

# Recruitment of MBD1 to target genes requires sequence-specific interaction of the MBD domain with methylated DNA

Thomas Clouaire, Jose Ignacio de las Heras, Cara Merusi and Irina Stancheva\*

Wellcome Trust Centre for Cell Biology, University of Edinburgh, Michael Swann Building, Mayfield Road, Edinburgh EH9 3JR, UK

Received January 21, 2010; Revised and Accepted March 18, 2010

## ABSTRACT

**MBD1, a member of the methyl-CpG-binding domain family of proteins, has been reported to repress transcription of methylated and unmethylated promoters. As some MBD1 isoforms contain two DNA-binding domains—an MBD, which recognizes methylated DNA; and a CXXC3 zinc finger, which binds unmethylated CpG—it is unclear whether these two domains function independently of each other or if they cooperate in facilitating recruitment of MBD1 to particular genomic loci. In this report we investigate DNA-binding specificity of MBD and CXXC3 domains *in vitro* and *in vivo*. We find that the methyl-CpG-binding domain of MBD1 binds more efficiently to methylated DNA within a specific sequence context. We identify genes that are targeted by MBD1 in human cells and demonstrate that a functional MBD domain is necessary and sufficient for recruitment of MBD1 to specific sites at these loci, while DNA binding by the CXXC3 motif is largely dispensable. In summary, the binding preferences of MBD1, although dependent upon the presence of methylated DNA, are clearly distinct from those of other methyl-CpG-binding proteins, MBD2 and MeCP2.**

## INTRODUCTION

DNA methylation at CpG dinucleotides is an abundant modification in vertebrate and plant genomes (1,2). Generally, DNA methylation associates with formation of heterochromatin in the genome and, when detected near transcription start sites (TSS) of genes, leads to stable transcriptional silencing (1,3). Although ~62% of

human gene promoters are CpG-rich and are usually free of DNA methylation, a fraction of these is methylated in differentiated tissues and a large number of promoters can be aberrantly methylated in human cancers (4–7). In addition, a significant proportion of CpG-poor and intermediate CpG-density promoters, which account for ~38% of human protein coding genes, are usually methylated in normal human somatic cells (7). Expression analyses of mouse and human cells deficient for the maintenance DNA methyltransferase enzyme Dnmt1 have identified a large number of misexpressed transcripts, suggesting that DNA methylation is essential for the maintenance of a transcriptionally inactive state of many genes (8,9).

Transcriptional silencing by DNA methylation operates in part via proteins that bind to methylated DNA and recruit co-repressor complexes containing histone deacetylases and histone methylase activities (10). Three families of proteins that bind to methylated DNA have been identified so far. These include: the MBD domain family; Kaiso and Kaiso-like proteins and the SRA domain proteins (11). The MBD family consists of MBD1, MBD2, MBD3, MBD4 and MeCP2 (12). Three of these proteins, MBD1, MBD2 and MeCP2, function as methylation-dependent transcriptional repressors (10,13–15). Mice null for these three proteins are viable and, with the exception of *Mecp2*-deficient animals, display relatively mild, but distinct, phenotypes (16–20). This is in stark contrast to mice deficient in DNA methyltransferase enzymes, which die early in development (21,22). As MBD proteins are ubiquitously expressed in all somatic tissues, although at varying levels, functional redundancy between MBD family members and, perhaps, with other methyl-CpG-binding proteins has been the most common explanation for the mild phenotypes of *Mbd1* and *Mbd2* null animals.

This hypothesis is supported in part by observations that in human cancer cell lines several MBD proteins

\*To whom correspondence should be addressed. Tel: +44 131 650 3728; Fax: +44 131 650 3714; Email: istancheva@ed.ac.uk

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

can be detected at the same methylated promoter (23,24). However, even in cancer cells a proportion of methylated promoters are occupied by a single MBD protein (23,25). Chromatin immunoprecipitation (ChIP)-and-clone experiments have shown that shared occupancy of methylated sites by several MBD proteins is rarely observed in primary human lung fibroblasts (26). Moreover, MeCP2 was unable to colonize methylated sites vacated after MBD2 knock down, while MBD2 could migrate into about half of the binding sites generated by knocking down MeCP2 (26). These experiments indicate that MeCP2 has binding preferences distinct from those of MBD2 and subsequently it was found that high affinity binding of MeCP2 to methylated DNA requires a run of four or more A/T bases adjacent to a methylated CpG (26). Whether MBD1 can recognize methylated DNA in some preferred sequence context is currently unknown.

MBD1 is the largest protein of the MBD family. It cooperates with a histone H3K9 methylase SETDB1 and its cofactor AM/MCAF to repress transcription (27,28). In addition to an MBD domain, MBD1 contains CXXC-type zinc fingers, and a transcriptional repression domain (TRD) located at the C-terminus (15,29). Several MBD1 isoforms with either two or three CXXC motifs have been identified in human and mouse cells (30,31). The third CXXC motif, CXXC3, is present in three of the five isoforms of human MBD1 and is highly homologous to the cysteine-rich zinc fingers of histone H3K4 methylase MLL, DNA-binding protein CGBP and maintenance DNA methyltransferase DNMT1 (29,31,32). The CXXC3 domain of mouse Mbd1 can bind unmethylated DNA *in vitro* and localizes to pericentric heterochromatin when expressed in cells that lack Dnmt1 (31). NMR and band-shift studies indicate that CXXC of MLL binds to a single CpG pair via amino acids located in an extended loop formed within a crescent-like structure stabilized by eight cysteine residues coordinating two zinc atoms (32). The other two CXXC motifs of MBD1, although similar to CXXC3, differ significantly in key amino acids within the positively charged DNA-binding loop and do not bind to DNA *in vitro* (31). Consistent with the presence of two DNA-binding domains, it has been reported that MBD1 isoforms carrying CXXC3 repress transcription from methylated and unmethylated promoters in transient reporter assays (31,33). However, it remains unclear whether cooperation between the MBD and CXXC3 domains of MBD1 is essential for recruitment of MBD1 to specific genomic loci *in vivo*.

In this report we investigate in detail the DNA-binding preferences and relative contribution of MBD and CXXC3 domains to stable binding of MBD1 to DNA *in vitro* and *in vivo*. We show that the MBD domain of MBD1 recognizes methylated CpGs within TCGCA and TGCGCA sequence context more efficiently than any other sequence we tested, while the CXXC3 motif requires a single unmethylated CpG and displays no detectable preference for particular flanking nucleotides. Using a transactivator fusion approach we identify specific target genes for MBD1 in several human cell lines and find that they often contain unmethylated CpG-rich promoters and methylated high affinity

MBD1-binding sites located in the vicinity of the TSS. We also demonstrate that stable binding of MBD1 to DNA *in vivo*, including at MBD1 targeted loci, requires a functional MBD domain but not DNA binding by CXXC3. Point mutations in CXXC3 that completely abolish binding to unmethylated DNA *in vitro* do not disrupt the recruitment of MBD1 to target genes. Taken together, this indicates that MBD1 functions primarily as a methyl-CpG-binding protein with a preference for specific methylated sites. Our data also suggest that MBD family proteins have evolved towards more specific recognition of methylated DNA by their MBD domains leading to silencing of a restricted subset of target genes by each of these proteins.

## MATERIALS AND METHODS

### Recombinant proteins

For recombinant protein expression and purification, all MBD1 fragments were cloned into EcoRI and XhoI sites of pGEX-4T-1. MBD1 deletions corresponding to residues 1–161 or 1–125 were used for methylated DNA-binding assays. For CXXC3 DNA-binding experiments, an MBD1 fragment corresponding to amino acids 252–344, was used. The GST-CXXC3-His construct was generated by ligating a double-stranded oligonucleotide encoding an hexahistidine tag into XhoI and NotI sites in the PGEX-4T-1 plasmid. pGEX-MeCP2 1–162 has been described (34). All recombinant proteins were produced in *Escherichia coli* BL21 strain according to standard procedures. GST fusion proteins were purified using glutathione sepharose (GE Healthcare), eluted with reduced glutathione and buffer was exchanged to 50 mM Tris-HCl pH 7.5, 10% glycerol, 150 mM NaCl, 1 mM DTT using NAP-10 columns (GE Healthcare). Purified proteins were concentrated using Millipore filter device and stored at  $-80^{\circ}\text{C}$ . The GST-CXXC3 fusion protein was cleaved with thrombin (GE Healthcare) prior to use in bandshift assays.

### SELEX

Methyl-SELEX was performed essentially as described (26). The following sequence: 5'ACCAGGAAGCTTCTGTATGTAGATCTG-N(16)-GCGC-N(16)-GAGATCTCCTAAGACTTCTAGATCCC3' was used to generate the starting methyl SELEX library by annealing a reverse primer 5'GGGATCTAGAAGTCTTAGGAGATCTC3' followed by extension by Klenow polymerase. The gel purified double-stranded DNA fragments were methylated with M.HhaI methyltransferase, end labelled with [ $\gamma$ - $^{32}\text{P}$ ] dATP by T4 polynucleotide kinase (PNK) and used in an electrophoretic mobility shift assay (EMSA) reaction with increasing amount of GST-MBD1 1–161 (10, 50 and 150 nM) in the presence of 1  $\mu\text{g}$  of poly(dG-dC)•poly(dG-dC) (GE Healthcare). Shifted bands were cut out, boiled for 10 min and the eluted material was used for PCR amplification using the following primers: F 5'ACCAGGAAGCTTCTGTATGTAGATCTG3' and R 5'GGGATCTAGAAGTCTTAGGAGATCTC3'. The resulting PCR product was

used for the next round of selection. SELEX for the CXXC3 domain was performed using the following oligonucleotide: ACCAGGAAGCTTGTTCCTCCAGTCACT ACA-N(5)-CG-N(5)-AGTCATAGCTGG TTCCTGCC TAAGACTTCTAGATCCC. GST-CXXC3-His fusion protein was cleaved with thrombin and immobilized on magnetic Ni<sup>2+</sup>-Beads (Dynal). SELEX assay was the performed as described (35). After the last round of selection, PCR products were digested with HindIII and XbaI then cloned in pBluescript for sequencing.

### EMSA

Oligonucleotides used for EMSA were synthesized in their methylated or unmethylated forms, and correspond to the sequence ACATGCCTCATGCCGACTTAACTGCA GCT with the indicated substitutions. Complementary probes were annealed, purified on polyacrylamide gels and end labelled with [ $\gamma$ -<sup>32</sup>P] dATP by T4 PNK. Purified proteins were mixed at the indicated concentrations with radiolabelled probes in EMSA-binding buffer (20 mM Hepes pH 7.9, 150 mM KCL, 5% glycerol, 0.1% Triton X-100, 0.2 mM EDTA, 2 mM DTT, 100  $\mu$ g/ml BSA including 250 ng poly(dA-dT)•poly(dA-dT) or 150 nG poly(dI-dC)•poly(dI-dC)). After 10 min of incubation at room temperature, the reaction mixtures were loaded onto 6 or 8% polyacrylamide –1×Tris–glycine–EDTA gels and run for 2 h, 180 V at 4°C. Gels were dried exposed to a PhosphorImager screen (Molecular Dynamics) or autoradiographed. To determine the relative  $K_D$  of MBD binding to different probes, poly(dA-dT)•poly(dA-dT) was reduced to 100 ng. Radioactivity was quantified by PhosphorImager from at least three independent experiments and the bound fraction (bound DNA/[bound DNA + free DNA]) was calculated for each protein concentration. Binding curves were fitted using Sigma Plot Systat Software.

### Plasmids, cells and transient transfections

To generate MBD-VP16 expression plasmids sequences containing amino acids 1–314 of MBD1 (PCM1 variant), 1–205 of MeCP2 and 1–214 of MBD2 were PCR amplified and cloned into NotI and EcoRI sites of pCMV-Tag4 (Promega). VP16 activation domain was PCR amplified from pTET-ON plasmid (Clontech) and cloned into EcoRI and XhoI sites between and in frame with MBD and C-terminal FLAG tag. MBD1 point mutations were introduced into MBD1<sup>PCM1</sup>-VP16 by mutagenic PCR. MBD1 shRNA and non-silencing control plasmids were purchased from SABioscience. MBD2 shRNA and non-silencing control were purchased from Open Biosystems. The plasmids were transfected into HeLa, NCI-H226, HCT116, HCT116 DNMT3B KO and DNMT1/DNMT3B DKO cells by electroporation using Nucleofection device and transfection reagents (Amaxa Biosystems). In all large scale experiments 5  $\mu$ g of plasmid DNA was used for  $1.3 \times 10^6$  cells.

### Western blots

Cells transfected with plasmids expressing wild-type or mutated MBD-VP16 fusions were collected 48 h

post-transfection. Two-third of each sample was used for RNA extraction and one-third was used to prepare nuclear extracts using modified Dignam protocol (27). Nuclear extracts (25–40  $\mu$ g) were run on 10% SDS-PAGE gels and the fusion proteins and HDAC1 control protein were detected by anti-FLAG M2 (Sigma) and anti-HDAC1 sc-7872 (Santa Cruz Biotechnology) antibodies followed by secondary IRDye 800CW donkey anti-mouse, IRDye 680 donkey anti-rabbit antibodies (LiCor Biosciences). Images were acquired on LiCor Odyssey Infrared Imager (LiCor Biosciences) and quantified by Odyssey V3.0 Software.

### RNA extraction, semi-quantitative and RT-qPCR

RNA was extracted using Trizol reagent (Invitrogen) according to manufacturer's instructions. cDNA was prepared from 4  $\mu$ g of total RNA using poly-dT primer and SuperScriptII reverse transcriptase (Invitrogen) according to standard protocols. For semi-quantitative RT-PCR 60-, 30-, 15- and 7.5-fold dilutions of each cDNA sample was amplified with control primers for gamma-Actin and primers specific for HBA1, HBA2, NGFR, RND2 and ASPP2. RT-qPCR reactions (25  $\mu$ l) were performed in quadruplicate (iCycler, Bio-Rad) using 2X SYBR Green Supermix (Bio-Rad Laboratories) and quantified by Bio-Rad iCycler iQ system software. Normalization and analysis for each target gene were carried out using ACTB as a reference gene, according to standard methods using the following equation:  $\text{Ratio} = \frac{(E_{\text{target}})^{\Delta C_t}}{(E_{\text{ref}})^{\Delta C_t}} \frac{(\text{control-target})}{(\text{control-sample})}$  (36). Primer sequences are available on request.

### Microarray experiments and data analyses

Each experiment was carried out in three biological replicas on human spotted cDNA arrays (CRUK22K, Cancer Research UK), with cDNA from untransfected cells and from cells transfected with either pCMV-MBD1-VP16-Tag4 or the double mutant R22A/2CA control. Each experiment included at least one dye swap. The cDNA was directly labelled with either [Cy3]dCTP or [Cy5]dCTP (GE Healthcare, UK) by reverse transcription of 80–100  $\mu$ g of total RNA. Reverse transcription was performed overnight according to standard protocols, using an anchored poly(dT) oligo primer and SuperScriptII reverse transcriptase (Invitrogen). After labelling, RNA was hydrolysed at 65°C for 20' in basic conditions, and the labelled cDNA was purified through a QiaQuick spin column (Qiagen). The incorporation of labelled nucleotides into cDNA was assessed with a Nanodrop 1000 spectrophotometer (Thermo Scientific) at 550 and 650 nm wave length, respectively. Hybridizations were performed at 45°C for ~20 h in hybridization buffer containing 5× SSC, 6× Denhart's solution, 60 mM Tris-HCl pH 8 and 48% deionized formamide. After hybridization and subsequent washing, the arrays were scanned and quantified using a Genepix 4200AL scanner (Molecular Devices). Raw data were background corrected using the 'Normexp' algorithm, loess normalized using the Limma package (37)

from the BioConductor project and analysed in R environment, using Limma and standard R/BioConductor tools (38). To identify possible targets for the MBD1-VP16 fusion protein, we subtracted the genes activated by the double mutant control from the ones activated in the MBD1-VP16 sample, and kept those with values  $\log_2(\text{transfected/untransfected}) > 1.5$  with a 0.05 false discovery rate (FDR) cut off. Microarray data are submitted to ArrayExpress, accession number: E-MTAB-103.

### Bisulphite DNA sequencing

Genomic DNA was phenol–chloroform extracted from cultured cells after RNase A and proteinase K digestion. Bisulphite treatment of genomic DNA was carried out as described (39), and prepared for sequencing as outlined in ref. (40). Genomic DNA (2 µg) was treated with sodium bisulphite and precipitated after the desulphonation step. The samples were resuspended in 1×Tris–EDTA buffer for subsequent PCR and sequencing reactions. Colony PCRs were performed, clones of the correct size were sequenced and sequences were analysed using BiQ Analyser (41). A list of primers used for bisulphite sequencing of HBA, NGFR and RND2 promoters can be found in the Supplementary Data.

### ChIP

ChIP experiments were performed essentially as described (42). Typically chromatin from  $3 \times 10^6$  cells and 2–5 µg of antibody were used for each IP. The antibodies used were anti-VP16AD, sc-7546 (Santa Cruz Biotechnology), anti-MBD1 sc-10751 (Santa Cruz Biotechnology), anti-H3 K9/K14ac 06–599 (Millipore). 1/100 of DNA purified from ChIP was used in each 20 µl qPCR reaction. All qPCRs were performed in triplicates on three independent ChIP samples. Primer sequences are available on request.

## RESULTS

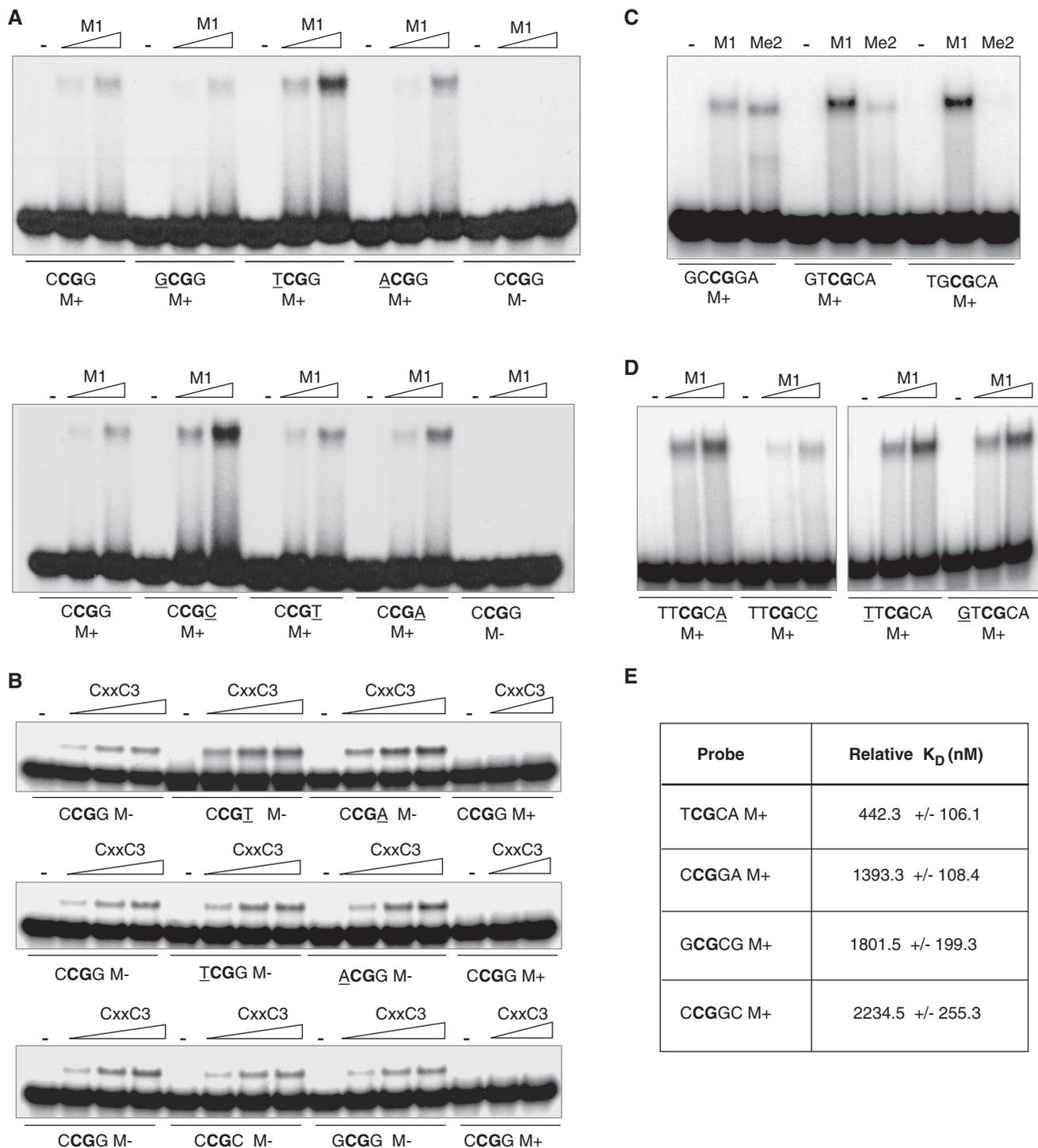
### Recognition of methylated and unmethylated CpGs by DNA-binding domains of MBD1

To investigate in detail DNA-binding properties of MBD1, we generated recombinant GST-tagged proteins that correspond to the MBD and the CXXC3 domains, respectively, (Supplementary Figure S1) and used them in EMSA experiments. We first investigated whether the efficiency of complex formation of these domains with double-stranded DNA probes is affected by the bases immediately adjacent to the CpG dinucleotide. To do so, we performed base substitution scanning mutagenesis at positions –1 and +1 relative to the methylated or the unmethylated CpG. Interestingly, when we tested the MBD domain of MBD1 against probes containing C, G, T or A nucleotides at positions either –1 or +1 relative to methylated CpG, we observed a marked preference of MBD domain for T at position –1 and C at position +1 (Figure 1A). When these two flanking bases were combined into a probe containing  $\underline{\text{TC}}^{\text{M}}\underline{\text{GC}}$  sequence, we detected an additive effect on MBD binding as evident by

the increased band-shift efficiency compared to probes containing either  $\underline{\text{TC}}^{\text{M}}\underline{\text{GG}}$  or  $\underline{\text{CC}}^{\text{M}}\underline{\text{GC}}$  (Supplementary Figure S2A). In similar experiments with unmethylated DNA, the CXXC3 domain of MBD1 bound all probes with approximately equal efficiency (Figure 1B). Thus the base pairs flanking methylated CpG affect significantly the affinity of the MBD domain for methylated DNA, but have little if any effect on binding efficiency of the CXXC3 zinc finger to unmethylated CpG.

In parallel with base substitution mutagenesis, we also undertook *in vitro* binding site selection (SELEX) using pools of DNA fragments containing methylated (methyl-SELEX) (26) or unmethylated CpGs to identify potential additional sequence requirements for the MBD and CXXC3 domains, respectively. The sequences recovered from the SELEX with CXXC3 enriched only for additional CpG pairs, further indicating that this domain has no extended recognition sequence (data not shown). In contrast, the methyl-SELEX assay performed with the MBD domain of MBD1 on a pool of double-stranded fragments containing a fixed GCGC sequence methylated with HhaI methylase produced a different outcome. After eight cycles of binding and amplification, the fragments recovered from the EMSA gels were cloned and sequenced. Interestingly, 82 out of 87 fragments contained one or more (97 in total) palindromic TGCGCA sequences (Supplementary Figure S3). In addition to TGCGCA sites, we also observed an expansion of GCGC sequences in the selected pool (Supplementary Figure S3). Thus the MBD domain of MBD1 was able to select simultaneously for a unique TGCGCA site containing a single methylated CG as well as a multiple methylated CGs within the extended GCGC sequence. As most methylated CpGs in the genome are not located within GC-rich stretches of DNA, we further focused on MBD selected sequences containing a single CG.

A subset of fragments derived from the methyl-SELEX assay with the MBD of MBD1 was further tested in independent band-shift experiments. Probes containing the methyl-SELEX-enriched  $\text{TGC}^{\text{M}}\text{GCA}$  motif were as efficient in supporting complex formation with the MBD domain of MBD1 as probes with  $\text{TC}^{\text{M}}\text{GC}$  derived from base substitution mutagenesis experiments (Figure 1C). As the GCGC core sequence was imposed by the use of HhaI methylase, it appeared that the only two selected bases in the enriched  $\text{TGC}^{\text{M}}\text{GCA}$  motif were T and A positioned at –2 and +2, respectively, relative to methylated CpG. Combining the information derived from base substitution scanning mutagenesis and the methyl-SELEX, we predicted and verified experimentally that  $\text{TTC}^{\text{M}}\text{GCA}$  constitutes a high affinity binding site for the MBD of MBD1 (Figure 1D). However, T at –2 and A at +2 did not contribute equally, as the MBD domain of MBD1 displayed significantly reduced binding when A at +2 was substituted to C, but not when T at –2 was replaced with G (Figure 1D). We conclude that the identified palindrome sequence  $\text{TGC}^{\text{M}}\text{GCA}$  reflects the selection of A at position +2 relative to methylated CpG. As this sequence can be read by the MBD domain on both DNA strands, this may explain why it was



**Figure 1.** Binding specificity of MBD and CXXC3 domains of MBD1 *in vitro*. (A) EMSA combined with base substitution scanning mutagenesis of nucleotides adjacent to methylated CpG detect preferential binding of MBD domain to probes with T in position -1 and C in position +1. Lanes labelled with '-' contain no protein. The triangle indicates increasing concentrations of MBD (75 and 200 nM). All probes used in these experiments happen to contain an A at position +2 from the CG. Therefore MBD1 binds seemingly identical probes such as T<sup>M</sup>CGG and C<sup>M</sup>CGA or G<sup>M</sup>CGG and C<sup>M</sup>CGC with different efficiency. (B) The CXXC3 domain binds with similar efficiency to all unmethylated probes containing a CG pair independently of the sequence context, but does not bind to methylated probes. Lanes labelled with '-' contain no protein. The triangle indicates increasing amounts of CXXC3 (50–200 nM). (C) The MBD domain of MBD1 (M1) but not the MBD domain of MeCP2 (Me2) efficiently binds TG<sup>M</sup>CGCA sequence enriched in methyl-SELEX experiment and probe containing T<sup>M</sup>CGC from base substitution scanning mutagenesis with A at position +2. 200 nM of M1 and Me2 were used in these EMSA experiments. (D) Replacement of A to C at position +2 in T<sup>M</sup>CGCA sequence reduces the affinity of MBD1 MBD domain to methylated probe. Replacement of T at position -2 with G does not affect the efficiency of binding. (E) Relative  $K_D$  of MBD1 MBD binding to probes containing the optimal (T<sup>M</sup>CGCA) or suboptimal methylated binding sites. Note that the MBD domain binds T<sup>M</sup>CGCA probe 3–5-fold more efficiently than any other methylated sequence.

efficiently enriched in the methyl-SELEX assays. Therefore, the shortest sequence containing a single methylated CG which supports high affinity binding of the MBD domain of MBD1 is TC<sup>M</sup>GCA followed by TGC<sup>M</sup>GCA. Overall, the affinity of MBD domain to methylated DNA was 3–5-fold higher when probes containing TC<sup>M</sup>GCA were compared to suboptimal sequences (Figure 1E and Supplementary Figure S4). Most importantly, probes that supported high affinity binding of the MBD domain of MBD1 formed weaker complexes with the MBD domain of MeCP2, irrespective of whether they contained an A/T run (Figure 1C and Supplementary Figure S2B). This indicates that these two MBD domains recognize methylated DNA within specific sequence contexts differently from each other.

In summary, from these experiments we conclude that the MBD domain of MBD1 binds more efficiently to a single methylated CpG within TC<sup>M</sup>GCA and TGC<sup>M</sup>GCA sequence context while the CXXC3 domain binds unmethylated CpG and shows no apparent preference towards particular base pairs flanking the CpG dinucleotide.

#### Point mutations in MBD and CXXC3 domains of MBD1 abolish DNA binding

Potentially, the two DNA-binding domains of MBD1 could either function independently of each other or cooperate in the recruitment of MBD1 to specific genomic loci. To investigate this, we introduced point mutations in the MBD or CXXC3 domains of MBD1 (Supplementary Figure S1) and tested *in vitro* whether they abolish binding to methylated or unmethylated DNA, respectively. The solution structure of the MBD domain of MBD1 has been solved by NMR and shown to consist of four beta-sheets, an alpha-helix and an extended loop L1 (Figure 2A) (43). Several residues, including arginine 22 (R22) at the base of L1 are involved in binding to symmetrically methylated CpGs. As shown previously (31,43) and confirmed by our band-shift experiments, a replacement of R22 with alanine (R22A) almost completely abolished binding of the MBD domain to methylated DNA (Figure 2B). To disrupt binding of the CXXC3 to DNA, we mutated either both zinc coordinating cysteines (C289,292A), which should perturb the folding of CXXC3 domain (32), or the conserved lysine residues (K310A, K312A and K319A) of the DNA-binding loop (Figure 2A). None of the mutant CXXC3 peptides could bind efficiently to DNA *in vitro* (Figure 2C and D). Based on these experiments we designed mutant MBD1 proteins to investigate the contribution of MBD and CXXC domains in the recruitment of MBD1 to specific loci *in vivo*.

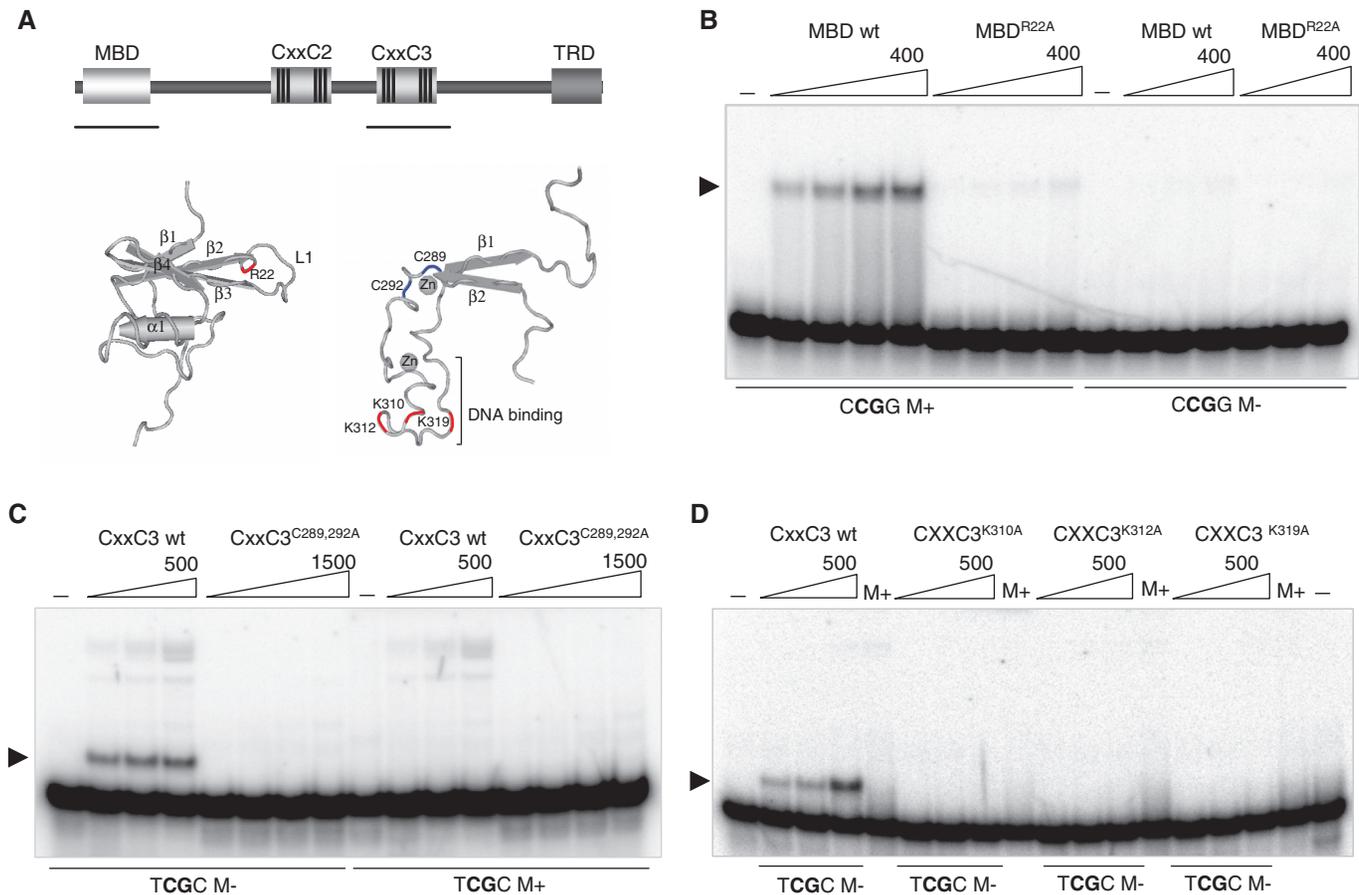
#### MBD1-VP16 transactivator fusion targets specific genes in HeLa cells

As a strategy to identify genes that are normally silenced by MBD1, we transfected HeLa cells with plasmid expressing MBD1-VP16 protein which contains amino acids 1–341 of MBD1 fused to the strong transactivation domain of the herpes virus protein VP16 (VP16AD)

followed by a C-terminal FLAG-tag (Figure 3A). Additionally, we transfected cells with plasmids expressing amino acids 1–205 of MeCP2 or amino acids 1–214 of MBD2 also fused to VP16AD-FLAG (Figure 3A). To control for non-specific effects resulting from introducing VP16AD into the cells, we also expressed at comparable levels a double mutant construct MBD1<sup>R22A/C289,292A</sup>-VP16 (DM-VP16) carrying point mutations in the MBD and CXXC3 DNA-binding domains. All fusion proteins expressed equally well when transiently transfected into HeLa cells or other human cell lines (Figure 3B and C, Supplementary Figure S6 and data not shown). We expected the MBD-VP16 fusion proteins to bind at their endogenous genomic locations and, if such binding occurs near or at a gene promoter, to activate the adjacent gene via recruitment of SAGA histone acetyltransferase (HAT) complex by the VP16AD (44). Thus we predicted that genes normally silenced by a particular MBD protein will be activated in this assay.

In order to identify such genes, we extracted RNA from cells transiently expressing MBD1-VP16 or controls. cDNAs were synthesized, labelled and hybridized to microarrays containing probes for ~10 000 human genes. Comparison of HeLa cells transfected with MBD1-VP16 to controls expressing DM-VP16 identified 34 transcripts that were up-regulated 3–16-fold (Figure 3D, left panel; and Supplementary Table S1). None of these transcripts were significantly induced by the MBD2-VP16 fusion and only four responded weakly to the expression of MeCP2-VP16 (Figure 3D, middle and right panels). Interestingly, most of the transcripts induced by MBD1-VP16 in HeLa represented tissue-specific genes, which are normally not expressed either in HeLa or in the normal cervix tissue. We did not detect mRNAs significantly down-regulated by MBD1-VP16 or any other MBD-VP16 fusion proteins (Figure 3D). Experiments performed with HCT116 and H226 cell lines produced similar results, although most of the transcripts as well as their numbers were different, perhaps reflecting cell line-specific DNA methylation patterns (Supplementary Table S1 and data not shown).

To validate independently the microarray results, we selected three of these genes, HBA, RND2 and NGFR for further analyses. HBA, haemoglobin alpha, is transcribed from a pair of almost identical genes, HBA1 and HBA2, which are located on chromosome 16. The two HBA transcripts differ only in their 3' untranslated regions, encoded by their third exons. The expression of HBA is normally restricted to the cells of the erythroid lineage and is regulated by GATA factors (45,46). RND2 is a small Rho GTPase expressed primarily in brain and testis. RND2 is essential for migration of differentiating neural progenitor cells into the brain cortex (47). NGFR/p75 (NTR) is a neural growth factor receptor normally expressed in neurons and weakly in other tissues. The NGFR and RND2 genes are located in on chromosome 17 (17q21–22 and 17q21, respectively). Additionally, we examined the expression of p53BP2/ASPP2, which was previously identified as a gene silenced by MBD1 in HeLa (27). When tested by semi-quantitative RT-PCR or real time RT-qPCR the transcripts for HBA, RND2,



**Figure 2.** Mutations in MBD and CXXC3 domains disrupt DNA binding. (A) Schematic representation of PCM1 variant of MBD1. The MBD domain, CXXC zinc fingers and transcriptional repression domain (TRD) are indicated. The solution structure of the MBD (43) and CXXC3 domains (32) is shown as well as the position of amino acids, which we have mutated, contributing to DNA binding (red) or coordination of zinc (blue). (B) Replacement of arginine (R) 22 to alanine (A) disrupts binding of MBD to methylated DNA. 50, 100, 200 and 400 nM of MBD were used in EMSA experiments with methylated probes and 100, 200 and 400 nM for EMSA with unmethylated probes. (C and D) Mutations of zinc coordinating cysteines (C289 and C292) to alanine and lysines (K310, K312 and K319) to alanine within the DNA binding part of CXXC3 disrupt binding of CXXC3 to DNA. The triangles represent 5-fold increments in CXXC3 concentration in EMSA experiments. In control lanes indicated with ‘–’ no protein was used. In lanes marked with M+ methylated DNA probes were used in EMSA experiments.

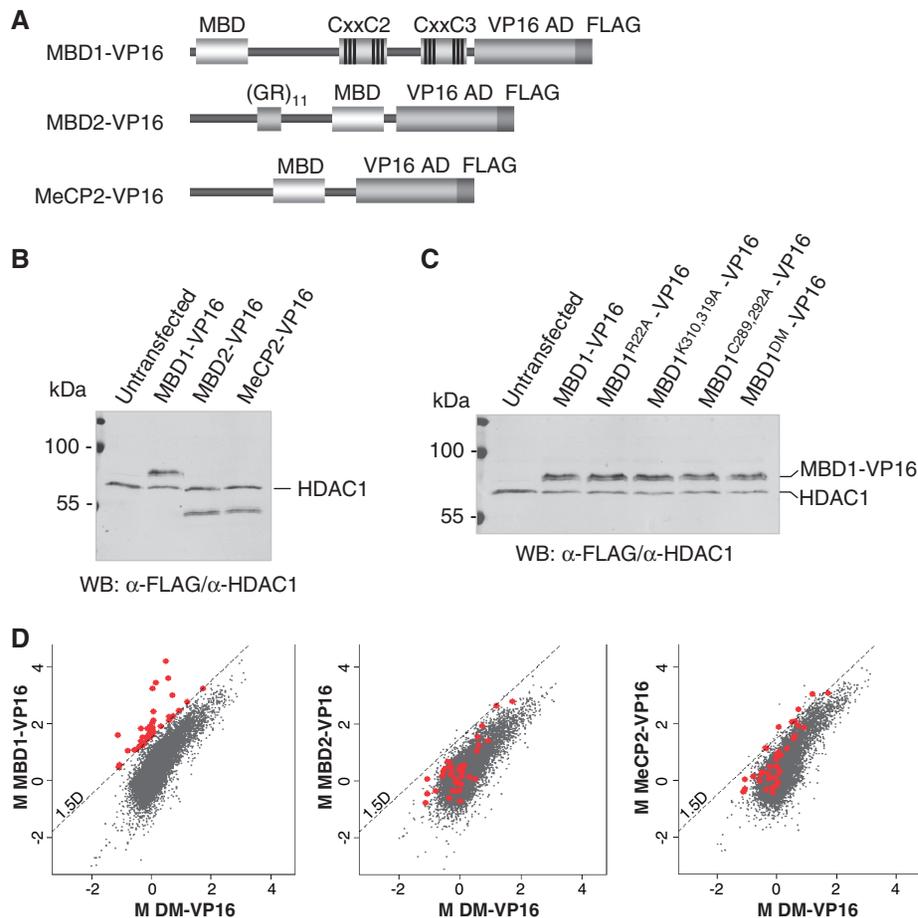
NGFR and p53BP2 were barely detectable in HeLa and in control cells transfected with either MBD2-VP16 or MeCP2-VP16 fusions. However, the same transcripts were highly up-regulated in MBD1-VP16 transfected cells (Figure 4A and B). These experiments indicate that the genes activated by MBD1-VP16 fusion protein are not targeted by other MBD-VP16 fusions and therefore may represent specific targets for MBD1 in HeLa.

#### Contribution of MBD and CXXC3 domains to binding of MBD1 at target genes

To investigate further whether the recruitment of MBD1 to specific loci is guided by DNA-binding preferences of MBD, CXXC3 or cooperative binding to DNA of both domains, we expressed in HeLa cells MBD1-VP16 proteins carrying mutations either in the MBD (R22A-VP16) or in the CXXC3 domain (K310, 319A-VP16; C289A,C292A-VP16) and monitored whether these proteins were capable of activating MBD1-VP16 target genes. All mutant proteins expressed

equally well in HeLa cells after transient transfection of the corresponding plasmids (Figure 3C).

Semi-quantitative or real time RT-qPCR detected a 4–10-fold weaker induction of HBA, RND2 and NGFR by peptides carrying the MBD mutation (R22A-VP16 and DM-VP16) compared to ‘wild-type’ MBD1-VP16 (Figure 4C and D). RT-qPCR analyses of six additional transcripts showed similar lack of induction by MBD1<sup>R22A</sup>-VP16 (Supplementary Figure S7). However, cells expressing MBD1-VP16 with mutations in DNA-binding lysines of CXXC3, K310,319A, activated HBA, NGFR and RND2 at levels comparable to the ‘wild-type’ MBD1-VP16 (Figure 4C and D). Interestingly, mutations in CXXC3 zinc binding cysteine residues, C289,292A, also compromised activation of HBA, RND2 and NGFR, suggesting that a misfolded CXXC domain close to VP16AD may have a general deleterious effect on the overall structure of the fusion protein (Figure 4C and D). Taken together, these experiments suggest that in most cases stable binding to methylated DNA via the MBD domain is essential for the recruitment of MBD1-VP16 to target



**Figure 3.** MBD1-VP16 fusion activates specific genes in HeLa cells. (A) The fusion proteins containing DNA binding domains of MBD1, MBD2 and MeCP2 and transactivation domain of herpes virus (VP16 AD). All proteins have a C-terminal FLAG peptide. (GR)<sub>11</sub> is a region of MBD2 rich in glycine and arginine. (B) Expression of MBD-VP16 fusion proteins in HeLa cells. Histone deacetylase (HDAC1) detected simultaneously with the fusion proteins serves as a loading control. (C) Expression of MBD1-VP16 and MBD1-VP16 proteins carrying point mutations in either the MBD (R22A) or CXXC3 (K310,319A and C289,292A) domains. MBD1<sup>DM</sup>-VP16 combines R22A with C289,292A mutations. HDAC1 serves as a loading control. (D) Log<sub>2</sub> plots show the effect of MBD-VP16 fusion proteins on transcription in HeLa cells. In all graphs the x-axis represents  $M = \log_2$  MBD-VP16/HeLa and the y-axis, the control,  $M = \log_2$  MBD1<sup>DM</sup>-VP16/HeLa. Transcripts up-regulated by MBD1-VP16 3-fold or more ( $\Delta M \geq 1.5$ ) are shown in red. Note that transcripts induced by MBD1-VP16 are not up-regulated after expression of either MBD2-VP16 or MeCP2-VP16.

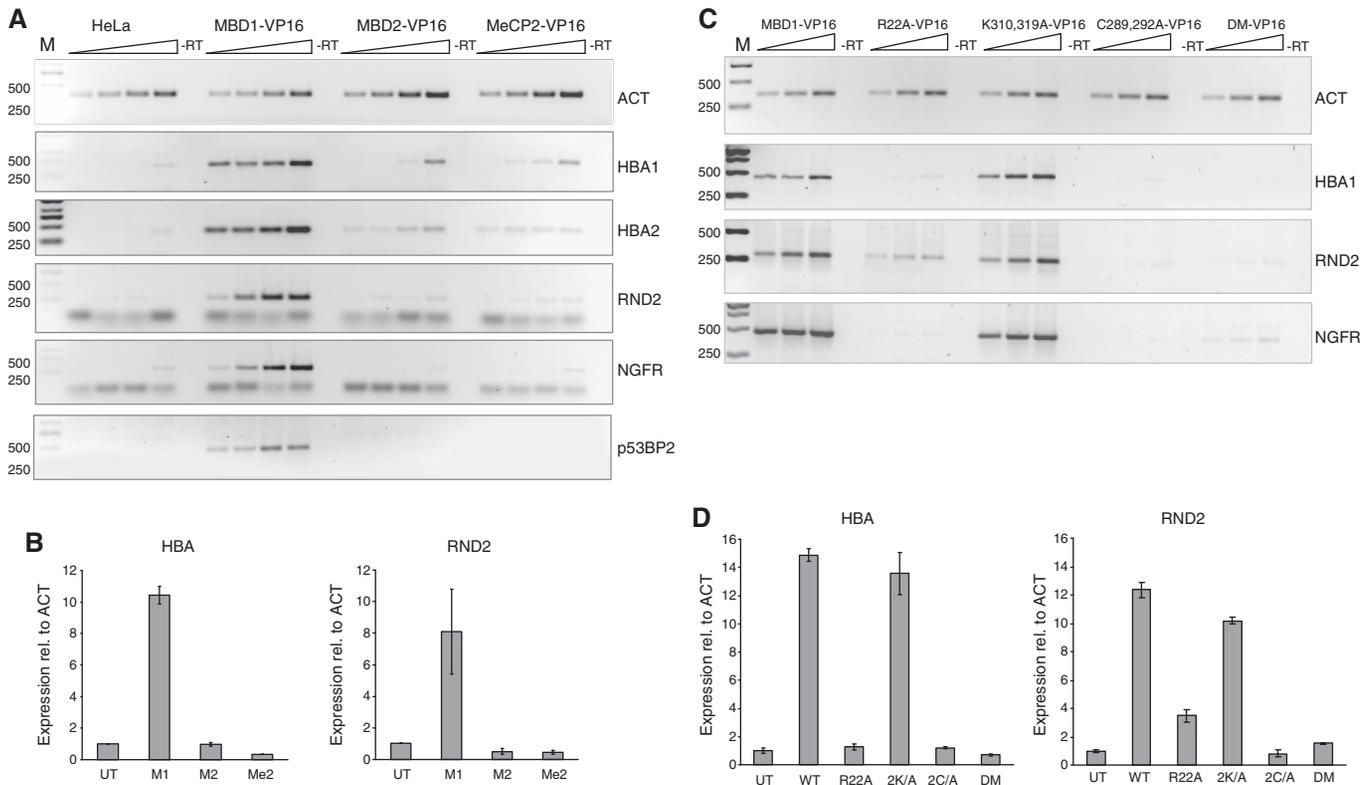
genes, while DNA binding by the CXXC3 domain is largely dispensable. Nevertheless, from these experiments we could not exclude that the CXXC3 stabilizes the interaction of MBD1 with DNA once the MBD domain is bound.

Alternatively spliced isoforms of human MBD1 include variants 3 and 4, which contain CXXC1 and CXXC2 domains but lack CXXC3 (30). Variants 1 and 2 carry CXXC1, CXXC2 and CXXC3, while PCM1 contains CXXC2 and CXXC3 (Figure 5A). Aiming to investigate further whether CXXC3 contributes to binding of MBD1 at target genes, we compared the ability of MBD1-VP16 (derived from the PCM1 isoform) and MBD1<sup>Var3</sup>-VP16 to activate HBA and RND2. When we expressed MBD1-VP16 and MBD1<sup>Var3</sup>-VP16 in HeLa cells at comparable levels, we found that they did not differ significantly in their capacity to activate target genes as detected by RT-qPCR (Figure 5B and C). These experiments indicate that the CXXC3 domain is largely dispensable for the recruitment of MBD1-VP16 to its target genes in HeLa cells.

### MBD1-VP16 and endogenous MBD1 bind to high affinity sites near the RND2 and NGFR promoters

In order to investigate if binding of MBD1-VP16 occurs at or near NGFR, RND2 and HBA promoters, we first determined their DNA methylation patterns by bisulphite sequencing. We found that the sequences flanking TSS of NGFR and HBA genes were not methylated, while the RND2 promoter was partially methylated at sequences immediately adjacent to the TSS (Figure 6A and C, and 7A). Notably, TCGCA sequences, which we identified as high affinity binding sites for MBD1 (Figure 1), are present near these promoters: at +801 from the TSS of NGFR gene and +1315, +2640 and +2746 from the TSS of RND2 (Figure 5A and C). Four TCGCA and one TGCGCA sequences (-545, -456, +505, +1518, +2204) flank the HBA promoter (not shown).

ChIP experiments with antibodies against VP16 activation domain showed enrichment over the high affinity binding site at +801 downstream of RND2 TSS and +2646 and +2746 downstream of NGFR TSS in cells transfected with MBD1-VP16 but not in control cells



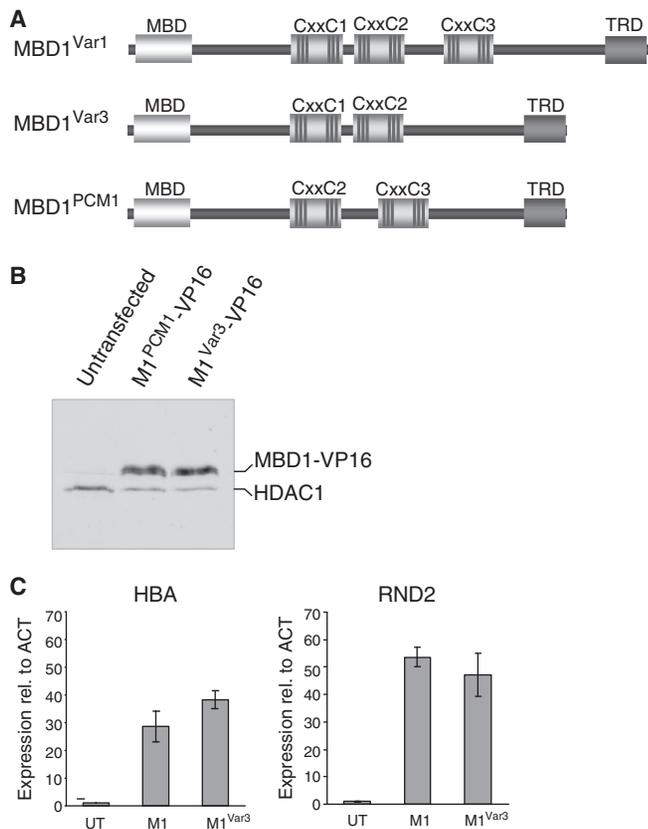
**Figure 4.** Validation of microarray data and characterization of transactivation potential of mutant MBD1-VP16 proteins. (A and B) Semiquantitative and quantitative RT-PCR experiments demonstrate that transcripts induced by expression of MBD1-VP16 in HeLa by ~10-fold are not up-regulated in cells expressing either MBD2-VP16 or MeCP2-VP16. Actin served as ubiquitously expressed control. The graphs in (B) represent triplicate RT-qPCRs performed on two independent transfection experiments for each fusion protein. (C and D) Semi-quantitative and quantitative RT-PCRs show that MBD1-VP16 carrying R22A a mutation in the MBD domain (R22A and DM) is unable to activate target genes compared to MBD1-VP16 carrying K310,319A mutations in CXXC3. Cysteine mutations (C289A,292A) also abolished transactivation by MBD1-VP16, presumably by affecting protein folding. Actin was used as a control. RT-qPCR experiments were performed in triplicates on two independent transfections for each protein. The expression of MBD1-, MBD2- and MeCP2-VP16 fusion proteins as well as expression of mutant forms of MBD1-VP16 is shown in Figure 3B and C.

transfected with DM-VP16 (Figure 6B and D, left). As expected, acetylation of histone H3 was high at RND2 and NGFR promoters in MBD1-VP16 transfected cells but not in the controls (Figure 6B and D, middle). Moreover, ChIP with antibodies against MBD1 detected enrichment of the endogenous MBD1 in the same location as MBD1-VP16 in transfected cells, suggesting that the fusion protein may compete with the endogenous MBD1 for binding to high affinity sites near RND2 and NGFR promoters (Figure 6B and D, right). We did not detect either MBD1 or MBD1-VP16 binding to unmethylated CpG-rich sequences of RND2 and NGFR promoters although these, especially in the case of NGFR, would potentially provide significant number of unmethylated CpGs for binding of CXXC3. Taken together, these experiments indicate that when binding of MBD1-VP16 occurs at methylated CpGs near gene promoters this leads to histone acetylation and strong expression of targeted genes.

#### DNA methylation and MBD1 are required for silencing of MBD1-VP16 targeted genes

Given that our ChIP experiments detected MBD1 binding near promoters that were highly induced by MBD1-VP16

transactivator, we wanted to know whether DNA methylation and the endogenous MBD1 are essential for silencing of these genes. Similar to HeLa cells, HBA and RND2 are not expressed in the colorectal carcinoma derived HCT116 cell line, although the CpG islands of both genes are largely unmethylated (Figure 7A and not shown). Methylated CpGs can be detected -650 bp upstream and +337 bp downstream from the HBA TSS in HeLa and in HCT116 cells (Figure 7A), but not in double knock out (DKO) HCT116 cells, which lack DNMT3B and express a N-terminally truncated hypomorphic allele of DNMT1 (48-50) (Figure 7A and not shown). DNA methylation was also absent from the RND2 promoter in DKO cells compared to the parental HCT116 cell line (not shown). This is consistent with the significant hypomethylation of the genome observed in DKO cells (49). Interestingly, HBA and RND2 transcripts were detectable in DNMT3B null (3B KO) and DKO cells, suggesting that DNA methylation near these promoters is important for the recruitment of MBD1 and silencing of HBA and NGFR in HeLa and HCT116 cells (Figure 7B). Consistently, expression of MBD-VP16 proteins in DKO cells did not lead to further activation of these genes (data not shown).



**Figure 5.** Lack of CXXC3 domain does not affect transactivation by MBD1-VP16 and binding of MBD1 to DNA in living cells. (A) Schematic representation of MBD1 isoforms Variant1 (MBD1<sup>Var1</sup>), Variant3 (MBD1<sup>Var3</sup>) and PCM1 (MBD1<sup>PCM1</sup>). Note that MBD1<sup>Var1</sup> carries three CXXC motifs, while MBD1<sup>Var3</sup> lacks DNA binding CXXC3 and MBD1<sup>PCM1</sup> lacks CXXC2. (B) Expression of MBD1<sup>PCM1</sup>-VP16 and MBD1<sup>Var3</sup>-VP16 in HeLa cells. HDAC1 serves as a loading control. (C) RT-qPCR experiments demonstrate that MBD1<sup>PCM1</sup>-VP16 (M1) and MBD1<sup>Var3</sup>-VP16 (M1<sup>Var3</sup>) do not differ significantly in their ability to induce HBA and RND transcription in HeLa.

To investigate whether endogenous MBD1 is required for the repression of genes that are targeted and activated by MBD1-VP16, we transiently knocked down MBD1 and MBD2 in HeLa cells by a plasmid-driven expression of a small hairpin (sh) RNAs (Figure 7C and D). Plasmids carrying a non-silencing shRNAs served as controls. We did not attempt to knock down MeCP2 as it is expressed at negligible levels in HeLa cells. MBD1 shRNA reduced MBD1 RNA and protein levels by ~60% compared to the control cells and led to a partial derepression of HBA, RND2 and NGFR (Figure 7C and data not shown). Notably, the expression of these genes in cells with reduced levels of MBD1 was much weaker than in cells transfected with MBD1-VP16, which is consistent with the strong activation properties of the VP16 domain (Figure 5C and Figure 7C). A comparable knock down of MBD2 in HeLa did not derepress HBA and NGFR genes suggesting that MBD2 does not bind at these loci (Fig 7D). However, we could detect a 2-fold up-regulation of RND2 in cells stably expressing MBD2 shRNA. Given that RND2 was not activated by MBD2-VP16 fusion

protein, it is likely that the effect of MBD2 knockdown on RND2 is indirect.

In summary, these experiments indicate that DNA methylation and MBD1 protein are required for silencing of a subset of genes with unmethylated promoters in HeLa and other cell lines.

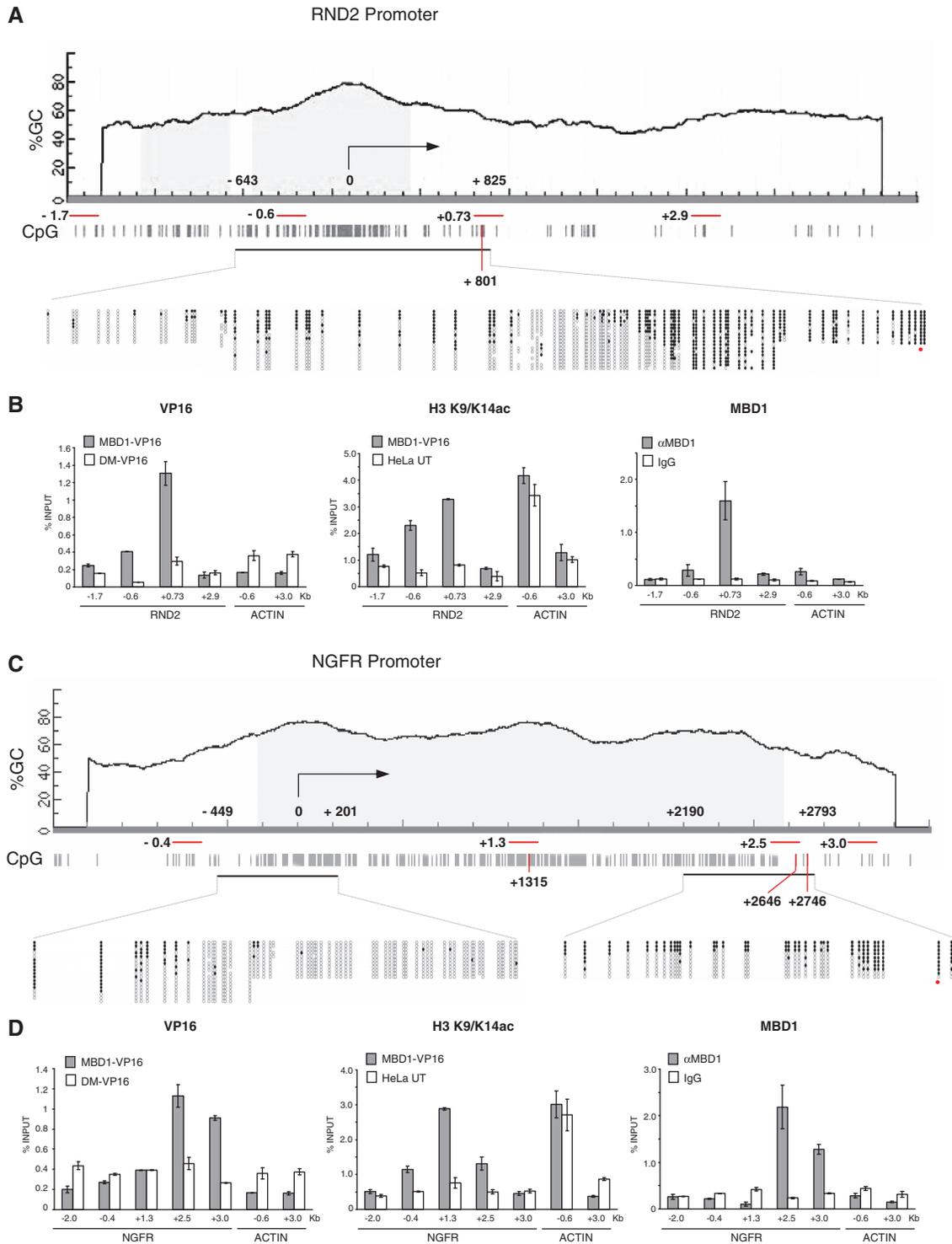
## DISCUSSION

Despite being viable and fertile, mice null for Mbd1, Mbd2 and Mecp2 proteins display distinct phenotypes (16–19,51). Deficiency or mutations in Mecp2 in humans result in a severe neurological disorder known as Rett syndrome. In Rett patients, as well as in mouse models of the disease, loss of Mecp2 function mostly affects brain physiology, as Mecp2 is particularly abundant in mature postmitotic neurons (18,19,52). Mbd2-null animals show defects in maternal behaviour and misexpression of genes in the intestine and in the immune system (17,53–55). In addition, *Mbd2*<sup>-/-</sup> mouse embryonic fibroblast display leaky silencing of non-coding RNA Xist resulting in inappropriate inactivation of the single active X chromosome in a proportion of Mbd2-null male cells (53). The phenotype of Mbd1-deficient mice is extremely mild. However, these animals display impaired spatial learning and defective neurogenesis (16,51). These phenotypes seem incompatible with high degree of functional redundancy between the MBD family members.

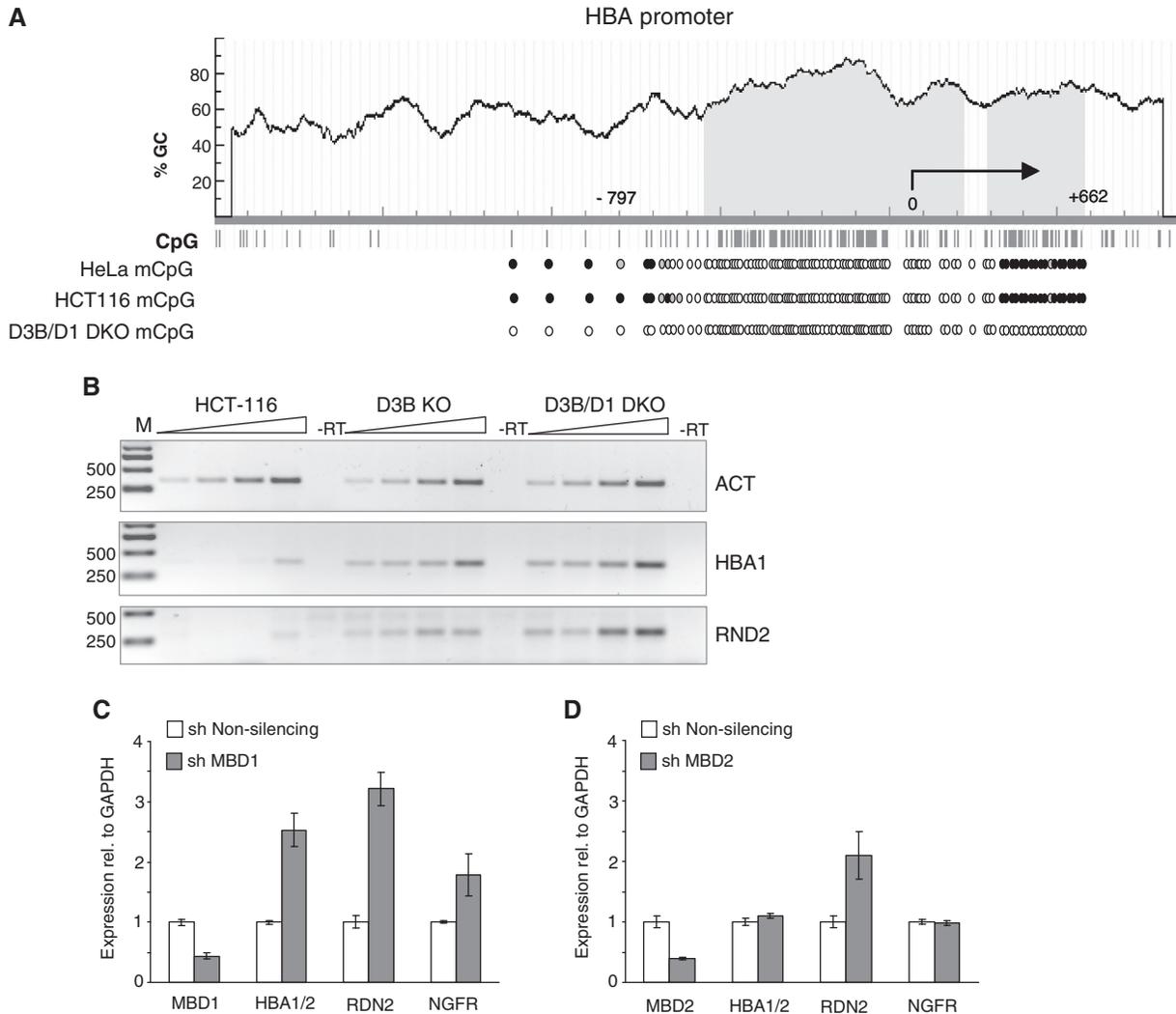
Moreover, investigations in primary human cells have rarely detected shared occupancy of methylated sites by MBD family proteins, although this may occur at densely methylated CpG islands in cancer cell lines (23,26). Studies aiming to determine DNA-binding preference of MeCP2 have shown that MeCP2 binds more efficiently to methylated CpG followed by an A/T<sub>≥4</sub> run (26). The recently solved crystal structure of the MBD domain of Mecp2 bound to a sequence derived from the BDNF promoter, a known target gene for Mecp2, suggests that the A/T<sub>≥4</sub> run causes narrowing of the minor groove of DNA and DNA bending which may stabilize the interaction of MBD domain with methylated CpGs (56). Additionally, the tandem Asx-ST motif in the MBD of Mecp2, which is not present in Mbd1 or Mbd2, contact the phosphate backbone at the start of the A/T run (56).

Consistently, we did not detect preferential binding of MBD domain of Mbd1 to methylated CpGs with an adjacent A/T run (Supplementary Figure S2B).

In this report we investigate how Mbd1 protein interacts with DNA *in vitro* and *in vivo*. We focused on the two DNA-binding domains of MBD1, the MBD and the CXXC3, aiming to determine if they bind to DNA in a sequence-specific manner. We report for the first time that the MBD domain of MBD1 binds with higher affinity to specific methylated sequences *in vitro*, namely TC<sup>M</sup>GCA and TGC<sup>M</sup>GCA. However, the *K<sub>D</sub>* of MBD1 MBD binding to these sequences was only 4–5-fold lower than the *K<sub>D</sub>* measured for the least efficient methylated sequence CC<sup>M</sup>GGC. Furthermore, we also detected selection for multiple methylated CpGs within the expanded GCGC sequence in the methyl-SELEX assay. Additional



**Figure 6.** Chromatin immunoprecipitation detects MBD1 binding near RND2 and NGFR promoters. (A) Schematic drawing of RND2 promoter with transcription start site (TSS) indicated by arrow. The GC content, position of CpGs and regions analysed by ChIP in (B) are shown. The promoter sequences from  $-643$  to  $+825$  were analysed for DNA methylation by bisulphite sequencing. The results are displayed below the line corresponding to the sequenced region. Filled circles represent methylated CpGs. The red bar in the CpG plot at  $+801$  and the red dot under the bisulphite sequencing plot indicate a high affinity MBD1 binding site (TCGCA). (B) qPCRs on chromatin immunoprecipitations (ChIP) with antibodies against VP-16AD and MBD1 detect binding of MBD1 and MBD1-VP16, but not DM-VP16, at high affinity MBD1 binding site downstream from RND2 TSS. Actin promoter and downstream sequence at  $+3$ Kb served as a negative control. ChIP with anti-acetyl H3 K9/K14 antibody indicates that RND2 promoter is acetylated in cells expressing MBD1-VP16 compared to untransfected HeLa. (C) Schematic representation of NGFR promoter. The regions from  $-449$  to  $+201$  and from  $+2190$  to  $+2443$  were analysed by bisulphite sequencing. The regions investigated by ChIP and qPCR in (D) are indicated as well as the position of putative MBD1 binding sites (TCGCA) at  $+1315$ ,  $+2646$  and  $+2746$ . (D) ChIP with anti-VP16AD and anti-MBD1 antibodies detect binding of MBD1 and the fusion protein at high affinity binding sites 2.5 Kb downstream of NGFR TSS. Acetylation of histone H3K9/K14 is high at NGFR promoter only in cells expressing MBD1-VP16. Beta Actin promoter was used as a control.



**Figure 7.** DNA methylation and endogenous MBD1 are required for silencing of HBA and RND2 genes. (A) Schematic representation of HBA promoter. DNA methylation of the region from  $-797$  to  $+662$  was analysed by bisulphite sequencing. The average methylation patterns in HeLa, HCT116 and HCT116 DNMT3B KO/DNMT1 hypomorph cells (D3/D1 DKO) are shown. (B) Semi-quantitative RT-PCRs detect HBA and RND2 transcripts in DNMT-deficient but not in wild-type HCT116 cells. (C) Quantitative RT-PCRs detect derepression of HBA, RND2 and NGFR genes relative to the GAPDH control after a partial 60% knock down of MBD1 in HeLa cells by shRNA. A vector carrying non-silencing shRNA sequence directed against MBD1 served as a negative control. (D) Comparable knock down of MBD2 by shRNA in HeLa cells has no effect on silencing of HBA and NGFR, but results in some derepression of RND2. A vector carrying non-silencing shRNA sequence directed against MBD2 served as a negative control.

bandshift experiments confirmed that multiple methylated CpGs can override the sequence preference of MBD1 at single methylated ones (Supplementary Figure S5). Therefore, MBD1 as well as MBD2 would be expected to bind with high affinity to densely methylated CpG islands. However, in most somatic cells the CpG-rich promoters are not methylated and binding of MBD1 should mostly occur at single methylated CpGs.

Are the affinities we measured *in vitro* likely to determine the patterns of MBD1 binding to methylated DNA *in vivo*? Potentially, the occupancy of methylated sites by MBD1 and other methyl-CpG-binding proteins may reflect the ratio between the total number of available binding sites and the number of MBD protein molecules present in the nucleus. Quantitative western blots indicated that there are  $\sim 1.2 \times 10^4$  molecules of MBD1

per nucleus in HeLa and normal diploid human fibroblasts [(15) and data not shown]. Given that on average there is one CpG/100 bp in the bulk genome (excluding CpG islands) and that  $\sim 70\%$  of all CpGs are methylated, the calculated number of methylated CpGs is  $\sim 2 \times 10^7$ . Therefore, the ratio of MBD1 to methylated CpGs is  $\sim 1/1670$ . If only 1% of all methylated CpGs are present within high affinity MBD1-binding sites, the number of such sites will be 16-fold higher than the number of MBD1 molecules in each cell. Therefore, it is likely that most MBD1 is bound to its preferred binding sites in HeLa, provided that they are not obstructed by nucleosomes or other DNA bound proteins.

On the other hand, we found that the CXXC3 zinc finger of MBD1 has no sequence specificity and can bind *in vitro* to a single unmethylated CG within any

sequence context. A SELEX experiment with CXXC3 domain led to enrichment for additional CG dinucleotides but not CXXC3-specific sequence motifs (data not shown). This is consistent with recent findings that CXXC3 domain of MBD1 can be used to purify CpG-islands from fragmented genomic DNA (4). However, we failed to detect binding of CXXC3 to unmethylated CpG island promoters *in vivo*. MBD1<sup>K310,319A</sup>-VP16 mutant deficient in binding to unmethylated CpGs could activate target genes as efficiently as the 'wild-type' MBD1-VP16 and neither the endogenous MBD1 nor the exogenous MBD1-VP16 were enriched at CpG island promoters of NGFR and RND2 genes. This indicates that the unmethylated CpG-rich promoters are somehow protected against binding of MBD1 presumably by other proteins that may compete with MBD1. In contrast, the R22A mutation in the MBD domain, which abolishes binding to methylated DNA *in vitro*, significantly reduced the transactivation activity of MBD1<sup>R22A</sup>-VP16 at all genes that we analysed. Taken together, these data demonstrate that the MBD and the CXXC3 domains do not contribute equally to stable binding of MBD1 to DNA *in vivo* and indicate the dominant role of the MBD domain and DNA methylation in determining the patterns of MBD1 distribution throughout the genome.

The physiological function of CXXC3 *in vivo* is yet to be determined. It is possible that this domain is functional only when the levels of DNA methylation in the genome are significantly reduced, for example in preimplantation mouse embryos (1,57). This is consistent with earlier observations that in the presence of non-functional MBD domain CXXC3 can support binding of MBD1 at major satellite DNA in *Dnmt1*<sup>-/-</sup> mouse embryonic fibroblasts but not in 'wild-type' cells (31).

Although the transactivation assays provide a useful tool to study the contribution of DNA-binding domains of MBD1 to targeting specific genes *in vivo*, it is important to acknowledge that the microarray experiments identified only a small number of significantly ( $\geq 3$ -fold) up-regulated genes by MBD1-VP16 in any of the examined cell lines. One may wonder why the number of 'activatable' promoters is small and whether they share any common features. Notably, all the 34 genes identified as MBD1 targets in HeLa contain 1–4 either TGCGCA or TCGCA sites in close proximity to the promoter. However, since a limited number of MBD1-VP16 activated genes are available for analysis, any features found at these promoters may not withstand rigorous statistical tests when compared to the rest of the promoters in the genome. In fact  $\sim 33\%$  (7743) of all protein coding genes contain one or more MBD1-preferred-binding sequences within a 1 kb region flanking TSS (from -500 to +500 bp). These promoters would be potential targets for MBD1 if methylated. As 82% of them are CpG islands and the majority of these are not methylated in any given cell type, this would reduce significantly the number of potential target genes. On the other hand, TC<sup>M</sup>GCA and TGC<sup>M</sup>GCA sequences are relatively underrepresented in the rest of the genome (on average one site/ 5kb). As the promoters of genes which we

examined in detail were not methylated, but binding of MBD1 was detectable at TC<sup>M</sup>GCA sites in the proximity of their CpG-islands, this may indicate that our transactivation assay was biased towards identification of genes, at which binding of MBD1 occurs sufficiently close to the promoter to allow activation. Further experiments designed to map binding of endogenous MBD1 in mammalian cells in comparison to DNA methylation patterns will be needed to fully understand how MBD1 functions *in vivo*.

In summary, we found that MBD1 binds more efficiently to a single methylated CpG in the context of TC<sup>M</sup>GCA and TGC<sup>M</sup>GCA sites *in vitro* and *in vivo*. We identified specific genes targeted for silencing by MBD1 in several human cell lines and demonstrated that recruitment of MBD1-VP16 transactivator to these loci occurs due to the preference of the MBD domain for methylated DNA in specific sequence context. Most importantly, other MBD domains did not display high affinity for MBD1-binding sites and were not recruited to the same genes when expressed as MBD-VP16 fusion proteins. Taken together with the sequence preference of MeCP2 (26), our data suggest that silencing by MBD1 and MeCP2 is restricted to subsets of genes that harbour methylated sequences supporting stable binding of these proteins to DNA.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

The authors thank Ken Sawin and Adrian Bird for critical reading and helpful comments during the manuscript preparation.

## FUNDING

Cancer Research UK Senior Fellowship (C7215/A8983 to I.S.); Wellcome Trust project grant (077691/Z/05/Z to I.S.). T.C. was supported by a long-term EMBO Fellowship. Funding for open access charges: Wellcome Trust.

*Conflict of interest statement.* None declared.

## REFERENCES

- Bird, A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev.*, **16**, 6–21.
- Chan, S.W., Henderson, I.R. and Jacobsen, S.E. (2005) Gardening the genome: DNA methylation in *Arabidopsis thaliana*. *Nat. Rev. Genet.*, **6**, 351–360.
- Goll, M.G. and Bestor, T.H. (2005) Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.*, **74**, 481–514.
- Illingworth, R., Kerr, A., Desousa, D., Jorgensen, H., Ellis, P., Stalker, J., Jackson, D., Clee, C., Plumb, R., Rogers, J. *et al.* (2008) A novel CpG island set identifies tissue-specific methylation at developmental gene loci. *PLoS Biol.*, **6**, e22.
- Costello, J.F., Fruhwald, M.C., Smiraglia, D.J., Rush, L.J., Robertson, G.P., Gao, X., Wright, F.A., Feramisco, J.D., Peltomaki, P., Lang, J.C. *et al.* (2000) Aberrant CpG-island

- methylation has non-random and tumour-type-specific patterns. *Nat. Genet.*, **24**, 132–138.
6. Keshet, I., Schlesinger, Y., Farkash, S., Rand, E., Hecht, M., Segal, E., Pikarski, E., Young, R.A., Niveleau, A., Cedar, H. *et al.* (2006) Evidence for an instructive mechanism of de novo methylation in cancer cells. *Nat. Genet.*, **38**, 149–153.
  7. Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Paabo, S., Rebhan, M. and Schubeler, D. (2007) Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.*, **39**, 457–466.
  8. Jackson-Grusby, L., Beard, C., Possemato, R., Tudor, M., Fambrough, D., Csankovszki, G., Dausman, J., Lee, P., Wilson, C., Lander, E. *et al.* (2001) Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. *Nat. Genet.*, **27**, 31–39.
  9. Robert, M.F., Morin, S., Beaulieu, N., Gauthier, F., Chute, I.C., Barsalou, A. and MacLeod, A.R. (2003) DNMT1 is required to maintain CpG methylation and aberrant gene silencing in human cancer cells. *Nat. Genet.*, **33**, 61–65.
  10. Klose, R.J. and Bird, A.P. (2006) Genomic DNA methylation: the mark and its mediators. *Trends Biochem. Sci.*, **31**, 89–97.
  11. Clouaire, T. and Stancheva, I. (2008) Methyl-CpG binding proteins: specialized transcriptional repressors or structural components of chromatin? *Cell Mol. Life Sci.*, **65**, 1509–1522.
  12. Hendrich, B. and Bird, A. (1998) Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol. Cell Biol.*, **18**, 6538–6547.
  13. Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N. and Bird, A. (1998) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature*, **393**, 386–389.
  14. Ng, H.H., Zhang, Y., Hendrich, B., Johnson, C.A., Turner, B.M., Erdjument-Bromage, H., Tempst, P., Reinberg, D. and Bird, A. (1999) MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. *Nat. Genet.*, **23**, 58–61.
  15. Ng, H.H., Jeppesen, P. and Bird, A. (2000) Active repression of methylated genes by the chromosomal protein MBD1. *Mol. Cell Biol.*, **20**, 1394–1406.
  16. Zhao, X., Ueba, T., Christie, B.R., Barkho, B., McConnell, M.J., Nakashima, K., Lein, E.S., Eadie, B.D., Willhoite, A.R., Muotri, A.R. *et al.* (2003) Mice lacking methyl-CpG binding protein 1 have deficits in adult neurogenesis and hippocampal function. *Proc. Natl Acad. Sci. USA*, **100**, 6777–6782.
  17. Hendrich, B., Guy, J., Ramsahoye, B., Wilson, V.A. and Bird, A. (2001) Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. *Genes Dev.*, **15**, 710–723.
  18. Guy, J., Hendrich, B., Holmes, M., Martin, J.E. and Bird, A. (2001) A mouse MeCP2-null mutation causes neurological symptoms that mimic Rett syndrome. *Nat. Genet.*, **27**, 322–326.
  19. Chen, R.Z., Akbarian, S., Tudor, M. and Jaenisch, R. (2001) Deficiency of methyl-CpG binding protein-2 in CNS neurons results in a Rett-like phenotype in mice. *Nat. Genet.*, **27**, 327–331.
  20. Allan, A.M., Liang, X., Luo, Y., Pak, C., Li, X., Szulwach, K.E., Chen, D., Jin, P. and Zhao, X. (2008) The loss of methyl-CpG binding protein 1 leads to autism-like behavioral deficits. *Hum. Mol. Genet.*, **17**, 2047–2057.
  21. Li, E., Bestor, T.H. and Jaenisch, R. (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, **69**, 915–926.
  22. Okano, M., Bell, D.W., Haber, D.A. and Li, E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, **99**, 247–257.
  23. Ballestar, E., Paz, M.F., Valle, L., Wei, S., Fraga, M.F., Espada, J., Cigudosa, J.C., Huang, T.H. and Esteller, M. (2003) Methyl-CpG binding proteins identify novel sites of epigenetic inactivation in human cancer. *Embo J.*, **22**, 6335–6345.
  24. Lopez-Serra, L., Ballestar, E., Ropero, S., Setien, F., Billard, L.M., Fraga, M.F., Lopez-Nieva, P., Alaminos, M., Guerrero, D., Dante, R. *et al.* (2008) Unmasking of epigenetically silenced candidate tumor suppressor genes by removal of methyl-CpG-binding domain proteins. *Oncogene*, **27**, 3556–3566.
  25. Koch, C. and Stratling, W.H. (2004) DNA binding of methyl-CpG-binding protein MeCP2 in human MCF7 cells. *Biochemistry*, **43**, 5011–5021.
  26. 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**, 667–678.
  27. Sarraf, S.A. and Stancheva, I. (2004) Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. *Mol. Cell*, **15**, 595–605.
  28. Fujita, N., Watanabe, S., Ichimura, T., Ohkuma, Y., Chiba, T., Saya, H. and Nakao, M. (2003) MCAF mediates MBD1-dependent transcriptional repression. *Mol. Cell Biol.*, **23**, 2834–2843.
  29. Cross, S.H., Meehan, R.R., Nan, X. and Bird, A. (1997) A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins. *Nat. Genet.*, **16**, 256–259.
  30. Fujita, N., Takebayashi, S., Okumura, K., Kudo, S., Chiba, T., Saya, H. and Nakao, M. (1999) Methylation-mediated transcriptional silencing in euchromatin by methyl-CpG binding protein MBD1 isoforms. *Mol. Cell Biol.*, **19**, 6415–6426.
  31. Jorgensen, H.F., Ben-Porath, I. and Bird, A.P. (2004) Mbd1 is recruited to both methylated and nonmethylated CpGs via distinct DNA binding domains. *Mol. Cell Biol.*, **24**, 3387–3395.
  32. Allen, M.D., Grummitt, C.G., Hilcenko, C., Min, S.Y., Tonkin, L.M., Johnson, C.M., Freund, S.M., Bycroft, M. and Warren, A.J. (2006) Solution structure of the nonmethyl-CpG-binding CXXC domain of the leukaemia-associated MLL histone methyltransferase. *Embo J.*, **25**, 4503–4512.
  33. Fujita, N., Shimotake, N., Ohki, I., Chiba, T., Saya, H., Shirakawa, M. and Nakao, M. (2000) Mechanism of transcriptional regulation by methyl-CpG binding protein MBD1. *Mol. Cell Biol.*, **20**, 5107–5118.
  34. 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**, 4886–4892.
  35. Clouaire, T., Roussigne, M., Ecochard, V., Mathe, C., Amalric, F. and Girard, J.P. (2005) The THAP domain of THAP1 is a large C2CH module with zinc-dependent sequence-specific DNA-binding activity. *Proc. Natl Acad. Sci. USA*, **102**, 6907–6912.
  36. Pfaffl, M.W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.*, **29**, e45.
  37. Wettenhall, J.M. and Smyth, G.K. (2004) limmaGUI: a graphical user interface for linear modeling of microarray data. *Bioinformatics*, **20**, 3705–3706.
  38. Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J. *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.*, **5**, R80.
  39. Feil, R., Charlton, J., Bird, A.P., Walter, J. and Reik, W. (1994) Methylation analysis on individual chromosomes: improved protocol for bisulphite genomic sequencing. *Nucleic Acids Res.*, **22**, 695–696.
  40. Suzuki, M.M., Kerr, A.R., De Sousa, D. and Bird, A. (2007) CpG methylation is targeted to transcription units in an invertebrate genome. *Genome Res.*, **17**, 625–631.
  41. Bock, C., Reither, S., Mikeska, T., Paulsen, M., Walter, J. and Lengauer, T. (2005) BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing. *Bioinformatics*, **21**, 4067–4068.
  42. Ren, B. and Dynlacht, B.D. (2004) Use of chromatin immunoprecipitation assays in genome-wide location analysis of mammalian transcription factors. *Methods Enzymol.*, **376**, 304–315.
  43. 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**, 487–497.
  44. Hall, D.B. and Struhl, K. (2002) The VP16 activation domain interacts with multiple transcriptional components as determined by protein-protein cross-linking in vivo. *J. Biol. Chem.*, **277**, 46043–46050.

45. Higgs,D.R., Sharpe,J.A. and Wood,W.G. (1998) Understanding alpha globin gene expression: a step towards effective gene therapy. *Semin. Hematol.*, **35**, 93–104.
46. Anguita,E., Hughes,J., Heyworth,C., Blobel,G.A., Wood,W.G. and Higgs,D.R. (2004) Globin gene activation during haemopoiesis is driven by protein complexes nucleated by GATA-1 and GATA-2. *Embo J.*, **23**, 2841–2852.
47. Nakamura,K., Yamashita,Y., Tamamaki,N., Katoh,H., Kaneko,T. and Negishi,M. (2006) In vivo function of Rnd2 in the development of neocortical pyramidal neurons. *Neurosci. Res.*, **54**, 149–153.
48. Egger,G., Jeong,S., Escobar,S.G., Cortez,C.C., Li,T.W., Saito,Y., Yoo,C.B., Jones,P.A. and Liang,G. (2006) Identification of DNMT1 (DNA methyltransferase 1) hypomorphs in somatic knockouts suggests an essential role for DNMT1 in cell survival. *Proc. Natl Acad. Sci. USA*, **103**, 14080–14085.
49. Rhee,I., Bachman,K.E., Park,B.H., Jair,K.W., Yen,R.W., Schuebel,K.E., Cui,H., Feinberg,A.P., Lengauer,C., Kinzler,K.W. *et al.* (2002) DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. *Nature*, **416**, 552–556.
50. Myant,K. and Stancheva,I. (2008) LSH cooperates with DNA methyltransferases to repress transcription. *Mol. Cell Biol.*, **28**, 215–226.
51. Allan,A.M., Liang,X., Luo,Y., Pak,C., Li,X., Szulwach,K.E., Chen,D., Jin,P. and Zhao,X. (2008) The loss of methyl-CpG binding protein 1 leads to autism-like behavioral deficits. *Hum. Mol. Genet.*, **17**, 2047–2057.
52. Moretti,P. and Zoghbi,H.Y. (2006) MeCP2 dysfunction in Rett syndrome and related disorders. *Curr Opin. Genet. Dev.*, **16**, 276–281.
53. Barr,H., Hermann,A., Berger,J., Tsai,H.H., Adie,K., Prokhortchouk,A., Hendrich,B. and Bird,A. (2007) Mbd2 contributes to DNA methylation-directed repression of the Xist gene. *Mol. Cell Biol.*, **27**, 3750–3757.
54. Berger,J., Sansom,O., Clarke,A. and Bird,A. (2007) MBD2 is required for correct spatial gene expression in the gut. *Mol. Cell Biol.*, **27**, 4049–4057.
55. Hutchins,A.S., Mullen,A.C., Lee,H.W., Sykes,K.J., High,F.A., Hendrich,B.D., Bird,A.P. and Reiner,S.L. (2002) Gene silencing quantitatively controls the function of a developmental trans-activator. *Mol. Cell*, **10**, 81–91.
56. Ho,K.L., McNae,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**, 525–531.
57. Kafri,T., Ariel,M., Brandeis,M., Shemer,R., Urven,L., McCarrey,J., Cedar,H. and Razin,A. (1992) Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. *Genes Dev.*, **6**, 705–714.