Binding sites in mammalian genes and viral gene regulatory regions recognized by methylated DNA-binding protein

Xian-Yang Zhang, Clement K. Asiedu, Prakash C. Supakar, Rana Khan, Kenneth C. Ehrlich¹ and Melanie Ehrlich*

Department of Biochemistry, Tulane Medical School, New Orleans, LA 70112 and ¹Southern Regional Research Center, US Department of Agriculture, New Orleans, LA 70179, USA

Received July 31, 1990; Revised and Accepted September 18, 1990

ABSTRACT

Methylated DNA-binding protein (MDBP), a ubiquitous mammalian protein, recognizes a variety of related DNA sequences. Some of these sequences require methylation of their CpG dinucleotides for binding and others do not. We report that MDBP binds, in a DNA methylation-independent fashion, to two sites in the mouse polyomavirus enhancer, one in the enhancer of the human hepatitis B virus, and to one in the long terminal repeat of equine infectious anemia proviral DNA. We have also found a number of MDBP sites in human and rodent DNAs which bind much better to MDBP when they are methylated at CpG dinucleotides within the recognition site. These include sites at the beginning of the human genes for hypoxanthine phosphoribosyl transferase, HLA-A2, -A3, and -A25 antigens, and α -galactosidase A. In the case of methylation-responsive MDBP sites, changes in their methylation status during differentiation or DNA replication could help drive development by modulating transcription.

INTRODUCTION

Methylated DNA-binding protein (MDBP) is a sequence-specific DNA-binding protein present in a wide variety of mammalian sources (1). It was originally isolated on the basis of its ability to bind better to human DNA nick-translated in the presence of m⁵dCTP instead of dCTP than to the analogous DNA nick translated with only the standard nucleoside triphosphates (2). It was subsequently shown (3) that MDBP binds specifically to five sites in pBR322 and M13mp8 replicative form DNA but only when C residues at CpG dinucleotides in these sites were methylated to give m⁵CpG-containing sequences (m⁵C, 5-methylcytosine). Because the predominant methylated sequence in vertebrate DNA is m⁵CpG and much evidence implicates certain m⁵CpG sites and tissue-specific DNA methylation in transcription control (4–6), we proposed that MDBP binds to methylated DNA sequences in vertebrate cells as a transcription

regulatory factor (2). We recently showed that certain recognition sites for MDBP contain T residues rather than m⁵C residues at analogous positions and so are bound by MDBP in a cytosine-methylation independent mode (7). The consensus sequence for 14-bp (base-pair) MDBP sites is 5'-RTm⁵YRYAm⁵YRGm⁵-YRAY-3' (R, A or G; m⁵Y, m⁵C or T; Y, C or T).

We now report that MDBP recognizes two high-affinity sites in the mouse polyomavirus (Py) enhancer which do not require C methylation for binding and one in the human hepatitis B virus (HBV) enhancer, which has been directly implicated in enhancer function (8). In addition to methylation-independent MDBP sites in viral enhancers, we found one in a long terminal repeat (LTR) of equine infectious anemia virus (EIAV) which may help negatively regulate transcription.

We also describe a number of sites in or around mammalian genes that bind to MDBP much better when methylated in both strands than when unmethylated. Several of the latter MDBP sites are associated with gene regions whose methylation has been implicated in negative transcription control. Based upon our findings, we propose a model for how certain methylation-responsive MDBP sites within or upstream of mammalian genes might help regulate transcription and differentiation in a manner dependent upon whether the sites are unmethylated or are methylated at CpG dinucleotides in both strands (bifilarly methylated) or only in one or the other strand (hemimethylated).

MATERIALS AND METHODS

MDBP preparation, band shift assays, and partial proteolysis

Nuclear extracts of HeLa S3 cells and mouse LTK⁻ cells were prepared as previously described except that the ammonium sulfate precipitation step was omitted for the HeLa cells (1). A nuclear extract from a human B cell lymphoblastoid line (Raji) was prepared similarly except that 0.42 M NaCl was used in the nuclear extraction step (9) instead of 0.3 M NaCl (2). The hydroxylapatite chromatography fraction of MDBP was obtained from 0.3 M NaCl extracts of term human placental nuclei by our standard procedure (2,3). For experiments with partially

* To whom correspondence should be addressed

proteolyzed MDBP, 0.5 units per μl of the hydroxyapatite fraction in 50–100 mM potassium phosphate, pH 7, 0.1 mM Na-EDTA, 20% glycerol, and 0.5 mM dithiothreitol was digested with 1 ng/ μl of proteinase K for 20 min at room temperature and proteolysis was stopped with 2 mM phenylmethylsulfonyl fluoride (10). Band shift (gel retardation) assays were performed on 10 or 20 fmol of ^{32}P -labeled oligonucleotide duplex and 1–2 units of MDBP with electrophoresis in 5% polyacrylamide gels as described previously (1). One unit of MDBP complexes 1 fmol of the 22-bp methylated pB site 1 duplex (Table 1) in 15 min under standard conditions. Dissociation of MDBP-DNA complexes was studied by previously described methods (10).

Preparation of oligonucleotide duplexes

Oligonucleotides were synthesized, end-labeled, and annealed as described previously (11,12). HBV site 1 was methylated at the CpG in one strand by using the 24-base complement of the strand shown in Table 1 as a template, 5'-TTACCCCGTTGCC-3' as a primer, and primer-extending, in a reaction catalyzed by the Klenow fragment of *E. coli* DNA polymerase I, with 1.5 μM m⁵dCTP for 10 min followed by the addition of 250 μM dCTP, dTTP, dGTP, and dATP for a further 30-min incubation. This protocol has been used to efficiently incorporate m⁵C residues into correctly sized duplexes at specific positions in methylation-dependent MDBP sites rendering them specific ligands (unpubl. results). Also, we replaced five of the C residues with m⁵C by primer-extending the above template/primer just in the presence of 250 μM m⁵dCTP, dGTP, dATP, and dTTP. The other oligonucleotides were methylated by using m⁵C phosphoramidite derivatives for their synthesis.

RESULTS

MDBP recognizes a site in the HBV enhancer and is the same as EF-C

We noticed that HBV contains in its enhancer a binding site (13) that differs from the MDBP site in phage λ DNA by only one T·A-to-C·G substitution (Table 1). This HBV site is recognized by a ubiquitous mammalian protein, detected in crude nuclear extracts, called EF-C or EP (8,14). Therefore, we tested a 24-bp oligonucleotide duplex containing this palindromic site and several surrounding base-pairs for binding to the hydroxyapatite fraction of human placental MDBP, whose DNA sequence-specificity we have studied extensively (3,7,11,15). A family of related, low-mobility complexes formed between MDBP and an oligonucleotide duplex containing this ^{32}P -labeled HBV sequence (HBV site 1). These exactly co-electrophoresed with those of our standard methylation-dependent MDBP ligand, pB site 1, a methylated pBR322 sequence (Fig. 1A and Table 1).

Co-electrophoresis of specific complexes containing HBV site 1 with those containing pB site 1 was also seen when crude nuclear extracts from HeLa cells (Fig. 1B) or a human B cell line (Fig. 1C and data not shown) were the source of MDBP. The MDBP-type complexes were the only specific complexes formed between these ligands and the nuclear extracts. When nuclear extracts from cultured cells were the source of MDBP, the family of specific complexes had a somewhat lower electrophoretic mobility and a less diffuse appearance than those from many of our preparations of the hydroxyapatite fraction of the placental extract (Fig. 1A). Also, the relative amount of complex in each band varied between certain nuclear extracts (Figs. 1B and 1C). We previously showed that some of this

Table 1. Relative binding of viral sites to native MDBP¹

Name of sequence	Location	Sequence ²	Extent of binding
pB site 1/m ⁵ C	pBR322	ctag ¹ A ⁵ T M G ¹⁰ T C A M G ¹⁴ G M G A T ctag	1.0
pB site 1/no m ⁵ C		ctag A T <u>C</u> G T C A <u>C</u> G G <u>C</u> G A T ctag	<0.02
λ site 1	λ bacteriophage DNA	actc G T T G C C - <u>C</u> G G T A A C aaca	0.73
HBV site 1	hepatitis B virus enhancer	ttacccc G T T G C C - <u>C</u> G G <u>C</u> A A C ggcc	0.80
HBV site 1/m ⁵ C		ttacccc G T T G C C - M G G C A A C ggcc	0.81
Py site 2	polyomavirus enhancer	ggca G T T G C C - T A G <u>C</u> A A C taat	1.90
EIAV site 1	equine infectious anemia virus LTR	ttta G T T G C C - T A G <u>C</u> A A C atga	1.79
EIAV site 2	equine infectious anemia virus LTR	aag G T T A T G A <u>C</u> A G <u>C</u> A T C agca	<0.02
EIAV site 3	equine infectious anemia virus LTR	atg A T T A T <u>A</u> A <u>C</u> A G T A <u>A</u> A aga	<0.02
Py site 1	polyomavirus enhancer/ori region	gct G T T G T C A T A G T A A T gat	1.18
LPV site 1	lymphotrophic papovavirus ori region	tagg G T T G C C A T A G T G A T ttgt	1.62

¹Band shift assays were performed with the hydroxyapatite fraction of human placental MDBP. The data for binding of unmethylated pB site 1 and λ site 1 were previously reported (3,7) and are given for comparison.

²The sequence at the MDBP site which most closely matches the partially palindromic, degenerate consensus sequence 5'-RTm⁵YRYAm⁵YRGm⁵YRAY-3' is shown and its positions are numbered. Residues which differ from the consensus sequence are indicated by an "x". A hyphen indicates, in effect, a 1-bp deletion that is required for some of these sites to match the consensus sequence. The starting positions of the mammalian viral sites in standard GenBank maps are as follows: HBV site 1, 1152; Py sites 1 and 2, 5267 and 5173 and the repeated LTR EIAV site 1, 111 and 8018 (all four sites are in the opposite strand to that in GenBank); and LPV site 1, 187 in the wild type LPV genome and also at 428 in the LPV K38 variant which has a duplication of a 178-bp repeat containing this region (16). pB site 1 is our diagnostic ligand for MDBP because it is recognized by MDBP only when methylated at CpG dinucleotides. pB site 1 was bifilarly methylated at its CpG sites.

variability is probably due to a small degree of proteolysis (despite the presence of protease inhibitors) during purification of MDBP, which does not interfere with the DNA sequence-specificity of this protein (1,10). Differences in the mobility, but not the DNA sequence-specificity, of complexes containing EF-C, dependent on the source of cells, were also seen by Ostapchuk *et al.* (8). Formation of all of these low-mobility complexes with ^{32}P -labeled HBV site 1 was competed by an excess of unlabeled, MDBP-specific oligonucleotide duplexes containing HBV site 1, Py site 1 (see below), or methylated pB site 1, but little or no competition was seen with duplexes containing unmethylated pB site 1 or a heterologous Sp1-binding site (Fig. 1C). Also, one strand of unlabeled HBV site 1 added in excess did not compete for binding of the duplex ligand (Fig. 1C).

The MDBP·oligonucleotide complexes were just as extraordinarily long-lived when HBV site 1 was the ligand as when pB site 1 was. In the presence of 3 pmol of unlabeled pB site 1 (methylated), the percentage dissociation of MDBP·HBV site 1 complexes and MDBP·pB site 1 complexes using a hydroxylapatite fraction of MDBP was as follows: only 17 and 19%, respectively, after 1 h, and 37 and 33%, respectively, after

2 h. No further dissociation was seen after an additional hour. The rate of dissociation of MDBP·DNA complexes seems to vary somewhat with the preparation of MDBP.

Because HBV site 1 contains a CpG dinucleotide equivalent to the important (1,12) m^5CpG dinucleotide at position 8 of pB site 1 (Table 1), we tested the effect of methylation at this C residue. Methylation of only one strand of pB site 1 suffices to increase binding (12) and so we studied HBV site 1 methylated in the one CpG dinucleotide in the strand of HBV depicted in Table 1. Methylation gave no significant increase in binding (Table 1 and Fig. 1A). This is consistent with the high affinity of MDBP for HBV site 1 in the absence of C methylation. As described below, the cytosine-methylation independence of the binding of HBV site 1 is probably attributable to the very high degree of recognition of the 5'-GTTGC-3' motif (with its two m^5Y residues) at either end of the HBV site 1 duplex (Table 1).

MDBP recognizes two sites in the polyomavirus enhancer B

Ostapchuk *et al.* (8) and Ben-Levy *et al.* (14) showed that a mouse polyomavirus enhancer sequence, which we refer to as Py site 2 (Table 1), cross-competed with the HBV sequence for binding to EF-C or EP. Similarly, Py site 2·MDBP complexes co-electrophoresed with those of HBV site 1 or methylated pB site 1 and the latter two sites competed with Py site 2 for MDBP binding while duplexes that were not specific for MDBP did not (Fig. 2). Similar results were obtained with crude extracts from HeLa cells, a human B cell line, and mouse LTK⁻ cells (data

Fig. 1. Band shift assays with MDBP and oligonucleotide duplexes containing a hepatitis B virus enhancer sequence (HBV site 1) or a pBR322 sequence (pB site 1). A, the ^{32}P -labeled ligands were methylated pB site 1 (pB_m) and unmethylated HBV site 1 (HBV_u) or HBV site 1 methylated at only one C residue (HBV_{m1}), as shown in Table 1, or at four additional C residues (HBV_{m2}). They were incubated with the hydroxylapatite fraction of human placental MDBP (HA) or the crude nuclear extract (N.E.) from a human B cell line. B and C, the following unlabeled oligonucleotide competitors (300 fmol) were added to a HeLa cell or B cell nuclear extract together with the standard nonspecific competitor, 400 ng of poly(dI-dC)·poly(dI-dC): methylated or unmethylated pB site 1 (pB_m or pB_u); a 41-bp duplex containing a binding site for the transcription factor Sp1 (50); HBV site 1 (unmethylated) either as a duplex (ds) or as just the single-stranded oligonucleotide (ss) containing the sequence shown in Table 1; or Py site 1 (Py 1). Five min later, the ^{32}P -labeled unmethylated HBV site 1 (10 fmol) was added. In panels B and C, the electrophoresis buffer was 25 mM Tris base, 192 mM glycine, 1 mM Na-EDTA, pH 8.3, instead of 22 mM Tris base, 22 μM boric acid, 0.5 mM Na-EDTA, pH 8.3, which was used for panel A and Figs. 2, 4, 5 and 6. The former buffer gave better resolution of the several bands of MDBP·DNA complexes from nuclear extracts but not from partially purified MDBP. The specific MDBP·DNA complexes are denoted by 'C' and free oligonucleotide duplex by 'F'. As described in the text, these complexes from the cultured cell nuclear extract migrated more slowly than those from the placental hydroxylapatite fraction. It should be noted that just as pB site 1 (methylated) competed with HBV site 1 for complex formation, unlabeled HBV site 1 efficiently competed with ^{32}P -labeled pB site 1 (methylated) for formation of complexes with MDBP (data not shown).

Fig. 2. Complex formation between MDBP and two sequences in the polyomavirus enhancer. Band shift assays were performed in the absence (first three lanes) or the presence (last six lanes) of 1 pmol of the indicated unlabeled competitor DNA preincubated with the hydroxylapatite fraction of human placental MDBP. The radiolabeled ligand was either pB site 1, Py site 1, or Py site 2. Abbreviations are as in Fig. 1 and HBV is unmethylated HBV site 1.

	5240	MDBP	T Ag	2
Py	GTTCACCAATCATCTACTATGACACAGCTGTTTTTTTASTATTAAAGCAGAGCCGG			
LPV	TGCAAGTCTGCAAAATCACTATGCAACCTAGTTTTTCTACCTGATAGAGGCCAG			
	214			155

Fig. 3. Comparison of DNA sequences around the polyomavirus B enhancer/ori core border with the corresponding region of the lymphotropic papovavirus genome. The MDBP binding sites, Py site 1 or LPV site 1, and the pentanucleotide motif recognized by the large T antigen (T Ag) are indicated. Note that the opposite strand of the MDBP sites is shown compared to that in Table 1. The run of T residues in the Py DNA and the illustrated sequences to the right of them are the beginning of the *ori* core. Vertical lines denote identical base-pairs in the Py and LPV DNA sequences.

not shown). By searching DNA databases (GenBank) for sequences homologous to the 14-bp MDBP consensus sequence, we found a second MDBP site in the Py enhancer. This new specific protein-binding site, Py site 1 (Table 1) is adjacent to the DNA replication origin (*ori*) core. Complexes of the same mobility were seen when Py sites 1 and 2 were incubated with MDBP; also, Py site 1 competed with Py site 2 for complex formation (Fig. 2).

During database searches, we located another 14-bp homologue of the MDBP consensus sequence in a different papovavirus genome, namely, lymphotropic papovavirus (LPV) DNA (16). This site, LPV site 1, which is adjacent to the putative *ori* core (Fig. 3), differs from Py site 1 only by transitions at two positions, both of which are compatible with the consensus sequence (Table 1). LPV site 1 formed complexes that co-electrophoresed, in band shift assays, with MDBP·pB site 1 complexes. Furthermore, Py site 1 or methylated pB site 1, but not unmethylated pB site 1, competed for this complex formation (data not shown). We conclude that LPV site 1 bound to MDBP and it did so even better than did Py site 1 (Table 1).

We also tested a mutant Py site 2 (5'-gGTTACCTAGC-AAC-3') because its transition mutations (underlined) were implicated by Tseng and Fujimura (17) in altering Py B enhancer activity and viral DNA replication in a mouse embryonal carcinoma cell line. One of these mutations (in lower case) is outside the 13-bp recognition element and the other mutation, which is within this 13-bp region, still matches the consensus sequence; therefore, these mutations would not be expected to affect the affinity for MDBP much. Only, a small (30%), but reproducible, decrease in binding of the hydroxylapatite fraction of placental MDBP was observed in comparing this mutant site to the wild type Py site 2. A somewhat larger decrease in binding (47%) at the mutation site was seen when a nuclear extract of human B cells was the source of MDBP.

MDBP binds specifically to a site in the EIAV LTR

Although they did not assay the site, Ben-Levy *et al.* (14) noticed in the LTR of equine infectious anemia virus a sequence similar to the Py site 2 region. We synthesized a 21-bp oligonucleotide duplex containing this sequence (EIAV site 1) which is identical to Py site 2 in its central 13 bps (Table 1). Upon binding to MDBP, EIAV site 1 and Py site 2 formed complexes which co-electrophoresed and these short DNA duplexes competed with each other for complex formation with MDBP (data not shown). In the EIAV LTR, there are two 14-bp sequences matching the consensus sequence for MDBP binding at 10 or 11 out of 14 positions but, unlike EIAV site 1, these sites, EIAV sites 2 and 3, which are 41 and 67 bps upstream of EIAV site 1 (18), were not recognized by MDBP (Table 1).

Methylation-responsive MDBP sites in mammalian DNA

Searches of the mammalian DNA database revealed many sequences matching the degenerate MDBP consensus sequence, which should exhibit methylation-independent binding as well as putative methylation-responsive MDBP sites. In this study, we tested nine of the latter sequences for binding to MDBP. All nine sites, which had at least 10 base-pairs out of 14 matching those of the consensus sequence, were recognized by MDBP and required C methylation at their CpG dinucleotides for optimal binding (Table 2). The DNA ligands had been synthesized so that either both strands were CpG-methylated (m/m duplexes), only one strand was methylated (m/u or u/m duplexes, depending

Fig. 4. Complex formation between MDBP and an oligonucleotide duplex containing a CpG dinucleotide pair in its human HPRT site 1 sequence. Band shift assays were performed with the 32 P-labeled 22-bp HPRT site 1 duplex shown in Table 2. The duplex was prepared with both strands methylated at their CpG dinucleotide (m/m), one strand methylated (m/u or u/m depending on whether the top strand or the bottom one, as shown in Table 2, was methylated) or with neither strand methylated (u/u). Bifilarly methylated pB site 1 was used as a standard. C, the family of MDBP-specific complexes; F, free DNA.

Fig. 5. Complex formation between MDBP and an oligonucleotide duplex containing two CpG dinucleotide pairs in its rat serum albumin site 1 sequence. Analysis of MDBP binding and the labeling of the figure was as described for Fig. 4. In the first panel the same gel was exposed for a much shorter time to reveal the sub-bands containing the pB site 1·MDBP complexes.

on which strand was methylated; Table 2), or both strands were unmethylated (u/u duplexes).

Figs. 4 and 5 illustrate how specific DNA·protein complex formation was visualized for hHPRT site 1, which is located ~170 bps upstream of the two major, adjacent cap sites of the human hypoxanthine phosphoribosyl transferase gene (19), and for rSA site 1, which is in the rat serum albumin gene's cDNA (20). Upon incubation with MDBP, these mammalian site-containing duplexes yielded a family of complexes that co-

Table 2. Relative binding to native MDBP of methylation-responsive MDBP sites

Name of site	Sequence of bifilarly methylated site ¹	Location in gene or cDNA ²	Extent of binding				
			Relative percent ³		m/m binding relative to pB binding ⁴		
			m/m	m/u	u/m	u/u	
pB site 1	5'-ctagATMGTCCAMGGGAT-3' 3'-TAGMAGTCMCGTAGatc-5'	in vitro-methylated pBR322 standard	100	50	14	2	1.0
hHPRT site 1	5'-gaaaATTCCAMGGCTACctag-3' 3'-ctttTAAGGGTGMCGATGgac-5'	human HPRT approx. -170	100	70	28	19	0.10
hHPRT site 2	5'-cggcTCMGTATGGMGACccgc-3' 3'-gccgAGGMAATACCGMTCggcg-5'	human HPRT approx. +85	100	33	66	17	0.06
hHPRT site 3	5'-ctgCTCACCAMGAGCCagg-3' 3'-gacGAGTGGTGTGTCGgtcc-5'	human HPRT approx. +110	100	ND	ND	37	0.06
hGal site 1	5'-ctgcATTGTCCAMGGTGACmga-3' 3'-gacgTAACAGTGMCACTGmct-5'	human α -gal. A +49	100	86	30	22	0.46
hHLA site 1	5'-gatgCGMGTTCATGCGGCCcga-3' 3'-ctacCGMAGTACCGMGGgct-5'	human HLA-A2 and -A25 approx. +30	100	13	15	5	0.30
rSA site 1	5'-gagtGCTGTCAMGGGACctgt-3' 3'-ctcaCGACAGTGMCGMTGgaca-5'	rat serum albumin approx. +840	100	13	21	6	0.52
rAFP site 1	5'-agtGCTGCCAMGGCAACgcg-3' 3'-tcaCGACGGTGMCGTTGcg-5'	rat α -fetoprotein approx. +850	100	96	47	30	0.47
hFur site 1	5'-ccagTGMGCCAMGGGATtata-3' 3'-ggtcCAGMGGTGMCGTAatat-5'	human furin +1581	100	46	6	6	0.67
hFer site 1	5'-gtmGCMGCCATGAMGACcgc-3' 3'-cagMCGMGTACTGTCg-5'	human apoferritin H +202	100	65	35	22	0.23

¹Positions corresponding to the 14-bp MDBP consensus sequence region for MDBP binding are capitalized and those agreeing with the consensus sequence underlined.

²The sequence positions are given with respect to the (major) transcription start site of the genomic DNA except for the rSA and rAFP sites where the position relative to the beginning of a cDNA is given and the hFur site for which the position in a 3' genomic fragment (26) is given.

³The given bifilarly methylated (m/m) site, the hemimethylated site with m⁵C residues in one or the other strand (m/u or u/m, where the first letter designates the upper strand as shown), and the completely unmethylated site (u/u) were tested. Binding is compared to that of the analogous m/m site. ND, not determined.

⁴The extent of binding of the given bifilarly methylated site divided by that of bifilarly methylated pB site 1 under standard conditions.

electrophoresed with that of the pB site 1 standard and this complex formation was essentially completely inhibited by an excess of methylated pB site 1 while a similar excess of unmethylated pB site 1 had little or no effect (data not shown).

When we used ~2 units of MDBP, as in our standard conditions, the increase in complex formation upon methylation of the CpG dinucleotides in both strands of the site ranged from ~3 fold for rAFP site 1 in the rat α -fetoprotein cDNA (21) to ~50 fold for pB site 1 (Table 2). However, the ratio of binding of a methylated site to that of the unmethylated form of the site increased as the MDBP concentration was lowered. For example, when the concentration of MDBP was decreased several fold, the methylated form of hGal site 1 was bound by MDBP ~20 times better than the unmethylated form instead of only ~5 times better.

Hemimethylated duplexes usually gave more complex formation than the corresponding unmethylated duplexes but generally less than the analogous bifilarly methylated duplexes (Table 2). However, hemimethylated sites were often bound by MDBP to very different extents depending on which strand was CpG-methylated (m/u or u/m in Table 2). Based on the data in Table 2, it is difficult to predict whether methylation of one strand or the other will be more favorable for MDBP binding especially because most of the sites exhibited the hyphenated dyad symmetry characteristic of high-affinity MDBP sites. The degree of methylation-dependence of MDBP binding was generally greater when the sequences contained more CpG dinucleotides within the 14-bp recognition region. However, certain favorable sequence combinations, such as G at both 5' ends of the 14-bp

Fig. 6. The formation of two types of complexes between various DNA ligands and partially proteolyzed MDBP. The hydroxylapatite fraction of MDBP fragmented by partial digestion with proteinase K was incubated with the indicated [³²P]oligonucleotide duplex ligand including the mutant Py site 2 described in the text. The reaction mixtures were electrophoresed on a 7% polyacrylamide gel. The MDBP-specific complexes containing one or two truncated MDBP subunits are marked as are lower-mobility nonspecific complexes and the free ligand.

region (10) of hFer site 1 in the human apoferritin H chain gene (22) plus a T·A base-pair at position 8 (1) of that 14-bp site apparently ameliorated the cytosine methylation requirement (Table 2).

Several of these methylation-responsive MDBP sites were close to the transcription initiation point of the gene and downstream from it (Table 2). Examples of such sites are hHPRT sites 2 and 3, hHLA site 1 in the human HLA-A2, -A3, and -A25 antigen genes (23,24), and hGal site 1 in the human α -galactosidase A gene (25). In contrast, hFur site 1 is far downstream from the beginning of the open reading frame of the putative *fur* gene. This open reading frame is transcribed in a tissue-specific manner and is immediately upstream from the *c-fes/fps* proto-oncogene but its function is unknown (26). Its MDBP site has a rather high affinity for MDBP and differs from pB site 1 by only two transitions (Table 2). The rSA and rAFP sites, both of which are far downstream from the beginning of the cDNA, are in highly homologous regions of their genes and encode the same amino acids (21).

Formation of DNA-protein complexes containing either one or two subunits of fragmented MDBP

Two types of complexes (one-subunit or two-subunit) with different electrophoretic mobilities form between partially proteolyzed MDBP and specific oligonucleotide duplexes and the relative amounts of these complexes vary greatly with the sequence of the MDBP site (10). In order to provide additional confirmation that MDBP is the protein recognizing the viral sites; to compare the ability of these sites to form 1-subunit or 2-subunit complexes; and to better understand the importance of C methylation in C methylation-dependent MDBP sites, we used a number of these sites as ligands for partially proteolyzed MDBP (Fig. 6; Table 3). As expected (10), characteristic 1-subunit complex formation was favored by low concentrations of fragmented MDBP in the binding reactions and high concentrations gave more of the 2-subunit complexes (Table 3).

Some correlation was seen between the presence of part or all of the 5'-GTTGCT-3' motif of HSV site 2 (left half, as shown in Table 3) and the formation of appreciable amounts of 1-subunit complexes at a low concentration of partially proteolyzed MDBP (Table 3). However, none of the ligands was able to complex with partially proteolyzed MDBP as well as HSV site 2 did, even though, for example, HBV site 1 and Py site 2 have a symmetrically placed 5'-GTTGC-3' motif in each half site. Presumably, differences in the base-pairs adjacent to this motif somehow account for the relatively large extent of total complex formation of fragmented MDBP with HSV site 2. However, all the viral ligands tested were able to form more 2-subunit complexes than was HSV site 2, which has three base differences from the consensus sequence in its right half (Table 3). In contrast to HSV site 2, but like the Py and HBV sites, the two half-sites of bifilarly methylated hHLA site 1 are apparently more equally involved in binding to each of the subunits of MDBP as seen in this site's more favorable formation of 2-subunit complexes than of 1-subunit complexes at a high concentration of partially proteolyzed MDBP in the standard reaction mixture (Table 3). Bifilar methylation of this site increased the ratio of 2-subunit complexes to 1-subunit complexes as well as the extent of complex formation with fragmented MDBP.

DISCUSSION

We have demonstrated that HBV contains in its enhancer a cytosine methylation-independent binding site for MDBP (Table 1; Fig. 1). This viral site is recognized by a ubiquitous mammalian protein, detected in crude nuclear extracts, called EF-C or EP (8,13,14). By co-electrophoresis of their DNA-protein complexes in band shift assays and ligand cross-

Table 3. Formation of one-subunit and two-subunit complexes with fragmented MDBP¹

Name of sequence	Sequence ²	Low concentration of MDBP ³			High concentration of MDBP ³		
		% complexes as 1-subunit complexes	% complexes as 2-subunit complexes	Total complex ⁴ (% of input)	% complexes as 1-subunit complexes	% complexes as 2-subunit complexes	Total complex ⁴ (% of input)
HSV site 2	<u>G T T G C T</u> A T G G C C G C	96	4	27	92	8	73
pB site 1	A T M G T C A M G G M G A T	19	81	2.5	1	99	20
Py site 1	<u>G T T G T C A T A G T A A T</u>	50	50	5.6	11	89	20
HBV site 1	<u>G T T G C C</u> - C G G C A A C	56	44	8.8	37	63	21
Py site 2	<u>G T T G C T A</u> - G G C A A C	66	34	9.0	13	87	55
Py site 2/mut 1	<u>G T T G C T A</u> - G G T A A C	74	26	6.0	21	79	58
hHLA site 1 m/m ⁵ m/u u/m u/u	G C M G T C A T G G M G C C	ND	ND	ND	16	84	11
		ND	ND	ND	34	66	4
		ND	ND	ND	39	61	4
		ND	ND	ND	51	49	3

¹MDBP was the hydroxylapatite fraction of human placental MDBP that had been partially digested with of proteinase K thereby releasing dissociated fragments containing its DNA-binding domain (10). Fig. 6 shows the 1-subunit complexes, the predominant ones with HSV site 2 (from the herpes simplex virus thymidine kinase gene), and the 2-subunit complexes with both half-sites of an MDBP binding site, characteristic of the predominant complex with pB site 1 (10). The 5'-GTTGCTA-3' motif seen in HSV site 2 and 3' truncated versions of this motif as well as its complement are underlined.

²The sequence of one strand at the MDBP site is given. Py site 2 is shown in the opposite orientation from that of Table 1. M, m⁵C.

³Fragmented protein from approximately 2 units of MDBP (low concentration) or 15 units of MDBP (high concentration) was used for each binding reaction with 20 fmol of [³²P]oligonucleotide duplex. ND, not determined.

⁴The sum of the radioactivity in the 1-subunit and the 2-subunit complexes was used to determine the total percentage of input ligand forming complexes with fragmented MDBP.

⁵Both strands (m/m) were methylated or only the indicated strand (m/u), the complement (u/m), or neither strand (u/u) contained m⁵C residues.

competition, using the same kind of assay conditions, nuclear extracts, and cell source (HeLa), EF-C/EP is indistinguishable from MDBP (Fig. 1; 8,14). We previously mapped, by protection and interference assays, several different MDBP sites, including λ site 1 that differs from HBV site 1 by only a C·G-to-T·A transition (7,11). DNase I footprinting revealed a 21- to 22-bp protected region centered around the highly homologous, partially palindromic 13-bp or 14-bp sites for both MDBP (7,11) and EF-C/EP (14). Furthermore, the seven G residues of λ site 1 contacting MDBP, as determined by dimethyl sulfate interference assays, have their counterparts implicated in HBV site 1/EF-C interaction by analogous assays (8). We could detect no HBV site 1-specific activity in crude nuclear extracts of mammalian cells other than MDBP, *e.g.*, no HBV site 1·protein complex formation was detected in the presence of excess methylated pB site 1 while such complexes formed in the presence of excess unmethylated pB site 1 (Fig. 1). The functional importance of HBV site 1 was demonstrated by Ostapchuk and coworkers (8) who found that deletion of 1 bp or insertions of 3 to 10 bps at the single 13-bp MDBP/EF-C site in the HBV enhancer reduces enhancer activity about 10 fold in transient expression assays as well as inhibiting binding of this factor. We have recently demonstrated that one of the two, high-affinity methylation-independent MDBP sites from an enhancer of human cytomegalovirus, when multimerized by itself, can weakly stimulate transcription in transient expression assays (submitted for publication). That the stimulation was only weak and was not seen in all the multimer constructs suggests that MDBP sites in enhancers resemble certain other DNA binding sites (enhancers) for sequence-specific DNA-binding proteins in requiring correctly positioned dissimilar enhancers to form a functional enhancer module (27). This may be especially true for MDBP because it is a large protein (10) and because it might act to down-modulate transcription in certain DNA sequence contexts (see below).

A site in the polyomavirus B enhancer (Py site 2, Table 1) that is closely related to HBV site 1, was shown by Ostapchuk *et al.* (8) to bind specifically to EF-C. We found, as expected, that Py site 2 was bound specifically by MDBP and that HBV site 1 competed efficiently with Py site 2 for binding to MDBP (Fig. 2). The hypothesis that Py site 2 plays a role in the activity of the B enhancer of polyomavirus is consistent with deletion analyses of Böhnlein *et al.* (28) and Muller *et al.* (29). We have also located a previously unknown Py enhancer site, namely, Py site 1 (Fig. 2). It differs at four out of 14 positions from Py site 2 including a central A·T base-pair that is, in effect, deleted from Py site 2 (Table 1). It too is found in the B enhancer but at the very end of the enhancer, only 5 bps away from the run of eight T residues that mark the beginning of the Py *ori* core sequence (Fig. 3). This region adjacent to the beginning of the *ori* core might affect the tissue-specificity or latency of viral infection *in vivo* (30). LPV, a papovavirus related to Py and infecting primate cell lines of B-lymphocyte origin, also contains an MDBP site near its putative *ori* core region (Table 1). The significance of the high degree of homology between *ori* core border regions of the Py and LPV genomes at Py site 1 and LPV site 1 (Fig. 3) remains to be determined.

We also showed that a 13-bp sequence in the LTR of a lentivirus, equine infectious anemia virus, was recognized by MDBP (Table 1). The occurrence of an MDBP site 50 bps upstream of a TATA box and 8 bps upstream of a potential CC-AAT box (18) suggests a role in promoter function. However,

transient expression assays indicate that the LTR promoter region begins shortly downstream of the MDBP site, EIAV site 1 (31). Deletion of EIAV site 1 increased promoter activity several fold in assays using feline cells although not in an analogous assay on a canine cell line (31). MDBP may, therefore, play a negative role in transcription at the EIAV LTR, which might be demonstrable only in cell lines with high concentrations of this protein.

We have proposed a negative transcription regulatory role for the highly repeated methylation-independent MDBP sites spread throughout the human apolipoprotein(a) gene (32). We also propose that the four identified methylation-responsive MDBP sites located only ~30 to ~110 bps after the transcription initiation site of mammalian genes (Table 2) are involved in methylation-dependent repression or attenuation. At these sites, MDBP may be able to interact with proteins binding to the nearby upstream promoter elements. Alternatively, it may interact with factors bound to immediate-downstream promoter elements (33) or with other parts of the transcription machinery (34). Several of the CpG methylation-responsive MDBP sites are in genes whose CpG methylation is implicated in transcription control. One methylation-dependent MDBP site was found ~30 bps downstream from the transcription initiation site of the human HLA-A2 gene (23 and K. Zachow and H. Orr, unpubl. results), which encodes a class I major histocompatibility antigen (Table 1). Several studies suggest that certain changes in methylation of HLA gene regions can control transcription (35,36). For example, human cells whose DNA was unable to genetically transform mouse cells to a human HLA class I antigen-producing phenotype were able to do so after treatment with 5-azacytidine, a potent DNA demethylating agent, prior to isolation of their DNA for transfection (37). Similarly, the three methylation-responsive HPRT sites at approximately -170, +85, and +110 (Table 2) might be involved in negative transcription control. Several studies have provided evidence for inverse correlations between methylation at certain sites in the human HPRT gene region and expression as well as for demethylation of this region activating expression of this X chromosome-linked gene (38-41). Our identification of a methylation-responsive MDBP site only 49 bps downstream from the cap site for the human α -galactosidase A gene (Table 1) may be a related phenomenon. This is a X chromosome-linked gene whose expression in females, like that of the HPRT gene, seems to be tightly controlled by X chromosome inactivation (42) and which can be activated by 5-azacytidine treatment (43).

We suggest that MDBP is one of a number of sequence-specific DNA-binding proteins that can bind to the 5' regions within vertebrate genes in a CpG methylation-dependent manner and thereby help to negatively regulate the expression of those genes. Consistent with this hypothesis is the report of Saluz and coworkers of another DNA-binding protein that is DNA sequence- and DNA methylation-specific and binds preferentially to a CpG-containing avian vitellogenin DNA sequence 10 bps downstream from the transcription initiation site (44). That site has been implicated in methylation-controlled down-regulation of expression of this gene. Although, as described above, MDBP is likely to function as a positive transcription factor in its binding to the HBV enhancer, several transcription factors have been shown to positively regulate transcription from some sites and negatively regulate it from others depending on which other proteins bind nearby (45-47). We identified MDBP's ability to distinguish between CpG-methylated sequences and unmethylated

sequences only because we specifically searched for such a protein (2). We hypothesize that some of the site-specific changes in DNA methylation correlated with transcription may be associated with recognition by other transcription factors whose CpG-methylation specificity has not yet been detected.

If MDBP or other sequence-specific DNA-binding proteins act as transcription repressors, stimulators, or terminators at CpG-containing sites requiring methylation for appreciable binding, then specific DNA-binding and transcription control could be disrupted by active demethylation of the site, in the absence of DNA replication. Alternatively, the temporary hemimethylation resulting from DNA replication might help initiate a round of differentiation. Based upon the differences observed in the binding of hemimethylated vs. bifilarly methylated CpG-containing MDBP sites (Table 2), we propose the following model relating DNA replication, cellular differentiation, and DNA methylation at certain transcription factor-binding DNA sites. Replication involves a transient single-stranded state of the site, which for MDBP precludes specific binding (11), and temporarily converts a bifilarly methylated site to a hemimethylated site that may bind the protein much more poorly than the parental duplex site (e.g., hHLA site 1, Table 2). Interestingly, even though MDBP sites have hyphenated dyad symmetry and the CpG dinucleotide is symmetrical, methylation of only one strand or the other at a methylation-responsive DNA site often gave very different extents of complex formation with MDBP (Figs. 4 and 5; Table 2). Therefore, the two progeny chromosomes could have different fates in terms of MDBP binding. Although usually both copies of the replicated site will eventually become bifilarly methylated, competition between sequence-specific DNA-binding proteins, histones, and other chromatin proteins for binding to the DNA immediately after replication of a DNA sequence may determine the long-term organization of chromatin around that sequence. If the two homologous in each of the daughter cells happens to have the same orientation of hemimethylation at this site (m/u or u/m), then, for certain kinds of cells this may allow the gene to be expressed in one of the two daughter cells but not in the other. MDBP could thereby serve as one of the links between DNA methylation and transcription control in a manner that would help explain how one parental cell can differentiate into two phenotypically different progeny cells (48). In accord with this model, evidence has been presented for DNA replication-associated hemimethylation at 5'-GATC-3' sites activating IS10 transposition in *E. coli* and doing so unequally depending upon which of the DNA strands is methylated (49).

ACKNOWLEDGEMENTS

We thank Lothar Hennighausen for the gift of the Raji nuclear extract. This research was supported in part by USPHS grant GM33999.

REFERENCES

- Supakar, P.C., Weist, D., Zhang, D., Inamdar, N., Zhang, X.-Y., Khan, R., Ehrlich, K.C. and Ehrlich, M. (1988) *Nucleic Acids Res.* **16**, 8029–8046.
- Huang, L.-H., Wang, R., Gama-Sosa, M.A., Shenoy, S. and Ehrlich, M. (1984) *Nature* **308**, 293–295.
- Wang, R.Y.-H., Zhang, X.-Y., Khan, R., Zhou, Y., Huang, L.-H. and Ehrlich, M. (1986) *Nucleic Acids Res.* **14**, 9843–9860.
- Ehrlich, M. and Wang, R.Y.-H. (1981) *Science* **212**, 1350–1357.
- Gama-Sosa, M.A., Midgett, R.M., Slagel, V.A., Githens, S., Kuo, K.C., Gehrke, C.W. and Ehrlich, M. (1983) *Biochim. Biophys. Acta* **740**, 212–219.
- Clawson, G.A., Willis, D.B., Weissbach, A. and Jones, P.A., eds., (1990) *Nucleic Acid Methylation*. John Wiley & Sons, New York.
- Zhang, X.-Y., Supakar, P.C., Khan, R., Ehrlich, K.C. and Ehrlich, M. (1989) *Nucleic Acids Res.* **17**, 1459–1474.
- Ostapchuk, P., Scheirle, G. and Hearing, P. (1989) *Molec. Cell Biol.* **9**, 2787–2797.
- Dignam, J.D., Lebovitz, R.M. and Roeder, R.G. (1983) *Nucleic Acids Res.* **11**, 1475–1489.
- Supakar, P.C., Zhang, X.-Y., Githens, S., Khan, R., Ehrlich, K.C. and Ehrlich, M. (1989) *Nucleic Acids Res.* **17**, 8611–8629.
- Khan, R., Zhang, X.-Y., Supakar, P.C., Ehrlich, K.C. and Ehrlich, M. (1988) *J. Biol. Chem.* **263**, 14374–14383.
- Zhang, X.-Y., Ehrlich, K.C., Wang, R.Y.-H. and Ehrlich, M. (1986) *Nucleic Acids Res.* **14**, 8387–8397.
- Karpen, S., Banerjee, R., Zelent, A., Price, P. and Acs, G. (1988) *Molec. Cell Biol.* **8**, 5159–5165.
- Ben-Levy, R., Faktor, O., Berger, I. and Shaul, Y. (1989) *Molec. Cell Biol.* **9**, 1804–1809.
- Wang, R.Y.-H., Zhang, X.-Y. and Ehrlich, M. (1986) *Nucleic Acids Res.* **14**, 1599–1614.
- Pawlita, M., Clad, A. and zur Hausen, H. (1985) *Virology* **143**, 196–211.
- Tseng, R.W. and Fujimura, F.K. (1988) *J. Virol.* **62**, 2890–2895.
- Derse, D., Dorn, P.L., Levy, L., Stephens, R.M., Rice, N.R. and Casey, J.W. (1987) *J. Virol.* **61**, 743–747.
- Kim, S.H., Moores, J.C., David, D., Respass, J.G., Jolly, D.J. and Friedmann, T. (1986) *Nucleic Acids Res.* **14**, 3103–3118.
- Sargent, T.D., Yang, M. and Bonner, J. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 243–246.
- Jagodzinski, L.L., Sargent, T.D., Yang, M., Glackin, C. and Bonner, J. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 3521–3525.
- Costanzo, F., Colombo, M., Staempfli, S., Santoro, C., Marone, M., Frank, R., Delius, H. and Cortese, R. (1986) *Nucl. Acids Res.* **14**, 721–736.
- Koller, B.H. and Orr, H.T. (1985) *J. Immunol.* **134**, 2727–2733.
- Ennis, P.D., Zemmour, J., Salter, R.D. and Parham, P. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 2833–2837.
- Quinn, M., Hantzopoulos, P., Fidanza, V. and Calhoun, D.H. (1987) *Gene* **58**, 177–188.
- Roebroek, A.J.M., Schalken, J.A., Bussemakers, M.J.G., van Heerikhuizen, H., Onnekink, C., Debruyne, F.M.J., Bloemers, H.P.J. and Van de Ven, W.J.M. (1986) *Molec. Biol. Rep.* **11**, 117–125.
- Dynan, W.S. (1989) *Cell* **58**, 1–4.
- Böhnlein, E. and Gruss, P. (1986) *Molec. Cell Biol.* **6**, 1401–1411.
- Muller, W.J., Dufort, D. and Hassell, J.A. (1988) *Molec. Cell Biol.* **8**, 5000–5015.
- Rochford, R., Campbell, B.A. and Villarreal, L.P. (1990) *J. Virol.* **64**, 476–485.
- Dorn, P.L. and Derse, D. (1988) *J. Virol.* **62**, 3522–3526.
- Ehrlich, K. and Ehrlich, M. (1990) *Molec. Cell Biol.* **10**, 4957–4960.
- Ayer, D.E. and Dynan, W.S. (1988) *Molec. Cell Biol.* **8**, 2021–2033.
- Buratowski, S., Hahn, S., Guarente, L. and Sharp, P.A. (1989) *Cell* **56**, 549–561.
- Levine, F. and Pious, D. (1985) *Immunogenetics* **22**, 427–440.
- Reitz, M.S., Mann, D.L., Eiden, M., Trainor, C.D. and Clarke, M.F. (1984) *Molec. Cell Biol.* **4**, 890–897.
- Alberti, S. and Herzenberg, L.A. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 8391–8394.
- Venolia, L. and Gartler, S.M. (1983) *Nature* **302**, 82–83.
- Wolf, S.F., Jolly, D.J., Lunnan, K.D., Friedmann, T. and Migeon, B.R. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 2806–2810.
- Yen, P.H., Patel, P., Chinault, A.C., Mohandas, T. and Shapiro, L.J. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1759–1763.
- Paterno, G.D., Adra, C.N. and McBurney, M.W. (1985) *Molec. Cell Biol.* **5**, 2705–2712.
- Desnick, R.J. and Sweeley, C.C. (1983) In Stanbury, J.S., Wyngaarden, J.S., Fredrickson, D.S., Goldstein, J.L. and Brown, M.S. (eds), *The Metabolic Basis of Inherited Diseases*. McGraw-Hill, New York, pp. 906–944.
- Mohandas, T., Sparkes, R.S., Bishop, D.F., Desnick, R.J. and Shapiro, L.J. (1984) *Am. J. Hum. Genet.* **36**, 916–924.
- Saluz, H.P., Feavers, I.M., Jiricny, J. and Jost, J.P. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6697–6700.
- Shore, D. and Nasmyth, K. (1987) *Cell* **51**, 721–732.
- Thierry, F. and Yaniv, M. (1987) *EMBO J.* **6**, 3391–3397.
- Glass, C.K., Holloway, J.M., Devary, O.V. and Rosenfeld, M.G. (1988) *Cell* **54**, 313–323.
- Hall, P.A. and Watt, F.M. (1989) *Development* **106**, 619–633.
- Marinus, M.G. (1987) *Ann. Rev. Genet.* **21**, 113–131.
- Zhang, D., Ehrlich, K.C., Supakar, P.C. and Ehrlich, M. (1989) *Molec. Cell Biol.* **9**, 1351–1356.