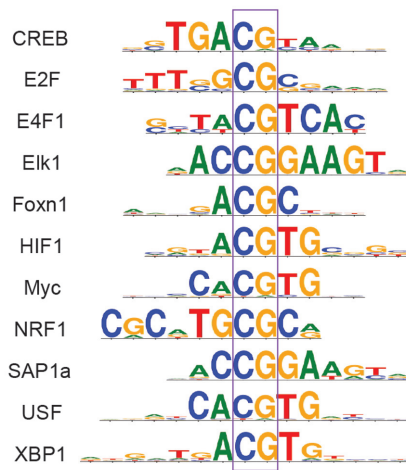


Figure 6. Some examples of human transcriptional activators that recognize CpG-containing sequences. Height of each letter represents likelihood of base type at each position within the recognition sequences. Adapted from the MotifMap database (<http://motifmap.ics.uci.edu/>).

that recognize CpG-containing sequences (some examples are shown in Figure 6) (27). The binding of these transcriptional activators to their unmethylated targets in CGI promoters may be facilitated by highly expressed MBD proteins that block numerous methylated decoys outside CGIs, as shown in Figure 1.

Potential mechanism for transactivation by MeCP2

Our hypothetical model (Figure 1) provides insight into the transcriptional activation by MeCP2. With an MBD and a transcriptional repression domain, this protein was originally regarded as a methyl CpG-dependent transcription repressor (47). However, despite the lack of any known transcriptional activation domains, MeCP2 activates more than 2,000 genes in neurons under the control of CGI promoters where CpG methylation is rare (48). In fact, MeCP2 activates ~6 times more genes than it represses. Chahrour *et al.* proposed that protein–protein interactions with the transcription factor CREB1 could be responsible for transactivation by MeCP2 (48). While this may explain the activation of some genes, it remains unclear why MeCP2 can activate so many genes. Our model (Figure 1) may provide an explanation. Because CpGs are highly methylated outside CGIs, highly expressed MeCP2 can effectively block non-functional high-affinity sites for transcriptional activators that recognize CpG-containing sequences. This will indirectly guide transcriptional activators to unmethylated targets within CpGs and thereby activate downstream genes. Our model does not require any direct protein–protein interactions with particular transcriptional activators and could explain why MeCP2 activates so many genes. This model is directly applicable to CREB1 as well because the CREB1 recognition sequence, TGACGTCa, contains a CpG. Our model is also consistent with the observation that the promoters of genes activated by MeCP2 are not methylated (48).

Relevance to ‘DNA methylation paradox’

Our model could also explain the so-called ‘DNA methylation paradox’ (46,49), which concerns DNA methylation yielding opposite effects depending on genomic contexts. In vertebrates, although DNA methylation within CGI promoters is typically associated with gene silencing, DNA methylation in the gene bodies is positively correlated with gene expression level (46,49). In our model (Figure 1), a high level of DNA methylation outside CGIs should increase CGI-binding of transcriptional activators that recognize CpG-containing sequences. Because the MBD proteins block methylated decoys outside CGIs, the transcriptional activators can occupy their functional sites within CGI promoters more easily, thereby increasing gene expression level. On the other hand, CpG methylation within CGIs should cause binding of MBD proteins to CGI promoters and reduce the expression of downstream genes. This repression could occur through excluding transcriptional activators from CGI promoters and recruiting co-repressor proteins via the transcriptional repressor domain of MBD proteins.

However, it should be noted that our model is not applicable for some transcriptional activators such as Klf4 and CEBPB, which preferentially bind to CpG-methylated DNA *in vitro* (27). Genome-wide ChIP-seq and bisulfite sequencing studies for Klf4 and CEBPB show that ~20–40% of the genomic regions bound by these proteins are methylated *in vivo* (50,51). It remains to be addressed why these proteins can occupy the CpG-methylated regions despite the high abundance of MBD proteins in the cells. Protein–protein interactions with other proteins might increase occupancies of Klf4 and CEBPB in these regions.

New perspective on target DNA search

Our current work provides a new perspective on the target DNA search mechanisms for transcription factors. Over the past four decades, the mechanisms allowing transcription factors to rapidly locate their specific targets in the genome have been the subject of considerable interest in biophysics and biochemistry (52–55). While recent studies using NMR and single-molecule techniques revealed great details on how proteins scan DNA efficiently (56–59), studies that focus on factors that impede the search process have been rare (60). Numerous non-functional high-affinity sites in the genome could serve as decoys that trap transcription factors and affect their functions (3,13,61–65). Our study shows that the target search process for transcription factors can be greatly accelerated when other proteins block such decoys. This mechanism does not require any direct protein–protein interactions, although the current paradigm for synergy between transcription factors typically assumes their direct interactions (66,67). In our current case, the MeCP2 MBD facilitates the association of the Egr-1 zinc-finger protein with the unmethylated target by blocking methylated decoys. In principle, other proteins could also provide the same acceleration mechanism, as long as they selectively block decoys but not targets. For example, highly expressed transcription factors with similar but different sequence specificity may enhance the function of other transcription factors by selectively blocking their decoys. Natural decoys

and their DNA methylation may play key roles in regulation of transcription factors.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Drs Alexandre Esadze, Abhijnan Chattopadhyay and Levani Zandarashvili for useful discussion and technical assistance.

FUNDING

National Institutes of Health [R01-GM107590 to J.I.]; National Science Foundation [CHE-1608866 to J.I.]. Funding for open access charge: National Institutes of Health [R01-GM107590 to J.I.].

Conflict of interest statement. None declared.

REFERENCES

- Alberts, B., Johnson, A., Lewis, J., Morgan, D., Martin, R., Roberts, K. and Walter, P. (2014) Control of gene expression. *Molecular Biology of the Cell*, 6th edn. Garland Science, NY, pp. 369–438.
- Slattery, M., Zhou, T., Yang, L., Dantas Machado, A.C., Gordan, R. and Rohs, R. (2014) Absence of a simple code: how transcription factors read the genome. *Trends Biochem. Sci.*, **39**, 381–399.
- Kemme, C.A., Nguyen, D., Chattopadhyay, A. and Iwahara, J. (2016) Regulation of transcription factors via natural decoys in genomic DNA. *Transcription*, **7**, 115–120.
- Todeschini, A.L., Georges, A. and Veitia, R.A. (2014) Transcription factors: specific DNA binding and specific gene regulation. *Trends Genet.*, **30**, 211–219.
- Wunderlich, Z. and Mirny, L.A. (2009) Different gene regulation strategies revealed by analysis of binding motifs. *Trends Genet.*, **25**, 434–440.
- Pavletich, N.P. and Pabo, C.O. (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science*, **252**, 809–817.
- Bozon, B., Davis, S. and Laroche, S. (2003) A requirement for the immediate early gene zif268 in reconsolidation of recognition memory after retrieval. *Neuron*, **40**, 695–701.
- Lee, J.L., Everitt, B.J. and Thomas, K.L. (2004) Independent cellular processes for hippocampal memory consolidation and reconsolidation. *Science*, **304**, 839–843.
- Khachigian, L.M., Lindner, V., Williams, A.J. and Collins, T. (1996) Egr-1-induced endothelial gene expression: a common theme in vascular injury. *Science*, **271**, 1427–1431.
- Yan, S.F., Fujita, T., Lu, J., Okada, K., Shan Zou, Y., Mackman, N., Pinsky, D.J. and Stern, D.M. (2000) Egr-1, a master switch coordinating upregulation of divergent gene families underlying ischemic stress. *Nat. Med.*, **6**, 1355–1361.
- Baumgartel, K., Tweedie-Cullen, R.Y., Grossmann, J., Gehrig, P., Livingstone-Zatchej, M. and Mansuy, I.M. (2009) Changes in the proteome after neuronal zif268 overexpression. *J. Proteome Res.*, **8**, 3298–3316.
- Fu, M., Zhu, X., Zhang, J., Liang, J., Lin, Y., Zhao, L., Ehrenguber, M.U. and Chen, Y.E. (2003) Egr-1 target genes in human endothelial cells identified by microarray analysis. *Gene*, **315**, 33–41.
- Kemme, C.A., Esadze, A. and Iwahara, J. (2015) Influence of quasi-specific sites on kinetics of target DNA search by a sequence-specific DNA-binding protein. *Biochemistry*, **54**, 6684–6691.
- Badis, G., Berger, M.F., Philippakis, A.A., Talukder, S., Gehrke, A.R., Jaeger, S.A., Chan, E.T., Metzler, G., Vedenko, A., Chen, X. *et al.* (2009) Diversity and complexity in DNA recognition by transcription factors. *Science*, **324**, 1720–1723.
- Chattopadhyay, A., Zandarashvili, L., Luu, R.H. and Iwahara, J. (2016) Thermodynamic additivity for impacts of base-pair substitutions on association of the Egr-1 zinc-finger protein with DNA. *Biochemistry*, **55**, 6467–6474.
- Hamilton, T.B., Borel, F. and Romaniuk, P.J. (1998) Comparison of the DNA binding characteristics of the related zinc finger proteins WT1 and EGR1. *Biochemistry*, **37**, 2051–2058.
- The ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.
- Koldamova, R., Schug, J., Lefterova, M., Cronican, A.A., Fitz, N.F., Davenport, F.A., Carter, A., Castranio, E.L. and Lefterov, I. (2014) Genome-wide approaches reveal EGR1-controlled regulatory networks associated with neurodegeneration. *Neurobiol. Dis.*, **63**, 107–114.
- Kubosaki, A., Tomaru, Y., Tagami, M., Arner, E., Miura, H., Suzuki, T., Suzuki, M., Suzuki, H. and Hayashizaki, Y. (2009) Genome-wide investigation of in vivo EGR-1 binding sites in monocytic differentiation. *Genome Biol.*, **10**, R41.
- Deaton, A.M. and Bird, A. (2011) CpG islands and the regulation of transcription. *Genes Dev.*, **25**, 1010–1022.
- Klose, R.J. and Bird, A.P. (2006) Genomic DNA methylation: the mark and its mediators. *Trends Biochem. Sci.*, **31**, 89–97.
- Bergman, Y. and Cedar, H. (2013) DNA methylation dynamics in health and disease. *Nat. Struct. Mol. Biol.*, **20**, 274–281.
- Dantas Machado, A.C., Zhou, T., Rao, S., Goel, P., Rastogi, C., Lazarovici, A., Bussemaker, H.J. and Rohs, R. (2015) Evolving insights on how cytosine methylation affects protein-DNA binding. *Brief. Funct. Genomics*, **14**, 61–73.
- Zandarashvili, L., White, M.A., Esadze, A. and Iwahara, J. (2015) Structural impact of complete CpG methylation within target DNA on specific complex formation of the inducible transcription factor Egr-1. *FEBS Lett.*, **589**, 1748–1753.
- Elrod-Erickson, M., Rould, M.A., Nekudova, L. and Pabo, C.O. (1996) Zif268 protein-DNA complex refined at 1.6 Å: a model system for understanding zinc finger-DNA interactions. *Structure*, **4**, 1171–1180.
- Hashimoto, H., Olanrewaju, Y.O., Zheng, Y., Wilson, G.G., Zhang, X. and Cheng, X. (2014) Wilms tumor protein recognizes 5-carboxylcytosine within a specific DNA sequence. *Genes Dev.*, **28**, 2304–2313.
- Blattler, A. and Farnham, P.J. (2013) Cross-talk between site-specific transcription factors and DNA methylation states. *J. Biol. Chem.*, **288**, 34287–34294.
- Schubeler, D. (2015) Function and information content of DNA methylation. *Nature*, **517**, 321–326.
- Skene, P.J., Illingworth, R.S., Webb, S., Kerr, A.R., James, K.D., Turner, D.J., Andrews, R. and Bird, A.P. (2010) Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. *Mol. Cell*, **37**, 457–468.
- Esadze, A. and Iwahara, J. (2014) Stopped-flow fluorescence kinetic study of protein sliding and intersegment transfer in the target DNA search process. *J. Mol. Biol.*, **426**, 230–244.
- Esadze, A., Kemme, C.A., Kolomeisky, A.B. and Iwahara, J. (2014) Positive and negative impacts of nonspecific sites during target location by a sequence-specific DNA-binding protein: origin of the optimal search at physiological ionic strength. *Nucleic Acids Res.*, **42**, 7039–7046.
- Takayama, Y., Sahu, D. and Iwahara, J. (2010) NMR studies of translocation of the Zif268 protein between its target DNA sites. *Biochemistry*, **49**, 7998–8005.
- Zandarashvili, L., Esadze, A., Kemme, C.A., Chattopadhyay, A., Nguyen, D. and Iwahara, J. (2016) Residence times of molecular complexes in solution from NMR data of intermolecular hydrogen-bond scalar coupling. *J. Phys. Chem. Lett.*, **7**, 820–824.
- Zandarashvili, L., Esadze, A., Vuzman, D., Kemme, C.A., Levy, Y. and Iwahara, J. (2015) Balancing between affinity and speed in target DNA search by zinc-finger proteins via modulation of dynamic conformational ensemble. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E5142–E5149.
- Zandarashvili, L., Vuzman, D., Esadze, A., Takayama, Y., Sahu, D., Levy, Y. and Iwahara, J. (2012) Asymmetrical roles of zinc fingers in dynamic DNA-scanning process by the inducible transcription factor Egr-1. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, E1724–1732.

36. Chattopadhyay, A., Esadze, A., Roy, S. and Iwahara, J. (2016) NMR scalar couplings across intermolecular hydrogen bonds between zinc-finger histidine side chains and DNA phosphate groups. *J. Phys. Chem. B*, **120**, 10679–10685.
37. Chen, C., Esadze, A., Zandarashvili, L., Nguyen, D., Pettitt, B.M. and Iwahara, J. (2015) Dynamic equilibria of short-range electrostatic interactions at molecular interfaces of protein-DNA complexes. *J. Phys. Chem. Lett.*, **6**, 2733–2737.
38. Esadze, A., Chen, C., Zandarashvili, L., Roy, S., Pettitt, B.M. and Iwahara, J. (2016) Changes in conformational dynamics of basic side chains upon protein-DNA association. *Nucleic Acids Res.*, **44**, 6961–6970.
39. Pollard, T.D. and De La Cruz, E.M. (2013) Take advantage of time in your experiments: a guide to simple, informative kinetics assays. *Mol. Biol. Cell*, **24**, 1103–1110.
40. Zandarashvili, L., Nguyen, D., Anderson, K.M., White, M.A., Gorenstein, D.G. and Iwahara, J. (2015) Entropic enhancement of protein-DNA affinity by oxygen-to-sulfur substitution in DNA phosphate. *Biophys. J.*, **109**, 1026–1037.
41. Lakowicz, J.R. (2006) *Principles of Fluorescence Spectroscopy*. 3rd edn. Springer, NY.
42. Lao, V.V., Darwanto, A. and Sowers, L.C. (2010) Impact of base analogues within a CpG dinucleotide on the binding of DNA by the methyl-binding domain of MeCP2 and methylation by DNMT1. *Biochemistry*, **49**, 10228–10236.
43. Bird, A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev.*, **16**, 6–21.
44. Robertson, K.D. (2005) DNA methylation and human disease. *Nat. Rev. Genet.*, **6**, 597–610.
45. Hashimoto, H., Zhang, X., Vertino, P.M. and Cheng, X. (2015) The Mechanisms of Generation, Recognition, and Erasure of DNA 5-Methylcytosine and Thymine Oxidations. *J. Biol. Chem.*, **290**, 20723–20733.
46. Jones, P.A. (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.*, **13**, 484–492.
47. Nan, X., Campoy, F.J. and Bird, A. (1997) MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. *Cell*, **88**, 471–481.
48. Chahrour, M., Jung, S.Y., Shaw, C., Zhou, X., Wong, S.T., Qin, J. and Zoghbi, H.Y. (2008) MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science*, **320**, 1224–1229.
49. Jones, P.A. (1999) The DNA methylation paradox. *Trends Genet.*, **15**, 34–37.
50. Hu, S., Wan, J., Su, Y., Song, Q., Zeng, Y., Nguyen, H.N., Shin, J., Cox, E., Rho, H.S., Woodard, C. *et al.* (2013) DNA methylation presents distinct binding sites for human transcription factors. *Elife*, **2**, e00726.
51. Mann, I.K., Chatterjee, R., Zhao, J., He, X., Weirauch, M.T., Hughes, T.R. and Vinson, C. (2013) CG methylated microarrays identify a novel methylated sequence bound by the CEBPB/ATF4 heterodimer that is active in vivo. *Genome Res.*, **23**, 988–997.
52. Halford, S.E. and Marko, J.F. (2004) How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.*, **32**, 3040–3052.
53. Kolomeisky, A.B. (2011) Physics of protein-DNA interactions: mechanisms of facilitated target search. *Phys. Chem. Chem. Phys.*, **13**, 2088–2095.
54. Mirny, L., Slutsky, M., Wunderlich, Z., Tafvizi, A., Leith, J. and Kosmrlj, A. (2009) How a protein searches for its site on DNA: the mechanism of facilitated diffusion. *J. Phys. A Math. Theor.*, **42**, 434013.
55. von Hippel, P.H. and Berg, O.G. (1989) Facilitated target location in biological systems. *J. Biol. Chem.*, **264**, 675–678.
56. Clore, G.M. (2011) Exploring translocation of proteins on DNA by NMR. *J. Biomol. NMR*, **51**, 209–219.
57. Clore, G.M. and Iwahara, J. (2009) Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chem. Rev.*, **109**, 4108–4139.
58. Gorman, J. and Greene, E.C. (2008) Visualizing one-dimensional diffusion of proteins along DNA. *Nat. Struct. Mol. Biol.*, **15**, 768–774.
59. Tafvizi, A., Mirny, L.A. and van Oijen, A.M. (2011) Dancing on DNA: kinetic aspects of search processes on DNA. *Chemphyschem*, **12**, 1481–1489.
60. Halford, S.E. (2009) An end to 40 years of mistakes in DNA-protein association kinetics? *Biochem. Soc. Trans.*, **37**, 343–348.
61. Burger, A., Walczak, A.M. and Wolynes, P.G. (2010) Abduction and asylum in the lives of transcription factors. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 4016–4021.
62. Burger, A., Walczak, A.M. and Wolynes, P.G. (2012) Influence of decoys on the noise and dynamics of gene expression. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, **86**, 041920.
63. Lee, T.H. and Maheshri, N. (2012) A regulatory role for repeated decoy transcription factor binding sites in target gene expression. *Mol. Syst. Biol.*, **8**, 576.
64. Liu, X., Wu, B., Szary, J., Kofoed, E.M. and Schaufele, F. (2007) Functional sequestration of transcription factor activity by repetitive DNA. *J. Biol. Chem.*, **282**, 20868–20876.
65. Wang, Z., Potoyan, D.A. and Wolynes, P.G. (2016) Molecular stripping, targets and decoys as modulators of oscillations in the NF-kappaB/IkappaBalpha/DNA genetic network. *J. R. Soc. Interface*, **13**, 20160606.
66. Smith, N.C. and Matthews, J.M. (2016) Mechanisms of DNA-binding specificity and functional gene regulation by transcription factors. *Curr. Opin. Struct. Biol.*, **38**, 68–74.
67. Wolberger, C. (1999) Multiprotein-DNA complexes in transcriptional regulation. *Annu. Rev. Biophys. Biomol. Struct.*, **28**, 29–56.