4886–4892 Nucleic Acids Research, 1993, Vol. 21, No. 21

Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2

Xinsheng Nan, Richard R.Meehan and Adrian Bird*
Institute of Cell and Molecular Biology, University of Edinburgh, Kings Buildings, Mayfield Road, Edinburgh EH9 3JR, UK

Received August 5, 1993; Revised and Accepted September 9, 1993

ABSTRACT

MeCP2 is a chromosomal protein which binds to DNA that is methylated at CpG. *In situ* immunofluorescence in mouse cells has shown that the protein is most concentrated in pericentromeric heterochromatin, suggesting that MeCP2 may play a role in the formation of inert chromatin. Here we have isolated a minimal methyl-CpG binding domain (MBD) from MeCP2. MBD is 85 amino acids in length, and binds exclusively to DNA that contains one or more symmetrically methylated CpGs. MBD has negligable non-specific affinity for DNA, confirming that non-specific and methyl-CpG specific binding domains of MeCP2 are distinct. *In vitro* footprinting indicates that MBD binding can protect a 12 nucleotide region surrounding a methyl-CpG pair, with an approximate dissociation constant of 10^{-9}M.

INTRODUCTION

DNA methylation in vertebrates is associated with alterations in chromatin structure and silencing of gene expression (1–3). The discovery of proteins with an affinity for methylated DNA raises the possibility that their binding mediates the effects of methylation on both transcription and chromatin. In line with this idea, Antequera *et al* (4) have shown that accessibility of CpGs in nuclei to endonucleases depends on their methylation status. Methylated CpGs in intact nuclei are resistant to digestion by MspI and TthI, whereas nonmethylated CpGs are comparatively accessible. Since both these enzymes are indifferent to CpG methylation in naked DNA, the simplest explanation for the observation is that there are factors in the nucleus that protect methylated sites.

Two factors have been identified that may mediate the biological effects of CpG methylation. One of these, MeCP1, binds to DNA containing multiple methyl-CpG sites (5). There is also evidence that MeCP1 can inhibit transcription from methylated promoters both *in vitro* and *in vivo* (6). The second factor, MeCP2, also binds to DNA containing methylated CpG, but it differs from MeCP1 in that binding is promoted by a single methyl-CpG pair (7, 8). *In situ* immunofluorescence in mouse cells has shown that the protein is present throughout the chromosome arms, but is concentrated in the pericentromeric heterochromatin. The latter is also the location of the mouse major satellite DNA, which contains about half of all 5-methylcytosine in the mouse genome (9). In the rat, which lacks a highly abundant methyl-CpG-rich satellite, MeCP2 is uniformly distributed in the genome. Thus the distribution of MeCP2 matches that of methyl-CpG in the genome.

At present only MeCP2 has been cloned, and its deduced amino acid sequence exhibits no extensive similarities with other DNA binding proteins, although it is rich in basic residues and contains short motifs [(R)GRP(K) (10, 11) and SPKK (12)] that have been implicated in binding to the minor groove of AT-rich DNA (see Fig1A). The presence of potential minor groove binding structures in MeCP2 is unexpected, as methyl groups on the 5 position of cytosine are located in the major groove of B-form DNA.

MeCP2 also binds weakly to non-methylated DNA. Since this non-specific binding can be competed easily by addition of an excess of non-methylated DNA without affecting methylated DNA binding, it seemed likely that specific and non-specific binding sites were located in different parts of the molecule (8). This possibility led us to try and isolate the minimal region required for binding to methylated DNA. The experiments described below show that this has been possible, and establish some characteristics of the resulting methyl-CpG binding domain (MBD).

MATERIALS AND METHODS

Construction of deletion mutants of MeCP2

The NotI/EcoRI fragment containing full length MeCP2 was blunt-ended and inserted into the SmaI site of pGEX-3X for construction of an in-frame GST fusion plasmid. Deletion mutant plasmids were generated by restriction endonucleases or by exonuclease III/S1 deletion. HindIII deletion generated truncated proteins that start or end at amino acid 108; Asp700 truncated at amino acid 206; PvuII truncated at amino acid 277; and SacI truncated at amino acid 392. Other deletions were produced by exonucleaseIII resection from the above sites, followed by treatment with SI nuclease and blunt-end cloning. The reading frame of all clones was checked by sequencing the junction of

* To whom correspondence should be addressed

5' truncated plasmids using the 18-mer primer TTCCCTCTA-GAAATAATT, which is complementary to glutathione S-transferase gene (20). Full length and truncated MeCP2-GST fusion proteins were expressed in *E.coli* XL-1 Blue cells. Protein samples were denatured in protein loading buffer, and resolved on a 10% SDS-polyacrylamide gel. DNA binding activity was determined by southwestern assay as described (7).

GST fusion proteins were purified on glutathione-Sephadex (Sigma) as follows (21). Log phase *E.coli* cells were induced with IPTG for 3 hours, cooled on ice and harvested by centrifugation at 5000 rpm for 10 min. The cell pellet was resuspended in extraction buffer (10mM sodium phosphate pH 6.8, 0.4M NaCl, 10mM mercaptoethanol and 0.1%Triton X-100), sonicated and centrifuged at 12,000 rpm for 10 min. Crude extract was mixed gently with hydroxyapatite (HAP) that had been pre-equilibrated in extraction buffer for 1 hour at 4°C. The HAP was then centrifuged at 3000 rpm for 5 min. Supernatant from the HAP step was diluted 4 fold with 50mM Tris−Cl pH 7.5, 0.1% Triton X-100 and GST fusion proteins were adsorbed by the addition of glutathione-Sephadex beads. The beads were washed with PBS 3 times and GST fusion proteins were eluted by addition of elution buffer (50mM Tris-Cl pH 8.0, 10mM reduced glutathione). Affinity-purified GST fusion protein was stored at −20°C.

To construct a non-fusion MBD expression plasmid, a PCR fragment containing MBD cDNA was inserted into pET-3c plasmid (22). MBD was overexpressed in *E.coli* BL21(DE3)LysS cells and was predominantly present in the insoluble fraction of the cell lysate. Most of bacterial protein was soluble in buffer A (50mM NaCl, 50mM sodium phosphate pH 7.0, 10% glycerol and 0.1% Triton-X100). The remaining bacterial protein was further extracted with 1M urea and 0.4M NaCl in buffer A. MBD was solubilized in 2M urea-0.8M NaCl in buffer A and renatured by dialysis against 3 successive 2-fold diluted buffers (1M urea-0.4M NaCl in buffer A, 0.5M urea-0.2M NaCl in buffer A and 0.25M urea-0.1M NaCl in buffer A) and buffer A alone. Partially purified MBD (up to approximately 10mg) was applied to a Mono-S HR 5/5 column (Pharmacia) (equilibrated with buffer A) at a flow rate of 0.5ml/min. The unbound protein was washed out with 5ml of buffer A. Then a 20ml linear gradient up to 0.6M NaCl was applied to the column, with a final step to 1M NaCl in buffer A, and fractions of 1ml were collected. MBD was eluted between 300 and 400 mM NaCl. The fraction containing MBD was dialyzed against 50mM NaCl-50mM sodium phosphate pH 7.0.

Bandshift assay

DNA-protein binding reactions were carried out in 20mM HEPES (pH 7.9), 1mM EDTA, 3mM $MgCl_2$, 10mM 2-mercaptoethanol, 10% glycerol and 0.1% Triton X-100. Labeled DNA was incubated with different amounts of protein on ice for 30min and then loaded on a 5% polyacrylamide gel buffered in 6.7mM Tris−HCl (pH 7.9), 3.3mM sodium acetate and 1mM EDTA for 4 hours at 100V in the cold.

DNase I footprinting

Plasmid pBR322 was methylated with M.HpaII (BioLabs) according to the manufacturers instructions. Methylated or nonmethylated plasmid was then digested with BamHI and labelled at its 3' end. Labelled DNA was either digested with SalI to generate a 318bp BamHI/SalI fragment that contains one HpaII site 158bp away from the 3' end, or digested with PstI to generate a 1125bp BamH I/PstI fragment which contains two HpaII sites that are 205 bp and 214 bp away from the 3'end. DNase I footprinting was carried out according to Brenowitz *et al.* (23). End-labeled DNA fragments were incubated with increasing amount of protein in 100 μl of bandshift buffer for 30 min at 0°C. DNase I (2μl of 0.1mg/ml) was added for a further 2.5 mins of incubation. Digestions were terminated by addition of 350μl of ethanol, 25μl of saturated ammonium acetate and 2.5μl of 1mg/ml tRNA. Precipitated DNA was washed with 70% ethanol and the dried DNA was resuspended in 10μl of loading buffer. Samples were heated to 80°C for 5 min before being applied to a 6% polyacrylamide-urea sequencing gel.

DNA methylation interference assay

Gel-purified single-stranded GAM1 (1μg) was labelled at its 5' end by T4 polynucleotide kinase and annealed with 4 μg of its complementary strand GTM1. Double stranded probe was modified with 1 μl of DMS for 2 min at room temperature in 200μl of DMS reaction buffer (50mM sodium cacodylate pH 8.0, 1mM EDTA) (24). The reaction was stopped by addition of 40 μl of DMS stop solution (1.5mM sodium acetate pH 7.0, 1M mercaptoethanol) and 600 μl of ethanol. DNA was precipitated, washed once with 70% ethanol and resuspended in TE. Bandshift gels were used to separate DNA-protein complexes from free probe. DNA was eluted from the acrylamide gel by incubation in 300 ml of elution buffer (0.5M ammonium acetate, 10mM $MgSO_4$, 1mM EDTA and 0.1% SDS) overnight. DNA was precipitated by ethanol, resuspended in 100 μl of 1M piperidine, and incubated at 90°C for 30 mins. Piperidine-cleaved DNA was lyophilized in a Speedvac evaporator. Lyophilization was repeated by adding 100μl of distilled water. Samples were dissolved in 10μl of formamide loading buffer and analyzed on a 6% polyacrylamide sequencing gel.

RESULTS

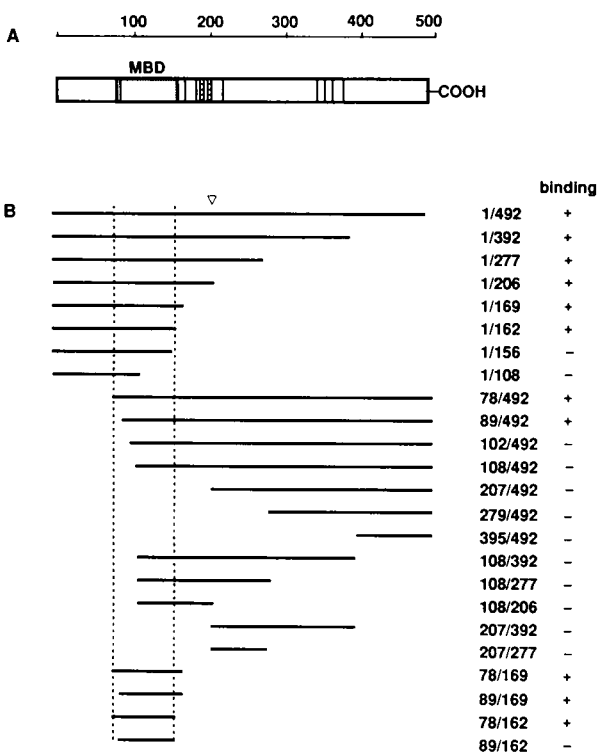
Delineation of the methyl-CpG binding domain by deletion analysis

A set of expression constructs was prepared by inserting intact and truncated versions of MeCP2 cDNA into the plasmid expression vector pGEX-3X. Truncated MeCP2 cDNA fragments were generated by restriction endonucleases and exonuclease III (see Methods) and are shown in Fig1B. The resulting MeCP2-glutathione transferase (GST) fusion constructs were transfected into *E.coli* XL-1 blue host bacteria for expression. Upon induction, most of the GST fusion proteins were well expressed as shown by Coomassie blue staining of whole cell protein after separation by SDS polyacrylamide gel electrophoresis (Fig 2A). The full length GST-MeCP2 fusion protein, however, was poorly expressed (Fig 2A, 1/492). The expression products migrate on gels more slowly than their molecular weights would predict. Slow migration is a feature of intact MeCP2, which has a molecular weight of 53kDa but co-migrates with markers of 81kDa (7).

To determine their DNA binding activity, proteins were electroblotted onto nitrocellulose membranes and incubated with radioactively labelled DNA probe (southwestern assay). Probes were 40bp in length and were either non-methylated (GAC), or methylated at multiple CpGs (GAM12; see Fig 3B; see reference 7). The southwestern assays showed that in the presence of excess non-methylated competitor DNA, binding to labelled GAM12 was unaffected by amino terminal deletions of 78 or 89 amino

acids (Fig 1; Fig 2B, lane16), but further truncation to amino acid 102 abolished binding (Fig 2B, lane17). Truncation from the carboxy terminus gave a boundary at position 162, beyond which binding to GAM12 was lost (Figure 2B, compare lanes 19 and 20). Neither intact nor truncated MeCP2's could bind to the non-methylated probe GAC under these conditions, confirming that the interaction under study is methylation-specific (data not shown).

It is apparent from Figure 2A and B that the Coomassie blue staining patterns of certain truncated fusion proteins do not match their southwestern assay patterns. Although the induced protein bands on polyacrylamide-SDS gels were sharp and of the predicted size, southwestern analysis in the presence of competitor DNA showed methylated DNA bound to a 48kDa protein band (open arrow), corresponding with a stained band that is common to lanes 1–3 (open arrow, Fig 2A). It appears that the intact GST fusion protein interacts relatively poorly with DNA, but



C SASPKQRRSI IRDRGPMYDD PTLPEGWTRK
LKQRKSGRSA GKYDVYLINP QGK***AFRSKV***E
LI***AYFEKV***GD TSLDPNDFDF TVTGR

Figure 1. Structure of MeCP2 and its deletions. **(A)** Diagrammatic structure of full-length MeCP2. MBD is represented by the shaded box. SPKK and (R)GRP(K) repeats are represented by single lines and shaded bars respectively. **(B)** Schematic summary of deletions of MeCP-2. Amino acids are numbered from the N terminus to the C terminus. Deletion numbers refer to the first and last amino acids of the remaining polypeptide. Methylated DNA binding activity is denoted by +, and absence of binding by – symbols. The region between the two vertical broken lines corresponds to the methyl-CpG binding domain (see deletion 78/162). An open arrow indicates the approximate position of a frequent break in the bacterially expressed protein (see text). **(C)** Amino acid sequence of the methyl-CpG binding domain (MBD). The sequence shows amino acids 78–162 inclusive, as defined by the deletion analysis (see text). One repeated motif is underlined and the other is italicized.

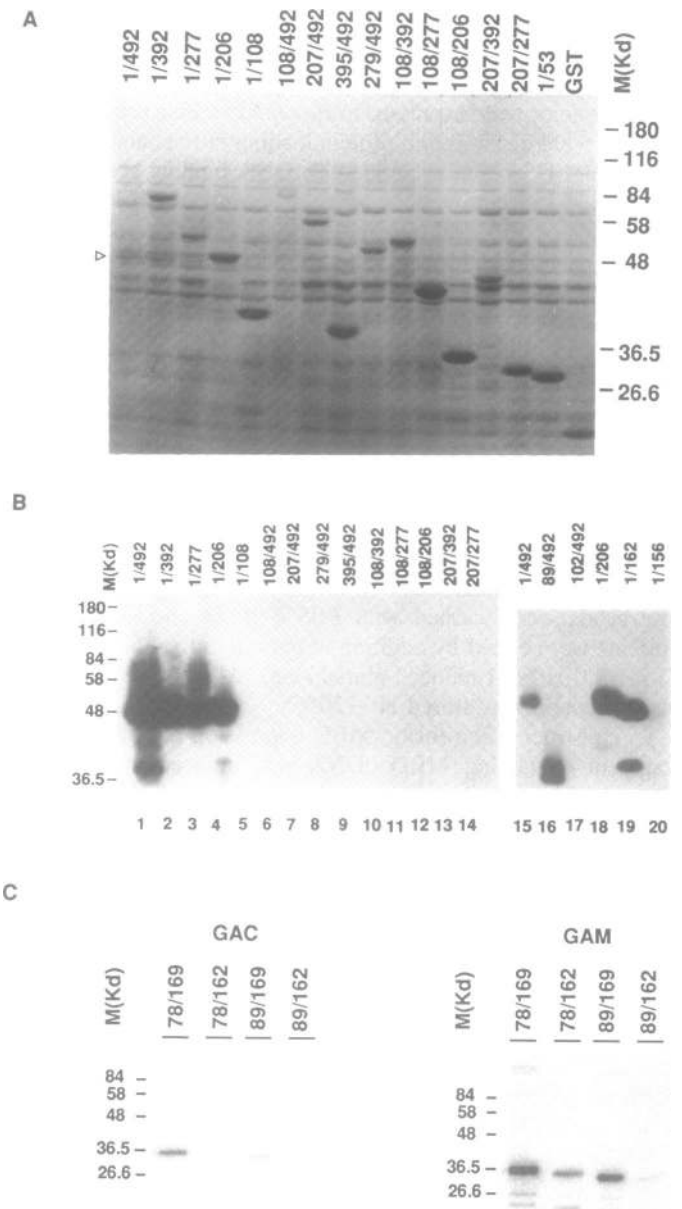


Figure 2. Expression and DNA binding of deleted MeCP2 molecules. **(A)** SDS-PAGE analysis of expressed GST fusion proteins. Cell lysates were prepared as described in Methods, separated by SDS polyacrylamide gel electrophoresis, and stained with Coomassie blue. Numbers above each lane correspond to the deletion numbers in Figure 1B. An open arrow shows the product of MeCP2 proteolysis in lanes 1–3 at approximately 48kDa. This product is responsible for most of the southwestern signal in lanes 1–3 of panel B (see text for explanation). **(B)** DNA binding analysis of truncated MeCP-2 fragments fractionated on a gel comparable, but not identical to the one shown in A. Induced cell lysates were separated by electrophoresis and electroblotted onto a nitrocellulose membrane. Binding reactions were carried out by using ^{32}P-labeled methylated GAM12 (see Fig 3B) in the presence of non-methylated competitor DNA. SDS-PAGE of the fusion proteins analysed in lanes 15–20 showed levels of expression comparable with those in (A) (data not shown). **(C)** Binding of N-terminal fragments of MeCP2 to methylated and non-methylated sequences in the absence of competitor. Fragments of MeCP2 as denoted above the lanes (see maps in Fig1B) were electroblotted as in (B) and incubated with multimers of ^{32}P-labeled GAM12 or GAC (see Fig 3B). Expression levels were comparable with those shown in (A) as judged by Coomassie Blue staining (not shown).

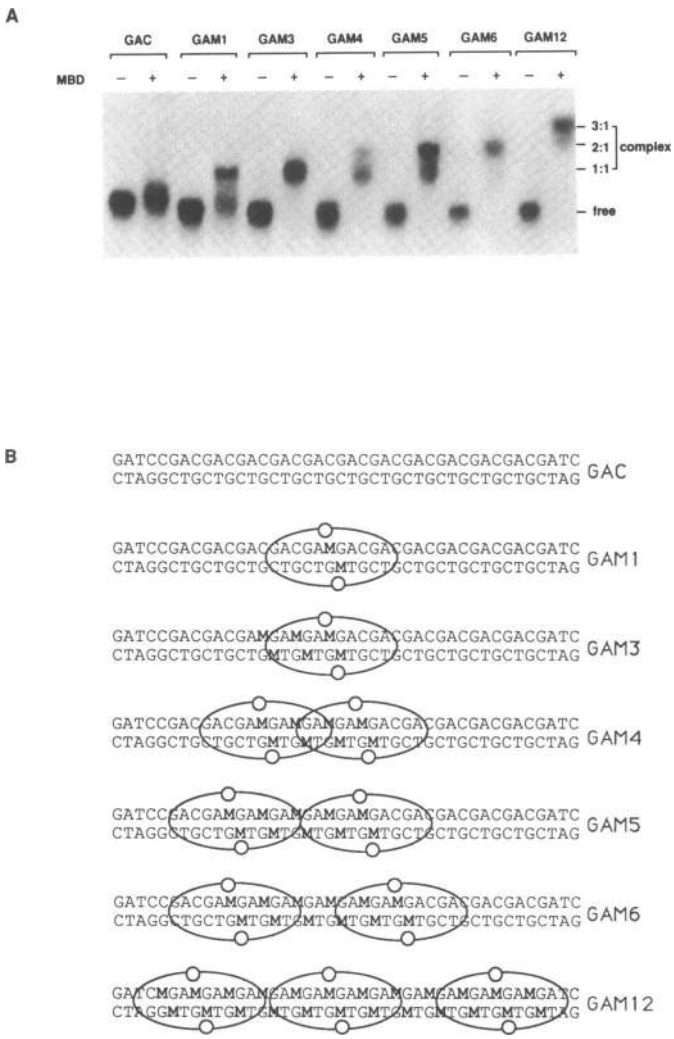


Figure 3. Methyl-CpG dependence of DNA-MBD binding. **(A)** Probes containing different numbers of methyl-CpG pairs (see below) were used in a bandshift assay. Complexes 1:1, 2:1 and 3:1 represent putative binding of one, two and three protein molecules to each molecule of DNA respectively. **(B)** A model of DNA-MBD binding to the GAM probes. M denotes 5-methylcytosine. The ellipsoid represents an MBD molecule and the small circles represents 5-methylcytosine binding sites on MBD. When methyl-CpG pairs are 5bp apart (e.g. GAM3), two molecules of MBD are unable to bind together. Limited double binding can occur when the CpGs are 8bp apart (GAM4), and the ternary complex (DNA plus two molecules of MBD) predominates when methyl-CpGs are more than 11bp apart (GAM5 and 6). The GAM12 probe has room for three molecules of MBD.

a degradation product can bind efficiently. From the migration of this fragment, and taking into account the abnormally high apparent molecular weights of these polypeptides (see above), we deduced that methylation-specific binding is associated with the N-terminal fragment generated by a reproducible break near to amino acid 200 of MeCP2. A protease-sensitive site close to this position has been seen previously in native MeCP2 (7).

Identification of a minimal methylated DNA binding domain

Deletion analysis indicated that binding to methylated DNA required the region between amino acids 89 and 162. A fragment corresponding to this interval (see Fig 1B, bottom line) was therefore expressed in *E.coli* cells as a GST fusion protein. Successful expression of the fusion protein was established by Coomassie blue staining after electrophoresis on SDS-polyacrylamide gels (not shown). DNA binding activity was tested by southwestern blot analysis in the absence of competitor DNA. The 89−162 polypeptide showed very weak binding to methylated DNA, and did not bind to non-methylated probe DNA (Fig 2C). Strong DNA binding activity was recovered when 11 amino acids from the N terminus (generating peptide 78−162) or 7 amino acids from the C terminus (generating peptide 89−169) were added back to the 89−162 peptide (Fig 1B and Fig 2C). Two of these fusion proteins (89−169 and 78−169) also bound to non-methylated DNA. The third (peptide 78−162), bound methylated DNA but not non-methylated DNA (Fig 2C). Thus peptide 78−162 corresponded to the minimal region of MeCP2 that could bind methylated DNA specifically without showing excessive non-specific DNA binding activity. This peptide was named MBD.

To facilitate structural and functional studies, we constructed a non-fusion MBD expression plasmid driven by the T7 promoter. A pair of oligonucleotides were designed that would amplify the MBD cDNA. The primers were also used to introduce an NcoI site plus an extra methionine codon at the 5′ end, and a BamHI site at the 3′ end. The NcoI/BamHI fragment was filled in and inserted into the blunted NdeI site of pET-3c. When expressed in *E.coli* cells, MBD was found predominantly in the insoluble fraction of the cell lysate. This material was collected by centrifugation, solubilised. Following renaturation, MBD was purified on a Mono-S column.

MBD binds DNA containing one methyl-CpG pair

The methyl-CpG dependent DNA binding of MBD was assayed by a bandshift experiment, using duplex probes containing different numbers of methyl-CpG pairs (Fig 3B). As shown in Fig 3A, MBD formed a complex with methylated DNA even when only one CpG pair was methylated (GAM1). A complex that migrated more slowly appeared when the number of methylated CpG pairs was increased to 4 (GAM4), and and a still slower complex was seen with a probe containing 12 methylated CpG pairs (GAM12). The likely interpretation of these results is that the three shifted bands correspond to DNA-MBD complexes containing one, two and three molecules of protein per molecule of DNA respectively. According to this model, one methyl-CpG pair is required for binding, and the protein covers approximately 12bp of DNA. Additional methyl-CpGs within this 12bp domain are not available to be bound by another molecule of MBD. This explains why GAM3 can only form a 1:1 complex although it contains 3 pairs of methyl-CpG. Fig 3B summarizes the interpretation of the bandshift results.

DNase I footprinting with MBD

DNase I footprinting was performed to locate the specific binding sites of MBD on DNA. An end-labelled fragment of the plasmid pBR322 was methylated at a CCGG site 158bp upstream of the 3′ end, and incubated with purified MBD. Fig. 4A shows that the addition of increasing amount of MBD resulted in the appearance of a protected region encompassing the methylated CpG (Fig. 4A lanes b−g). No footprint was seen on non-methylated DNA (Fig. 4A i−j). A similar protection pattern was seen when the complementary strand was labelled (data not shown; see diagram Fig 4C). The protected region of 12−14 nucleotides for each methyl-CpG pair fits with the model suggested by the bandshift data (compare Figs 3B and 4C). This

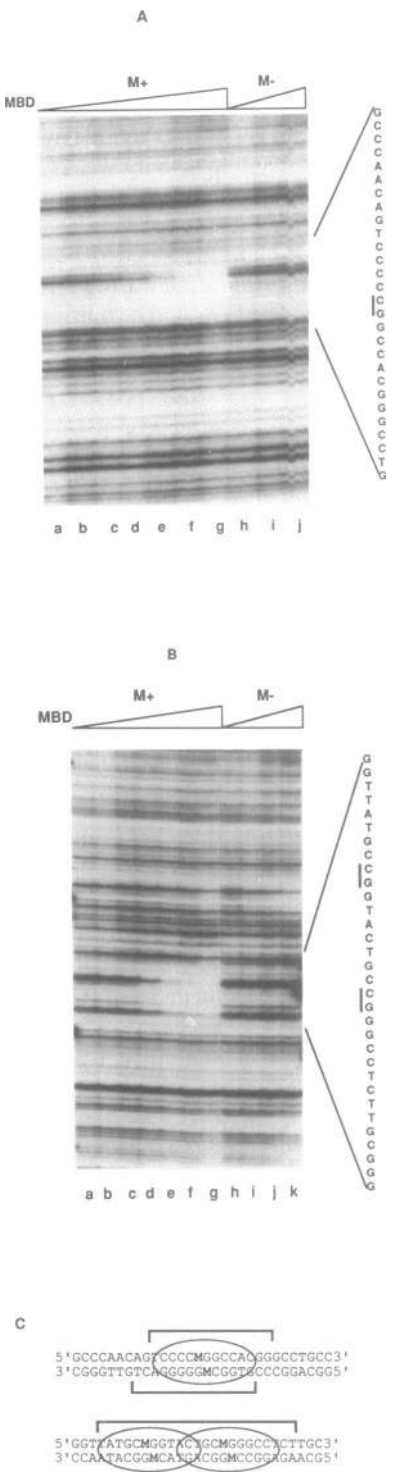


Figure 4. DNase I footprinting of MBD. **(A)** A BamHI/SalI fragment of pBR322 containing one HpaII site was used as probe. M+ and M− indicate that HpaII site is methylated or non-methylated respectively. Lanes a−g contain 0, 0.312, 0.625, 1.25, 2.5, 5 and 10ng of MBD. Lanes h-j contain 0, 5 and 10ng of MBD. **(B)** A BamHI/PstI fragment of pBR322 containing two HpaII sites was used as probe. Lanes a−g contain 0, 0.312, 0.625, 1.25, 2.5, 5 and 10ng of MBD. Lanes h−j contain 0, 2.5, 5 and 10ng of MBD. **(C)** Schematic summary of footprinting results. Ellipsoids represent MBD molecules, and brackets show the extent of nuclease protection. The fragment containing two methyl-CpGs was footprinted on one strand only.

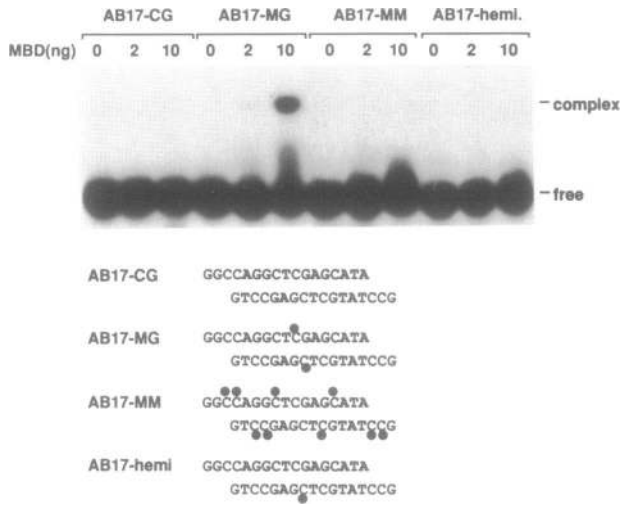


Figure 5. Binding specificity of MBD. DNA binding activity was detected by band-shifts using probes with the same nucleotide sequence, but different configurations of methylation. AB17-CG, unmethylated; AB17-MG, methylated at a single CpG pair; AB17-MM, containing multiple methylcytosine residues *except* at the CpG pair; AB17-hemi, methylated at one cytosine of the CpG pair. Black dots indicate the presence of 5-methylcytosine. Bandshifts were carried out in the absence of competitor using the amounts of MBD shown above the lanes.

was confirmed by footprinting over a BamHI/PstI fragment of pBR322 containing two HpaII sites 9bp apart. In this case, the footprint was 23bp in length (Fig 4B), again consistent with a protected region of 12 nucleotides at each CpG (see diagram Fig 4C).

We estimated the dissociation constant of the protein-DNA complex using data from the DNaseI footprinting experiment. The methyl-CpG site was approximately 50% protected against DNaseI at an MBD concentration of 1.25×10^{-9}M, indicating a dissociation constant close to this value (Fig 4A, lane d; and Fig 4B, lane d). Non-methylated DNA showed no protection against DNaseI at an MBD concentration 1.0×10^{-8}M. We conclude that binding to non-methylated DNA is at least an order of magnitude less strong than that for methylated DNA.

MBD binding requires symmetrically methylated CpG

Purified intact MeCP2 binds to DNA molecules containing symmetrically methylated CpG, but 5-methylcytosine in other sequence contexts does not bind to MeCP2 (8). To investigate whether the MBD peptide retains specificity for methyl-CpG pairs, a set of probes were obtained that had the same base sequence, but different patterns of methylation. The probes were non-methylated (AB17-CG); methylated only at one CpG pair (AB17-MG); methylated at all cytosine's *except* the CpG pair (AB17-MM); and hemimethylated at CpG (AB17-hemi). As shown in Figure 5, only AB17-MG gives a complex with MBD. Thus MBD retains the specificity for symmetrical methyl-CpG that is characteristic of the intact protein.

Methylation of G in the CpG site does not interfere with DNA binding

A methylation interference assay was carried out to determine the effect of G modification on MBD binding (13). Guanines in the probe GAM1 (Fig 3B) were partially methylated by DMS, and the modified probe was incubated with MBD. Bound and

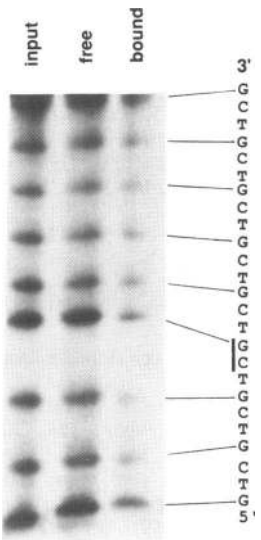


Figure 6. Methylation of G does not interfere with MBD binding. ^{32}P-labeled GAM-1 (see Fig 3) was treated with DMS to partially methylate guanine at the N7 position. Bound and unbound probe were separated by a bandshift experiment and, together with input modified DNA, were treated with piperidine to cleave DNA at DMS-modified guanine. The resulting labeled fragments were analysed on a DNA sequencing gel.

free probes were separated on a polyacrylamide gel and treated with piperidine to cleave the phosphodiester backbone at 7N-methyl-guanine. The cleavage patterns of free probe and bound probe were then compared by electrophoresis on a DNA sequencing gel. No difference was found between bound and free probes. In particular, DNA modified at the G that is base-paired with 5-methyl cytosine in the binding site was not under-represented in the bound fraction (Fig 6). An alternative assay by DMS footprinting did not detect methyl-CpG protection by MBD either (data not shown). The results of both DMS methylation interference and footprinting assays indicate no direct contact on the major groove between MBD residues and the 7N position of the guanine of methyl-CpG. However, the possibility that MBD interacts with 5-methylcytosine via the major groove is not excluded.

MBD binds DNA as a monomer

Proteins that bind DNA often do so as homo- or hetero-dimers. Dimerization is particularly common when the binding site shows twofold rotational symmetry. The methyl-CpG pair, though short, is symmetrical. To determine whether MBD binds as a dimer, a GST-MBD fusion protein and an unfused MBD were used separately and together in a bandshift assay. Using GAM1 as probe (one pair of methyl-CpGs; see Fig 3B), the fusion protein gave a complex with low mobility on an acrylamide gel (lanes 2 and 3), whereas the non-fusion protein gave a high-mobility complex (lane 4). When GST-MBD and MBD were mixed, the fast and slow complexes were both produced (lanes 5 and 6), but no intermediate complex, indicative of a heterodimer between the two proteins, was seen. Similar results were obtained when GST-MBD fusion protein was partially cleaved with factor Xa. Once more only two complexes were formed, presumably corresponding to GST-MBD and cleaved MBD respectively (data not shown). The results indicate that MBD binds DNA as a monomer.

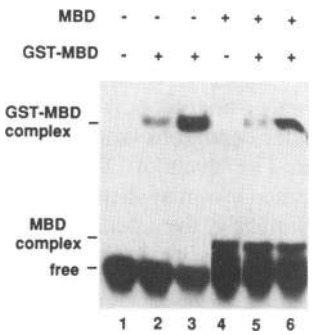


Figure 7. MBD binds to DNA as a monomer. A probe containing one methyl-CpG pair (GAM1; see Fig 3) was used in a band-shift assay in the presence (+) or absence(−) of GST-fusion MBD (GST-MBD) and non-fusion MBD. Lanes 2 and 5 contain 3ng of GST-MBD; lanes 3 and 6 contain 6 ng of GST-MBD; lane 4, 5 and 6 contain 1 ng of non-fusion MBD. The absence of a complex of intermediate size when both MBD and GST-MBD are used together indicates that each form binds as a monomer.

DISCUSSION

The methyl-CpG binding domain of MeCP2

MeCP2 is a chromatin-associated nuclear protein that binds to DNA containing methyl-CpG in a variety of sequence contexts. The purified protein also has a significant non-specific affinity for DNA. By constructing deletions of MeCP2, we found that amino acids 78−162 of MeCP2 contain the minimal region (MBD) required for binding to methylated DNA, while lacking its non-specific affinity for DNA. The amino acid sequence of MBD contains no recognisable DNA binding motifs. Intact MeCP2, however, contains several short motifs that have been seen in proteins that bind to the minor groove of AT-rich DNA sequences (reference 9; see Fig 1A). These motifs, which are clustered in the central region of the protein, are excluded from MBD (the SPKK motif at the extreme N-terminal end is not essential for specificity; data not shown). It is likely that they contribute to non-specific DNA binding, as the central region of MeCP2 binds DNA regardless of methylation (data not shown).

The methyl-CpG binding characteristics of MBD are similar to those of intact MeCP2. Duplex DNA with only one methyl-CpG pair is bound, but there is negligible binding to DNA containing hemimethylated CpG, or 5-methylcytosine at non-CpG sites (Fig 5). By these criteria, it appears that MBD folds appropriately in the absence of the remainder of the protein. Bandshifts and DNase I footprinting indicate that a 12 nucleotide region surrounding the methyl-CpG pair is covered by MBD. At least 4bp of DNA are required flanking the methyl-CpG (data not shown), suggesting that MBD contacts the adjacent DNA backbone as well as the methylated bases themselves.

Absence of footprints *in vivo*

MeCP2 is a DNA binding protein *in vitro*, and it co-localises with chromosomes, but as yet no *in vivo* footprints comparable with those shown in Fig 4 have been detected. Sequences that might have been expected to show binding are the methylated human PGK1 promoter (14, 15) and purified mouse satellite chromatin (16). Methylated sites in the PGK promoter did not show obvious footprints, and MeCP2 did not appear to co-purify with satellite chromatin. At face value, these results suggest that the protein is not bound. It is possible, however, that bound

MeCP2 is lost under the conditions used in these experiments. Meehan et al (reference 8) showed that a large fraction of chromatin-bound MeCP2 is released by mild treatment with restriction enzymes or micrococcal nuclease. Extensive release occurred even under conditions where the DNA remained high molecular weight. The levels of DNaseI used for footprinting of the PGK1 region *in vivo* may therefore have rapidly displaced bound MeCP2. Similarly, the extensive nuclease digestion used to prepare mouse satellite chromatin would be expected to cause loss of MeCP2 binding (16). Thus the location of bound MeCP2 in chromatin remains an open question.

The molecular basis of MBD binding

Proteins that bind to DNA sequences with twofold rotational symmetry often bind as dimers, with each monomer contacting one half of the dyad (17). The symmetry of the methyl-CpG pair raised the possibility that MeCP2 also requires dimerization for recognition of its binding site. The apparent absence of helix-loop-helix, helix-space-helix or leucine-zipper motifs in MeCP2 did not preclude this possibility, but our results indicate that no heterodimer complex is formed when MBDs of distinguishable size are mixed with probe DNA (Fig 7). Our ability to use the southwestern assay for MeCP2, in which proteins are immobilised following SDS gel electrophoresis, also argues that dimerisation is not required for binding to methylated sites.

In view of the symmetry of the methyl-CpG pair, we have looked for motifs within the MBD monomer that also showed potential symmetry. The sequence contains two copies of the sequence KQRR(K)S and two copies of AF(Y)XXKV (Fig1C). Deletion of the first N terminal copy of KQRRS abolished the methylated DNA binding activity, but activity could be restored by addition of the 7 amino acid sequence GSPSRRE to the C terminus of MBD. We conclude that the highly basic KQRRS repeats do not contribute to the specific recognition of the methyl-CpG pair, but may be involved in binding to the sugar phosphate backbone of DNA. The two AF(Y)XXKV repeats are separated by three amino acids within a 15 amino acid sequence that has potential to form a-helix. Whether this region indeed forms α-helix, and whether it contains side chains that interact with methyl cytosine residues remain to be determined.

In considering the molecular basis of binding specificity, it is of interest that contacts with the 5-methyl group of thymidine are frequently mediated by alanine. For example, structural analysis of the yeast transcriptional activator GCN4 showed that the alanines at positions 238 and 239 are in van der Waals contact with thymine 5-methyl groups at position +3 and +1 of DNA (18, 19). It is not yet known whether the three alanines in MBD are required for specificity. Current efforts to construct point mutations in MBD and to determine its 3-dimensional structure should to shed light on this matter.

ACKNOWLEDGEMENTS

We thank Frank Johnson for photography, Jillian Charlton for expert technical assistance, the ICRF DNA synthesis service for oligonucleotides, Sally Cross and Stuart MacNeill for advice on expression and purification of fusion proteins. Sally Cross, Francisco Antequera and Eric Selker suggested improvements to the manuscript. This work was supported by The Wellcome Trust and by the Imperial Cancer Research Fund. X.N. is a Wellcome Postdoctoral Fellow, and R.R.M. is a member if the ICRF Epigenetics Laboratory. A.B. is a Howard Hughes International Scholar.

REFERENCES

1. Selker, E. (1990) TIBS *15*, 103–107.
2. Lewis, J. and Bird, A. (1991) FEBS *285*, 155–159.
3. Meehan, R., Lewis, J., Cross, S., Xinsheng Nan, Jeppessen, P. and Bird, A. (1992) J. Cell Sci. Suppl *16*, 9–14.
4. Antequera F., Macleod D. and Bird A. (1989) Cell *58*, 509–517.
5. Meehan, R., Lewis, J., Mckay, S., Kleiner, E.L. and Bird, A. (1989) Cell *58*, 499–507.
6. Boyes, J. and Bird, A. (1991) Cell *64*, 1123–1134.
7. Lewis, J., Meehan, R., Henzel, W., Maurer-Fogy, I., Jeppesen, P., Klein, F.and Bird, A. (1992) Cell *69*, 905–914.
8. Meehan R., Lewis J. and Bird A. (1992) Nucleic Acid Research *20*, 5085–5092.
9. Miller, O.L., Schnoedl, W., Allen, J. and Edlanger, B.F. (1974) Nature, *251*, 636–637.
10. Johnson, K., Lehn, D. and Reeves, R. (1989) Mol. Cell. Biol. *9*, 2114–23.
11. Reeves, R. and Nissen, M.S. (1990) J. Bio. Chem. *265*, 8573–8582.
12. Churchill, M.E.A. and Suzuki, M. (1989) EMBO J. *8*, 4189–95.
13. Hendrickson,W. and Schleif, R. (1985) Proc. Natl. Acad. Sci. U.S.A. *82*, 3129–33.
14. Pfeifer,G.P.and Riggs, A.D.(1991) Genes Dev 5,1102–1113.
15. Pfeifer,G.P., Tanguay,P.L., Steigerwald, S.D. and Riggs, A.D. (1990) Genes Dev. 4, 1277–1287.
16. Zhang, X-Y. and Horz, W. (1984) J. Mol. Biol. 176, 105–129.
17. Lamb, P. and McKnight, S.L. (1991) TIBS *16*, 417–22.
18. Blatter, E.E., Ebright, Y. W. and Ebright, R.H. (1992) Nature *359*, 650–652.
19. Ellenberger, T.E., Brandl, C. J. Struhl, K. and Harrison, S.C. (1992) Cell *71*, 1223–1237.
20. Smith, D.B., Davern, K. M., Board, P.G., Tiu, W.U., Garcia, E.G. and Mitchell, G. F. (1986) Proc. Natl. Acad. Sci. USA 83, 8703–8707.
21. Smith, D.B. and Corcoran, L.M. (1990) In Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (eds), Current Protocols in Molecular Biology. John Wiley and Sons, New York, pp 16.7.1–8.
22. Studier, F.W., Rosenberg, A. H., Dunn, J. J. and Dubendorff, J. W. (1990) Methods Enzymol. 185, 60–89.
23. Brenowitz, M., Senear, D.F., Kingston, R. E. (1991) In Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (eds), Current Protocols in Molecular Biology. John Wiley and Sons, New York, pp12.4.1–16.
24. Baldwin, A.S. (1991) In Ausubel, F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl, K. (eds), Current Protocols in Molecular Biology. John Wiley and Sons, New York, pp12.3.1–6.