



## PAPER

Expression regulation by a methyl-CpG binding domain in an *E. coli* based, cell-free TX-TL systemRECEIVED  
13 September 2016REVISED  
9 January 2017ACCEPTED FOR PUBLICATION  
31 January 2017PUBLISHED  
1 March 2017M Schenkelberger<sup>1</sup>, S Shanak<sup>2,4</sup>, M Finkler<sup>1</sup>, E G Worst<sup>1</sup>, V Noireaux<sup>3</sup>, V Helms<sup>2</sup> and A Ott<sup>1</sup><sup>1</sup> Department of Experimental Physics, Saarland University, Saarbrücken, 66041, Germany<sup>2</sup> Center for Bioinformatics, Saarland University, Saarbrücken, 66041, Germany<sup>3</sup> School of Physics and Astronomy, University of Minnesota, Minneapolis, MN 55455, United States of America<sup>4</sup> Arab American University-Jenin, Jenin, PO Box 240, PalestineE-mail: [albrecht.ott@physik.uni-saarland.de](mailto:albrecht.ott@physik.uni-saarland.de)**Keywords:** *Escherichia coli* cell extract, cell-free expression, MeCP2, brain-derived neurotrophic factor (BDNF), methyl-CpG binding domain (MBD), epigenetic regulation, molecular dynamics simulationsSupplementary material for this article is available [online](#)**Abstract**

Cytosine methylation plays an important role in the epigenetic regulation of eukaryotic gene expression. The methyl-CpG binding domain (MBD) is common to a family of eukaryotic transcriptional regulators. How MBD, a stretch of about 80 amino acids, recognizes CpGs in a methylation dependent manner, and as a function of sequence, is only partly understood. Here we show, using an *Escherichia coli* cell-free expression system, that MBD from the human transcriptional regulator MeCP2 performs as a specific, methylation-dependent repressor in conjunction with the BDNF (brain-derived neurotrophic factor) promoter sequence. Mutation of either base flanking the central CpG pair changes the expression level of the target gene. However, the relative degree of repression as a function of MBD concentration remains unaltered. Molecular dynamics simulations that address the DNA B fiber ratio and the handedness reveal cooperative transitions in the promoter DNA upon MBD binding that correlate well with our experimental observations. We suggest that not only steric hindrance, but also conformational changes of the BDNF promoter as a result of MBD binding are required for MBD to act as a specific inhibitory element. Our work demonstrates that the prokaryotic transcription machinery can reproduce features of epigenetic mammalian transcriptional regulatory elements.

**Introduction**

Methylation of DNA plays an essential role in the epigenetic regulation of mammalian gene expression. The enzymatic addition of a methyl group to the DNA base cytosine takes place at CpG positions. About 40% of mammalian genes contain CpG islands (CGIs), genomic regions with a higher frequency of CpG dinucleotides than elsewhere, in their promoters and exonic regions. The CGI promoters are mostly non-methylated [1]. Cytosine methylation is predominant in eukaryotic genomes while prokaryotes mainly exhibit adenosine methylation. A recent study [2] showed that cytosine methylation can be associated with stationary phase prokaryotic gene expression and has a weak influence on the exponential growth phase.

The human protein MeCP2 [3] belongs to a family of proteins that contain the conserved methyl-CpG

binding domain (MBD). MBD specifically binds to methylated CpG sites, which enables MBD proteins to modulate gene expression as a function of their concentration levels. In binding assays, hyper-methylated promoters become enriched with MBD protein with promoter specific distributions [4]. MeCP2 consists of six domains. Upon its discovery, MeCP2 was understood as a transcriptional repressor, its action based on two domains, MBD and TRD (transcriptional repression domain) [5]. However, it is now well established that MeCP2 has a multifunctional role beyond suppression or activation of transcription [6–8]. MeCP2 was studied by means of a purified *in vitro* transcription system [9]. MeCP2 exhibits high affinity against the promoter III of the brain-derived neurotrophic factor (BDNF), which contains a single central CpG pair [10–12]. Moreover, MeCP2 transcription control is of interest in connection with many neuropsychiatric disorders, among them the Rett syndrome [13].

Physiological or higher salt concentrations are necessary for the MBD to discriminate methylation with high specificity [14]. MBD interacts with the DNA methyl group through strong electrostatic interactions with a tendency to form dimers. Binding assays have shown a clear tendency to bind to one or more symmetrically methylated CpGs, while the non-specific affinity to DNA is negligible [15]. Zou and colleagues [16] conducted conventional molecular dynamics (MD) simulations as well as alchemical free energy perturbation calculations to suggest that the increased hydrophobic interaction in the methylated MBD protein/DNA complexes strengthens binding. While hydrophobic attraction has been shown to play a major role in methylation recognition of the methylation-binding domain of MBD1 [17], crystallographic investigations of the MBD from MeCP2 bound to the methylated BDNF promoter suggest that it makes contact and recognizes methylation through the hydration pattern of the major groove [11]. The recognition process of methylated DNA by MBD involves only little direct interaction at the CpG dinucleotide while interactions with flanking bases remain limited to the DNA backbone including hydrating water molecules [11]. An extensive binding study on MeCP2 revealed the requirement of A/T sequences adjacent to the CpG motif for MBD recognition [12]. This property was traced to amino acids 78–90 of MBD. In contrast, structural studies suggest that this is caused either by an Asx-ST motif at the C terminus, or an AT hook of MeCP2 that is not part of MBD [11]. The increased propeller twist of the AT run could also play a role. However, the AT-run as a prerequisite for MBD binding could not be firmly established on structural grounds [18].

*Escherichia coli* cell-free expression systems are useful for the synthesis of human proteins, antibiotic peptides, and the incorporation of non-canonical amino acids [19, 20], among many other applications. The molecular machinery of the cytoplasmic extract performs transcription and translation in a single test tube. The molecular composition of a cell-free reaction, DNA template stoichiometry as well as transcriptional elements, can be more tightly controlled and more easily varied than *in vivo*. In particular, these extracts give the experimentalist the opportunity to combine functional biomolecular elements in ways difficult to achieve otherwise. The expression of fluorescent protein enables real time, quantitative measurement of expression dynamics.

Noireaux and colleagues developed a new system that utilizes the endogenous *E. coli* RNA polymerase and sigma factor 70 [21, 22]. This expression system produces recombinant proteins in the micromolar range within a few hours. Compared to the common bacteriophage based transcription systems, such as T7, it presents the advantage of preserving the bacterial transcription toolbox.

In this work, we investigate the role of the methyl-CpG binding domain of MeCP2 as a possible

transcriptional modulator for prokaryotic gene expression in conjunction with the BDNF promoter. To this aim, we use an *E. coli* cell-free expression system with the endogenous *E. coli* RNA polymerase and sigma factor 70 [21, 22]. We combine our experiments with MD simulations that study how MBD binding and methylation affect the structure of the wild-type DNA compared to a mutant.

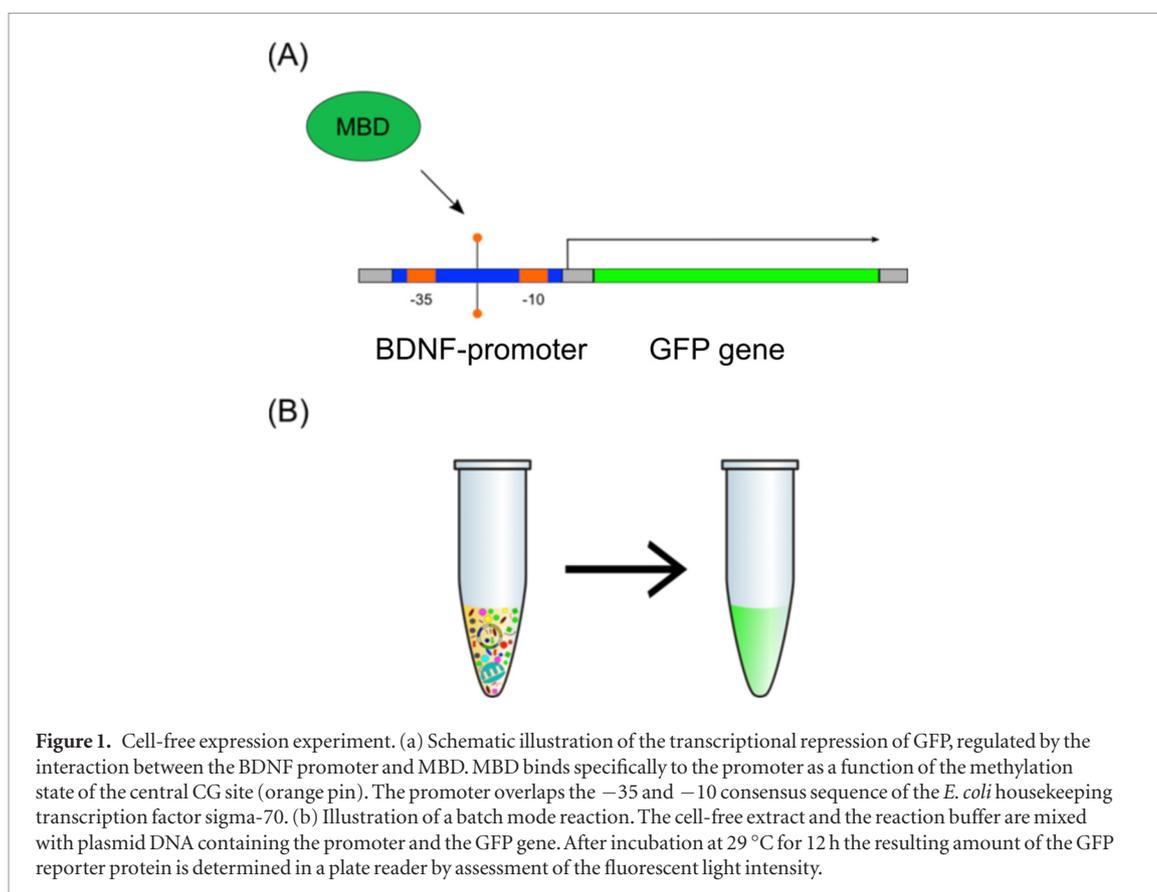
## Methods

Cell-free expression is illustrated in figure 1. The BDNF promoter (see table 1) is cloned into a plasmid in such a way that it overlaps the –35 and –10 consensus sequences of the *E. coli* promoter, specific to the housekeeping transcription factor sigma-70 present in the cell-free reaction. As a function of the methylation of the central CpG site, the BDNF promoter regulates the transcription of a gene that codes for green fluorescent protein (GFP) as a reporter.

### Cell-free extract and plasmid preparation

The crude cell-free extract is prepared according to an established protocol [21, 22].

All plasmids stem from pBEST-OR2-OR1-Pr-UTR1-deGFP-T500 (Addgene #40019) [22]. Its ColE1 origin of replication is replaced by a p15A. To avoid toxicity due to an over-expression of the recombinant protein during plasmid amplification, the strong lambda phage promoter Pr is flanked by the operons OR1 and OR2 [22]. This enables the temperature sensitive lambda repressor Cl857 to stabilize the plasmid during cell growth and avoid toxicity (*E. coli* strain KL740, Yale CGSC#:4382). The plasmid design involves cloning of the genes and their regulatory parts into the above described plasmid according to standard procedures of molecular biology. The wild type (wt) and several mutant versions of the BDNF promoter are cloned in between restriction sites NheI and SphI. The sequences overlap the –35 and –10 consensus sequence of the promoter specific to the *E. coli* housekeeping transcription factor sigma-70 present in the extract. Here, GFP refers to the enhanced green fluorescent protein (GenBank: CAD97424.1), truncated and modified in the N- and C-terminal [23]. MBD refers to the methylation binding domain of MeCP2 (amino acid 78-162, GenBank: 4202), UTR1 to the untranslated region containing the T7 g10 leader sequence for highly efficient translation initiation (GeneBank: M35614.1 [24];), T500 to the transcription terminator [25], BDNF to the promoter III of the mouse-brain-derived neurotrophic factor in the wild type (wt) and mutated (M2) version, and OR2-OR1-Pr to the lambda repressor Cro promoter (GenBank: J02459.1). Plasmid concentrations are determined using either QuantiFluor (Promega, USA), or Nanodrop 2000c UV-Vis Spectrometer (Thermo Fisher Scientific GmbH, Dreieich, Germany). Plasmid sequences are determined by a sequencing service (Microsynth AG, Switzerland or



**Figure 1.** Cell-free expression experiment. (a) Schematic illustration of the transcriptional repression of GFP, regulated by the interaction between the BDNF promoter and MBD. MBD binds specifically to the promoter as a function of the methylation state of the central CG site (orange pin). The promoter overlaps the  $-35$  and  $-10$  consensus sequence of the *E. coli* housekeeping transcription factor sigma-70. (b) Illustration of a batch mode reaction. The cell-free extract and the reaction buffer are mixed with plasmid DNA containing the promoter and the GFP gene. After incubation at  $29\text{ }^{\circ}\text{C}$  for 12 h the resulting amount of the GFP reporter protein is determined in a plate reader by assessment of the fluorescent light intensity.

**Table 1.** Promoter sequences of the studied BDNF promoters, and the plasmid coding for MBD. All promoters contain a central CpG motif (bold), which can either be methylated or non-methylated. In the case of the four mutations of the BDNF promoter (M1–M4), the modified bases are marked as minuscule.

Promoter	Sequence (5'-3')
BDNF-wt	TCTG-GAA- <b>CGG</b> -AAT-TCT-TC
BDNF-M1	TCTG-GAA- <b>CGc</b> -AAT-TCT-TC
BDNF-M2	TCTG-GAA- <b>CGg</b> -Agc-cCT-TC
BDNF-M3	TCTG-Ggg- <b>CGG</b> -AAT-TCT-TC
BDNF-M4	TCTG-Ggg- <b>CGG</b> -Agc-cCT-TC

SEQLAB Sequence Laboratories Göttingen GmbH, Germany). Enzymatic CpG methylation of all plasmids is performed *in vitro* using methyltransferases, *M.SssI* or *HhaI* (New England Biolabs, USA). Complete methylation of the plasmids is analyzed by restriction digest with the methylation dependent endonuclease *HhaI* (New England Biolabs, USA) and subsequent gel electrophoresis (figure S1 ([stacks.iop.org/PhysBio/14/026002/mmedia](http://stacks.iop.org/PhysBio/14/026002/mmedia))).

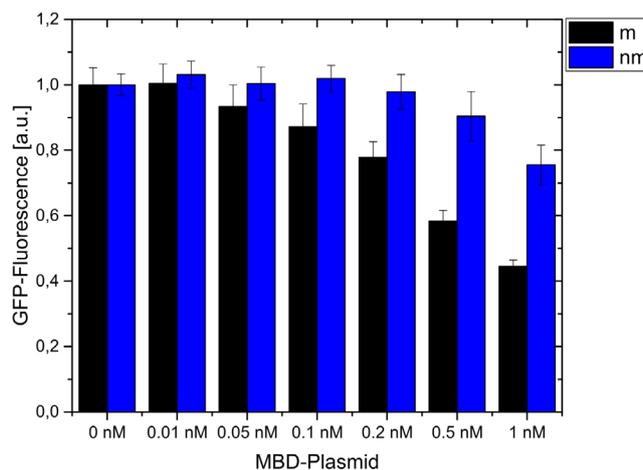
#### **In vitro expression experiment**

A typical cell-free reaction consists of 33% (v/v) extract and 67% (v/v) reaction buffer including the DNA templates. The reaction buffer is composed of 50 mM HEPES (pH value 8), 1.5 mM ATP and GTP, 0.9 mM CTP and UTP, 0.2 mg ml<sup>-1</sup> tRNA *E. coli*, 0.26 mM coenzyme A, 0.33 mM NAD, 0.75 mM cAMP, 0.068 mM folinic acid, 1 mM spermidine, 30 mM 3-phosphoglyceric acid (PGA), 1 mM DTT, and 2% (v/v) PEG8000 (all Sigma-Aldrich, Switzerland). Addition of amino

acids (0.5 mM of each of the 20 canonical amino acids, except for Leu that is 0.42 mM), magnesium glutamate (0–10 mM), and potassium glutamate (0 – 120 mM) are adjusted depending on the batch of extract and plasmid concentration. The system is prepared in such a way that the concentrations of these three components can be adjusted independently for any reaction. A typical batch mode reaction is prepared on ice by thoroughly mixing extract, buffer and plasmid DNA at specified concentrations. The solution is split into aliquots of 6  $\mu\text{l}$ . In this case, diffusion of oxygen into the extract is not a limiting factor [22]. The reaction tubes are incubated at  $29\text{ }^{\circ}\text{C}$  for at least 12 h. After incubation the aliquots are transferred into a single vessel. 10  $\mu\text{l}$  of the solution is transferred into a 384 multi well plate and the fluorescence intensity of the synthesized GFP is determined via plate reader (Polar Star Optima, BMG Labtech, Germany).

#### **Prokaryotic in vitro expression of CpG methylated DNA**

In bacteria the methylation sensitive restriction system mcr is believed to cleave CpG methylated DNA. Mcr acts as a primitive immune system [26] in *E. coli*. Figure S2 reveals that the mcr system is also present in our cell free extract and degrades CpG methylated reporter plasmids carrying mcr recognition motifs. In this case we do not detect any expression of the reporter gene. All CpG methylated reporter plasmids used in this study are devoid of any mcr recognition motifs. Their stability is confirmed by the successful expression of GFP from these reporter plasmids (figure S6).



**Figure 2.** Methylation response of MBD mediated repression of GFP using the wt BDNF promoter. The bars refer to the GFP fluorescence as a function of MBD plasmid concentration. m (black) refers to the fully CpG methylated promoter, nm (blue) to the non-methylated one. Data is normalized by the amount of expressed GFP in the absence of MBD. Only in the case of the CpG methylated reporter, adding MBD plasmids leads to an almost complete repression of GFP at the highest plasmid concentrations.

Based on the results in figure S6, we adjusted the reporter plasmid concentration to 5 nM for all experiments. For higher plasmid concentrations, sharing effects of the expression machinery are common [26]. Further, we checked that the presence of recombinant MBD from a different source, or MBD expressed from plasmid DNA within the cell free reaction system lead to a comparable result (figure S3).

### MD simulations

As a structural reference for the MeCP2:DNA complex, we use the x-ray structure of the BDNF promoter bound to the methyl-binding domain (PDB: 3C2I). The MD simulations are performed with the GROMACS 4.5.5 package [27] using the CHARMM27 force field [28] and the TIP3P water model [29]. The parameters for 5-methyl-cytosine are used as defined in the CHARMM force field. Systems with unbound DNA duplex strands or protein:dsDNA complexes are placed in a dodecahedral water box of 16 nm box dimensions with  $0.10 \text{ mol l}^{-1}$  of KCl added, so that the system has an overall zero electrostatic charge. The total size of the simulated systems is 56190 atoms for DNA solvated in a water box, and 56370 atoms for the solvated protein-DNA complex. Boundary conditions are periodic. Long-ranged Coulombic interactions beyond a cutoff of 13 Å are computed by the particle-mesh Ewald (PME) summation method [30]. The non-bonded Lennard-Jones interactions are computed using a smooth cutoff of 13 Å. The integration time step is set to 1 fs.

At first, each simulated system is energy-minimized for 50000 steps using the steepest descent algorithm followed by a second energy minimization for 10000 steps using a quasi-Newtonian algorithm with the low-memory Broyden-Fletcher-Goldfarb-Shanno approach. The tolerance is set to  $1.0 \text{ kJ mol}^{-1} \text{ nm}^{-1}$ . After that, the system is heated to 310 K during 4 ps. Then, each system is subjected to 2.0 ns-equilibration in the NVT

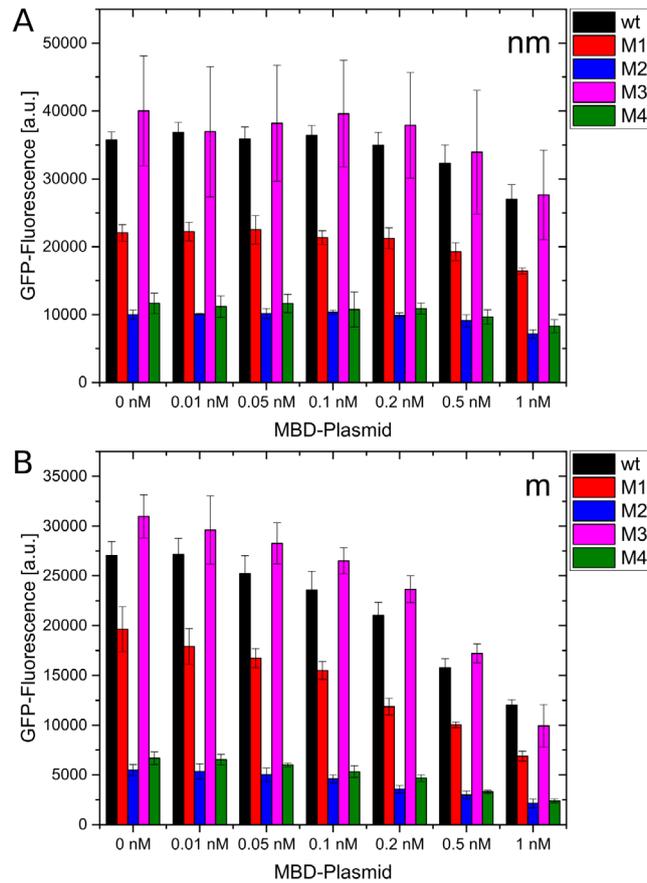
ensemble with harmonic restraints applied to all protein and DNA heavy atoms. The temperature is kept at 310 K by applying leap-frog stochastic dynamics forces with a damping coefficient of  $0.1 \text{ ps}^{-1}$ . With restraints kept, each system is further equilibrated for 0.5 ns in the NPT ensemble, and then for another 1.0 ns without restraints. MD simulations of 100 ns length are performed for the fully methylated DNA (the wild type), the non-methylated DNA, and methylated and non-methylated versions of the experimentally studied M2 mutant (modeled with 3DNA software [31]), which shows the strongest deviations from the wild type among the four studied mutants. Simulations are conducted in two replicates each. The results of the two replicates are almost indistinguishable. Thus the results for the second replicate are only shown in one table, but not in the figures. For the computation of RMSF fluctuations, the trajectories of the two replicates (100 ns each) are concatenated and fitted to the initial structure.

## Results

### MBD mediated repression of GFP with the wt version of the BDNF promoter

For the wt BDNF promoter we determine the overall amount of expressed GFP of the cell free reaction mix as a function of MBD plasmid DNA concentration (figure 2).

In our experiments, the final concentration of recombinant proteins attains the micromolar range [22] (figure S3). In case of the CpG methylated reporter, adding MBD plasmids entails repression of GFP expression. In case of the non-methylated reporter the same level of repression requires about  $5 \times$  higher plasmid concentrations. If BDNF is replaced by a different promoter that contains a repetition of the CpG recognition motif, this does not modulate GFP expression in a methylation dependent manner (figure S5). We conclude that it is the BDNF promoter that is responsi-



**Figure 3.** Amount of expressed GFP as a function of MBD concentration in conjunction with the wt BDNF promoter and mutations M1–M4. (a) Non-methylated plasmid: the expression is weakly repressed only for the highest MBD concentrations. M1, M2 and M4 present a lower expression level than the wt regardless of MBD concentration. (b) Methylated plasmids: the overall expression levels appear reduced compared to (a). All promoters exhibit stronger repression in presence of MBD compared to the non-methylated plasmid.

ble for the observed methylation-dependent repression and that MBD recognition is specific to this promoter.

### Sequence mutations of the BDNF promoter

We experimentally test four mutations in the BDNF promoter (table 1) for changes in transcriptional regulation. M1 and M3 are designed to study the influence of the flanking bases of the central CpG motif where methylation is discriminated. M2 breaks the run of the four AT base pairs downstream of the CpG motif [12]. M4 combines the mutations M2 and M3 within a single promoter. The central CpG motif of the BDNF promoter remains untouched in any case.

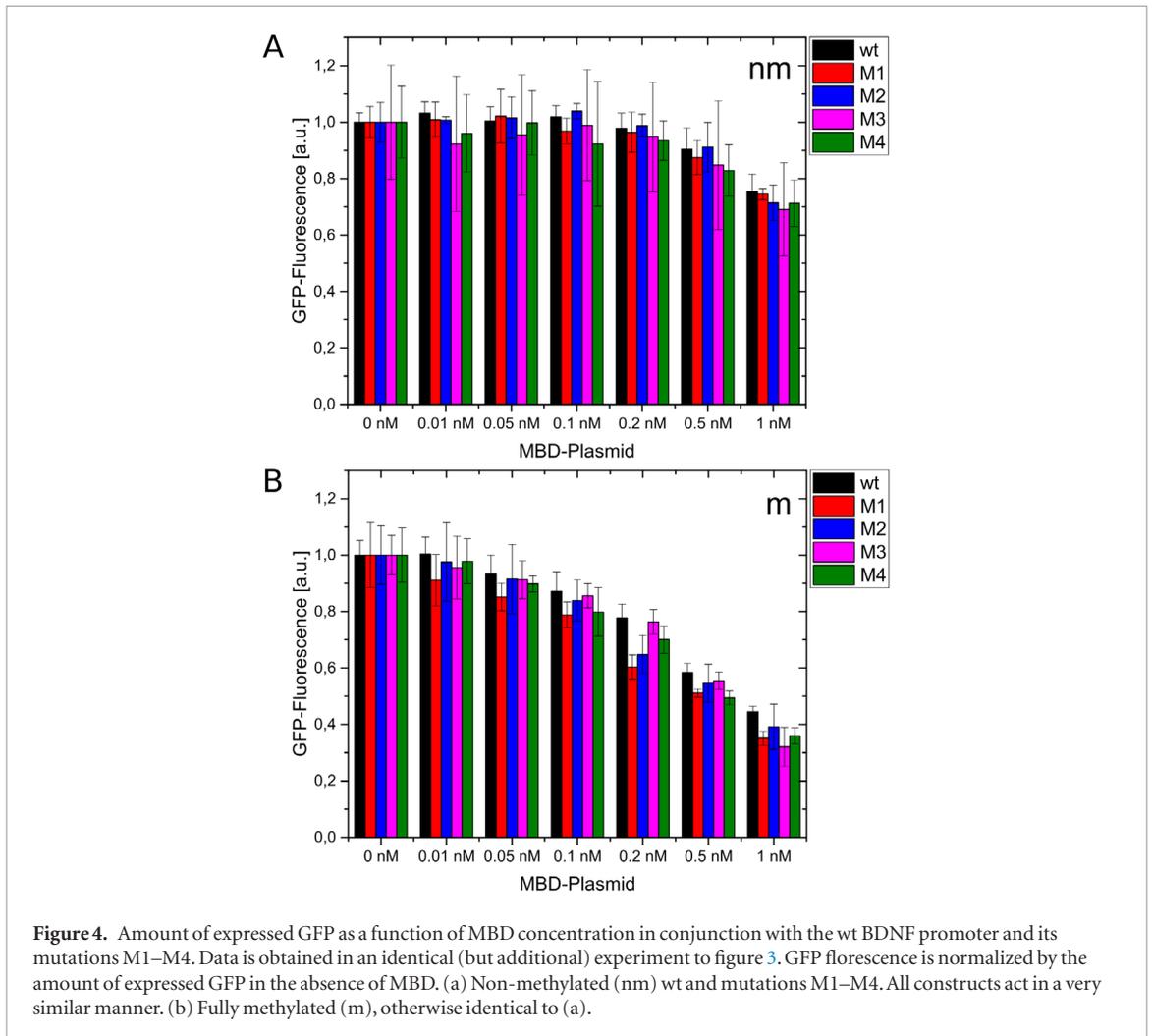
Figure 3 shows the expressed GFP concentrations as a function of plasmid methylation state and the type of mutation. Of the non-methylated plasmids, M1 generates lower expression levels than the wt, while M3 tends to exceed this level. M2 and M4 exhibit a strongly reduced expression level to about 20% of wt. If methylated, in presence of the MBD plasmid the expression levels of all constructions appear reduced. In an additional but otherwise identical experiment we normalize the measured GFP fluorescence to the value in absence of the MBD plasmid (figure 4). The results reveal a comparable, relative repression of all constructions

(wt, M1–M4) in the case of CpG methylated and non-methylated reporter plasmids.

### Simple kinetic model

For expression regulation, both, MBD and the sigma factor 70 bind to the BDNF-promoter. MBD binding ( $k_M^{\text{on}}$ ) is limited to the CpG pair that occurs in the wt and M1–M4 promoter only once. In contrast the sigma factor 70 recognizes the  $-35$  and  $-10$  region flanking the promoter. Conformational changes of the promoter by different mutations will influence the association constant of the sigma factor 70 ( $k_\sigma^{\text{on}}$ ) which has multiple contact points within the promoter. Binding also depends on the presence of MBD because of structural hindrance. When binding to the BDNF promoter in presence of MBD ( $k_{\sigma,+}^{\text{on}}$ ) or in absence ( $k_{\sigma,-}^{\text{on}}$ ), the sigma factor 70 initiates transcription and dissociates, ( $k_{\sigma,-}^{\text{on}} \gg k_{\sigma,+}^{\text{on}}$ ). This is fast, and we can neglect the influence of the presence of the sigma factor 70 on MBD binding. The promoter can occur in 4 states (figure 5).

The expression rate is given by  $E = A_- E_- + A_+ E_+$  where  $A_- = \frac{[(3)]}{[(3)] + [(4)]}$  and  $A_+ = \frac{[(4)]}{[(3)] + [(4)]}$  are the expression rates without ( $E_-$ ) or with MBD bound ( $E_+$ ) ( $E_- > E_+$ ).  $[(x)]$  denotes the concentration of



**Figure 4.** Amount of expressed GFP as a function of MBD concentration in conjunction with the wt BDNF promoter and its mutations M1–M4. Data is obtained in an identical (but additional) experiment to figure 3. GFP fluorescence is normalized by the amount of expressed GFP in the absence of MBD. (a) Non-methylated (nm) wt and mutations M1–M4. All constructs act in a very similar manner. (b) Fully methylated (m), otherwise identical to (a).

the promoter in state  $x$ . Using  $A_- + A_+ = 1$ , and  $K_{D,M} = \frac{[(3)][MBD]}{[(4)]}$  yields:

$$E = E_+ + \frac{(E_- - E_+) \cdot K_{D,M}}{K_{D,M} + [MBD]} \quad (1)$$

Figure 6 shows the GFP-Fluorescence as a function of MBD-Plasmid concentration in case of the methylated wt BDNF promoter and the fit with equation (1). The fits of the experimental data of the mutants M1–M4 show similar accuracy (supplementary material, figure S8).

In all cases (wt, M1–M4)  $E_-$  corresponds to the measured value  $E([MBD-Plasmid] = 0 \text{ nM})$ . The values of  $E_+$  and  $K_{D,M}$  were obtained from a fit (Origin). While  $E_+$  is in the range of approximately 1/8 to 1/5 of  $E_-$ ,  $K_{D,M}$  is between 0.29 nM and 0.55 nM for the different mutants (supplementary material, S9). However, here the calculated dissociation constant is given as a function of MBD-Plasmid concentration. Figure S3 relates MBD-Plasmid to MBD-Protein concentrations (0.5 nM MBD-Plasmid  $\approx 1 \mu\text{M}$  MBD protein). This leads to an effective dissociation constant of 0.5  $\mu\text{M}$  to 1  $\mu\text{M}$ .

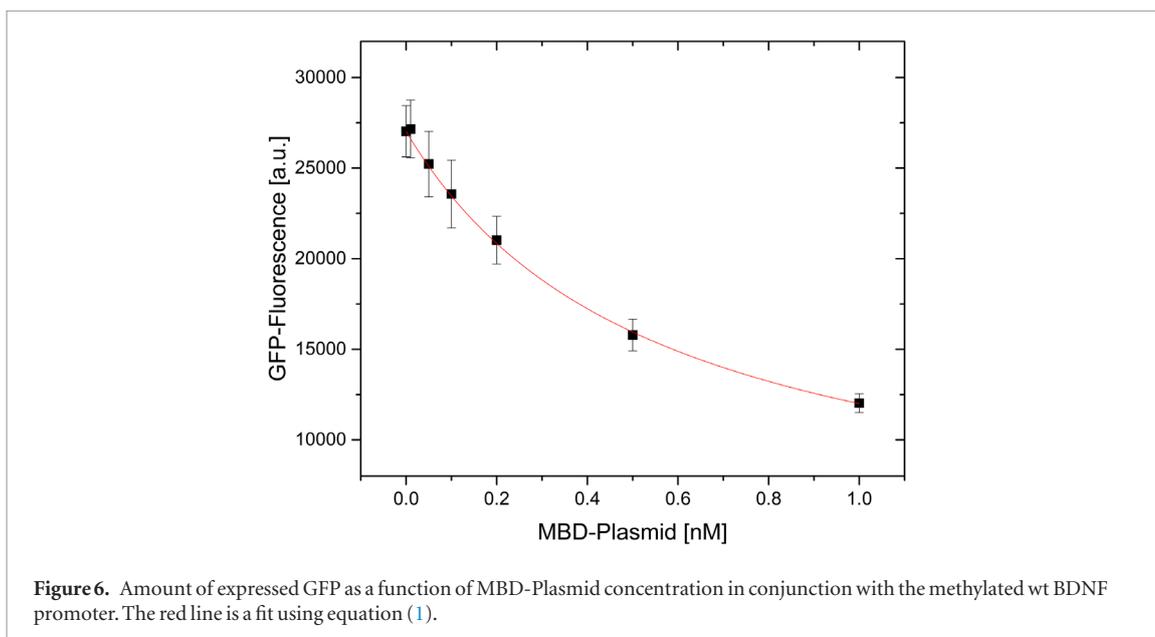
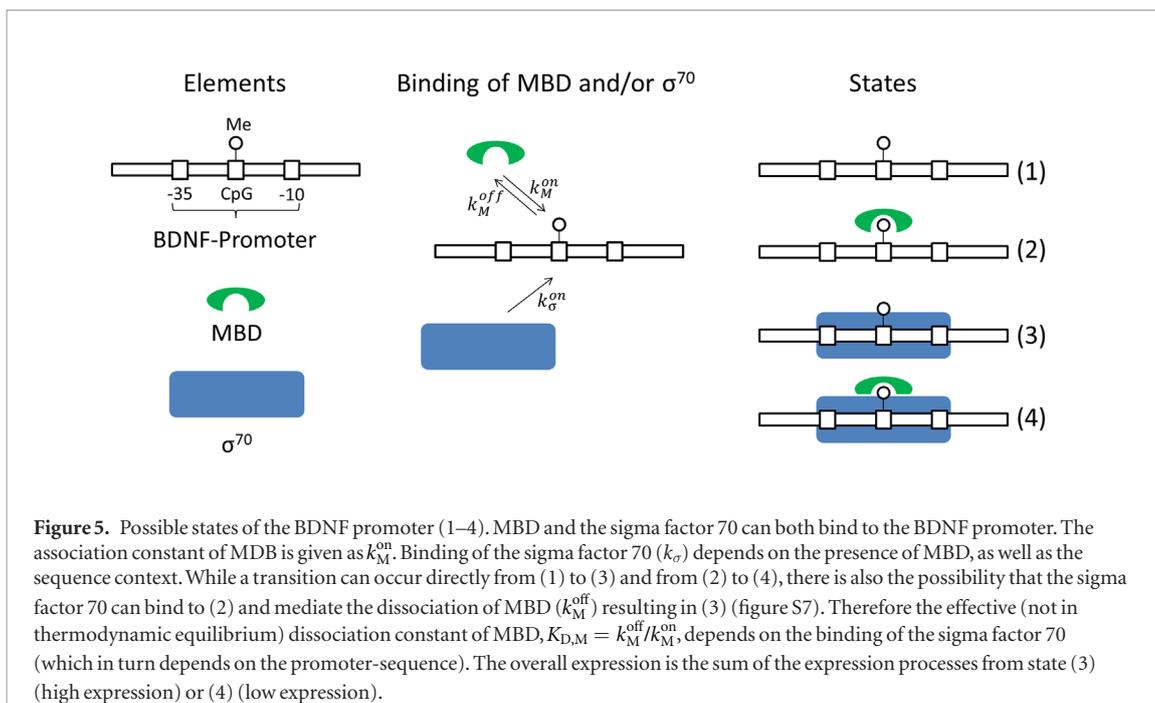
In the case where the BDNF promoter was replaced by the promoter with the CpG-repetitions there is no repression of the expression with increasing MBD-Plasmid concentration. In terms of the model, this is achieved by a higher  $K_{D,M}$  and a decreased  $E_-/E_+$ -ratio.

## MD simulations

MD simulations of free promoter DNA as well as the MBD:DNA complex reveal stable conformations for methylated and non-methylated DNA (RMSD  $\sim 2.0 \text{ \AA}$  from the average structure). Methylated, MBD-bound promoter DNA presents the smallest root mean square fluctuations (RMSF) of the heavy atoms in its backbone whereas non-methylated MBD bound DNA exhibits the largest ones (figure 7), exceeding even those of the unbound DNA, methylated or not.

The simulations show that binding of methylated DNA (wild-type and M2) to MBD induces a widening of the major DNA groove at the binding interface from 15  $\text{\AA}$  to about 20  $\text{\AA}$  (figure 8). This change is less pronounced when MBD binds to non-methylated DNA (wild-type and M2).

Figure 9 shows the fraction of DNA base pairs adopting a B-DNA conformation (also known as B-fiber ratio). For unbound wt DNA, both in the methylated and non-methylated states, roughly 80% of the base-pairs are in B-DNA conformation on average. Unbound M2-DNA is characterized by lower ratios of 70% (mDNA<sub>M2</sub>) and 65% (nDNA<sub>M2</sub>). Upon binding to MBD, the B-fiber ratios either decrease by about 10% to 70% (mDNA<sub>wt</sub> and nDNA<sub>wt</sub>) or by about 5% to 65% (mDNA<sub>M2</sub>). In the case of non-methylated

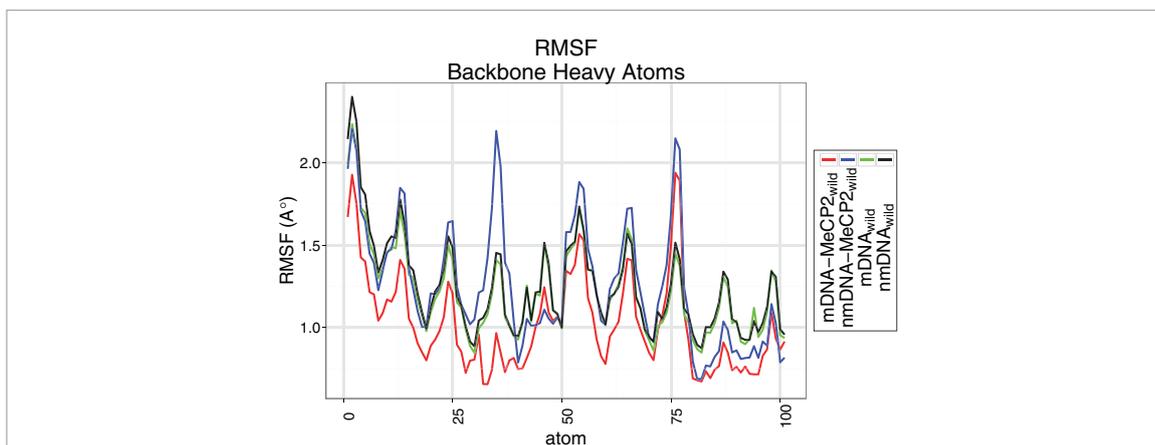


M2, the B-fiber ratio decreases insignificantly to a low value of 62%.

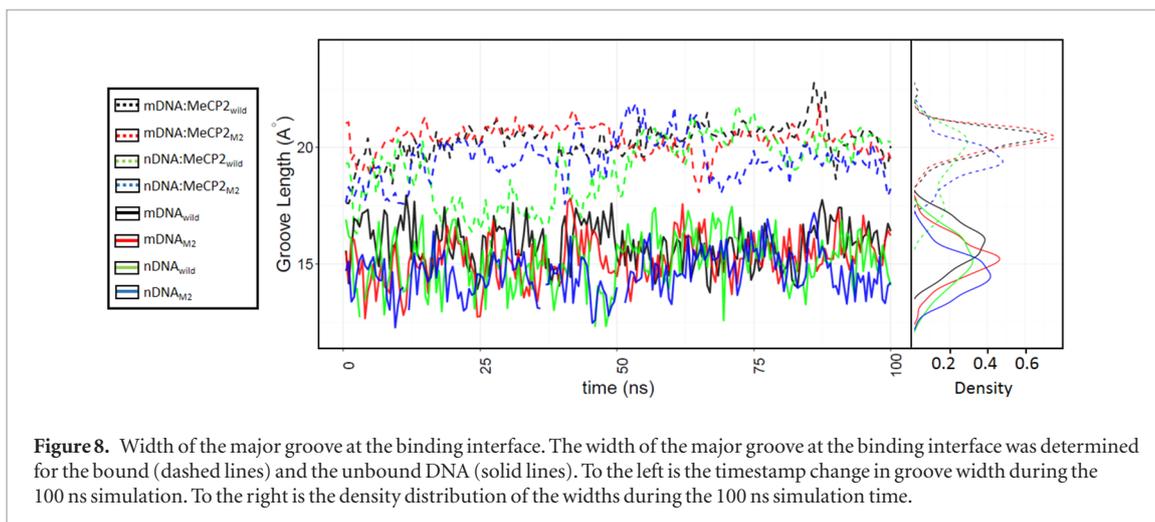
The handedness (see supplementary material and figure S4 for further details on handedness) describes the helical twisting of DNA. It can be used to check for transitions from compact, right-handed B-DNA to the relatively extended, left-handed Z-DNA. The handedness values indicate that upon binding of MBD to wt DNA the CpG recognition motif untwists partially from the perfect B-DNA fiber (table 2, figure 10). This is more pronounced if the DNA is methylated (from 0.552 to 0.446 on average) than for the non-methylated form (from 0.548 to 0.480). M2, however, displays a handedness that is practically independent of methylation while fluctuations are larger in the unbound state. The value of handedness leans towards the methylated wt complex.

## Discussion

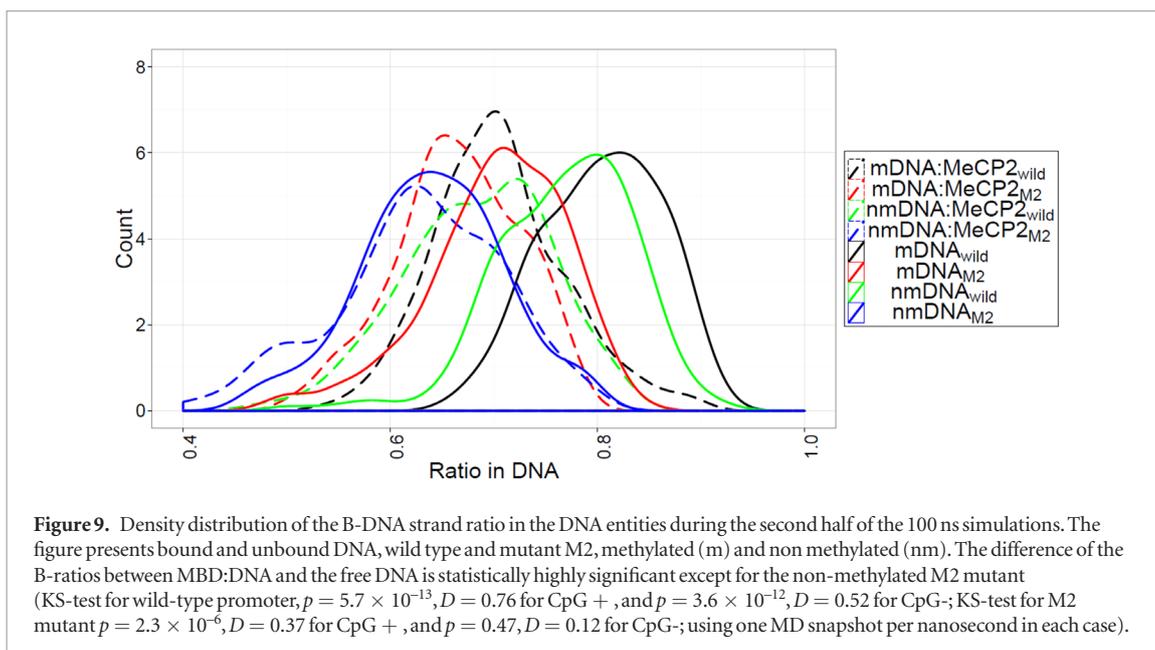
Since the molecular environment of *E. coli* is tuned for prokaryotic expression, one might suspect eukaryotic transcriptional elements not to function in our experiments. Adenine methylation of the DNA sequence motif 5'-GATC-3' by the enzyme deoxyadenosine methylase (Dam) plays an important role in the timing of initiation of DNA replication in *E. coli*, as well as in the coordination of cellular events, DNA mismatch repair, and gene regulation [32]. At the same time, cytosine methylation is easily recognized as foreign. However, we find that our experiments are not affected by any of the above. We use solely MBD, the recognition domain, instead of the entire MeCP2 protein and find that the expression downstream of the BDNF promoter is strongly repressed by MBD only if



**Figure 7.** Root mean squared fluctuations (RMSF) of the heavy atoms in the DNA backbone, methylated (m) and non methylated (nm), free (green and black, respectively) and after binding to MBD (red and blue, respectively). Trajectories of the two replicates (100 ns each) were concatenated and fitted to the initial structure. According to the RMSF analysis, the DNA in the methylated complex exhibits the smallest conformational fluctuations. This suggests that the methyl group structurally stabilizes the protein: DNA complex.



**Figure 8.** Width of the major groove at the binding interface. The width of the major groove at the binding interface was determined for the bound (dashed lines) and the unbound DNA (solid lines). To the left is the timestamp change in groove width during the 100 ns simulation. To the right is the density distribution of the widths during the 100 ns simulation time.



**Figure 9.** Density distribution of the B-DNA strand ratio in the DNA entities during the second half of the 100 ns simulations. The figure presents bound and unbound DNA, wild type and mutant M2, methylated (m) and non-methylated (nm). The difference of the B-ratios between MBD:DNA and the free DNA is statistically highly significant except for the non-methylated M2 mutant (KS-test for wild-type promoter,  $p = 5.7 \times 10^{-13}$ ,  $D = 0.76$  for CpG +, and  $p = 3.6 \times 10^{-12}$ ,  $D = 0.52$  for CpG-; KS-test for M2 mutant  $p = 2.3 \times 10^{-6}$ ,  $D = 0.37$  for CpG +, and  $p = 0.47$ ,  $D = 0.12$  for CpG-; using one MD snapshot per nanosecond in each case).

this promoter is methylated (figure 2). This happens at MBD concentrations that are typical for transcription factors in cells [33], however, they are well below the

MeCP2 concentrations found in cell nuclei [34]. An unrelated promoter, inserted at the place of the BDNF sequence, does not exhibit any methylation dependence

**Table 2.** Handedness of the central dinucleotide steps, namely the (36T-7A-35T-85CM) and the (7DA-35DT-85CM-34DG) in bound and unbound states for wild-type BDNF promoter and for the M2 mutant. The 2 values given in each field belong to 2 independent replicate simulations of 100 ns length each.

	DNA in complex		Unbound DNAs	
	Central step 1 (36T-7A-35T-85CM)	Central step 2 (7DA-35DT-85CM-34DG)	Central step 1 (36T-7A-35T-85CM)	Central step 2 (7DA-35DT-85CM-34DG)
Wild type	0.408 ± 0.057	0.459 ± 0.045	0.524 ± 0.043	0.582 ± 0.038
Methylated	0.441 ± 0.067	0.476 ± 0.063	0.521 ± 0.042	0.580 ± 0.037
Mutant 2	0.402 ± 0.052	0.462 ± 0.038	0.523 ± 0.042	0.579 ± 0.041
Methylated	0.403 ± 0.052	0.462 ± 0.038	0.522 ± 0.044	0.576 ± 0.042
Wild type	0.492 ± 0.055	0.540 ± 0.048	0.518 ± 0.044	0.575 ± 0.042
Non-methylated	0.411 ± 0.084	0.477 ± 0.058	0.523 ± 0.047	0.576 ± 0.043
Mutant 2	0.405 ± 0.071	0.459 ± 0.055	0.526 ± 0.046	0.577 ± 0.039
Non-methylated	0.408 ± 0.048	0.451 ± 0.038	0.452 ± 0.200	0.500 ± 0.214

although it contains a repetition of symmetrical CpG motifs (figure S5) that were shown to considerably enhance MBD affinity in binding assays [15]. All of the above exhibits a high degree of specific repression that is astonishing, given that we are in a prokaryotic molecular environment, and MBD represents a rather small portion of MeCP2 that, to our knowledge, has not been described as a functional eukaryotic regulator *per se*.

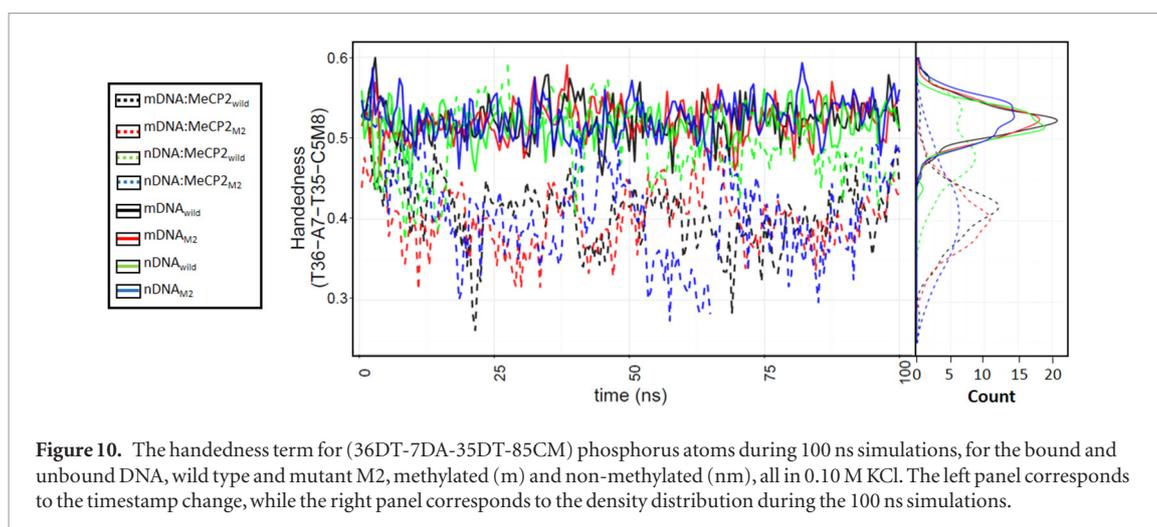
We included four mutations into the promoter and examined them for expression efficiency (figures 3 and 4). Both, M1 and M3 carry a mutation of one of the bases flanking the central CpG recognition motif. Based on structural data (14), Ho *et al* showed that direct MBD binding with respect to the flanking bases is limited to the DNA backbone. In agreement, in our experiments the flanking bases only change the expression level of the target gene in the absence of MBD but they do not change the relative repression that is caused by MBD.

Mutant M2 had a broken run of AT bases downstream of the CpG motif. This AT-run was reported to be essential for MBD binding to BDNF [12]. We observe that M2 leads to a lower expression level of the target GFP gene even in the absence of MBD (figure 3). We conclude that the AT-run of the BDNF promoter enhances the expression by the *E. coli* polymerase. For M2, methylation produces repression of comparable magnitude as the wild type. In agreement with the structural analysis [11], the AT-run does not appear as a requirement for MBD binding in our study.

The observations made within the experiments are well described by a simple kinetic model assuming that the BDNF promoter can occur in four states. This Michaelis–Menten type description fits the data well. While the value for the effective dissociation constant of MBD,  $K_{D,M}$ , from fitting the experimental data is estimated between 0.5  $\mu\text{M}$  and 1  $\mu\text{M}$ , the value found in the literature in equilibrium is about 8.5 nM [35]. However, in our experiment there is non-equilibrium, as well as competition between MBD and the sigma factor 70, and possibly many other molecules from the *E. coli* extract. As a result it may not be very useful to compare both values.

In MD simulations, the width of the major groove of the promoter increases upon binding to MBD (figure 8), as with most DNA binding proteins [36]. This improves the accessibility of the functional groups of DNA and favors the specificity of protein-DNA contacts. In parallel to this, the fraction of DNA base pairs in B-DNA conformation decreases (figure 9) as well as the handedness (table 2, figure 10). Both observables are indications of a subtle conformational transition toward Z-DNA. For wtBDNF promoter, MD simulations predict that its conformational fluctuations are lower for the methylated MBD complex as compared to unbound DNA, or to the non-methylated complex. Also, binding of the methylated BDNF promoter induces a larger opening of the major groove than for the non-methylated form. These observations suggest that MBD forms tighter contacts to methylated BDNF (as evidenced by low conformational fluctuations, see RMSF analysis in figure 7). This matches well the efficiency for repression reported in figure 3, as well as the computational results of Zou and co-workers [16].

The degree of opening of the DNA major groove for methylated M2-DNA and the change in handedness of the central dinucleotide steps, are comparable to that for wt DNA (figures 8 and 10, table 2). This suggests that MBD also binds to the mutant M2 although the AT-run is lacking. On the other hand, in simulations of the unbound M2 promoter, a lower fraction of DNA bases adopt B-DNA conformation (figure 9) than for the unbound wt DNA. Also when bound to MBD, the fraction of B-DNA is smaller for M2-DNA than for wt DNA. Moreover, the distribution of handedness of M2 is distinctly broader than for the wt, regardless of methylation. Both observables point to a more floppy DNA conformation of the M2 promoter. At the same time the data above suggests that the decreased expression level of M2 is due to a DNA conformation that, without being methylated, retains structural features of the repressing, methylated wt MBD complex. We suggest that the AT stretch helps to stabilize in particular the non-methylated DNA conformation in such a way that a high expression-level is achieved.



As a general observation, the structural transitions upon MBD binding are stronger for the methylated promoter than for the non-methylated one. They extend until the AT stretch, where the mutation of M2 is located. *In vitro* footprinting revealed protection that extends 6 bases from the central motif [15]. We hypothesize that the high degree of cooperativity is a reason for the highly specific BDNF motif recognition that MBD presents in our experiments—in spite of its very localized interaction. In agreement with previous knowledge, our results support the idea that MBD binds to the central CpG pair as a recognition motif. However, we suggest that in order to act as a suppressor, not only steric hindrance plays a role, but also MBD modifies the DNA conformation in such a way that the polymerase binds with reduced affinity. This idea is corroborated by our observation that a repetition of CpGs does not act as an inhibitor, although MBD binds with even higher affinity to this type of sequence [15]. It is clear that a high degree of cooperativity along the DNA BDNF sequence in conjunction with a very localized recognition site implies a certain fragility towards mutations. We speculate that the highly cooperative nature of molecular recognition may contribute to the loss/gain of functions of MBD-proteins, which has been discovered as a result of point mutations in connection with the Rett syndrome [37], a considerable fraction of them identified in the MBD domain. However, a much more detailed study would be required to confirm this hypothesis.

## Conclusion

In this study we report the first epigenetic regulation due to a eukaryotic protein and eukaryotic cytosine methylation that is realized in a prokaryotic bacterial cell-free transcription translation system. At first sight, the eukaryotic polymerase and its expression machinery seem to have nothing much in common with prokaryotes. However, the fact that in our experiments MBD alone, as opposed to the entire MeCP2, performs as a specific repressor in conjunction with the BDNF promoter sequence, shows that in our case the situation

is actually quite clear. Whereas binding assays simply detect the formation of MBD-DNA contacts, our cell-free expression system reveals repression of the target gene by MBD that is probably mediated by conformational changes of the connecting DNA stretch. MD simulations suggest subtle cooperative transitions of the DNA conformation. Given the highly localized interaction of MBD, this may well explain the high degree of specificity to the BDNF sequence as observed in our experiments.

## Acknowledgments

This work was supported by the collaborative research center SFB 1027 funded by Deutsche Forschungsgemeinschaft (DFG). SS acknowledges a doctoral fellowship from the German Academic Exchange Service (DAAD).

## Author contributions

VN, VH, AO, MS, designed research, MF, MS, SS performed research, analyzed data with the help of AO and VH, MF, EGW performed replicate experiments, SS, MS, VH and AO wrote the paper with help from EGW and MF.

## References

- [1] Fatemi M, Pao M M, Jeong S, Gal-Yam E N, Egger G, Weisenberger D J and Jones P A 2005 Footprinting of mammalian promoters: use of a CpG DNA methyltransferase revealing nucleosome positions at a single molecule level *Nucleic Acids Res.* **33** e176
- [2] Kahramanoglou C, Prieto A I, Khedkar S, Haase B, Gupta A, Benes V, Fraser G M, Luscombe N M and Seshasayee A S N 2012 Genomics of DNA cytosine methylation in *Escherichia coli* reveals its role in stationary phase transcription *Nat. Commun.* **3** 886
- [3] Lewis J D, Meehan R R, Henzel W J, Maurer-Fogy I, Jeppesen P, Klein F and Bird A 1992 Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA *Cell* **69** 905–14
- [4] Lopez-Serra L, Ballestar E, Fraga M F, Alaminos M, Setien F and Esteller M 2006 A profile of methyl-CpG binding domain protein occupancy of hypermethylated promoter CpG

- islands of tumor suppressor genes in human cancer *Cancer Res.* **66** 8342–6
- [5] Nan X, Campoy F J and Bird A 1997 MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin *Cell* **88** 471–81
- [6] Long S W, Ooi J Y Y, Yau P M and Jones P L 2011 A brain-derived MeCP2 complex supports a role for MeCP2 in RNA processing *Biosci. Rep.* **31** 333–43
- [7] Chahrour M, Jung S Y, Shaw C, Zhou X, Wong S T C, Qin J and Zoghbi H Y 2008 MeCP2, a key contributor to neurological disease, activates and represses transcription *Science* **320** 1224–9
- [8] Guy J, Cheval H, Selfridge J and Bird A 2011 The role of MeCP2 in the brain *Annu. Rev. Cell Dev. Biol.* **27** 631–52
- [9] Kaludov N K and Wolffe A P 2000 MeCP2 driven transcriptional repression *in vitro*: selectivity for methylated DNA, action at a distance and contacts with the basal transcription machinery *Nucleic Acids Res.* **28** 1921–8
- [10] Chen W G, Chang Q, Lin Y, Meissner A, West A E, Griffith E C, Jaenisch R and Greenberg M E 2003 Derepression of BDNF transcription involves calcium-dependent phosphorylation of MeCP2 *Science* **302** 885–9
- [11] Ho K L, McNaie I W, Schmiedeberg L, Klose R J, Bird A P and Walkinshaw M D 2008 MeCP2 binding to DNA depends upon hydration at methyl-CpG *Mol. Cell* **29** 525–31
- [12] Klose R J, Sarraf S A, Schmiedeberg L, McDermott S M, Stancheva I and Bird A P 2005 DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG *Mol. Cell* **19** 667–78
- [13] Li W and Pozzo-Miller L 2014 BDNF deregulation in Rett syndrome *Neuropharmacology* **76** 737–46
- [14] Khrapunov S, Warren C, Cheng H, Berko E R, Grealley J M and Brenowitz M 2014 Unusual characteristics of the DNA binding domain of epigenetic regulatory protein MeCP2 determine its binding specificity *Biochemistry* **53** 3379–91
- [15] Nan X, Meehan R R and Bird A 1993 Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2 *Nucleic Acids Res.* **21** 4886–92
- [16] Zou X, Ma W, Solov' yov I A, Chipot C and Schulten K 2012 Recognition of methylated DNA through methyl-CpG binding domain proteins *Nucleic Acids Res.* **40** 2747–58
- [17] Ohki I, Shimotake N, Fujita N, Jee J, Ikegami T, Nakao M and Shirakawa M 2001 Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA *Cell* **105** 487–97
- [18] Ho K L 2009 Structural studies of MeCP2 in complex with methylated DNA *Thesis* The University of Edinburgh
- [19] Worst E G, Exner M P, De Simone A, Schenkelberger M, Noireaux V, Budisa N and Ott A 2015 Cell-free expression with the toxic amino acid canavanine *Bioorg. Med. Chem. Lett.* **25** 3658–60
- [20] Worst E G, Exner M P, De Simone A, Schenkelberger M, Noireaux V, Budisa N and Ott A 2016 Residue-specific incorporation of noncanonical amino acids into model proteins using an *Escherichia coli* cell-free transcription-translation system *J. Vis. Exp.* **114** e54273
- [21] Sun Z Z, Hayes C A, Shin J, Caschera F, Murray R M and Noireaux V 2013 Protocols for implementing an *Escherichia coli* based TX-TL cell-free expression system for synthetic biology *J. Vis. Exp.* **79** e50762
- [22] Shin J and Noireaux V 2010 Efficient cell-free expression with the endogenous *E. Coli* RNA polymerase and sigma factor 70 *J. Biol. Eng.* **4** 8
- [23] Li X, Zhang G, Ngo N, Zhao X, Kain S R and Huang C C 1997 Deletions of the *Aequorea victoria* green fluorescent protein define the minimal domain required for fluorescence *J. Biol. Chem.* **272** 28545–9
- [24] Olins P O, Devine C S, Rangwala S H and Kavka K S 1988 The T7 phage gene 10 leader RNA, a ribosome-binding site that dramatically enhances the expression of foreign genes in *Escherichia coli* *Gene* **73** 227–35
- [25] Larson M H, Greenleaf W J, Landick R and Block S M 2008 Applied force reveals mechanistic and energetic details of transcription termination *Cell* **132** 971–82
- [26] Shin J and Noireaux V 2012 An *E. coli* cell-free expression toolbox: application to synthetic gene circuits and artificial cells *ACS Synth. Biol.* **1** 29–41
- [27] Hess B, Kutzner C, van der Spoel D and Lindahl E 2008 GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation *J. Chem. Theory Comput.* **4** 35–447
- [28] Foloppe N and MacKerell A D Jr 2000 All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data *J. Comput. Chem.* **21** 86–104
- [29] Jorgensen W L, Chandrasekhar J, Madura J D, Impey R W and Klein M L 1983 Comparison of simple potential functions for simulating liquid water *J. Chem. Phys.* **79** 926
- [30] Darden T, York D and Pedersen L 1993 Particle mesh Ewald: an  $N \cdot \log(N)$  method for Ewald sums in large systems *J. Chem. Phys.* **98** 10089
- [31] Lu X-J and Olson W K 2008 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures *Nat. Protocols* **3** 1213–27
- [32] Braaten B A, Nou X, Kaltenbach L S and Low D A 1994 Methylation patterns in pap regulatory DNA control pyelonephritis-associated pili phase variation in *E. coli* *Cell* **76** 577–88
- [33] Milo R, Phillips R and Orme N 2015 *Cell Biology by the Numbers* 1st edn (London: Taylor and Francis)
- [34] Skene P J, Illingworth R S, Webb S, Kerr A R W, 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** 457–68
- [35] Ghosh R P, Nikitina T, Horowitz-scherer R, Gierasch L M, Uversky V N, Hite K, Hansen J C and Christopher L 2010 Unique physical properties and interactions of the domains of methylated DNA binding protein 2 (MeCP2) *Biochemistry* **49** 4395–410
- [36] Pabo C O and Sauer R T 1984 Protein-DNA recognition *Annu. Rev. Biochem.* **53** 293–321
- [37] Kucukkal T G, Yang Y, Uvarov O, Cao W and Alexov E 2015 Impact of rett syndrome mutations on MeCP2 MBD stability *Biochemistry* **54** 6357–68

## Supplementary Material

### Handedness

The collective variable of handedness was used to characterize the conformational changes of DNA upon MeCP2 binding. It is a good choice for detecting the helical twisting from the right-handed B-DNA to the left-handed Z-DNA. According to the definition by Moradi et al [1], given a set of basepairs, starting at basepair  $n$  and ending at basepair  $m$ ; the following sequence of atoms is used to describe the collective variable:  $P_n^1, P_n^2, P_{n+1}^1, P_{n+1}^2, \dots, P_m^1, P_m^2$ ; where  $P_n^1$  is the atom starting from the 5' position in the  $n^{\text{th}}$  basepair. As such, the sum can also be started starting at the 5' nucleotide triphosphate from the other end. The total collective variable of handedness for the DNA strand is thus the polynomial sum of the handedness terms, starting at one phosphorus atom each, and ending three bases thereafter (e.g;  $P_n^1, P_n^2, P_{n+1}^1, P_{n+1}^2 + P_{n+2}^1, P_{n+2}^2 + \dots + P_{m-1}^1, P_{m-1}^2, P_m^1, P_m^2$ ). Thus, the position of these atoms defines the handedness via:

$$H(p_1 p_2 p_3 \dots p_n) = \sum_{k=1}^{n-3} H(p_i p_{i+1} p_{i+2} p_{i+3})$$

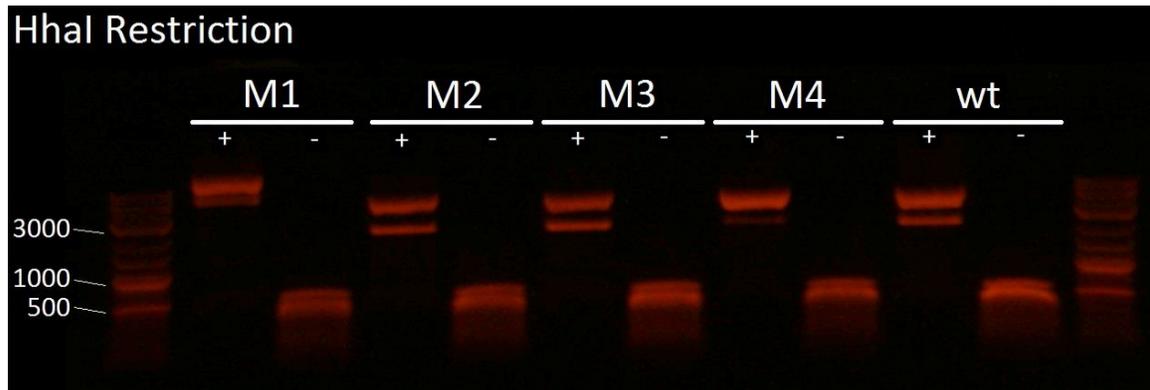
Given a sequence of points  $A, B, C$ , and  $D$ ; the single handedness term is defined as (figure S4):

$$H(ABCD) = \frac{\overrightarrow{AB}}{|\overrightarrow{AB}|} \times \frac{\overrightarrow{CD}}{|\overrightarrow{CD}|} \cdot \frac{\overrightarrow{EF}}{|\overrightarrow{EF}|}$$

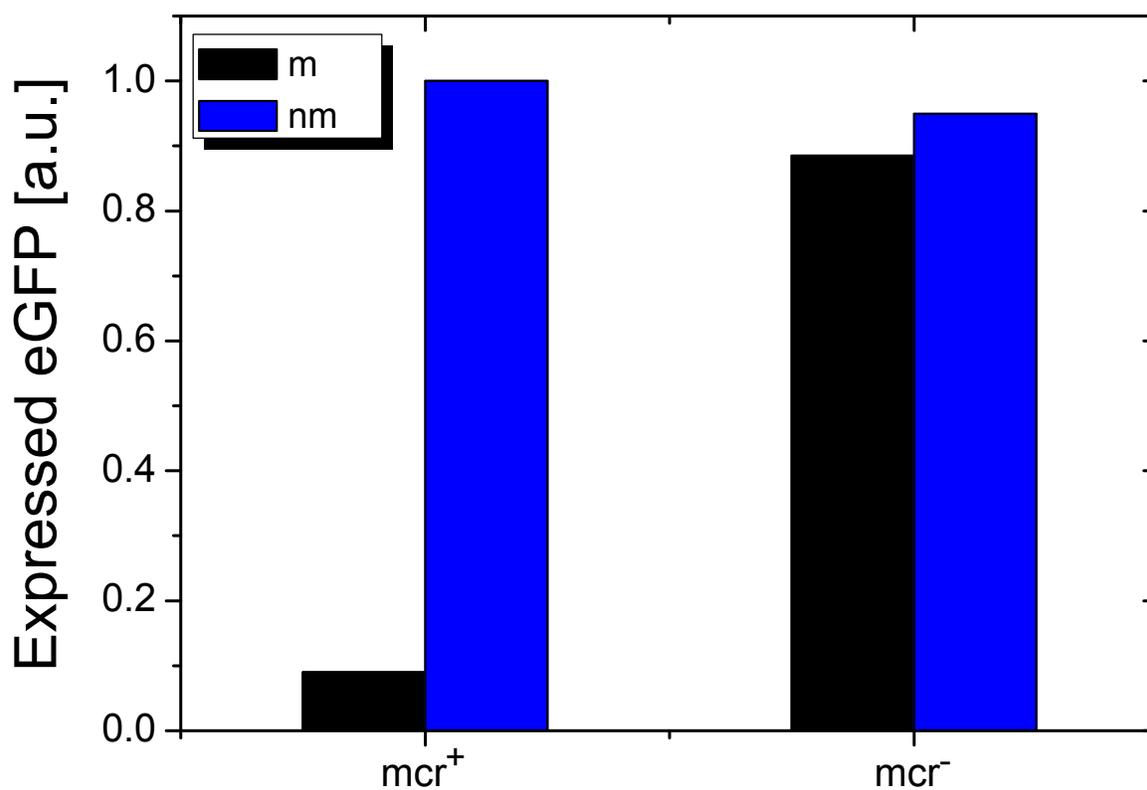
where the points define the vectors  $\overrightarrow{AB}$  and  $\overrightarrow{CD}$ . The vector  $\overrightarrow{EF}$  defines the vector matching the midpoints of  $\overrightarrow{AB}$  and  $\overrightarrow{CD}$ .

To detect possible features of B-Z transitions in the DNA fibers, the collective variable of handedness was investigated during the 100 ns simulation for the DNA in the bound and free forms (see figure S4). The range of the collective variable for the unbound DNA is between 52-58°. Trajectories used for analysis include the wild type DNA in the bound and unbound forms (CpG+ and CpG-) and the M2 mutant in the bound and unbound forms (CpG+ and CpG-).

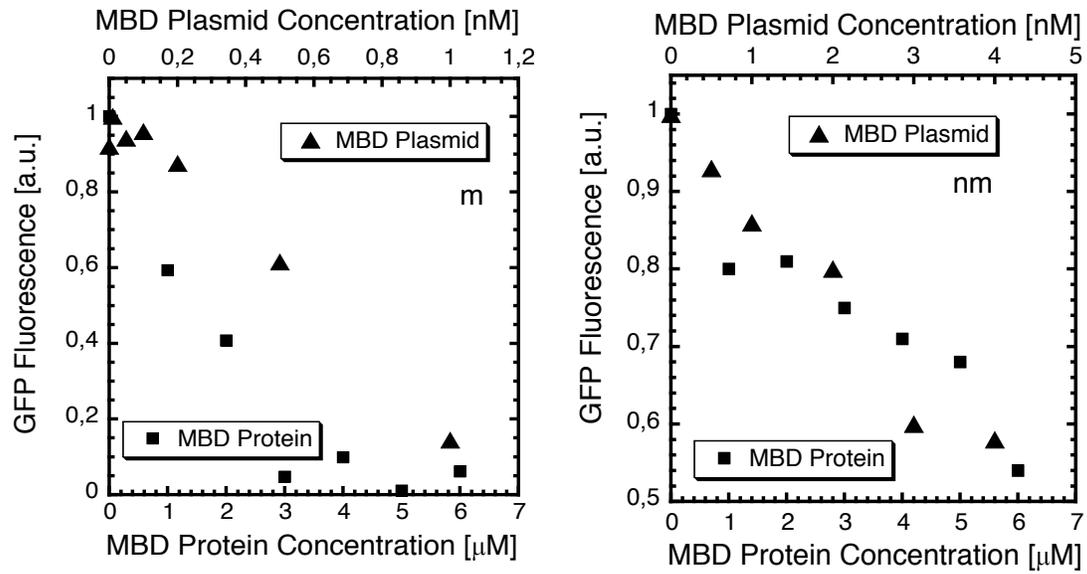
## Supplementary Figures



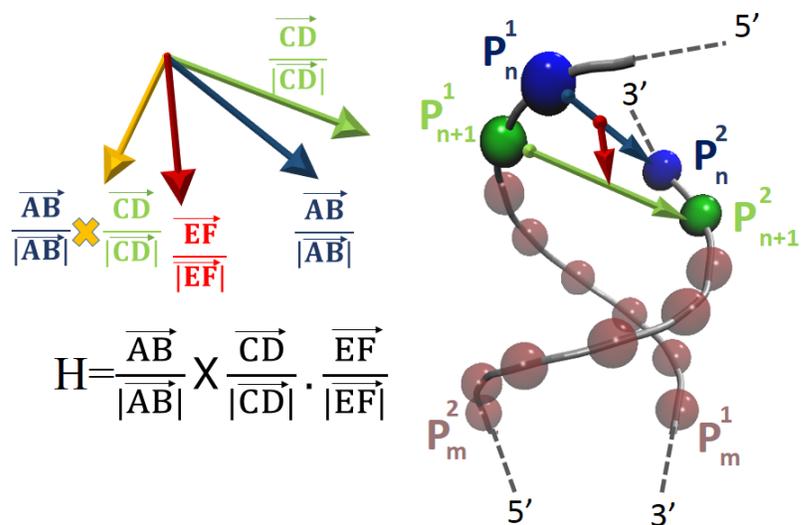
**Figure S1.** Assessment of the methylation state of the plasmid DNA. Restriction digest of the plasmids pBEST-BDNF-UTR1-GFP-T500 (wt and mutations M1-M4) using the methylation dependent restriction endonuclease HhaI. “+” refers to DNA plasmids previously incubated for three hours with the methyltransferase M.SssI which methylates all CpG sites of the DNA. “-” refers to DNA that has not been incubated with M.SssI. The first and the last lane correspond to a DNA ladder. The size of the three major fluorescent bands are given (number of bases).



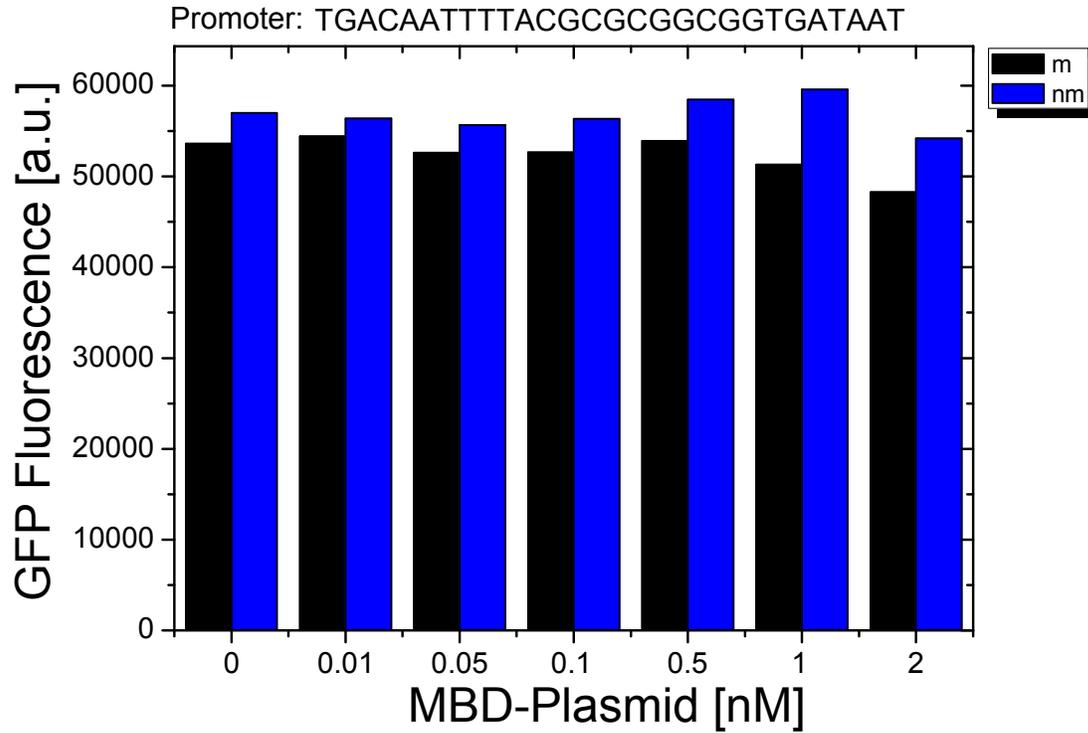
**Figure S2.** Degradation of CpG methylated DNA in the cell free reaction. The graph shows the GFP expression level using 5 nM of a CpG methylated (black) and unmethylated (blue) mcr positive reporter plasmid. Both reporters are expressed with the standard BL21 cell free extract (mcr<sup>+</sup>) and a NEB-10-beta extract (mcr<sup>-</sup>) not containing the mcr restriction system. Data is normalized to the mcr<sup>+</sup>/CpG<sup>-</sup> case. The concentration of amino acids, magnesium glutamate and potassium glutamate added to the reaction is 0.5 mM, 3 mM, and 50 mM, respectively. CpG methylated DNA is degraded in the mcr<sup>+</sup> extract.



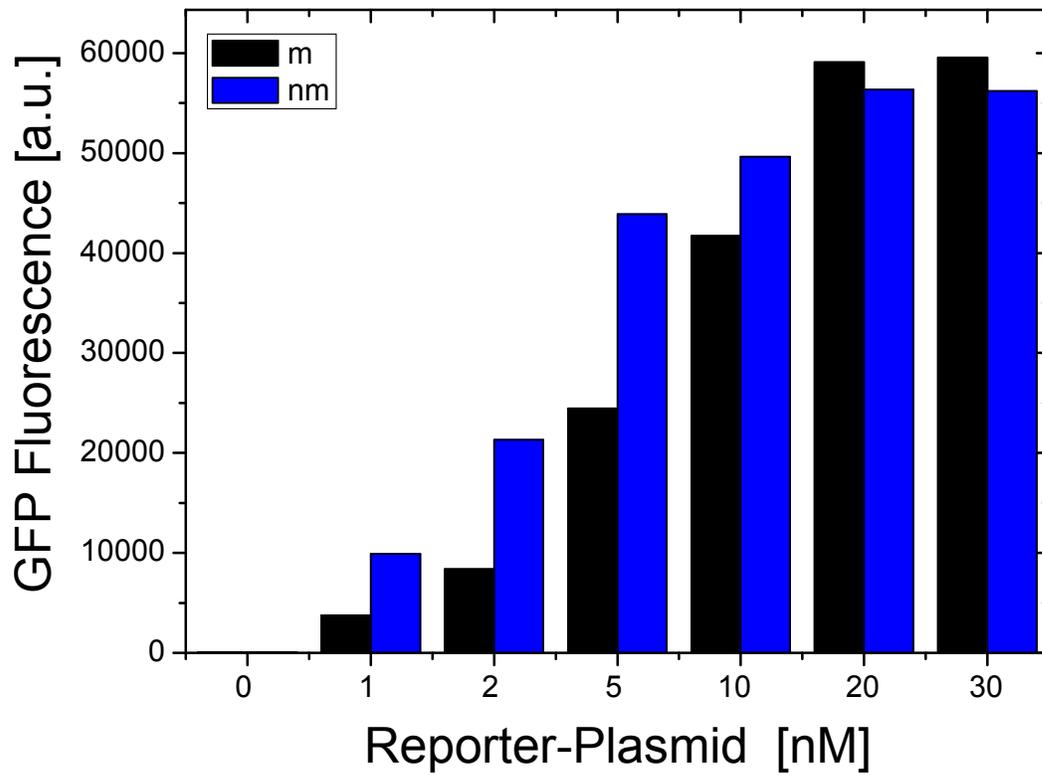
**Figure S3.** Comparison of GFP expression in the case of the plasmid pBEST-BDNF-wt-UTR1-GFP-T500 for either recombinant MBD protein fragment added to the reaction at a given concentration, or instead MBD expressed from plasmid DNA at a given concentration within the reaction. All data is normalized to the respective amount of GFP expressed in absence of MBD. left: methylated case, right non-methylated. Note the differences in scales.



**Figure S4.** A schematic representation for the collective variable of handedness [1]. On the right-hand side, the figure shows the vectors involved in this definition. P atoms are named in context of the residue number and the strand, with  $P_n^1, P_n^2$  atoms representing the  $n^{\text{th}}$  basepair P atoms; with 1 and 2 representing the Watson and Crick strands, running in the  $5' \rightarrow 3'$  direction, and the  $3' \rightarrow 5'$  direction, respectively. For the handedness term including the four atoms  $P_n^1, P_n^2, P_{n+1}^1, P_{n+1}^2$ , represented in the (ABCD) order here as well, the vectors connecting the atoms and contributing to handedness are defined on the right-hand side, with the blue vector connecting the first two atoms and the green vector connecting the second two atoms. The red vector defines the connection between the two midpoints connecting either vector, in the given direction. On the left-hand side, the definition of handedness is given, in terms of the units vectors of the vectors defined on the right-hand scheme. Vector multiplication, followed by a dot product is assumed, and the final handedness term retrieved. For more details about the definition of the global handedness term in terms of a long run of bases, please refer to the text.

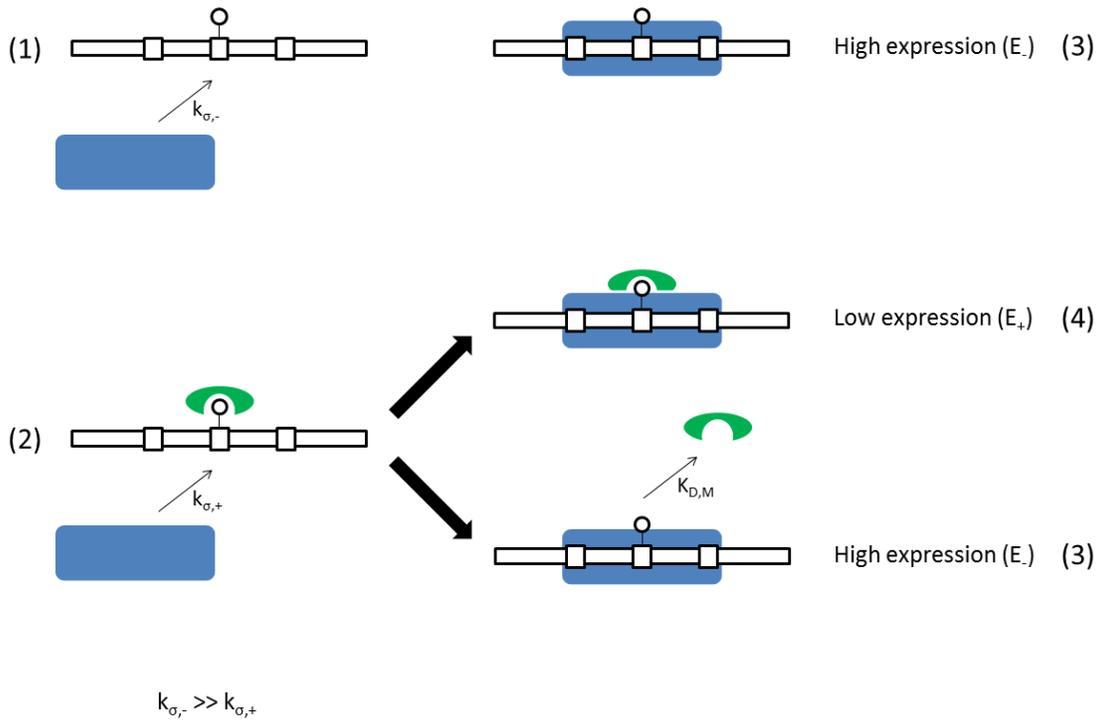


**Figure S5.** MBD mediated repression of GFP with a different promoter that contains a repetition of the CpG recognition motif. The promoter sequence is given in the figure. The graph shows GFP fluorescence as a function of MBD plasmid concentration for the CpG methylated promoter (black) and the non-methylated one (blue). GFP expression is not modulated in a methylation dependent manner.

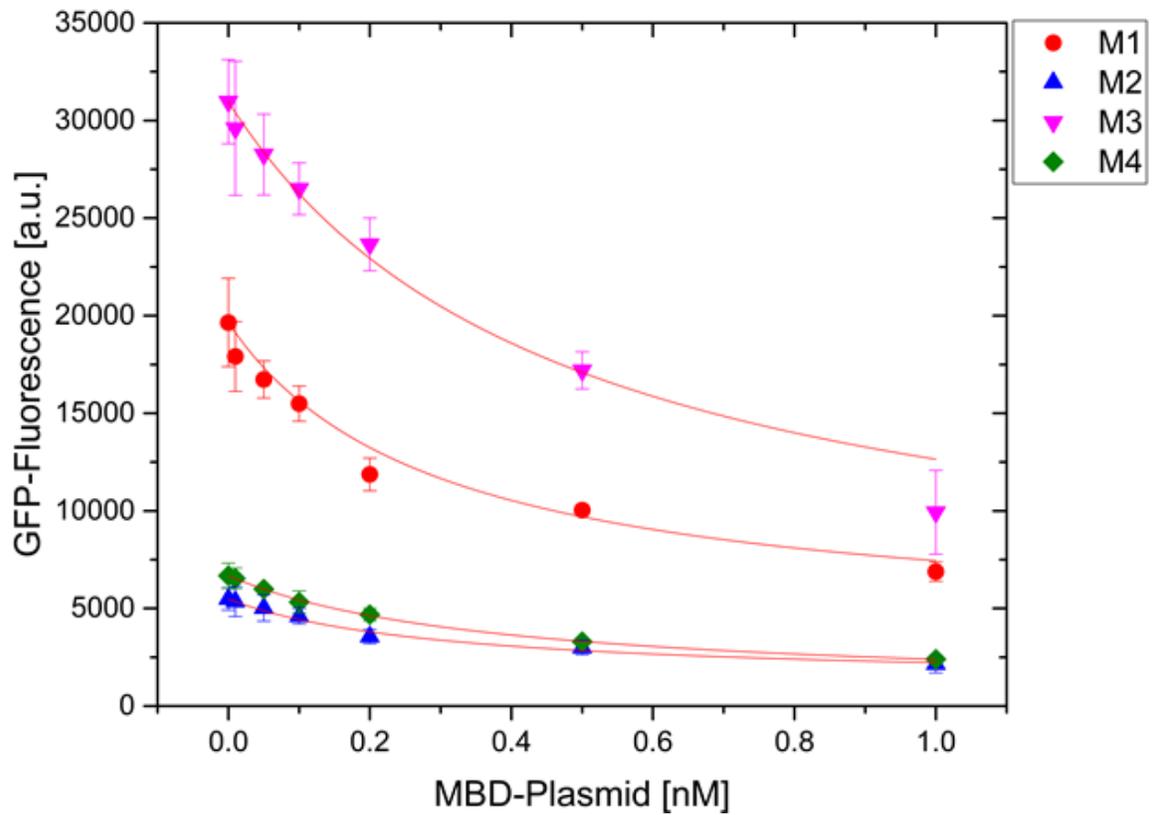


**Figure S6.** Concentration of recombinant GFP as a function of the concentration of the reporter plasmid pBEST-BDNF-wt-UTR1-GFP-T500 (m: CpG methylated (black); nm: unmethylated, blue). The concentration of amino acids, magnesium glutamate and potassium glutamate added to the reaction is 0.5 mM, 3 mM, and 50 mM, respectively. Expression levels of CpG+ and CpG- plasmids are comparable within a standard error of about 10 %. CpG methylated plasmids are not subject to degradation.

### Binding of $\sigma^{70}$ to state (1) or (2)



**Figure S7.** Transition from state [(1) or (2)] to [(3) or (4)], respectively. There are different association constants of the sigma factor 70: either for binding to the free promoter ( $k_{\sigma,-}$ ), or for binding to the promoter with bound MBD ( $k_{\sigma,+}$ ). The difference stems from structural hindrance mediated by MBD ( $k_{\sigma,-} \gg k_{\sigma,+}$ ). If bound to the promoter, the sigma factor 70 can facilitate the dissociation of MBD, resulting in an increased dissociation constant ( $K_{D,M}$ ). At the same time the sigma factor will initiate the expression. Because of this, the transition from (3) to (4) can be neglected.



**Figure S8.** Amount of expressed GFP as a function of MBD-Plasmid concentration in conjunction with the methylated mutants of the BDNF promoter (M1-M4). The red lines represent a fit using equation (1). For M3 the fitting parameter for  $E_+$  became negative, so we set it to a fixed value comparable to wt and the other mutants (see table S9). This results in a curve that fits the data about equally well.

**Table S9.** Fitting parameters of the plots in figure 6 and figure S8. The values where no error is listed were fixed. E<sub>-</sub> was fixed to the measured value E([MBD-Plasmid] = 0 nM). E<sub>+</sub> of M3 had to be fixed to a value that gives a K<sub>D,M</sub> and a ratio of E<sub>-</sub>/E<sub>+</sub> comparable to wt, M1, M2 and M4 to avoid negative parameters.

	wt	M1	M2	M3	M4
E <sub>-</sub> [a.u.]	27033.33	19649	5483.33	30968	6680.67
E <sub>+</sub> [a.u.]	3656.23 ± 650.58	3879.71 ± 1490.62	1225.65 ± 395.66	4000	779.45 ± 72.74
E <sub>-</sub> /E <sub>+</sub>	7.39	5.06	4.47	7.74	8.57
K <sub>D,M</sub> [nM]	0.555 ± 0.038	0.292 ± 0.079	0.306 ± 0.069	0.472 ± 0.029	0.376 ± 0.012

#### References

1. Moradi, M., Babin, V., Roland, C. and Sagui, C. (2013) Reaction path ensemble of the B-Z-DNA transition: a comprehensive atomistic study. *Nucleic Acids Research*, **41**, 33-43.