5786–5793 Nucleic Acids Research, 1993, Vol. 21, No. 24

A nuclear protein with enhanced binding to methylated Sp1 sites in the AIDS virus promoter

Peteranne Joel, Wei Shao and Karen Pratt*
Department of Microbiology and Molecular Genetics, University of Vermont, Burlington, VT 05405, USA

Received March 9, 1993; Revised and Accepted October 20, 1993

ABSTRACT

We report here the discovery of HMBP, a protein in nuclei of human T-helper lymphocytes and other human cell types, which binds with enhanced affinity to a promoter element in the HIV-1 long terminal repeat when that element is methylated at CpGs, the target site of the human DNA methyltransferase. This promoter element contains three (degenerate) binding sites for Sp1, a general activator of transcription. Gel shift assays and footprinting experiments indicate that HMBP binding overlaps two of these methylated Sp1 sites. Although HMBP binds these methylated Sp1 sites, it does not bind consensus Sp1 sites. Competition studies, differences in binding site specificities, binding conditions, and, in some cases, chromatographic separation further distinguish HMBP from Sp1 and from each of four previously identified methylated-DNA binding proteins. HMBP binds hemi-methylated DNA in a strand dependent manner. These binding characteristics suggest that HMBP may recognize newly replicated DNA and thereby play a role in differentiation. If HMBP is able to compete with Sp1 for binding at methylated, non-consensus Sp1 sites *in vivo* and repress transcription, it may play a role in AIDS latency.

INTRODUCTION

Vertebrate DNA is methylated at cytosines in the relatively rare dinucleotide CG. Although CGs are scarce in coding sequences, gene promoters are usually enriched in CGs. There is overwhelming correlative data between the presence of methylated promoter sites and transcriptional repression. Furthermore, experimental demethylation by 5'azacytidine has been shown to activate gene expression. The clearest demonstrations that DNA methylation represses transcription have been transfection experiments where expression of genes with methylated promoters was greatly repressed relative to those with unmethylated promoters. (The role of DNA methylation in gene repression is reviewed in 1–7).

DNA methylation is thought to repress transcription by one or both of two mechanisms. In one mechanism, DNA methylation masks the recognition site of a sequence-specific binding protein which acts as a trancriptional activator. This has been shown for a number of *trans*-activating factors including the myc/myn heterodimer (8), the cAMP responsive element binding protein CREB (9), NFκB (10), AP2 (11), EF2 (12, 13), MLTF (14), EBP-80 (15), and a factor required for tyrosine aminotransferase expression (16). In the second mechanism, DNA methylation creates a recognition site for a transcriptional repressor. Such a repressor might prevent binding of a trancriptional activator by competing for the same binding site or prevent binding of the initiation complex. Transcription decreases with methylation density at proximal promoters (17–19). DNA methylation at some promoters, including the HIV-1 LTR (19), represses transcription indirectly consistent with the requirement for a methylated-DNA binding protein (17–21). Thus far, two methylation-specific binding factors have been shown to repress transcription, MeCP1 (20) and MDBP-2-H1, a species of avian histone H1 (21).

DNA methylation may mediate repression of HIV-1 in latently infected CD4+ cells. In the early stages of AIDS, many of these cells are latently infected with HIV-1. Viral DNA is present but is not being transcribed (22, 23). Transcription of HIV-1 is controlled by promoter sequences contained in its 5' long terminal repeat (LTR). In cells stably transfected with plasmid constructs where the HIV-1 promoter controls expression of the reporter gene for chloramphenicol acetyl transferase (CAT), demethylation of the HIV-1 promoter by the methylase inhibitor 5-azacytidine activates transcription of the associated CAT gene (24). Methylation of the LTR in LTR-CAT constructs by prokaryotic methylases represses transcription in transfection assays (25). Moreover, methylation at the CG between the binding sites for NFκB, precludes binding by NFκB (10), an activated transcription factor in lymphocytes. The NFκB sites and the three Sp1 sites comprise the HIV-1 proximal promoter, a region of predominant importance in the activation of HIV-1 transcription (26).

The transcription factor Sp1 is required for the expression of many genes including those under the control of the HIV-1 promoter (27). The highest affinity binding sites for Sp1 contain a CG dinucleotide (28, 29). The three Sp1 sites in the HIV-1 promoter each contain a CG dinucleotide. Mutation of all three

* To whom correspondence should be addressed

sites greatly decreases transcription (27). Although Sp1 binds DNA and activates transcription *in vitro* even when the binding site is CG methylated (30,31,32), transcription *in vivo* is inhibited when Sp1 sites are methylated (31, 33).

In this paper we report the identification and binding characteristics of a putative repressor protein which preferentially binds sites in the Sp1 region of the HIV-1 promoter when the CGs in this sequence are methylated. HMBP (HIV-1 methylated DNA binding protein) is present in nuclear extracts prepared from cultured human T-helper cells and HeLa cells and can be distinguished from Sp1 and each of the four previously characterized methylated DNA binding proteins.

MATERIALS AND METHODS

Cell culture and nuclear extracts

HuT 78 cells (obtained from the ATCC and from Dr Robert Gallo through the AIDS Research and Reference Reagent Program, NIAID) and HPB-ALL cells (34) were cultured as suppliers recommend or in AIM V (Bethesda Research Laboratories). Nuclear extract was prepared as described in (35) or as described in (34, 36) but in buffer HDIE: 30 mM HEPES (pH 7.5), 0.25 % myoinositol, 0.25 mm EDTA, 5mm PMSF, 0.8 μg/ml pepstatin A, 1 mM DTT and 10 % glycerol as described (30, 32), dialysed in buffer HDIE with 0.2 M KCl and fractionated by Bio-Rex 70 (BioRad) chromatography. The 0.2M KCl flow through fraction contained HMBP. This flow-through was further fractionated for gel shifts with monomethylated probes: For the gel shift assay shown in Figure 4, a 25 mM-200mm KCl fraction containing HMBP was used. (The 25 mM wash contained a methylated DNA binding activity that did not bind fully methylated 43-mer.) HeLa cell nuclear extract was obtained from G. Gilmartin prepared as described (37). Although HMBP is only 2 fold purified by Bio-Rex chromatography over crude nuclear extract, this step effectively removes other proteins which bind the 43-mer on a gel shift assay. Fractions eluted with buffer containing 0.4 M KCl contained Sp1. Nuclear extract was precipitated with 50% ammonium sulfate and used for wheat germ agglutinin affinity chromatography as described (38).

CAT assay

The HIV-1 LTR-CAT plasmid, pBenn-CAT, was obtained from Malcolm Martin (39). After restriction by Bam HI and Pst I, a 3300 bp fragment containing the entire LTR and CAT gene was methylated by Steven Smith and Julie Kan (City of Hope Medical Center, Duarte CA) with methylase they purified from human placenta (40). We transfected the LTR-CAT 3300 bp fragment by electroporation into Hut 78 cells as described (41), and assayed for CAT activity after 72 h. as described (42) except that enzyme reaction was allowed to proceed overnight. Overnight assay allows detection of CAT activity in the absence of HIV tat activity. CAT activity was quantified using a Betascope model 603 blot analyser (Betagen Corp.) and is reported as % of total chloramphenicol converted to acetylated products.

Oligonucleotides, gel-shift probes, and proteins

Oligonucleotides were synthesized on an Applied Biosystems model 392 synthesizer using 5-methyldeoxycytosine and the other four bases as phosphoramidite derivatives. After gel purification, oligonucleotides were labeled at the 5′ end of either the top or bottom strand with T4 polynucleotide kinase and annealed with 50% excess of unlabeled, complementary strand. This ensured that gel shift bands did not contain single stranded DNA probe. Dimerized and multimerized competitor DNAs were prepared by ligation of phosphorylated, annealed, sticky-ended, oligonucleotides using T4 DNA ligase with buffer supplied by the manufacturer (US Biochemicals). DNA samples were analysed by gel electrophoresis before use. Histone H1 was purchased from Boehringer Mannheim and shown by SDS-PAGE to be intact. Sp1(15 ng protein/μl) was purchased from Promega.

Gel shift analyses

Binding reactions contained 0.07−0.7 ng kinase-labeled, double-stranded methylated, hemi-methylated, or unmethylated 43-mer, 80−500 fold excess poly (dI-dC) (Pharmacia), 75 mM KCl, 10 mM Tris−HCl, pH 7.5, 0.25 mM EDTA, 1mM DTT, 10% glycerol, varying amounts of competitor DNA, and 2−4 μg of dialysed nuclear extract. Protein concentration was determined by the Bradford assay (BioRad) with IgG as standard. After 20′ on ice, binding reactions were electrophoresed through 5% PAGE in 0.5×TBE for 1 hr at 15 mA at RT. Gels were dried and autoradiographed with an intensifying screen at −80°C. Radioactivity in bands in dried gels was quantified by a Betascope model 603 blot analyser (Betagen Corp.). For quantifying competition experiments, after subtracting background, the radioactivity in competed HMBP complexes was computed as a percentage of the radioactivity in non-competed HMBP complexes in the same experiment. Competition is plotted as a function of the mass excess of competitor DNA.

Two-dimensional DNase 1 footprinting

94-mer top strand: 43-mer was ligated into the Sma I site in the polylinker of pBS plasmid (Stratagene) and cloned in JM83 cells. Purified plasmid DNA was cut with Eco RI, methylated with excess SssI methylase (New England Biolabs) in buffer containing EDTA as recommended by the supplier and an aliquot analysed for completeness of methylation by digesting with Hpa II. Plasmid DNA was then cut with Hind III to release a 94 bp fragment containing the 43-mer. Labeling at the 3′ end of the top strand (5′ to 3′ reading towards the start of transcription in the HIV 1 LTR) was done at the Hind III site with Klenow and [α]^{32}P dCTP. 43-mer top or bottom strand was kinase labeled before annealing. After scaled-up gel shift binding reactions were performed, reactions were made 1 mM in $CaCl_2$ and treated with 2.9 μg DNase I (5 ng/μl) for 1′ at RT for the 94-mer and at 10−80 ng/μl for 43-mer and stopped with EDTA before gel electrophoresis. DNAs from both the HMBP shifted bands and from the free probe bands were isolated from the gel and purified. Equal cpm of DNAs were electrophoresed through a 10% (94-mer) or a 17 % sequencing gel (43-mer) before autoradiography.

Two-dimensional DMS footprinting

43-mer was kinase-labeled in the top strand before annealing in a HEPES buffer. (The bottom strand is too G-poor for DMS cleavage analysis.) After a scaled-up binding reaction, 3 μl DMS was added to the 150 μl vol binding reaction containing 250 μg protein and incubated at 0°C for 2.5′. The reaction was stopped with 0.2 M 2-mercaptoethanol before electrophoresis. DNAs from the HMBP shifted band and from the free probe band were isolated from the gel and purified. Equal cpm of DNAs were electrophoresed through a 10% sequencing gel before autoradiography.

RESULTS

DNA methylation by the human methylase represses transcription regulated by the HIV-1 LTR

An LTR-CAT reporter construct was methylated *in vitro* by the laboratory of S.Smith using DNA methylase they purified from human placenta (40). Approximately 12% of the CGs became methylated. After duplicate transfections into cultured Human T helper lymphocytes in parallel with unmethylated constructs and mock transfections without plasmid, cell extracts were analysed for CAT activity. Figure 1 shows that methylated constructs expressed very low levels of CAT activity not much more than mock transfection controls, whereas unmethylated constructs expressed high levels of CAT activity in the same experiment. Since methylation of coding sequences has no effect on transcription (1−7), these transfection experiments suggest that methylation of the HIV-1 LTR by the human methyltransferase represses transcription of an associated gene.

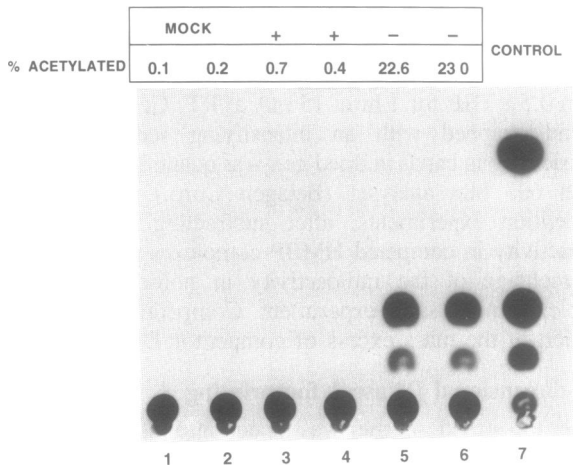

Figure 1. Methylation of an HIV-1 LTR-CAT reporter construct represses expression of CAT activity. HuT 78 cells were transfected in duplicate with HIV-1 LTR-CAT reporter constructs and assayed for CAT activity: mock transfections, lanes 1 and 2; methylated with human methylase *in vitro*, lanes 3 and 4; unmethylated, lanes 5 and 6; positive control with CAT enzyme, lane 7.

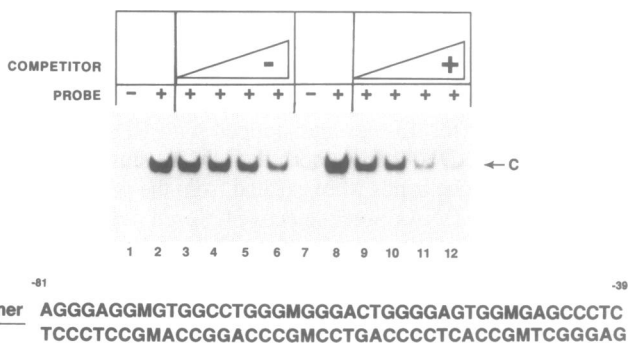

Figure 2. A nuclear factor preferentially binds the HIV-1 Sp1 region when CGs are methylated. Gel mobility band shift assay of HMBP binding to 43-mer methylated at all three CGs. lanes 1 and 7 show binding to unmethylated 43-mer (−). Gel shift assays in lanes 3, 4, 5, and 6 contain unlabeled, unmethylated 43-mer as competitor in 5.6, 11, 22, and 44 fold weight excess to probe, respectively. Lanes 9 through 12 show reactions similar to those in lanes 3−6 but with methylated 43-mer as competitor. Shifted complex (<C); methylated 43-mer probe, (+) lanes; unmethylated 43-mer probe (same sequence but with unmethylated cytosines, (−) lanes. The sequence of methylated 43-mer probe is shown at bottom; M = methylated cytosine.

A nuclear factor preferentially binds the HIV-1 Sp1 region when Sp1 sites are methylated

We used gel shift assays to determine if nuclear extracts from cells capable of being infected by HIV-1 contained factors which discriminate between methylated and unmethylated HIV-1 promoter sequences. We found a methylation-specific binding factor, HMBP, in nuclear extract of CD4+ T-helper cells and HeLa cells. Other cell types have not yet been assayed for HMBP activity. Column chromatography, using a weak cation exchange resin (BioRad's Bio-Rex 70) results in two-fold purification of binding activity and separates HMBP from other factors which bind the same probe regardless of its methylation state. Gel shift assays using this partially purified nuclear extract show increased binding to a 43 base pair region of the HIV-1 promoter when the three CGs in this sequence are methylated (Figure 2, compare lanes 1 and 7 to lanes 2 and 8, respectively). There is strong competition by cold methylated 43-mer as specific competitor DNA (Figure 2, lanes 9−12) suggesting that binding of HMBP to methylated 43-mer is sequence specific. There is weak competition by unmethylated 43-mer relative to competition by methylated 43-mer for the shifted DNA/protein complex formed with methylated 43-mer (Figure 2, compare lanes 3−6 with lanes 9−12). In this gel shift experiment, competition by a 44-fold excess of unmethylated 43-mer, in the presence of 550 fold excess poly dI·dC, decreased binding to 42% of control, whereas 44-fold excess of methylated 43-mer competitor reduced binding to 16% of control (Figure 2, lane 6 versus lane 12). Poly dI·dC was used as non specific competitor DNA. The sequence of the 43-mer is shown at the bottom of Figure 2.

With calf thymus DNA or *E.coli* DNA as non specific competitor DNA, there is no HMBP gel band (not shown). This implies that calf thymus DNA and *E.coli* DNA contain (methylated) sequences which bind HMBP. HMBP is protease sensitive and RNase resistant (not shown). Quantitation of HMBP binding in this and other gel shift experiments shows 4−7 fold preference for methylated 43-mer versus unmethylated 43-mer.

HMBP footprints the Sp1 region of the HIV-1 LTR

In order to more precisely determine the DNA sequence bound by HMBP, DNase I and DMS footprinting experiments were used to map DNA sequences in the LTR which are protected from these reagents by HMBP binding. After treatment with either DNase 1 or DMS, binding complexes were isolated from gels before the DNAs were analysed by sequencing gels. Two-dimensional footprinting insures that only complexes due to HMBP binding are analysed. DNase 1 was used to footprint HMBP binding to both the top and bottom DNA strands (5′ to 3′ going toward the transcription start site in the HIV-1 LTR as shown in Figure 2) of the methylated Sp1 region of the HIV-1 LTR. For Figure 3A, the 43-mer was cloned into the multicloning region of a plasmid, methylated at every CG, cut out of the plasmid as a 94-mer and the top strand 3′ end-labeled at a site distal to the end of the 43-mer. This fragment was used as the binding probe for the Bio Rex70 flow through fraction of nuclear extract containing HMBP. After binding, the reaction was treated with DNase 1 and then gel shifted. As shown in Figure 3A, there is a DNase 1 hypersensitive site near the 5′ methylated cytosine

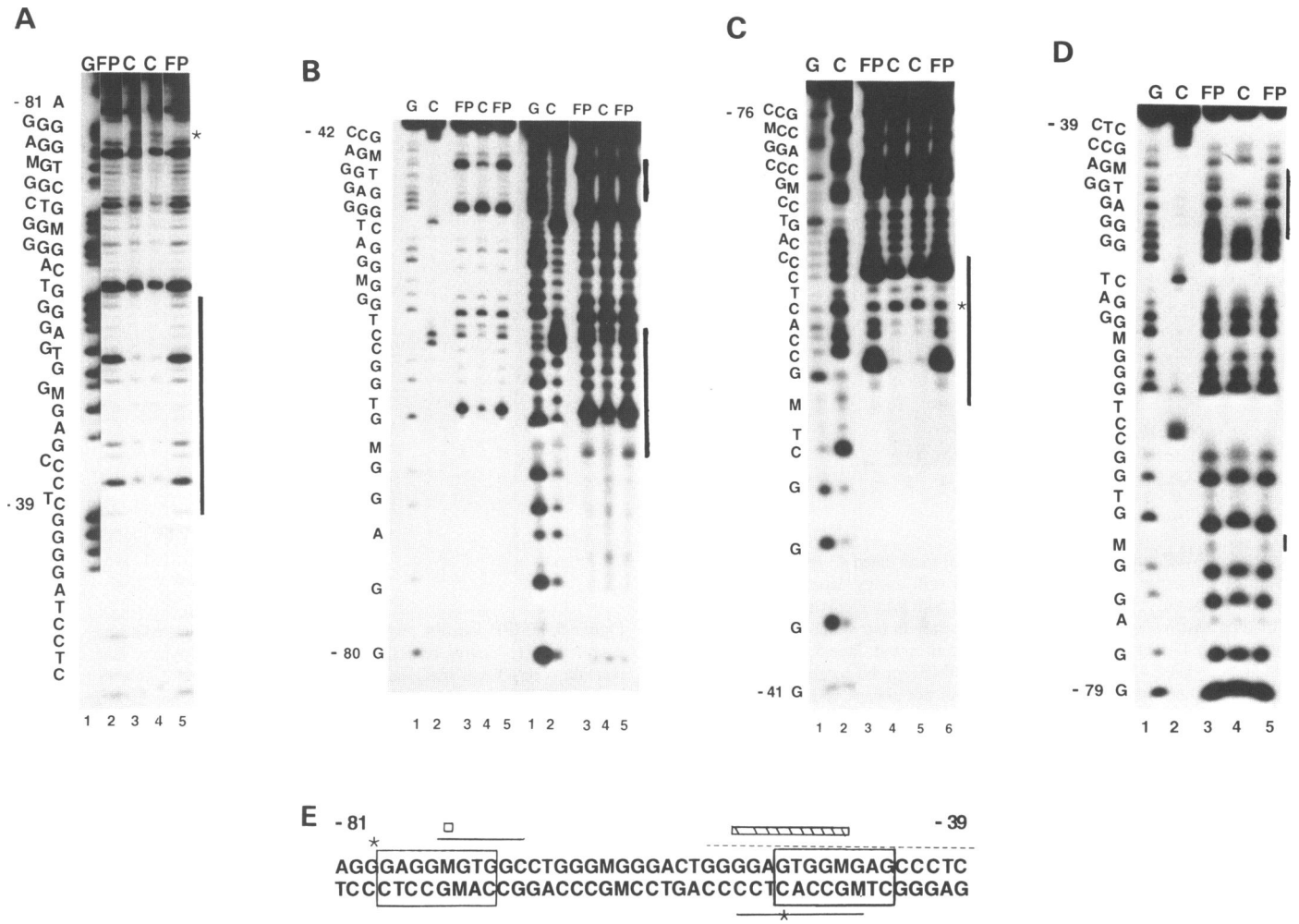

Figure 3. Two-dimensional DNase I and DMS footprinting of HMBP to the methylated Sp1 region of HIV-1 LTR. Binding reactions were subjected to DNase I digestion or DMS before gel shift assays. Bands corresponding to the HMBP/probe complex (C) and to the free probe (FP) were cut out and equal amounts of the purified DNAs were electrophoresed through sequencing gels along with G or G and C sequencing reactions (shown at the left of each figure). The sequence of methylated 43-mer is indicated next to the corresponding region of each gel. **A**: Footprint of 3′ labeled, top strand of 94-mer by Dnase 1. **B:** Footprint of 5′ labeled, top strand of 43-mer by Dnase 1.**C**: Footprint of 5′ labeled, bottom strand of 43-mer by DNase 1. **D**: Footprint of 5′ labeled, top strand of 43-mer by DMS. Heavy lines indicate regions protected in the complex relative to free probe. Asterisks indicate hypersensitive sites. DNase 1 is unable to efficiently cleave the terminal nucleotides in a DNA molecule. In order to see where the footprint ends, a longer exposure of B is shown in the righthand lanes. **E**: Diagram illustrating HMBP protected regions in the 43-mer. The direct repeat GA/TGGMGT/AG is boxed; asterisks indicate hypersensitive sites; the dotted line indicates the DNase 1 footprint in the 94-mer; the heavy lines indicate DNase 1 footprints in the 43-mer; the rectangles indicate DMS footprints; the shaded rectangle indicates the same footprint obtained by both DNase 1 and DMS.

(5′ M), and a protected region encompassing the 3′ M and extending into plasmid sequences. Other two-dimensional footprinting experiments were done with the methylated 43-mer. The top strand of the 43-mer was 5′ end-labeled before the binding reaction and DNase 1 footprinting shown in Figure 3B. In the lighter exposure (the left side), there is partial protection of sequences at the 5′ M and a footprinted region encompassing the 3′ M which is best observed in the darker exposure (right side of Figure 3B). Although DNase 1 cuts are fainter near the end of the fragment, protection of the complex relative to the free DNA terminates after the 3′ M. Figure 3 C shows DNase 1 footprinting of the bottom strand of the 5′ end-labeled 43-mer at the 3′ M. HMBP protects the 43-mer from DNase I cutting at sites near the 5′ methylated CG and the 3′ methylated CG in the top strand (Figure 3, A and B). Figure 3C shows that the methylated CG closest to the TATA box is the methylated site that is best protected from DNase I digestion in the bottom strand. DMS protection experiments were used to map the HMBP protected sites in the G-rich, top strand (Figure 3D). HMBP protects the 43-mer from DMS cleavage most clearly at the 3′ methylated CG in the top strand although a difference at the 5′ methylated cytosine can be detected (Figure 3D).

There appears to be no footprinting of the middle methylated CG in either the top or bottom strand. These results are diagramed in Figure 3E. When we used commercially obtained Sp1 to footprint this region, DNase I protection was detected over the entire 43-mer (not shown) as has been shown by others (27).

HMBP preferentially binds hemi-methylated DNA when the bottom strand is methylated

We tested, by gel shift assays, whether HMBP could bind the 43-mer when only the top strand or only the bottom strand is

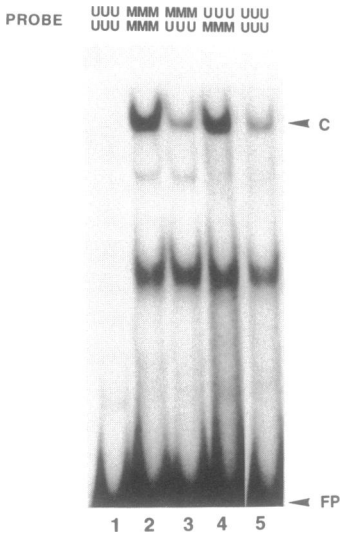

Figure 4. HMBP binds preferentially to hemi-methylated 43-mer only when the bottom strand is the methylated strand. Gel shift assays: lane 1, unmethylated 43-mer, probe only; lane 2, binding reaction with fully methylated 43-mer; lane 3, binding reaction with 43-mer methylated only in top strand; lane 4, binding reaction with 43-mer methylated only in bottom strand; lane 5, binding reaction with unmethylated 43-mer. < C indicates HMBP/probe complex; FP indicates free probe. Other bands are non-specific and not always visible on gel shifts.

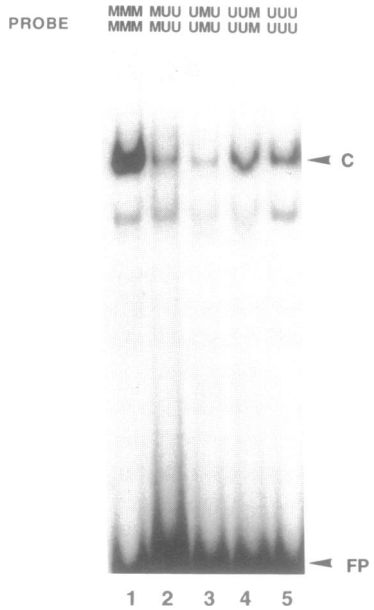

Figure 5. HMBP does not preferentially bind monomethylated 43-mer. Gel shift assays: lane 1, binding reaction with fully methylated 43-mer; lane 2, binding reaction with 43-mer methylated at only the 5′ site; lane 3 only in top strand; lane 4, binding reaction with 43-mer methylated only in bottom strand; lane 5, binding reaction with unmethylated 43-mer. < C indicates HMBP/probe complex; <FP indicates free probe. Other bands are non-specific and not always visible on gel shifts.

methylated. HMBP binds hemi-methylated 43-mer when the bottom strand is methylated as well as it binds fully methylated 43-mer (Figure 4, compare lanes 2 and 4). In contrast, if only the top strand is methylated, HMBP binds as poorly as it binds unmethylated 43-mer (Figure 4, compare lanes 3 and 5). Nonetheless, HMBP must contact the top strand when both strands are methylated, since HMBP protects sequences in this strand from DNase I digestion (See Figure 3).

HMBP fails to preferentially bind the 43-mer if only one site is methylated

We tested whether HMBP could bind the 43-mer containing only a single methylated site by gel shift assays. Figure 5 shows that HMBP binds no better to 43-mer when any one of the CGs is methylated than when no sites are methylated. Compare lanes 2, 3, and 4 to lane 1 (all three sites methylated) and to lane 5 (no sites methylated) in Figure 5.

The HIV-1 methylated DNA binding factor is not Sp1

Sp1, unlike HMBP, does not preferentially bind Sp1 sites when they are methylated. However, since the 43-mer used for these assays contains three binding sites for Sp1, and since the HMBP footprint overlaps with at least two Sp1 sites (Figure 3), it was important to determine whether this binding factor is a species of Sp1. We found that column chromatography of T-cell nuclear extract through a weak cation exchange resin (Bio-Rex 70) separates Sp1, and other proteins which bind the 43-mer, from HMBP (Figure 5, compare lane 2 to lane 7). Moreover, while a methylated Sp1 consensus binding site, GGGGMGGGG, in monomer or dimer form, competes for Sp1 binding, it does not compete for binding by HMBP (Figure 6A). Wheat germ agglutinin affinity (WGA) chromatography (38) also separates HMBP from Sp1 (Figure 6B). WGA fails to retain HMBP. We conclude from these differences that HMBP and Sp1 are different proteins. Moreover, the failure of the consensus Sp1 site to bind HMBP explains the lack of footprinting by HMBP over the middle Sp1 site In the HIV1 LTR since this site is most similar to the consensus SP1 sequence.

HMBP is not MDBP-2-H1, an avian histone H1 which binds methylated DNA

MDBP-2-H1, a methylated-DNA binding protein, has been shown to be a chicken histone H1 subspecies (43). Purified, unfractionated chicken histone H1 also preferentially binds methylated DNA sequences (43). However, mammalian histone H1 does not preferentially bind methylated DNA sequences but when denatured, mammalian histone H1 binds sequences irrespective of their methylation status (44−46). We were unable to detect binding by purified, undegraded, native bovine histone H1 to methylated 43-mer under gel-shift conditions in which HMBP binds (not shown). We estimate that less than 0.1 ng of HMBP complex is present in a gel shift band. Since up to 10 ngs of H1 were used in this experiment and no detectable binding was discerned, we conclude that HMBP is not histone H1 nor an H1 subspecies.

HMBP is distinct from methylated-DNA binding proteins MDBP-1, MeCP1, and MeCP2

MDBP-1 is a methylated DNA binding protein which binds a family of related methylated sequences and also binds some unmethylated sequences (47−51). One high affinity binding site for this factor is a 14 base pair sequence with three methylated cytosines and a 4 base pair overhang (see Figure 7). As a monomer, the 14-mer competes poorly with the HIV-1 LTR

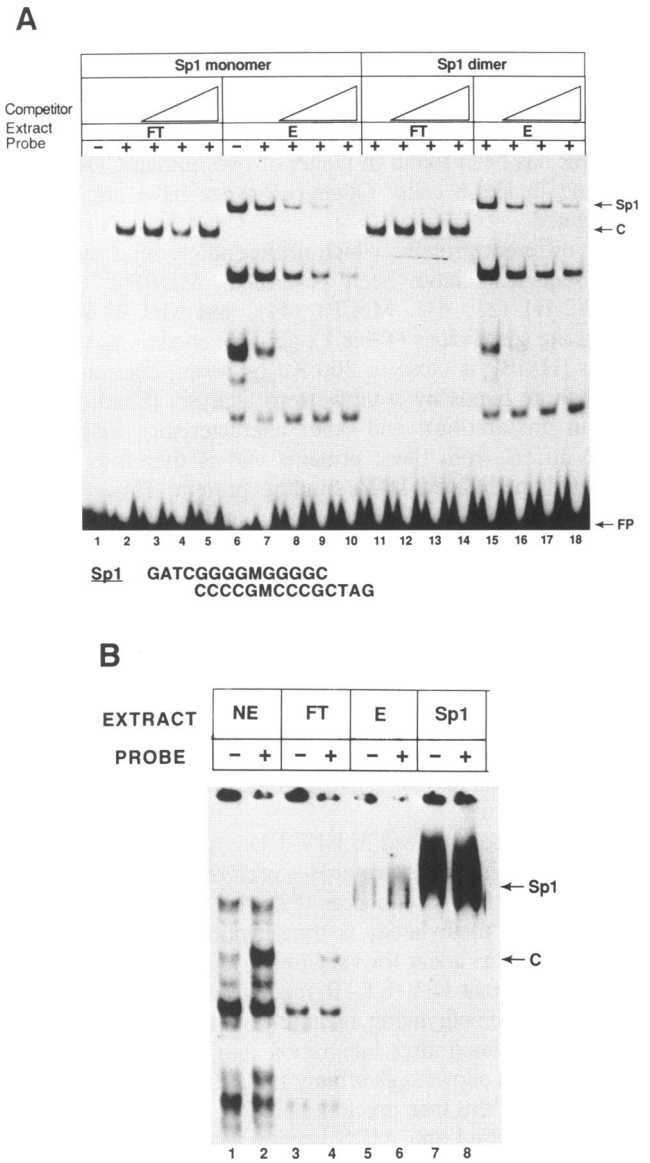

Figure 6. HMBP binding to methylated 43-mer is not competed for by methylated Sp1 consensus sites. HMBP can be separated from Sp1 by chromatography on Bio-Rex 70 and on WGA. **A.** Gel shift assays of Bio-Rex 70 chromatography fractions. Reactions without specific competitor: lanes 1, 2, 6, 7, 11, 15. Assays with increasing amounts of monomer or dimer Sp1 consensus binding site as competitor: lanes 3, 8, 12, and 16, were with 44 fold competitor DNA relative to probe; lanes 4, 9, 13, and 17 were with 88 fold competitor DNA; lanes 5, 10, 14, and 18 were with 167 fold competitor DNA. The flow-through fraction (FT) contains HMBP; the eluted fraction (E) contains Sp1. **B.** Gel shift assays of WGA chromatography fractions: lanes 1 and 2, crude nuclear extract (NE); lanes 3 and 4, flow through fraction (FT); lanes 5 and 6, eluted fraction (E); lanes 7 and 8, purified Sp1. The flow-through fraction (FT) contains HMBP; the eluted fraction (F) contains Sp1. Unmethylated 43-mer probe, (−) lanes; methylated 43-mer probe, (+) lanes. ←Sp1 indicates one of the Sp1/43-mer complexes. ←C =HMBP/43-mer complex. ←FP=Free probe. The methylated Sp1 consensus binding site used as competitor monomer is shown.

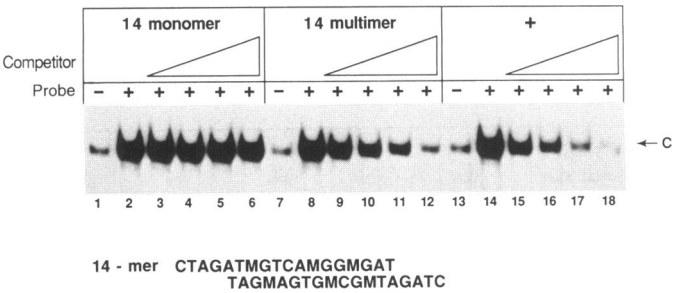

Figure 7. HMBP is not MDBP-1 or MeCP1. A binding site for MDBP-1 (38) was used as competitor for binding by HMBP to methylated 43-mer. Gel shift assays of binding reactions with increasing amounts of methylated, monomeric 14-mer as competitor are shown in lanes 3−6. Gel shift assays with increasing amounts of methylated, multimeric 14-mer (average of 10 repeats), a binding sequence for MeCP1, as competitor are shown in lanes 9−12. Lanes 15−18 show competition with increasing amounts of unlabeled methylated 43-mer. Lanes 1, 2, 7, 8, 13, 14 show binding reactions without specific competitor. Lanes 3, 9, and 15 show binding reactions with 11 fold more competitor than probe; lanes 4, 10, and 16 show binding reactions with 22 fold more competitor than probe; lanes 5, 11, and 17 show binding reactions with 44 fold more competitor than probe; lanes 6, 12, and 18 show binding reactions with 88 fold more competitor than probe. Unmethylated 43-mer used as labeled probe, (−) lanes; methylated 43-mer used as labeled probe, (+) lanes. The sequence of the MDBP-1 binding site (monomer) used as competitor is shown.

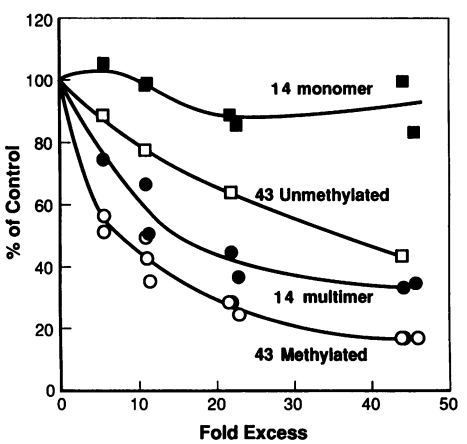

Figure 8. Quantitation of competition between various DNA binding sites and methylated 43-mer for binding of HMBP. Binding of HMBP to methylated 43-mer in the presence of various amounts of competitor DNA was quantified after gel shift assay by a blot analyser. Control is binding to methylated 43-mer without competitor. 14 monomer and 14 multimer are the competitors used in Figure 7. 43 Unmethylated and 43 Methylated refer to unmethylated and methylated 43-mer. Gel shift binding reactions were performed with 5.5, 11, 22, and 44 fold excess (in ng relative to probe), unlabeled, competitor DNAs. These results are a compilation of at least two experiments for each competitor DNA. All gel shift assays also contained poly (dI-dC) double stranded DNA in 500 fold mass excess as non-specific competitor DNA.

43-mer for binding to HMBP (Figure 8). Therefore HMBP is unlikely to be the same protein as MDBP-1.

When multimerized, methylated 14-mer becomes a high affinity binding site for MeCP1, a methylated DNA binding protein which binds sites containing at least 12 methylated CGs regardless of their sequence (44). Figure 8 shows that when the 14-mer is multimerized (to a weight-average length of 180 base pairs), it competes for HMBP binding but less well than methylated 43-mer competes for HMBP binding (Compare Figure 8, lanes 9−12 with lanes 15−18). Competition by methylated, multimerized 14-mer is probably due to the sequence, GGMG-A, which is present in the 14-mer and is part of a direct repeat in the 43-mer (See Figure 3E). Several other lines of evidence

Table 1. Binding site characterization distinguishes HMBP from the other methylated DNA binding proteins

	MDBP 1	MDBP-2-H1	MeCP 1	MeCP 2	HMBP
binds a single methylated site	yes	yes	no	yes	no
binding is sequence specific	yes	no	no	no	yes
binds hemi-methylated sites	yes	no	no	no	yes

Data for this table was obtained from references cited in the discussion and from results reported here.

indicate that HMBP is probably not the same protein as MeCP1. MeCP1 requires at least 12 methylated CGs for binding (44) while HMBP binds to a sequence with three methylated CGs (Figure 2B). Furthermore, MeCP1 forms such a large complex with methylated DNA that binding reactions must be resolved on agarose gels (44). In addition, MeCP1 binds DNA in 10 mM $MgCl_2$ (44), whereas HMBP binding is inhibited in the presence of Mg^{++}.

MeCP2, another recently described methylated DNA binding protein, like MeCP1, fails to discriminate among methylated DNA sequences (45, 46). Furthermore, unlike HMBP (Figure 4, lane 2), neither MeCP1 nor MeCP2 bind hemi-methylated DNA significantly (44, 45). Table 1 summarizes the binding site differences among the known methylated DNA binding proteins.

The competition by various methylated binding sites and by both unmethylated and methylated 43-mer for binding to HMBP is quantified in Figure 7. Competition by methylated monomeric and dimeric Sp1 consensus sequences show no significant HMBP binding (Figure 5A) and are not quantified.

DISCUSSION

We describe here the binding characteristics of a newly discovered protein, HIV-1 methylated DNA binding protein (HMBP). HMBP preferentially binds sites in the Sp1 region of the HIV-1 promoter when the CGs are methylated. However, on the basis of differing DNase I footprints, differing affinities for unmethylated binding sites, separation by two different chromatographic fractionation schemes (WGA and Bio-Rex 70), and differences in binding site specificities as revealed by competition with methylated Sp1 consensus binding sites, we conclude that HMBP and Sp1 are different DNA binding proteins. HMBP binding protects two sites in the HIV-1 LTR that overlap direct repeats of G A/TGGMGA/TG. Since these sites differ greatly from the consensus Sp1 site, GGGC/MGGG, the inability of the methylated Sp1 consensus site to compete for HMBP binding is explained. Whether HMBP will bind to other methylated, non-consensus Sp1 sites or other sequences awaits determination with a purified protein by a target detection assay such as has been used with Sp1(28, 29).

Since HMBP binds at two sites overlapping direct repeats in the 43-mer, we propose that HMBP binds to these direct repeats as a dimer. This would explain our finding that HMBP activity is separable from a methylated DNA binding activity that binds the 43-mer when only one site is methylated (Shao and Pratt, in preparation). In our binding studies, HMBP binds methylated DNA with 7 fold increased affinity over unmethylated DNA. Methylation specificity is equivalent to a conservative change of only one base pair per binding site. Other binding conditions, or *in vivo* conditions, may change the binding differential between methylated and unmethylated substrates. It may be that these differences in binding efficiency are sufficient for the functions served by HMBP binding *in vivo*.

HMBP has been found in nuclei of two human $CD4^+$ T cell lines and in HeLa cells. Other cell types have not yet been investigated.

Four different proteins which preferentially bind methylated DNA sequences have been described: MDBP-1 (47–51), MDBP-2-H1 (21, 43), MeCP1 (44), and MeCP2 (45–46). Binding site specificities (Table 1), gel-shift conditions, molecular weights [HMBP is close to 300 Kd by sizing column and has several large bands by southwestern analysis (Shao, Joel and Pratt, in preparation)] and other characteristics indicate that HMBP differs from these proteins and is therefore a newly discovered methylated DNA binding protein. However, until these proteins are purified and tested under exactly the same conditions, we cannot be sure they are unrelated.

HMBP preferentially binds hemi-methylated DNA only when the bottom strand is methylated. This suggests the intriguing possibility of a role for HMBP in cell differentiation. Immediately after DNA replication, a high affinity, methylated recognition site for HMBP would be present at only one side of the replication fork. This offers a way for differentiating mechanisms to discriminate between two daughter DNA molecules immediately after DNA synthesis if HMBP binds before maintenance methylation occurs.

HMBP may be involved in HIV-1 latency. This is suggested by studies showing that retroviruses preferentially integrate into methylated host DNA sequences (52); that integrated retroviral DNA becomes methylated, is transcriptionally repressed and thereafter remains latent for varying periods of time in infected cells (53); and that HIV-1 LTR regulated transcription can be activated by a demethylating agent (24). If HIV-1 proviral DNA becomes methylated after integration into the host genome, if methylation then allows high affinity HMBP binding to the HIV-1 promoter, and if this then precludes binding by Sp1, transcription would be repressed and AIDS latency would result.

ACKNOWLEDGEMENTS

We thank S.Smith and J.Kan for methylating LTR-CAT with human methylase, G.Gilmartin for HeLa cell nuclear extract, M.Martin for LTR-CAT, the NIH AIDS Research and Reference Reagent Program for HuT78 cells, R.Fishel for HPB-ALL cells and nuclear extract protocol, the laboratory of J.Burke for oligonucleotide synthesis, Greg Magin for cloning 43-mer, Adrienne Beddow for methylating plasmid DNA, and D.Pederson for helpful discussions and for comments on the manuscript. This research was supported by NIAID grant No. R29 AI 29354 and NSF grant No. DMB-8909279 (K.P.), by NRSA Fellowship No. F33 AI08461 (P.J.) and by the Markey Center for Molecular Genetics. P.J. and W.S. contributed equally to this work.

REFERENCES

1. Bird, A. (1992) *Cell*, **70**, 5–8.
2. Doerfler, W. (1991) *Biol. Chem.*, **372**, 557–564.
3. Doerfler, W., Toth, M., Hermann, R., Lichtenberg, U. andHoeveler, A. (1990) In: Clawson, G.A., Willis, D.B.,Weissbach, A. and Jones,P.A. (ed.), *Nucleic Acid Methylation*. John Wiley & Sons, Inc., New York, p. 329–349.

4. Holliday, R. (1990) *Biol. Rev.*, **65**, 431–471.
5. Dynan, W. S. (1989) *Trends Genet.*, **5**, 35–36.
6. Holliday, R. (1989) *Sci. Amer.*, **June**, 60–73.
7. Holliday, R. (1987) *Science*, **238**, 163–170.
8. Prendergast, G.C., Lawe, D. and Ziff, E.B. (1991) *Cell*, **65**, 395–407.
9. Iguchi-Ariga, S.M.M. and Schaffner, W. (1989) *Genes Dev.*, **3**, 612–619.
10. Bednarik, D.P., Duckett, C., Kim, S.U., Perez, V., Griffis, K., Guenthner, P. and Folks, T. (1991) *New Biol.*, **3**, 969–976.
11. Comb, M. and Goodman, H.M. (1990) *Nucleic Acids Res.*, **18**, 3975–3982.
12. Yee, A.S., Reichel, R., Kovesdi, I., and Nevins, J.R. (1987) *EMBO J.*, **6**, 2061–2068.
13. Kovesdi, I., Reichel, R. and Nevins, J.R. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 2180–2184.
14. Watt, F. and Molloy, P.L. (1988) *Genes Dev.*, **2**, 1136–1143.
15. Falzon, M. and Kuff, E.L. (1991) *Mol. Cell. Biol.*, **11**, 117–125.
16. Becker, P.B., Ruppert, S. and Schâtz, G. (1987) *Cell*, **51**, 435–443.
17. Levine, A., Cantoni, G.L. and Razin, A. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 10119–10123.
18. Boyes, J. and Bird, A. (1992) *EMBO J.*, **11**, 327–333.
19. Gutekunst, K.A., Kashanchi, F., Brady, J.N., and Bednarik, D.P. (1993) *J. Acquired Immune Def. Syndromes*, **6**, 541–549.
20. Boyes, J. and Bird, A. (1991) *Cell*, **64**, 1123–1134.
21. Pawlak, A., Bryans, M. and Jost, J. (1991) *Nucleic Acids Res.*, **19**, 1029–1034.
22. Embretson, J., Zupancic, M., Ribas, J.L., Burke, A., Racz, P., Tenner-Racz, K., Haase, A.T. (1993) *Nature*, **362**, 359–362.
23. Patterson, B.K., Till, M., Otto, P., Goolsby, C., Furtado, M.R., McBride, L.J. and Wolinsky, S.M. (1993) *Science*, **260**, 976–979.
24 Bednarik, D.P., Mosca, J.D. and Raj, N.B.K. (1987) *J. Virol.*, **61**, 1253–1257.
25. Bednarik, D.P., Cook, J.A. and Pitha, P.M. (1990) *EMBO J.*, **9**, 1157–1164.
26. Rosen, C.A., Sodroski, J. and Haseltine, W. (1985) *Cell*, **41**, 813–823.
27. Jones, K.A., Kadonaga, J.T., Luciw, P.A. and Tjian, R. (1986) *Science*, **232**, 755–759.
28. Thiesen, H. and Bach, C. (1990) *Nucleic Acids Res.*, **18**, 3203–3209.
29. Letovsky, J. and Dynan, W. (1989) *Nucleic Acids Res.*, **17**, 2639–2653.
30. Harrington, M.A., Jones, P., Imagawa, M. and Karin, M. (1988) *Proc. Natl. Acad. Sci.* USA, **85**, 2066–2070.
31. Ben-Hattar, J., Beard, P. and Jiricny, J. (1989) *Nucleic Acids Res.*, **17**, 10179–10190.
32. Höller, M., Westin, G., Jiricny, J. and Schaffner, W. (1988) *Genes Dev.*, **2**, 1127–1135.
33. Buschhausen, G.B., Wittig, B., Graessmann, M. and Graessmann, A. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 1177–1181.
34. Moore, S.P. and Fishel, R. (1990) *J. Biol. Chem.*, **265**, 11108–11117.
35. Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983)*Nucleic Acids Res.*, **11**, 1475–1489.
36. Wold, M.S. and Kelly, T. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2523–2527.
37. Gilmartin, G.M., McDevitt, M.A., and Nevins, J.R. (1988) *Genes Dev.*, **2**, 578–587.
38. Jackson, S. and Tjian, R. (1989) *Proc. Natl. Acad. Sci.USA*, **86**, 1781–1785.
39. Gendelman, H.E., Phelps, W., Feigenbaum, L., Ostrove, J.M.,Adachi, A., Howley, P.M., Khoury, G., Ginsberg, H.S., and Martin, M.A. (1986) *Proc. Natl. Acad. Sci.* **83**, 9759–9763.
40. Zucker, K.E., Riggs, A.D., and Smith, S.S. (1985) *J.Cellular Biochem.* **29**, 337–349.
41. Cann, A.J., Koyanagi, Y., Chen, I.S.Y. (1988) In B.R. Franza, B.R. Cullen, F. Wong-Staal, (Eds.). The Control Of Human Retrovirus Gene Expression. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., pp. 127–134.
42. Gorman, C.M., Moffat, L.F., and Howard, B.H. (1982) *Mol. Cell. Biol.* **2**, 1044–1051.
43. Jost, J. and Hofsteenge, J. (1992) *Proc. Natl. Acad. Sci.* USA, **89**, 9499–9503.
44. Meehan, R.R., Lewis, J.D., McKay, S., Kleiner, E.L. and Bird, A.P. (1989) *Cell*, **58**, 499–507.
45. Lewis, J.D., Meehan, R.R., Henzel, W.J., Maurer-Fogy, I., Jeppesen, P., Klein, F. and Bird, A. (1992) *Cell*, **69**, 905–914.
46. Meehan, R.R., Lewis, J. and Bird, A.P. (1992) *Nucleic Acids Res.*, **19**, 5085–5092.
47. Huang, L.-H., Wang, R., Gama-Sosa, M.A.,Shenoy, S. and Ehrlich,M. (1984) *Nature*, **308**, 293–295.
48. Khan, R., Zhang, X., Supakar, P., Ehrlich, K.C. and Ehrlich, M. (1988) *J. Biol. Chem.*, **263**, 14374–14383.
49. Supakar, P.C., Weist, D., Zhang, D., Inamdar, N., Zhang, X.,Khan, R., Ehrlich, K.C. and Ehrlich, M. (1988) *Nucleic Acids Res.*, **16**, 8029–8044.
50. Zhang, X., Supakar, P.C., Khan, R., Ehrlich, K.C. and Ehrlich, M. (1989) *Nucleic Acids Res.*, **17**, 1459–1474.
51. Ehrlich, K.C. and Ehrlich, M. (1990) *Mol. Cell. Biol.*, **10**, 4957–4960.
52. Kitamura, Y., Lee, Y.M.H. and Coffin, J. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 5532–5536.
53. Toth, M., Lichtenberg, U. and Doerfler, W. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 3728–3732.