Related sites in human and herpesvirus DNA recognized by methylated DNA-binding protein from human placenta

Xian-Yang Zhang, 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 December 5, 1988; Revised and Accepted January 16, 1989

ABSTRACT

Methylated DNA-binding protein (MDBP) from mammalian cells binds specifically to six pBR322 and M13mp8 DNA sequences but only when they are methylated at their CpG dinucleotide pairs. We cloned three high-affinity MDBP recognition sites from the human genome on the basis of their binding to MDBP. These showed much homology to the previously characterized prokaryotic sites. However, the human sites exhibited methylation-independent binding apparently because of the replacement of m⁵C residues with T residues. We also identified three other MDBP sites in the herpes simplex virus type 1 genome, two of which require in vitro CpG methylation for binding and are in the upstream regions of viral genes. A comparison of MDBP sites leads to the following partially symmetrical consensus sequence for MDBP recognition sites: 5'-R T m⁵Y R Y Y A m⁵Y R G m⁵Y R A Y-3'; m⁵Y (m⁵C or T), R (A or G), Y (C or T). This consensus sequence displays an unusually high degree of degeneracy. Also, interesting deviations from this consensus sequence, including a one base-pair deletion in the middle, are sometimes observed in high-affinity MDBP sites.

INTRODUCTION

Despite extensive evidence for the involvement of DNA methylation at CpG dinucleotides in controlling the expression of a number of vertebrate and viral genes (1-6), there is very little information about how this regulation is effected. We have isolated a novel eukaryotic protein shown to bind to certain DNA sequences only when they are methylated at C residues in CpG dinucleotide pairs to yield 5-methylcytosine (m⁵C) residues (7). This protein, MDBP, is present in various human, rodent, and bovine tissues and cultured cell lines but was not detected in cultured mosquito cells (8). We have characterized six different but partially homologous recognition sites for MDBP from E. coli plasmid and bacteriophage DNAs (9-12). These sites contain two to six CpG dinucleotides which have to be methylated in vitro for recognition by MDBP. Nonetheless, one site from pBR322 DNA, the 14-bp pB site 1, was shown to no longer require cytosine methylation for binding upon mutation of one specific CpG at this site to a TpG or TpA (8).

We describe here the cloning and analysis of three human DNA sites recognized by MDBP. These sites show considerable homology to pB site 1, which is the highest affinity site that we had found for MDBP (9-11). However, MDBP binds to these human sites even in the absence of DNA methylation apparently as a result of T residues substituting for several m^5C residues in the analogous positions of pB site 1. Three MDBP binding sites were also identified in herpes simplex virus (HSV) type 1 DNA. One of these in a putative promoter region of a gene of unknown function is almost identical to pB site 1 and similarly requires methylation for its recognition by MDBP. Of the other two HSV sites, one is upstream of the α -trans-inducing factor gene and dependent upon cytosine methylation for binding and the other is in the thymidine kinase gene and does not require cytosine methylation. A comparison of these sites, a newly determined and unusual site in bacteriophage λ DNA, and previously determined (11) phage and plasmid sites leads to a consensus sequence notable for its mixture of degeneracy and conservation.

MATERIALS AND METHODS

DNA ligands. Three human DNA inserts (hu-48, hu-23, and hu-40) from an MboI digest of nuclease S1-treated human brain DNA were cloned into pUC118 (13) on the basis of their affinity for MDBP. This was done by the method of Gronostajski et al. for NF1 (14) except that standard conditions for incubation with MDBP (10) were used for two cycles of binding followed by dissociation of the DNA-protein complexes and elution of the DNA with 0.3 M sodium acetate, 10 mM EDTA, 20 mM Tris-HCl, pH 8.0, for 3 h at 65°C. The Escherichia coli host for cloning was MV1193 (13). These human inserts were sequenced on both strands. Covalently closed circular replicative form (RF) DNA from the recombinant single-stranded DNAs (13) were prepared by primer extension (12). A pUC12 recombinant, pENKAT-12 (15), which contains human preproenkephalin DNA sequences and part of the HSV type 1 thymidine kinase gene (position +54 to +340 relative to the transcription initiation site; 16) and the E. coli chloramphenicol transacetylase gene, was also used. A pBR322/ λ DNA recombinant (pB λ) was constructed from an MDBP site-containing BamHI/NdeI fragment from positions 27,972 to 29,884 of the λ genome and a 2439-bp BamHI/NdeI fragment of pBR322.

Complementary oligonucleotides, one of which was labeled with [γ - ^{32}P]ATP, were synthesized and annealed (9, 12). In some cases, 5'-overhangs on oligonucleotide duplexes were filled in with m^5dCTP or dCTP

plus dGTP, dTTP, and dATP in a reaction catalyzed by the Klenow fragment of *E. coli* DNA polymerase I. For example, W17/C17, a 14-bp m⁵CpG-containing duplex composed of pB site 1 (Table 1) and *Xba*I 5'-overhangs (9), was end-filled with the standard deoxynucleoside triphosphates to give a 22-bp duplex. W9/C9 contains the same methylated pB site 1 as W17/C17 plus 21 adjacent base-pairs from pBR322 to give a 35-bp duplex (9).

Assays for MDBP binding sites. MDBP-specific DNA sequences or restriction fragments were identified, as previously described, by the gel retardation assay for visualizing complexes between MDBP and a ³²P-labeled duplex or by the fragment selection assay for determining which restriction fragment in a small restricted DNA is preferentially retained by MDBP (9, 10). In the first assay, the DNA-protein complexes are electrophoresed and in the second, these complexes are eluted from a nitrocellulose membrane after filtration and the dissociated DNA fragments are electrophoresed. These analyses utilized the hydroxylapatite fraction of MDBP (7). One unit of MDBP is defined as the amount needed to bind 1 fmol of W17/C17 (pB site 1) under standard conditions in the gel retardation assay (9, 11). Using the fragment selection assay, the relative binding of pB site 1, λ site 1, and HSV site 2 was quantitated by comparing the percentage retention of an MDBP-specific fragment in a *Hinf*I digest of human DNA methyltransferase-methylated pBR322 DNA (~85% methylation of CpG sites in the 298-bp MDBP-specific fragment; 17), a *Taq*I digest of p λ (729-bp MDBP-specific fragment), and a *Hpa*II digest of pENKAT-12 (560-bp MDBP-specific fragment), respectively. Newly identified MDBP binding sites were localized by DNase I protection assays (footprinting) and dimethyl-sulfate protection experiments (9, 10).

Copy number determination. Oligonucleotides containing hu sites 1 or 2 plus some surrounding sequence (5'-GGTTATTACATGTCATTCCTGGTAACC-3' and 5'-AAGAGGCATTGGGTGCCATGGTGACGG-3') were used as hybridization probes. These ³²P-labeled probes were hybridized to 10 μ g of human brain DNA digested with *Mbo*I or *Eco*RI and transferred to nitrocellulose membranes after electrophoresis (9). Blots were prehybridized and then, upon addition of the ³²P-labeled probe, hybridized at 42°C in 6X SSPE (1X SSPE: 180 mM NaCl, 10 mM sodium phosphate, 1 mM EDTA, pH 8) containing 1% sodium dodecyl sulfate (SDS), 10X Denhardt's solution and 100 μ g/ml *E. coli* DNA. The blots were washed with 6X SSPE, 1% SDS at 50°C and then 0.5X SSPE, 1% SDS at 40°C for 1 hr and subjected to autoradiography. Human DNA-pUC118 recombinants were used as standards to quantitate the hybridization signal from genomic DNA digests.

TABLE 1. Comparison of the sequences of MDBP binding sites

MDBP binding site	Sequence of binding site ^a				Matches in central 14 bp (%)	Symmetrical base-pairs in central 14 bp (No.)
Position No.	1	5	10	14		
pB site 1	g c t g <u>A T M G T C A M G G M G A T</u> t t a t				100	10
HSV site 1	a t c c <u>A T M G C C A M G G M G A T</u> t g g a				100	12
hu site 1	t g t c <u>A T T C C C A T G G T A A C</u> c a c a				93	10
hu site 2	t t t g <u>G G T G C C A T G G T G A C</u> g g g a				93	8
hu site 3	c t a t <u>G T A G T C A T A G A G A T</u> a g g g				86	4
λ site 1 ^b	a c t c <u>G T T G C C</u> - C G G T A A C a a c a				93 ^b	10
HSV site 2	t c t c <u>G C G G C C A T A G C A A C</u> c g a c				79	8
HSV site 3	c g a c <u>G M G G C C A M G G M G A C</u> t c g a				86	8
pB site 2	c t a m <u>G M G A T C A T G G M G A C</u> c a c a				86	6
M13 site 2	t t c m <u>G T T A T T A T T G M G T T</u> t c c t				86	2
M13 site 3	g c t a C T G A T T A M G G T G C T g c t a				79	2
Consensus ^c	- t c G T G C C A T G M G A C c - - -					
	c t m A T M A T T M A G T A T t					
Matches at 11 sites ^c	9 7 7 10 11 10 11 11 11					
	6 10 6 11 10 10 9 9 7					
MDBP ⁺ pB site 1 variants ^d	A T M G T C A M G G M G A T t					
	G T T G C T A A T G C c					
MDBP ⁻ mutations ^d at pB site 1	G A C A G A G					
M13 site 4 ^e	m g g g <u>A T M G T M A M M M T M A G</u> m a g m g				71	2

^aThe sequence of one strand (5' to 3') is shown with the central 14-base region analogous to the minimal recognition sequence of pB site 1 given in capital letters. Residues in this 14-bp region matching the consensus sequence are underlined. Although four sites have a G in the third position, it is not included in the consensus sequence because these sites all have a somewhat lower affinity than almost all of the other sites and this base destroys the dyad symmetry at that position. M denotes 5-methylcytosine residues. Localization of pBR322 (pB) and M13mp8 (M13) sites was previously described (10,11). HSV site 1 is in the promoter region of an HSV type 1 gene of unknown function, HSV site 2 is in the beginning of the thymidine kinase gene of HSV type 1, and HSV site 3 is upstream of the α -trans-inducing factor gene (see Results). The human sites (hu sites 1, 2 and 3) are in previously undescribed human DNA sequences. M (⁵mC) residues were introduced *in vitro* by methylation catalyzed by human DNA methyltransferase, incorporation of ⁵mC residues into synthetic oligonucleotides or, in the case of M13 site 4, incorporation of ⁵m³dCTP. For all these ⁵mC-containing sequences, binding was not observed in the absence of methylation and partial methylation can still give an MDBP site albeit one of lower affinity (Table 2B; 10, 12). M13 site 1, which also requires CpG methylation for binding to MDBP is not listed because it has not been precisely localized and might consist of several nearby weak binding sites (11; R.Y.-H. Wang and M. Ehrlich, unpubl. results).

^bThis homology, which is at position 28,249 to 28,261 of the λ genome, requires a 1-bp gap in λ site 1.

^cIn capital letters is the 14-bp core consensus sequence. Where alternate bases are at a given position, the most prevalent one is shown on top. The number of matches to the consensus sequence in the 11 sequences is shown.

^dSingle-site mutants of pB site 1, which are still recognized by MDBP (MDBP⁺) or are recognized very poorly or not at all (MDBP⁻) were described previously (8, 9, 12).

^eM13 site 4 is the only one of these sites whose recognition by MDBP depends on extensive substitution with m⁵C residues in this region rather than methylation just at CpG dinucleotides (9) and so was not used in deriving the consensus sequence because its pattern of methylation is unnatural.

RESULTS

Three MDBP recognition sites in herpes simplex viral DNA

We used the 14-bp pB site 1 sequence (Table 1) and its MDBP-binding variants (9, 12) to search for highly homologous sequences in the mammalian and mammalian viral DNA database in GENBANK. Only one exact homologue was found. This site from the HSV type 1 genome is referred to as HSV site 1 (Table 1). We synthesized it as a 22-bp oligonucleotide with three, one, or no m⁵C residues (Table 2). Fig. 1A shows that both m⁵CpG-containing constructs gave complexes with MDBP that coelectrophoresed with those of methylated wild-type pB site 1 (duplex W17/C17) and the more methylated duplex formed much more complex just as more methylation of pB site 1 gives more complex formation (12).

A second MDBP site in the HSV type 1 genome, HSV site 2 (Table 1), was discovered by serendipity. The GENBANK search had uncovered two sequences in the human preproenkephalin gene promoter (5'-CCCGGGACGGCGAG-3' and 5'-AGCCCGCCGGCGAT-3'; 18) with moderate homology to pB site 1. We, therefore, tested an *in vitro*-methylated, recombinant pUC DNA, pENKAT-12, having these sequences as well as some herpes DNA (18), by the fragment selection assay. MDBP did not preferentially retain any preproenkephalin sequence but did recognize a 560-bp *Hpa*II fragment containing HSV type 1 sequences and bound to this fragment whether or not it was methylated (Fig. 2A). This fragment competed much better than nonspecific DNA fragments for MDBP binding to methylated pB site 1 in fragment selection and gel retardation assays (data not shown). DNase I footprinting (Fig. 3A) indicated that the site was at about position +168 to +181 (16) of the HSV thymidine kinase gene (Table 2). Dimethylsulfate interference analyses and the use of oligonucleotide duplexes of different lengths in gel retardation assays confirmed the location of this MDBP binding site (Table 2 and Fig. 1A). The latter assays with various overlapping oligonucleotide duplexes indicate that a few extra base pairs

TABLE 2. Localization of sequences binding to MDBP by DNase I footprinting, dimethylsulfate interference, and oligonucleotide primer extension experiments.

A. DNase I footprinting and dimethylsulfate interference ^a			
Name of Site	Sequence		
HSV site 2	5'-tctcGCGGCCATAGCAACcgac-3'		
hu site 1	5'-tgtcATTCCCATGGTAACcaca-3' 3'-acagTAAGGGTACCATTGgtgt-5'		
hu site 2	5'-tttgGGTGCCCATGGTGACggga-3' 3'-aaacCCACGGTACCCTGcct-5'		
λ site 1	5'-actcGTTGCCCGGTAACaaca-3' 3'-tgagCAACGGGCCATTGttgt-5'		
B. Oligonucleotide primer extension ^b			
Name of Site	Duplex No.	Sequence	Binding relative to pB site 1 (%)
HSV site 1	I	5'-atccATCGCCACGGCGATtgga-3' 3'-taggTAGCGGTGCCGCTAacct-5'	-
HSV site 1 (1 m ⁵ C)	II	5'-atccATMGCCACGGCGATtgga-3' 3'-taggTAGCGGTGCCGCTAacct-5'	5
HSV site 1 (3 m ⁵ C's)	III	5'-atccATMGCCAMGGMGATtgga-3' 3'-taggTAGCGGTGCCGCTAacct-5'	48
HSV site 2	IV	5'-tctcGCGGCCATAGCA-3' 3'-agagCGCCGGTATCGTTGgctg-5'	<2
HSV site 2	V	5'-tctcGCGGCCATAGCAACc-3' 3'-agagCGCCGGTATCGTTGgctg-5'	5
HSV site 2	VI	5'-tctcGCGGCCATAGCAACcgac-3' 3'-agagCGCCGGTATCGTTGgctg-5'	26
HSV site 2	VII	5'-GCGGCCATAGCAACcgac-3' 3'-agagCGCCGGTATCGTTGgctg-5'	<2
λ site 1	VIII	5'-ctcGTTGCCCGGTAAC-3' 3'-gagCAACGGGCCATTGttgtcg-5'	11
λ site 1	IX	5'-ctcGTTGCCCGGTAACaacagc-3' 3'-gagCAACGGGCCATTGttgtcg-5'	73
λ site 1	X	5'-TGCCCGGTAACaacagc-3' 3'-gagCAACGGGCCATTGttgtcg-5'	<2
HSV site 3	XI	5'-tcgacGMGGCCAMGGMGAC-3' 3'-gCGCCGGTGCCGCTGagct-5'	13
HSV site 3	XII	5'-tcgacGCGGCCACGGGCAC-3' 3'-gCGCCGGTGCCGCTGagct-5'	<2

^aDNase I footprinting and dimethylsulfate interference experiments were conducted on restriction fragments as described in Materials and Methods. The lower strand of HSV site 2 was not examined. Base-pairs homologous to the central 14-bp region of dyad symmetry of pB site 1 (Table 1) are shown in capital letters and the adjacent regions in lower case letters. G residues whose methylation interfered with MDBP binding are indicated by a dot and regions protected by MDBP against DNase I are underlined or overlined. Note that both the right and left half of λ site 1 cannot be aligned simultaneously with homologous bases in the other sites unless a 1-bp gap is introduced in the middle as in Table 1.

^bTo determine the relative binding of synthetic oligonucleotide duplexes, gel retardation assays were performed with 10 fmol of duplex and 1.5 units of MDBP as described in Materials and Methods. Binding was compared to that of end-filled W17/C17 containing pB site 1, which was taken as 100%. The sequence of the upper strand of this bifilarly methylated, blunt-ended oligonucleotide duplex with its pB site 1 sequence capitalized and XbaI ends in lower-case letters is 5'-ctagATMGTCAMGGMGATctag-3'. Duplexes IV, VIII, XI and XII were formed just by annealing oligonucleotides. Duplexes I, II, VI, VII, IX, and X were the products of a standard primer extension reaction on annealed oligonucleotides. Duplex I was formed using 5'-ATCCATCGCC-3' as a primer and duplex II with 5'-ATCCATMGCC-3' as a primer. Only for synthesis of duplex III (with the same primer used for duplex II) was m⁵dCTP substituted for dCTP. Duplex V was formed by a primer extension reaction of duplex IV in which dGTP was omitted.

beyond the central 14-bp (Table 1) are necessary for binding (duplex VII vs. duplex VI, Table 2).

A third herpes MDBP site, HSV site 3, was found by searching the GENBANK database for homology to HSV site 2. MDBP bound a 15-bp oligonucleotide duplex containing this site when there were three m⁵CpG dinucleotides in one strand but showed negligible binding to it when the site was unmethylated (Table 2). Bifilar methylation would probably increase the extent of binding several times (12) and filling in the 5' overhangs might further increase binding (Table 2). This site is located 326 bp upstream of the α -trans-inducing factor gene, which places it at the end of a neighboring gene coding for a 42 kDa protein (19).

Three MDBP recognition sites cloned from human DNA

Partially purified MDBP (7) was used to select for recognition sites in an MboI digest of human brain DNA in order to determine whether the sequences of the prokaryotic MDBP sites previously identified (11) are representative of mammalian MDBP sites. Such cloning should reveal mostly high-affinity sites with MDBP as the selection agent against a background of $\sim 10^7$ restriction fragments per haploid human genome. The MboI fragments (5 μ g) were incubated with MDBP and filtered through a nitrocellulose membrane. The fragments that were retained by the membrane (0.4% of the total) were eluted, purified, reincubated with MDBP in the presence of 100 ng of linearized

Fig. 1. Visualization of complexes of MDBP with DNA fragments or oligonucleotide duplexes containing prokaryotic or eukaryotic MDBP recognition sites. Panel A shows MDBP complexes with 10 fmol of ^{32}P -labeled 22-bp, blunt-ended oligonucleotide duplexes containing HSV site 1 with 1 m^5C residue (m_1) or 3 m^5C residues (m_3), HSV site 2, λ site 1, or pB site 1 with 6 m^5C residues (Table 2B) in a 5% polyacrylamide gel. The free oligonucleotide ligands were run off the gel to afford better resolution of the very low-mobility complexes. Panel B shows the same type of assay with 3.5 fmol of purified ^{32}P -labeled 309-bp or 245-bp *Hind*III/*Eco*RI DNA fragments containing hu site 1 or hu site 2, respectively, cloned from human DNA. In addition to the standard 100 ng each of poly(dI)·poly(dC) and *Mbo*I-digested *M. luteus* DNA present in the preincubation mixture before addition of the ^{32}P -labeled ligand (9), 1 μg of *Hind*III-digested, unmethylated pBR322 (pBR), 1 μg of *Taq*I-digested pB λ , 2 pmol of the methylated 35-bp duplex containing pB site 1 (W9/C9), or 2 pmol of the analogous unmethylated duplex (W9u/C9u) were present. Samples which did not contain this additional competitor DNA are indicated. As seen in panel B, only the very low-mobility complexes formed with the hu site 1-containing DNA fragment have the specificity of MDBP. The uncomplexed DNA fragments are seen at the bottom of the gel.

pUC118 DNA, filtered, and the nitrocellulose-retained fragments (12% of the input DNA fragments) cloned in pUC118. The *in vitro* methylated DNA from twelve clones was examined by the fragment selection assay. Seven of these did not show specific binding of their insert fragments to MDBP. Five of the clones did demonstrate such binding but their binding was independent of *in vitro* methylation (Fig. 2B and data not shown); however, the cloning involved a *Mcr*A⁺, *Mcr*B⁺ *E. coli* host (20), which could have selected against propagation of methylated MDBP sites.

Two of the above clones, hu-48 and hu-23, showing methylation-independent binding to MDBP were examined further. Sequencing revealed that

Fig. 2. Detection of MDBP sites in herpes viral DNA, cloned human DNA sequences, and λ phage DNA by the fragment selection assay. Recombinant DNAs containing HSV site 2 (A), hu sites 1 (B, lanes 4-6) and 2 (B, lanes 1-3), and λ site 1 (C) were restricted with *Hpa*II (A) or *Taq*I (B and C). Radiolabeling was by filling in 5'-overhangs with [α - 32 P]dCTP plus the other standard deoxynucleotide triphosphates (A & C) or by primer extension on recombinant viral strand templates with [α - 32 P]dATP, dGTP, dTTP, and m⁵dCTP (B, lanes 1, 2, 5 & 6) or dCTP (B, lanes 3 & 4) to prepare RF DNA, which was purified (12) and then digested with *Taq*I. The autoradiogram shows fragments retained by the MDBP preparation except for lanes 1 and 6, which contained an aliquot of the restriction digests electrophoresed without prior incubation with MDBP or nitrocellulose-filtration. The size in base-pairs of insert fragments preferentially retained by MDBP is indicated. In panel B, lanes 2 and 5, the top band is a vector (pUC18) fragment which binds well to MDBP only when substituted with m⁵C at C residues throughout one strand rather than just at CpG sites (9, 12).

these are previously undescribed sequences having a site (hu site 1 or 2) highly homologous to pB site 1 (Fig. 4). Therefore, this previously identified prokaryotic MDBP site, which was studied by extensive site-directed mutagenesis (9, 12), does offer a good model for MDBP sites in general. Labeled restriction fragments containing these human sites formed very-low-mobility complexes with MDBP in the presence of excess unlabeled pB site 1 duplex only if the competing pB site 1 sequences were unmethylated (W9u/C9u or pBR, Fig. 1B). The location of hu sites 1 and 2 was confirmed by DNase I footprinting and dimethylsulfate interference experiments (Fig. 3 and Table 2). These experiments indicated that a sequence from position 16 to 30 of hu-23, which is 71% homologous, allowing a 1-bp loop-out, to the nearby hu site 2 (Fig. 4), does not bind to MDBP. Hybridization, under stringent

Fig. 3. DNase I footprint analysis of MDBP binding to HSV site 2 (A), hu site 1 (B), hu site 2 (C), or λ site 1 (D). DNase I footprinting was performed on an 153-bp *HpaII*/*Sau3A* fragment from pENKAT-12 (A), a 114-bp *TaqI*/*HinfI* fragment from pB λ (D), and the fragments from cloned human DNA described in Fig. 1 (B). These purified fragments were single end-labeled and subjected to limited DNase I digestion (9) in the presence of the hydroxyapatite fraction of MDBP (lane 1) or in its absence (lane 2) and electrophoresed on a standard sequencing gel. DNA partially degraded at G, G + A, or C residues was co-electrophoresed as sequencing markers. Brackets denote the region protected by MDBP against DNase I digestion. These footprints show the lower strand (see Table 2) of all the sites examined except for panel A, which shows the upper strand. For the human and λ sites, both strands were examined (Table 2A).

conditions, of *EcoRI* or *MboI* digests of human DNA to an oligonucleotide probe containing hu site 1 gave a single band in each digest with the intensity of a single-copy sequence. Similar analysis with the hu site 2 probe indicated

hu-48

hu site 1

GATCTTGAAGTCAAATTGATATCAGATTAATAATAGATTGTTATAACTTTAGGTTATTACATGTCATTCCCATGGTAACCA 80
 CAGAGAAAATATCTATAGGATATACACAAAAGAAATATGAGAAATGCATAAAAACGAGTGACTACAAAACCCAGCTAAGT 160
 GTAAAGAAAACAGTGAAGGAGGAATAAGGGATAAAAGTTATGACATATTAAAAAAATACAGGGGTGGACAAGATGGCC 240
 AACTAGAACAGTGCAGCAGAGCTCCATCCAAAGAACCATAATAAGCTTGTGAATCCTTCACCGGCAAGGTATCATGTTT 320
 TCTCATCAGAACTGACTAGGAAGTGGCATGATACACGGAGAGGAAGGAGCAGTGTGGTGCAGTACCCACCTGAAA 400
 GCCACATGGGACAGGGGAGTCTCCACCTCTGGCCAAGGGAGGTGGTGAGTGAGCATGCTACCCACCTGGGGAACTGTG 480
 CTCTCTCCATGGAAGTGTGAACCCATGGATC 512

hu-23

hu site 2

GATCTTAGCAGCCCAGAACTCACTGGTGATTTTGAGAAGAGGCATTGGGTGCCATGGTGACGGGAAAAAAATCTAAA 80
X.....
 ATGCACAGGAGCTAAAGAAATAGAAGATCCAAGAATTTTGGTAGTTTTTGTGTTTTTGTGTTTCTTAAAGTGAGTG 160
 AAAACAGCTTGTCTGTAGACTAAGCAAAATGATC 194

hu-40

hu site 3

GATCTAAAGAGAGTGTATGCAACAGTGCCAAATACTGCAGCAGGATTAAGTAGGATGAAAACAAAACCCAGGGATTGG 80
 ATTATTGATGAGGAATTCATTAGTAATTGTAGCAAAAGCTCTTCTATGTAGTCATAGAGATAGGGACCAGATTTAAGG 160
 CATTAGGGAATCAGTGGAATAGAGACTTGACATTACTGTTTCAGTTGATGAATGC 216

FIG. 4. DNA sequences of three cloned human DNA fragments containing MDBP recognition sites. The sequence of one strand, 5' to 3', is indicated. The MDBP binding sites of these cloned human DNA fragments of unknown function are underlined. The region in hu-23 which is partly homologous to hu site 2 but not recognized by MDBP is indicated by dots underneath the bases and an X under the base that must be looped out for optimum alignment. Only the indicated 216 bp of the ~400-bp hu-40 insert were sequenced.

that it is a very-low-copy sequence (data not shown).

In addition, we recently analyzed a cloned fragment of ~400-bp (hu-40) containing a third MDBP site, hu site 3. This site, from an anonymous human sequence was localized to a 170-bp *HinfI* subfragment by fragment selection assays conducted as previously described (11). Sequencing (Fig. 4) revealed only one site in this 170-bp region with considerable homology to other MDBP sites. We synthesized a 22-bp oligonucleotide duplex containing the putative MDBP site (Table 1) and showed that this site, hu site 3, is bound by MDBP with approximately the same affinity as hu site 2 (data not shown) despite the fact that it has an A residue instead of a 5-methylated pyrimidine at

position 11. However, it does have the symmetrical TpA dinucleotide at positions 8 and 9 providing methylated pyrimidine residues on both strands (Table 1).

An unusual MDBP-specific site in λ DNA

In the course of this work, we discovered a methylation-independent MDBP recognition site in λ DNA. It was localized to a 729-bp TaqI λ DNA fragment from recombinant p λ (Fig. 2C) by fragment selection assays (11). It was the only site detected in methylated or unmethylated λ DNA that displayed a high affinity for MDBP. The λ site was localized by DNase I footprinting and dimethylsulfate interference experiments (Fig. 3D and Table 2A). Gel retardation assays with a 22-bp duplex containing this λ site 1 revealed MDBP-type complexes (Fig. 1A) and the hydroxylapatite chromatography peak of activity for binding this λ site and for binding pB site 1 coincided (data not shown). The λ site shows the best homology to pB site 1 and its MDBP binding variants by introducing an imaginary 1-bp gap to give 5'-GTTGCC_^CGGTAAC-3' (Table 1). Without allowing such a gap, a much more limited extent of homology is observed with the overlapping sequence 5'-TGCCCGGTAACAAC-3' (Table 2B). In a test of three partially overlapping oligonucleotide duplexes containing this region, duplex VIII (with the former sequence) and not duplex X (with the latter sequence) was bound by MDBP (Table 2). We conclude that the λ site is missing 1-bp present in the other MDBP sites (Table 1). As for HSV site 2, a large increase in binding of the λ sequences was obtained if a few extra base-pairs were present in the oligonucleotide around the 14-bp region of hyphenated dyad symmetry (duplex IX compared to duplex VIII, Table 2B).

Comparison of affinity of MDBP for various specific sites

The extent of binding of λ site 1 and HSV site 2 in unmethylated pBR322 or pUC recombinants relative to that of pB site 1 in CpG-methylated pBR322 DNA was approximately 0.8 and 0.5, respectively, as determined by the fragment selection assay in the presence of 40, 80, or 110 mM NaCl. A similar hierarchy was seen when pB site 1, λ site 1, and HSV site 2 were analyzed as 22-bp oligonucleotide duplexes by the gel retardation assay (Table 2B). HSV site 1 similarly assayed as a 22-bp duplex hemimethylated with three or one ⁵C residues showed 48% or 5%, respectively, as much binding as bifilarly methylated pB site 1 (Table 2B). However, bifilar methylation would probably increase the affinity of this HSV site up to several fold (12).

We competed these radiolabeled 22-mers with simultaneously added unlabeled W9/C9, a 35-bp duplex containing the same methylated pB site 1 found in the above 22-bp pB site 1 duplex (9). With 3 units of MDBP, and

10 fmol of radiolabeled 22-bp ligands, a 50% reduction in binding of pB site 1, λ site 1, HSV site 1 (3 m⁵C's), or HSV site 2 required approx. 65, 60, 54, and 32 fmol of unlabeled W9/C9, respectively (data not shown). Competition of the radiolabeled pB site 1 (W17/C17) with the unmethylated form of W9/C9 (W9u/C9u) in a standard nitrocellulose filter-binding assay (7) indicated that more than a 1000-fold excess of W9u/C9u was needed for the same extent of competition obtained from methylated W9/C9. With 2.5 fmol of radiolabeled TaqI/HinfI or TaqI DNA fragments containing hu site 1 (312 bp) or hu site 2 (224 bp), about 350 and 270 fmol, respectively, of methylated W9/C9 was required to halve the amount of binding of the labeled ligand (data not shown). Therefore, hu sites 1 and 2 have similar high affinities for MDBP. Probably the larger size of these ligands or the nearby sequences, such as the runs of A residues, gave them a higher affinity than that of the pB site 1-containing oligonucleotide duplexes. Consistent with this explanation, TpG-containing unmethylated mutant derivatives of pB site 1 similar to hu sites 1 and 2 did not show more binding than did methylated, wild-type pB site 1 (8).

DISCUSSION

From four MDBP binding sites that were previously localized in pBR322 or M13mp8 RF DNA (11) and seven more MDBP binding sites identified in this report in human, HSV type 1, or phage λ DNA, a consensus sequence for recognition by MDBP can now be deduced (Table 1). This sequence contains alternative but related bases at most positions, namely, m⁵C or T, C or T, or either purine. A high degree of dyad symmetry over 14 bp (Table 1) is seen in many, but not all, of the sites (Table 1). The 14-bp core region suffices for specific binding to pB site 1 (9, 12) although several more adjacent bases seem to be required for much binding to HSV site 2 and λ site 1 (Table 2B). This is consistent with some conservation of a few residues surrounding the core 14-bp region (Table 1).

Despite the relatedness of the MDBP recognition sites, there are some remarkable dissimilarities (Table 1) as has been previously observed for other mammalian sequence-specific DNA-binding proteins (21-23). Among these moderate to high-efficiency MDBP sites a highly conserved motif is the following distribution of pyrimidine and purine residues in the 14-base core region of one strand: 5'-RmYmYRYR·mYRRmYRRY-3', where mY represents a 5-methylated pyrimidine and the dot indicates the axis of dyad symmetry. However, much deviation from that pattern is seen for the right half of

M13 site 4 and some deviation is found in hu sites 1, 2, and 3 (Table 1). Furthermore, there is an A residue which is present at position 7 in the middle of 11 of the 12 different MDBP recognition sites but is absent from λ site 1 (Table 1). Nonetheless, this site, effectively cross-competes with the other sites for MDBP binding (Fig. 1B). It can be aligned well with the consensus sequence only by envisioning a 1-bp gap at the position corresponding to the missing A·T base-pair, which might reflect some flexibility in the conformation of MDBP (24). However, when a 14-bp duplex was synthesized analogous to W17/C17's methylated pB site 1 but without the A·T base-pair at position 7 and with the extra T·A base pair at position 15 (Table 1), no binding at all was detected (unpubl. results). Therefore, in the sequence context of the λ site, but not in that of pB site 1, the absence of the middle A·T base-pair is tolerated by MDBP. Similarly, there are several single base-pair mutations which largely abolish binding by pB site 1 but which are acceptable in the sequence context of other MDBP sites (Table 1).

The function of the human MDBP sites and the herpes viral DNA sites remains to be determined. These anonymous human sequences, which we cloned on the basis of their recognition by MDBP, do not exhibit considerable homology to each other except for the common MDBP sites (Fig. 4). Two of them have an unusually high frequency of short (2 to 9 base) runs of T's or A's (Fig. 4). The MDBP recognition sites in these inserts do not require methylation for high-affinity binding apparently because they contain enough homology to the consensus MDBP sequence and the necessary number of TpG or TpA dinucleotides replacing m⁵CpG's of the methylation-dependent MDBP sites (Table 1). Given the propensity of m⁵C residues to undergo heat-induced or spontaneous deamination to T residues (25-28), it is possible that methylation-independent MDBP recognition sites have evolved from methylation-dependent ones. However, it should be noted that replacement of several of the m⁵C residues of pB site 1 with T residues as found in pB site 2 or M13 sites 2 or 3 did not allow specific binding of MDBP without methylation of C residues at the remaining CpG dinucleotides (Table 1). This may be due to the sequence context within the recognition site. The importance of the sequence context in determining whether a base substitution allows binding is also seen in the case of hu site 1, which is a high-affinity MDBP site despite having two base-pairs which inhibit MDBP binding if introduced into pB site 1 (positions 4 and 12, Table 1; 9, 12).

One of the MDBP-specific HSV sites, HSV site 2 (Table 1), is methylation-independent. It is present at positions +168 to +181 of the HSV type 1 thymidine kinase gene (16), a region which can be deleted without altering transcription rates from this gene (29, 30) and so may have no transcriptional role for this viral DNA. Another HSV site analyzed in these experiments, HSV site 1 (Table 1), is recognized by MDBP only when methylated at one or more of its three CpG dinucleotide pairs (Table 2B). It is located in a promoter region in an ~20-bp sequence several base-pairs upstream of the putative transcription initiation site for a gene encoding a 40 kDa viral protein of unknown function (31). Methylation of this site is required for MDBP binding (Table 2B). The third MDBP site in HSV type 1 DNA is similarly methylation-dependent in its binding to MDBP. This site, HSV site 3, is located 326 bp upstream of the transcription initiation site of the trans-acting regulatory gene encoding the α -trans-inducing factor (18). It is identical to HSV site 2 over the core 14-bp region except that the sequence 5'-TAGCA-3' from position 9 to 13 (on the upper strand; Table 1) is replaced by 5'-CGGCG-3', which makes it dependent on DNA methylation for binding (Table 2B). Evidence has been presented for methylation of HSV type 1 DNA during latency (32) but not in mature virions (33). It is, therefore, possible that methylation of HSV sites 1 and 3 helps repress expression of the adjacent genes in cells latently infected with HSV type 1.

ACKNOWLEDGMENTS

We thank Shannon Moore for assistance characterizing hu site 3, Michael Comb for recombinant plasmid pENKAT-12, Jeff Vieira for pUC118 and *E. coli* MV1193, and Edward Wagner for unpublished results.

This research was supported in part by Public Health Service Grant GM33999 and USDA Grant #85-CRCR-1-1751.

REFERENCES

1. Doerfler, W. (1983) *Ann. Rev. Biochem.* 52, 93-124.
2. Ehrlich, M. and Wang, R.Y.-H. (1981) *Science* 212, 1350-1357.
3. Guntaka, R.V., Gowda, S., Wagner, H. and Simon, D. (1987) *FEBS Let.* 221, 332-336.
4. Lassar, A.B., Paterson, B.M. and Weintraub, H. (1986) *Cell* 47, 649-656.
5. Saluz, H.P., Jiricny, J. and Jost, J.P. (1986) *Proc. Natl. Acad. Sci. USA* 83, 7167-7171.
6. Yisraeli, J., Adelstein, R.S., Melloul, D., Nudel, U., Yaffe, D. and Cedar, H. (1986) *Cell* 46, 409-416.
7. Huang, L.-H., Wang, R., Gama-Sosa, M. A., Shenoy, S. and Ehrlich, M. (1984) *Nature* 308, 293-295.

8. Supakar, P.C., Wiest, D., Zhang, D., Inamdar, N., Zhang, X.-Y., Khan, R., Ehrlich, K.C. and Ehrlich, M. (1988) *Nucleic Acids Res.* 16, 8029-8044.
9. Khan, R., Zhang, X.-Y., Supakar, P.C., Ehrlich, K.C. and Ehrlich, M. (1988) *J. Biol. Chem.* 263, 14374-14383.
10. Wang, R. Y.-H., Zhang, X.-Y. and Ehrlich, M. (1986) *Nucleic Acids Res.* 14, 1599-1614.
11. Wang, R. Y.-H., Zhang, X.-Y., Khan, R., Zhou, Y., Huang, L.-H. and Ehrlich, M. (1986) *Nucleic Acids Res.* 14, 9843-9860.
12. Zhang, X.-Y., Ehrlich, K. C., Wang, R. Y.-H. and Ehrlich, M. (1986) *Nucleic Acids Res.* 14, 8387-8397.
13. Vieira, J. and Messing J. (1987) *Meth. Enzymol.* 153, 3-11.
14. Gronostajski, R.M., Nagata, K., and Hurwitz, J. (1984) *Proc. Natl. Acad. Sci. USA* 81, 4013-4017.
15. Comb, M., Birnberg, N. C., Seasholtz, A., Herbert, E. and Goodman, H. M. (1986) *Nature* 323, 353-356.
16. McKnight, S. L. (1980) *Nucleic Acids Res.* 8, 5949-5964.
17. Wang, R. Y.-H., Huang, L.-H. and Ehrlich, M. (1984) *Nucleic Acids Res.* 12, 3473-3490.
18. Comb, M., Rosen, H., Seeburg, P., Adelman, J. and Herbert, E. (1983) *DNA* 2, 213-229.
19. Pellet, P.E., McKnight, J.L.C., Jenkins, F.J. and Roizman, B. (1985) *Proc. Natl. Acad. Sci. USA* 82, 5870-5874.
20. Raleigh, E.A., Murray, N.E., Revel, H., Blumenthal, R.M., Westaway, D., Reith, A.D., Rigby, P.W.J., Elhai, J. and Hanahan, D. (1988) *Nucleic Acids Res.* 16, 1563-1575.
21. Evans, T., DeChiara, T. and Efstratiadis, A. (1988) *J. Mol. Biol.* (199, 61-81.
22. Hwung, Y.-P., Crowe, D.T., Wang, L.-H., Tsai, S.Y. and Tsai, M.-J. (1988) *Molec. Cell. Biol.* 8, 2070-2077.
23. Pfeifer, K., Prezant, T. and Guarente, L. (1987) *Cell* 49, 19-27.
24. Gidoni, D., Dynan, W.S. and Tjian, R. (1984) *Nature* 312, 409-413.
25. Cooper, D.N. and Youssoufian, H. (1988) *Hum. Genet.* 78, 151-155.
26. Ehrlich, M., Norris, K.F., Wang, R.Y.-H., Kuo, K.C. and Gehrke C.W. (1986) *Bioscience Reports* 6, 387-393.
27. Lindahl, T. (1979) *Nucleic Acids Res.* 22, 135-192.
28. Wang, R. Y.-H., Kuo, K.C., Gehrke, C.W., Huang, L.-H. and Ehrlich, M. (1982) *Biochim. Biophys. Acta* 697, 371-377.
29. Coen, D.M., Weinheimer, S.P. and McKnight, S.L. (1986) *Science* 234, 53-59.
30. Halpern, M.E. and Smiley, J.R. (1984) *J. Virol.* 50, 733-738.
31. Costa, R. H., Draper, K. G., Kelly, T. J. and Wagner, E. K. (1985) *J. Virol.* 54, 317-328.
32. Youssoufian, H., Hammer, S.M., Hirsch, M.S. and Mulder, C. (1982) *Proc. Natl. Acad. Sci. USA* 79, 2207-2210.
33. Sharma, S and Biswal, N. (1977) *Virology* 82, 265-274.