

Sequence-specific DNA binding by the *MspI* DNA methyltransferase

Ashok K. Dubey⁺ and Richard J. Roberts^{*}

Cold Spring Harbor Laboratory, PO Box 100, Cold Spring Harbor, NY 11724, USA

Received February 5, 1992; Revised and Accepted May 18, 1992

ABSTRACT

The *MspI* methyltransferase (*M.MspI*) recognizes the sequence CCGG and catalyzes the formation of 5-methylcytosine at the first C-residue. We have investigated the sequence-specific DNA-binding properties of *M.MspI* under equilibrium conditions, using gel-mobility shift assays and DNaseI footprinting. *M.MspI* binds to DNA in a sequence-specific manner either alone or in the presence of the normal methyl donor S-adenosyl-L-methionine as well as the analogues, sinefungin and S-adenosyl-L-homocysteine. In the presence of S-adenosyl-L-homocysteine, *M.MspI* shows the highest binding affinity to DNA containing a hemimethylated recognition sequence ($K_d = 3.6 \times 10^{-7} \text{M}$), but binds less well to unmethylated DNA ($K_d = 8.3 \times 10^{-7} \text{M}$). Surprisingly it shows specific, although poor, binding to fully methylated DNA ($K_d = 4.2 \times 10^{-6} \text{M}$). *M.MspI* binds approximately 5-fold more tightly to DNA containing its recognition sequence, CCGG, than to nonspecific sequences in the absence of cofactors. In the presence of S-adenosyl-L-methionine, S-adenosyl-L-homocysteine or sinefungin the discrimination between specific and non-specific sequences increases up to 100-fold. DNaseI footprinting studies indicate that 16 base pairs of DNA are covered by *M.MspI*, with the recognition sequence CCGG located asymmetrically within the footprint.

INTRODUCTION

The interaction of proteins with DNA is of fundamental importance for many biological phenomena. Many regulatory systems that control gene expression depend upon specific binding of proteins such as repressors and transcription factors to regions flanking genes. Such interactions have been studied extensively and many examples of DNA-binding motifs have been documented (1). Two of the most common structures are based on zinc-fingers (2) and helix-turn-helix motifs (3,4) and have been demonstrated to mediate the tight binding necessary to regulate DNA transcription. A great deal of information is now available, both at the biochemical and structural levels, about repressor-

operator interactions in prokaryotes and transcription factor binding in eukaryotes (3,4). Much less is known about protein-DNA interactions in systems, such as restriction enzymes or integrases, where binding is accompanied by catalysis. In such cases the biological function requires sequence-specific DNA binding, but the strength of that binding must be tempered by the need for enzymatic action after binding and the subsequent release of the protein from the DNA. It is to be expected that the binding characteristics necessary for such interactions will be significantly different from the classical systems that typically regulate transcription and gene expression.

Restriction-modification systems present an interesting case of sequence-specific recognition of DNA. Usually the restriction enzyme and the cognate methyltransferase recognize the same DNA sequence although each catalyzes a quite different reaction (5). Examination of the primary structures of corresponding restriction enzymes and methylases show no obvious similarities and yet both proteins usually interact with the same DNA sequence (6). How is this recognition achieved? Most restriction enzymes recognize double-stranded DNA sequences that contain a dyad axis of symmetry (7). Since the normal substrate is unmethylated and symmetric, it is reasonable that the restriction enzymes would function as homo-dimers, as has been found (8). In contrast, the cognate methyltransferase usually acts as a monomer (5) and its normal substrate is hemimethylated DNA present immediately after replication of a fully methylated chromosome. In this case, because the normal substrate is asymmetric, there can be no requirement for symmetry in the interaction. These observations immediately suggest that the fundamental interactions between a restriction enzyme and a methyltransferase with their respective DNA substrates will be quite different. The DNA binding properties of the *EcoRI* restriction enzyme have been reported (9–11) and the interactions are quite different from those found in other more-classical systems. None of the usual binding motifs are involved and yet exquisite specificity is achieved. In the case of another restriction enzyme, *EcoRV*, it has recently been found that DNA binding occurs equally well between specific and non-specific sequences (12). The specificity in that case occurs during cleavage. So far no common themes of DNA recognition have been found among restriction enzymes.

* To whom correspondence should be addressed

⁺ Present address: Biochemical Engineering Research Centre, Indian Institute of Technology—Delhi, Hauz Khas, New Delhi 110016, India

In contrast to the restriction enzymes, no comparable studies of DNA binding by methylases have been reported. It is not known if common features might be present, although none of the traditional DNA binding motifs have been detected by inspection of the protein sequences. In the case of those methylases that form 5-methylcytosine in DNA many common features are found within their sequences. Five highly-conserved and five less well-conserved sequence motifs are present (13,14) and it seems likely that these enzymes possess similar overall structures. The region responsible for sequence-specificity has been localized to a variable region of 80–120 amino-acids in the mono-specific m5C-methylases (15). We have previously cloned the gene for the *M.MspI* methylase (16) and have recently overexpressed and purified the protein to homogeneity (17). We now report studies of the interaction of the purified *MspI* methylase with DNA.

MATERIALS AND METHODS

Chemicals and Reagents

Homogeneous *M.MspI* was prepared as described previously (17). Bacteriophage λ -DNA and the Klenow fragment of *E.coli* DNA polymerase I were obtained from New England Biolabs. Bovine pancreatic DNaseI was from Worthington. *E.coli* tRNA was from Boehringer-Mannheim. Reagents for quantitative assay and SDS-polyacrylamide gel electrophoresis of protein were from BioRad and used as recommended. S-adenosyl-L-methionine (AdoMet), sinefungin and S-adenosyl-L-homocysteine (AdoHcy) were obtained from Sigma. Deoxyadenosine 5'-[α -³²P] triphosphate and deoxycytidine 5'-[α -³²P] triphosphate (specific activities 3000 Ci/mmmole) were from Amersham. Deoxynucleoside triphosphates were from USB. Scintillation fluid, CytoscintTM, was from ICN Biomedicals, Inc. All other chemicals used were of the highest purity reagent grade.

Oligodeoxynucleotides and Radiolabelling

Oligonucleotides were either synthesized on an Applied Biosystem 380A/380B DNA synthesizer at the CHSL Oligonucleotide Facility or were obtained from New England Biolabs or the Midland Certified Reagent Company. The following oligonucleotides were used:

Specific 29-mer

5'-TGCAGTCGGATGCCGGGTACCTTGAGG-3'
3'-GTCAGCGCTACGGCCAGTGGAACTCCGT-5'

Nonspecific 29-mer

5'-TGCAGTCGGATGCAGGGTACCTTGAGG-3'
3'-GTCAGCGCTACGTCCAGAGGAAGTCCGT-5'

Specific 69-mer

5'-GGAGTAGCAGAGGTCGCTAGGCCCAACGTGTTCCGGTGAGTGG-
ATCGACAGAGTTTCGAGTCGCGAT-3'
3'-TCATCGTCTCCAGCGATCCGGGTTGCACAAAGGCCACTCAC-
CTAGCTGTCTCAAAGCTACGCGCTATT-5'

Two additional versions of the 69-mer oligonucleotides were synthesized in which the cytosine residues at position 35 of the top strand or position 34 of the lower strand (in both cases these are the outer cytosines in the 5'-CCGG-3' recognition sequence) were substituted with 5-methylcytosine. By appropriate hybridization between methylated and unmethylated strands both hemimethylated and fully methylated substrates were prepared.

Basic procedures for purification and labelling were from (18). In brief, following synthesis, each oligonucleotide was recovered by butanol extraction and ethanol precipitation, annealed and labelled with [α -³²P] dATP and/or [α -³²P] dCTP by completing the strands using the Klenow DNA-polymerase to give radiolabelled fully double-stranded oligonucleotides of 31 and 71 base pairs. These labelled oligonucleotides were then gel purified under native conditions on 15% or 12% polyacrylamide gels, eluted with 100 mM NaCl, 50 mM Tris-HCl (pH 8.0), 10 mM Na₂EDTA and stored in the same buffer at 4°C.

For DNaseI footprinting experiments, the single-stranded oligonucleotides were purified by gel electrophoresis on a 12% polyacrylamide/8.3 M urea gel, eluted with 0.5M NH₄-acetate, 1 mM Mg-acetate, 1% SDS and recovered by butanol extraction and ethanol precipitation. The purified oligonucleotides were dissolved in 50 mM Tris-HCl (pH 8.0), 10 mM Na₂EDTA and stored at 4°C. Appropriate strands were then annealed to give double stranded substrates which were then labelled at their 3'-ends using [α -³²P] dCTP. They were then filled in with unlabelled dNTPs to render the oligonucleotides fully double stranded. Oligonucleotide concentrations were determined from the measured A₂₆₀ and the calculated molar extinction coefficients.

Gel Retardation Assay

The standard methylase-DNA binding mixture contained 100 mM NaCl, 50 mM Tris.HCl (pH 8.0), 10 mM Na₂EDTA, 7 mM β -mercaptoethanol, 0.26 mg/ml BSA, 13% glycerol and appropriate amounts of *M.MspI* and [α -³²P]-labelled oligonucleotides in a reaction volume of 10 μ l. When necessary, cofactors were included in the incubation mixtures as indicated in the Figure legends. The reactions were incubated at 22°C for 15 to 25 min and were analyzed by electrophoresis on native polyacrylamide gels (12% gel for 31-mer and 10% for 71-mer substrates) run in 1×TBE buffer (89 mM Tris-borate, pH 8.0, 2 mM Na₂EDTA). The gels were pre-run at 100 volts for 60 min., samples were applied and electrophoresis was continued at 120 volts for 60 min. Following electrophoresis the gels were dried and analyzed using a Molecular Dynamics Phosphor-Imager and also by autoradiography. The amount of oligonucleotide bound by protein was calculated as the difference between the known total input concentration of oligonucleotide and the amount detected in the position for free oligonucleotide on the mobility shift gels.

DNaseI Protection Assay

1 μ M *M.MspI* and 0.25 μ M of the 71-mer unmethylated, double-stranded oligonucleotide, labelled with [α -³²P] dCTP, were incubated for 10 minutes at room temperature in a 10 μ l reaction containing 50 mM Tris.HCl (pH 8.0), 50 mM NaCl, 7 mM β -mercaptoethanol, 0.26 mg/ml BSA, 10 μ M AdoHcy and 13% glycerol. Then 10 ng of DNaseI (freshly diluted from 2.5 mg/ml stock in 50 mM Tris.HCl (pH 8.0), 50 mM NaCl, 60 mM MgCl₂, 15 mM CaCl₂ and 10% glycerol) in a 2 μ l volume was added to the binding reaction. Digestion was terminated after 60 seconds incubation at room temperature by adding 85 μ l of stop solution (200 mM NaCl, 20 mM Na₂EDTA, 1% SDS and 100 μ g/ml tRNA). Control reactions were performed as above except that *M.MspI* was omitted. The completed reaction mixture was extracted with phenol/chloroform and the partially digested DNA fragments were recovered by ethanol precipitation. The products

were analyzed by electrophoresis on a 12% polyacrylamide/8.3M urea gel. A Maxam-Gilbert A+G reaction was performed and run alongside the DNaseI digestion products to identify the sequence protected.

RESULTS

***M. MspI* – oligonucleotide complexes**

We have used a gel mobility-shift assay (19,20) to test the sequence-specific binding properties of homogeneous *M. MspI*. In early experiments we attempted to measure the initial rate of binding between *M. MspI* and specific oligonucleotides, but equilibrium was achieved so rapidly at 22°C that no differences could be detected by mobility shift analysis even at the shortest time points. This prevented us from using this assay to measure the kinetics of binding and so we focussed our analysis on the sequence-specific effects, under equilibrium conditions, since these could be monitored readily by the mobility shift assay.

Initially we monitored the extent and type of complex formation by using a fixed concentration of a 31-mer oligonucleotide containing the specific recognition sequence 5'-CCGG-3' and varying concentrations of *M. MspI*. The reaction was carried out in the absence of any cofactor or inhibitor. A control reaction, to demonstrate specificity, was carried out with another 31-mer oligonucleotide that differed only at a single position within the recognition sequence (CCGG to CAGG). This oligonucleotide

is referred to as the non-specific oligonucleotide. As can be seen from the left panels in Figure 1, two complexes are visible with the specific 31-mer. A faster-moving complex is due to specific binding and a slower-moving complex is due to non-specific binding. With a 15-fold molar excess of *M. MspI* non-specific binding increases to a point where it almost eliminates the signal due to specific binding. At constant *M. MspI* concentration (lower panels) increases in oligonucleotide concentration lead to some increase in specific complex formation, but the effect is not dramatic. Because of the propensity of *M. MspI* to form dimers (17) and the high concentrations of *M. MspI* necessary to observe binding, we have not included a calculation of the equilibrium binding constant from these data since it is likely to be suspect.

Since the normal interaction between *M. MspI* and DNA, leading to catalysis, requires the cofactor AdoMet we also examined DNA-binding in the presence of either AdoMet, AdoHcy, the end product of the reaction, or sinefungin, an analog

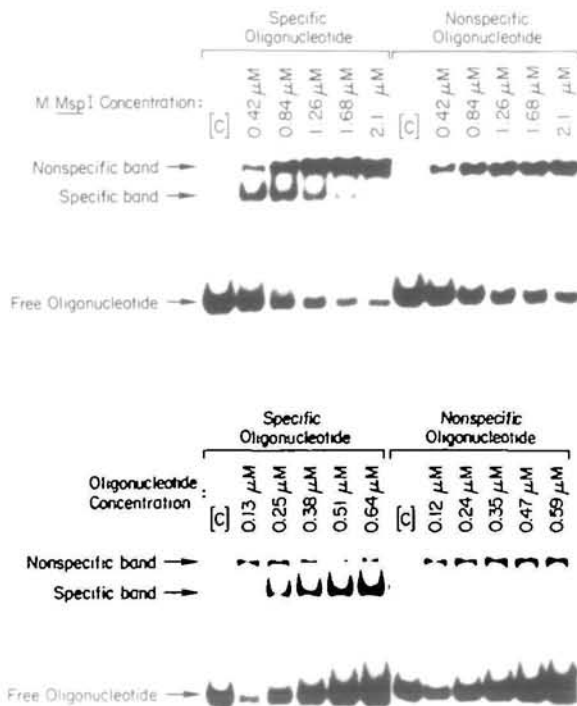


Figure 1. Specific vs. nonspecific binding of DNA by *M. MspI*. Upper panel: 0.13 μM of the 31-mer [α - 32 P]-labelled oligonucleotides and the indicated concentrations of *M. MspI* were incubated in a binding reaction as described in Methods. The lane marked [C] is a control lane containing the free oligonucleotide. Lower Panel: Binding reactions were performed as above except that the *M. MspI* concentration was held constant at 0.56 μM and the 31-mer oligonucleotide concentration was varied as indicated.

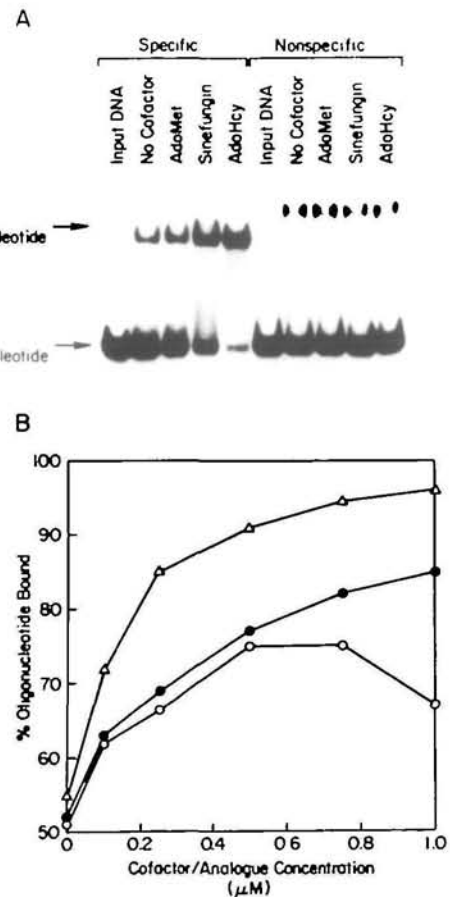


Figure 2. Effect of AdoMet, AdoHcy and sinefungin on *M. MspI* binding to DNA. Panel A: *M. MspI* (0.56 μM) and the 31-mer [α - 32 P] labelled oligonucleotides (0.25 μM) were incubated in a binding reaction as described in Methods. 2 μM AdoMet, AdoHcy or sinefungin were included as indicated. The binding was quantitated by measuring the free oligonucleotide as indicated in Methods. Specific binding led to 64% free oligonucleotide in the absence of cofactors and 44% with AdoMet, 20% with sinefungin and 6% with AdoHcy. The comparable values for free non-specific oligonucleotide were 72% (no cofactor), 75% (AdoMet), 80% (sinefungin) and 77% (AdoHcy). Panel B: *M. MspI* binding to the specific 31-mer oligonucleotide was quantitated as described in Methods and in the text. Open circles show binding in the presence of AdoMet, closed circles show sinefungin and open triangles show AdoHcy.

of AdoMet that is a known inhibitor of m5C-methylases (21). For these experiments, the specific 31-mer oligonucleotide was incubated with an excess of AdoMet, AdoHcy or sinefungin. Figure 2, panel A shows that while nonspecific binding was fairly constant, specific binding was significantly enhanced by both AdoMet and its analogues. The greatest effect was seen with AdoHcy, with lesser effects from sinefungin and AdoMet. In panel B of Figure 2 it can be seen that with increasing concentrations of AdoHcy and sinefungin the extent of specific binding increases and is still not maximal at 1.0 μM . However, at AdoMet concentrations above 0.75 μM the extent of binding drops presumably due to methylation and decreased affinity for the methylated oligonucleotide (see below).

Electrostatic interactions have been shown to play a significant role in many protein-DNA interactions (22). Examples include the *lac*-repressor-operator interaction (23), and several transcription factors such as MLTF (24) and fragments containing the yeast MAT α 2 homeodomain (25) or the *Drosophila* engrailed homeodomain (26). A simple method to assess the influence of electrostatic interactions is to measure binding at different ionic strengths. We therefore carried out *M.MspI*-DNA binding reactions in the presence of increasing amounts of NaCl up to 400 mM both in the absence and in the presence of AdoHcy. The results (not shown) indicated that binding was not significantly affected by NaCl concentrations up to 200 mM. A small reduction in binding was observed at higher NaCl concentrations, but even at 400 mM NaCl the binding was still 70% of that seen at 200 mM NaCl. These observations suggest that electrostatic interactions contribute little to the binding.

Binding to methylated DNA

The enhanced binding observed in the presence of the methyl donor AdoMet (Figure 2) raised an important question about the nature of the DNA substrate being bound. Under the conditions of that experiment, methylation of the substrate would be expected and so any complexes observed could involve hemi-methylated or fully-methylated DNA. To examine these possibilities, a binding reaction was set up that contained a 20-fold molar excess of AdoMet, sufficient for complete methylation. Following binding the products were checked for the sequence specific band shift, sensitivity to the *MspI* restriction enzyme (*R.MspI*) and rebinding in a second reaction. As seen in Figure 3, following binding the unmethylated oligonucleotide becomes resistant to *R.MspI* digestion after dissociation and is still able to rebind *M.MspI* in a second binding reaction.

From the above experiment it is clear that *M.MspI* is able to bind both unmethylated and methylated DNAs. In that experiment it is likely that the substrate becomes fully modified since a large excess of AdoMet was used. However, the experiment does not clearly differentiate between binding to hemi-methylated DNA or fully modified DNA, since *R.MspI* is unable to cleave fully methylated DNA and nicks hemi-methylated DNA so slowly (27) that it would not be detected under our experimental conditions. To look at this question directly and to determine the relative affinities of *M.MspI* for these three substrates, we prepared 71-long oligomers containing the CCGG recognition sequence either unmodified, methylated on one strand or fully methylated. Binding reactions were carried out at methylase concentrations varying from 0.38 μM to 1.27 μM while the concentration of 71-mer was kept constant at 0.25 μM . Binding conditions were the same as described before except that the buffer contained 10 μM AdoHcy. A gel-shift analysis is shown in Figure 4A. It

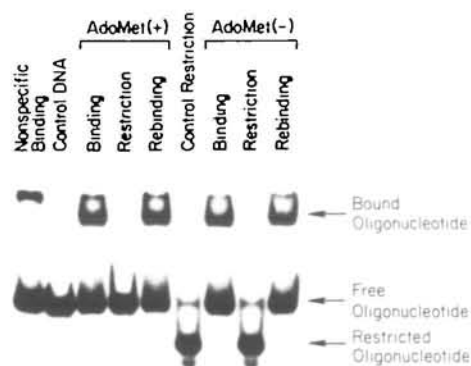


Figure 3. Binding of *M.MspI* to methylated substrate. 0.56 μM *M.MspI* and 0.25 μM 31-mer oligonucleotide were incubated in a binding reaction with or without AdoMet (10 μM) as described in Methods. Non-specific binding: *M.MspI* binding to nonspecific 31-mer. Control DNA: Free specific 31-mer. Binding lanes: *M.MspI* binding to the specific 31-mer in the presence or absence of AdoMet. Restriction lanes: Reactions as in the binding lanes were heated at 70° for 5 mins, MgCl_2 added to 10 mM and incubated with *R.MspI* (10 units) for 45 mins at 37°C prior to electrophoresis. Rebinding lanes: Reactions as in the binding lanes were heated at 70° for 5 minutes and fresh *M.MspI* added. Control restriction: The 31-mer specific oligonucleotide (0.25 μM) was digested with *R.MspI*.

should be noted that in the presence of AdoHcy there is a much better discrimination between specific and non-specific sequences and by keeping the concentrations of *M.MspI* lower than those used in the experiment of Figure 1 the amount of non-specific binding was much less.

We attempted to calculate dissociation constants for the specific interaction between *M.MspI* and its various specific DNA substrates using Scatchard plots (Figure 4, panel B). For this analysis we used the observed values of the free DNA concentrations and derived values for the concentrations of free and bound *M.MspI*. We made the assumption that each complex contained one molecule of *M.MspI* bound to one molecule of DNA. The bound *M.MspI* concentration was then calculated as the difference between the total DNA concentration input into the reaction minus the free DNA observed. This method will overestimate the bound DNA, because as can be seen from Figure 4A both specific and non-specific complexes are present. However, because of the high K_d values observed, attempting to measure the bound DNA concentration directly from the gel will lead to an underestimate of the complex because of dissociation during the experiment. Neither method is completely satisfactory and so the values presented must be viewed with caution. The K_d values obtained from this analysis are shown in Table 1, where they are compared with values for other well-characterized DNA binding proteins. It can be seen that binding is strongest to hemimethylated DNA (K_d : $3.6 \times 10^{-7}\text{M}$) 2-fold lower for unmethylated DNA (K_d : $8.3 \times 10^{-7}\text{M}$) and weakest for fully methylated DNA (K_d : $4.2 \times 10^{-6}\text{M}$).

Specificity of binding

A key feature of specific DNA binding proteins is their ability to discriminate between specific and non-specific sequences. They vary widely in that ability, which is clearly a parameter of mechanistic significance. We therefore examined the ability of *M.MspI* to discriminate between the bona-fide recognition sequence, CCGG, and other non-specific sequences. To address this question we performed a competition experiment in which

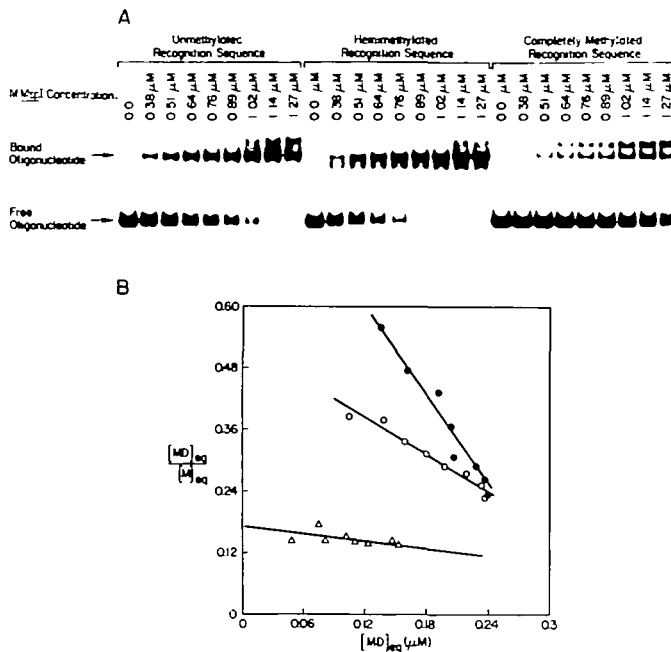


Figure 4. Equilibrium binding of *M.MspI* to DNA. The 71-mer [α - 32 P]-labelled oligonucleotides (0.25 μ M), either unmethylated, hemi-methylated (top strand—5'-GGAGT...mCCGG...CGCGAT-3' methylated) or fully methylated, were incubated with the indicated concentrations of *M.MspI* and analyzed as described in Methods. The mobility shift data are shown in the upper panel and their quantitative analysis using a Scatchard Plot is shown in the lower panel. $[MD]_{eq}$ is the equilibrium concentration of bound complex, $[M]_{eq}$ is the equilibrium concentration of *M.MspI* in the binding reaction. Open triangles show data for fully-methylated substrate: open circles show unmethylated substrate and closed circles show hemi-methylated substrate.

Table 1. Binding Constants for DNA-protein interactions

Protein	K_d (M)	Reference
lambda repressor	3×10^{-14}	28
lac repressor	1×10^{-13}	23
transcription factor NF- κ B	$10^{-12} - 10^{-13}$	29
Antennapedia homeodomain	1.6×10^{-9}	30
Engrailed homeodomain	1.2×10^{-9}	26
CAP protein (0.2 μ M cAMP)	1.6×10^{-9}	31
CAP protein (5 μ M cAMP)	1.2×10^{-11}	31
MLTF transcription factor	1×10^{-10}	24
<i>R.EcoRI</i>	1.0×10^{-11}	32
<i>R.EcoRI</i>	7.1×10^{-12}	33
<i>R.RsaI</i>	2.8×10^{-11}	34
<i>R.EcoRV</i>	6.3×10^{-8}	12
<i>M.MspI</i> (unmethylated DNA)	8.3×10^{-7}	this paper
<i>M.MspI</i> (hemi-methylated DNA)	3.6×10^{-7}	this paper
<i>M.MspI</i> (fully methylated DNA)	4.2×10^{-6}	this paper

binding of *M.MspI* to the 31-mer oligonucleotide containing the recognition sequence was analysed in the presence of increasing amounts of a similar 31-mer that differed only at a single position within the recognition sequence (CCGG to CAGG). For a competitive reaction containing only methylase and DNA, a 5-fold excess of unlabelled competitor is necessary to reduce the specific binding to 50% of maximal levels (Figure 5, upper panel). In a parallel experiment, in which AdoHcy is present, a 100-fold excess of competitor was needed to reduce the specific

binding by 50% (Figure 5, lower panel). Thus the presence of AdoHcy causes a 20-fold increase in the ability of *M.MspI* to discriminate between specific and non-specific sequences.

Footprint Analysis

Having demonstrated that *M.MspI* shows specific binding to an oligonucleotide containing its recognition sequence, it was of interest to examine the nature of the interaction in more detail. *M.MspI* and the unmethylated double-stranded 71-mer oligonucleotide, containing the recognition sequence, were incubated together in the presence of AdoHcy and then treated with DNaseI to generate a footprint. A sequence ladder (A+G) was run alongside to enable the identification of the bases protected. We had expected to see the average of two complexes, one positioned to methylate the upper strand of the substrate and a second positioned to methylate the lower strand of the substrate. However, Figure 6 shows the surprising result that the CCGG recognition sequence is located highly asymmetrically within the footprint. This suggests that the complex detected is only one of the two possible complexes that might have been formed between the monomer methylase and the unmethylated substrate, which is symmetric at the recognition sequence, but asymmetric in the flanking sequences. Sixteen base pairs are protected from DNaseI digestion, which must represent a maximum distance over which interactions between *M.MspI* and its substrate take place.

DISCUSSION

Most studies of sequence-specific DNA-protein interactions have been carried out on proteins whose primary function is DNA binding and which lack enzymatic activity. It is not surprising that they exhibit tight binding to their recognition sequences since they have no need for rapid dissociation so as to continue a catalytic cycle. In contrast, enzymes that act on specific DNA sequences, such as restriction enzymes and methylases must temper their DNA binding with the need to dissociate after the enzymatic reaction is complete. This moderation in the strength of binding is clearly evident in the case of *M.MspI* where the maximum value for K_d , observed in the presence of AdoHcy, is 8.3×10^{-7} . It should be noted that the K_d values reported in this paper should be viewed as approximate because of the experimental difficulties in obtaining accurate values by the gel-shift method for proteins that exhibit weak binding. We were also unable to determine K_d 's in the presence of AdoMet by this method because its inclusion in the binding reaction leads to methylation. The K_d value obtained above is in marked contrast to the K_d values of 10^{-13} for the lac repressor (23) and 10^{-9} to 10^{-12} for typical transcription factors (29,30). The dramatic differences of $10^3 - 10^6$ between *M.MspI* and a typical DNA binding protein undoubtedly reflects the different biology of the two systems. A protein designed mainly to bind DNA is expected to form a tight complex, whereas a methylase only binds DNA as an intermediate step prior to base methylation. However, we cannot exclude the possibility that the K_d we have determined for *M.MspI* may reflect the high concentrations (μ M) necessary for the assay. It has been shown previously that K_d values can vary by as much as three orders of magnitude depending upon the protein concentration range used (23). Finally the temperature at which binding reaction is performed can also affect K_d values (24).

Many other features of the binding by *M.MspI* also reflect the requirements or results of enzymatic function. For instance, in the presence of the cofactor, AdoMet, or the end product of the

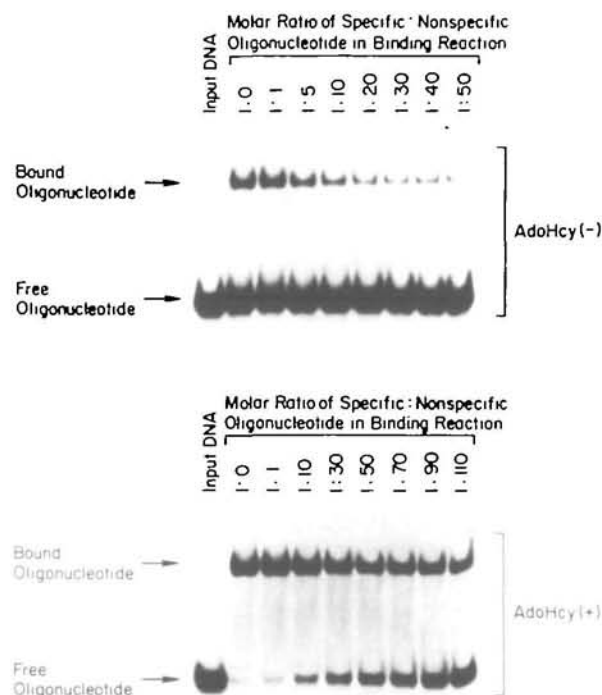


Figure 5. Competitive binding of *M.MspI* between specific and nonspecific substrates. *M.MspI* (0.56 μ M) and the specific α - 32 P labelled 31-mer oligonucleotide (0.25 μ M) were incubated in a standard binding reaction (see Methods) in the presence of increasing concentrations of the unlabelled nonspecific 31-mer oligonucleotide. Binding was carried out in the absence (Upper panel) or the presence (lower panel) of AdoHcy (2 μ M).

reaction, AdoHcy, the specificity of the binding increases. This suggests that some conformational adjustment accompanies cofactor binding and enhances the discrimination between specific and non-specific sequences (31). As can be seen from Figure 2B with increasing concentrations of cofactors or inhibitors the extent of binding increases. However, at high levels of AdoMet, binding decreases presumably as a result of methylation of the substrate. Surprisingly, the affinity does not drop to non-specific levels and *M.MspI* is still able to form a specific complex with fully methylated substrate (Figures 2, 3 and 4). While it was expected that the enzyme would bind to hemimethylated DNA, which is the normal substrate *in vivo*, there appears to be no functional necessity to bind fully methylated DNA. Perhaps this property is merely a consequence of the need to bind hemi-methylated DNA. In any case, the reduced binding affinity for fully methylated DNA is sufficient to allow product release and enzyme turnover.

M.MspI shows more than a 2-fold preference for binding hemi-methylated over unmethylated DNA as might be expected. It has been shown in several other systems that hemi-methylated DNA is the preferred substrate when overall catalytic efficiency is tested. Examples include the more complicated interactions of a Type I enzyme (35) as well as other methylases of Type II restriction-modification systems such as *M.BsuRI* (36) and *M.SssI* (37). However, this is not found in all systems, since similar reaction rates have been observed for hemi-methylated and unmethylated substrates by *M.EcoRI* (38). It should be noted that in these other cases no direct measurements of DNA-binding



Figure 6. DNaseI footprint of *M.MspI*. The substrate for these experiments was the 71-mer oligonucleotide which had been labelled with $[\alpha$ - 32 P] dCTP. Lane 1 shows a Maxam-Gilbert (A + G) reaction. Lane 2 shows a partial DNaseI digestion. Lane 3 shows the footprint obtained when *M.MspI* is included in the reaction of Lane 2. Reactions were carried out as described in Methods.

affinity were made and an increased overall catalytic efficiency does not necessarily reflect increased binding affinity.

An important parameter for any sequence-specific DNA interaction is the ability to discriminate the specific recognition sequence from nonspecific sequences. The competition experiments shown in Figure 5 show that in the absence of any cofactors or inhibitors *M.MspI* shows a 5-fold difference in binding to specific rather than non-specific sequences. Because of the complications introduced by the methylation reaction it was not possible to measure this discrimination in the presence of AdoMet. When the competition experiment was performed in the presence of AdoHcy, the specificity increased to 100-fold. This suggests that some comparable increase in discrimination might be provided by the normal cofactor AdoMet since in our other experiments the presence of AdoMet stimulated binding. Obviously this would be of advantage *in vivo* where the methylase is faced with a vast excess of nonspecific sequences and AdoMet is usually abundant. The AdoHcy which is formed as the end-product of methylation does not accumulate since it is broken down to homocysteine and adenosine by the normal metabolic processes of the cell.

The footprinting experiments have demonstrated that binding of *M.MspI* covers its recognition sequence within a stretch of 16 base pairs. This is a value comparable to that observed for other DNA binding proteins. For instance, several restriction enzymes have been footprinted with 17 base pair protection observed for *EcoRI* (39), 15 base pairs for *HaeIII* (39), 13 base pairs for *HinPI* (39) and 12 base pairs for *RsrI* (34). Probably the most interesting aspect of the footprint is the asymmetric location of the recognition sequence which shows that *M.MspI* interacts preferentially with one orientation of the recognition sequence in the unmethylated 71-mer substrate. Since this substrate is asymmetric, except for the four bases of the recognition site, the flanking sequences must play an important role in the binding. It would also be consistent with a two step reaction on unmethylated DNA. During the first step, the

methylase binds and methylates the first strand. Dissociation of the methylase then takes place before rebinding in a second step that leads to full methylation.

Close inspection of the banding pattern, obtained on the mobility-shift gel in Figure 4 (panel A) for the hemi-methylated substrate, shows that two bands are present in the position assigned to specific complex formation. The major band presumably corresponds to the complex giving the footprint in Figure 6, while the minor band could represent binding to the methylated strand in the other orientation. This is expected to be a much weaker interaction, based on the binding data to fully methylated DNA (Figure 4, far right of Panel A).

ACKNOWLEDGEMENTS

We thank S.Bell, D.Helfman, S.Klimasauskas, A.Krainer and G.Stormo for helpful comments on the manuscript and M.Ockler, J.Duffy and P.Renna for art and photography. This work was supported by grants from the NIH (GM 46127) and NSF (DMB-8917650). Centralized facility support for oligonucleotide synthesis was from a Cancer Center Support grant from the NCI (CA 45508).

REFERENCES

- Harrison, S.C. (1991) *Nature* **353**, 715–719.
- Miller, J., McLachlan, A.D. and Klug, A. (1985) *EMBO J.* **4**, 1609–1614.
- Pabo, C.O. and Sauer, R.T. (1984) *Ann. Rev. Biochem.* **53**, 293–321.
- Aggarwal, A.K. and Harrison, S.C. (1990) *Ann. Rev. Biochem.* **59**, 933–969.
- Modrich, P. and Roberts, R.J. (1982) In *Nucleases* (S.M. Linn, R.J. Roberts, Eds.) pp 109–153, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Chandrasegaran, S. and Smith, H.O. (1988) In *Structure and Expression. From Proteins to Ribosomes* (R.H. Sarma and M.H. Sarma, eds) pp 149–156, Adenine Press, NY.
- Roberts, R.J. and Macelis, D. (1991) *Nucleic Acids Res.* **19**, 2077–2109.
- Bennett, S.P. and Halford, S.E. (1989) *Curr. Top. Cell. Regulation* **30** 57–104.
- Terry, B.J., Jack, W.E., Rubin, R.A. and Modrich, P. (1983) *J. Biol. Chem.* **258**, 9820–9825.
- Jen-Jacobson, L., Kurpiewski, M., Lesser, D., Grable, J., Boyer, H.W., Rosenberg, J.M. and Greene, P.J. (1983) *J. Biol. Chem.* **258**, 14638–14646.
- Becker, M.M., Lesser, D., Kurpiewski, M., Baranger, A. and Jen-Jacobson, L. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6247–6251.
- Taylor, J.D., Badcoe, I.G., Clarke, A.R. and Halford, S.E. (1991) *Biochemistry* **30**, 8743–8753.
- Lauster, R., Trautner, T.A. and Noyer-Weidner, M. (1989) *J. Mol. Biol.* **206**, 305–312.
- Posfai, J., Bhagwat, A.S., Posfai, G. and Roberts, R.J. (1989) *Nucl. Acids Res.* **17**, 2421–2435.
- Klimasauskas, S., Nelson, J. and Roberts, R.J. (1991) *Nucl. Acids Res.* **19**, 6183–6190.
- Lin, P.M., Lee, C.H. and Roberts, R.J. (1989) *Nucl. Acids Res.* **17**, 3001–3011.
- Dubey, A.K., Mollet, B. and Roberts, R.J. (1992) *Nucl. Acids Res.* submitted.
- Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning, A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
- Garner, M.M. and Revzin, A. (1981) *Nucl. Acids Res.* **9**, 3047–3060.
- Fried, M. and Crothers, D.M. (1981) *Nucl. Acids Res.* **9**, 6505–6525.
- Friedman, S. (1986) *Nucl. Acids Res.* **14**, 4543–4556.
- Lohman, T.M. (1986) *CRC Crit. Rev. Biochem.* **19**, 191–245.
- Riggs, A.D., Suzuki, H., and Bourgeois, S. (1970) *J. Mol. Biol.* **48**, 67–83.
- Chodosh, L.A., Carthew, R.W., and Sharp, P.A. (1986) *Mol. Cell. Biol.* **6**, 4723–4733.
- Wolberger, C., Vershon, A.K., Liu, B., Johnson, A.D. and Pabo, C.O. (1991) *Cell* **67**, 517–528.
- Kissinger, C.R., Liu, B., Martín-Blanco, E., Kornberg, T.E. and Pabo, C.O. (1990) *Cell* **63**, 579–590.
- Butkus, V., Petrauskienė, L., Maneliene, Z., Klimasauskas, S., Laucys, V. and Janulaitis, A. (1987) *Nucl. Acids Res.* **15**, 7091–7102.
- Chadwick, P., Pirrotta, V., Steinberg, R., Hopkins, N. and Ptashne, M. (1970) *Cold Spring Harb. Symp. Quant. Biol.* **35**, 283–294.
- Zabel, U., Schreck, R. and Baeuerle, P.A. (1991) *J. Biol. Chem.* **266**, 252–260.
- Affolter, M., Smith, A.P., Müller, M., Leupin, W., and Gehring, W.J. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 4093–4097.
- Fried, M.G. and Crothers, D.M. (1984) *J. Mol. Biol.* **172**, 241–262.
- King, K., Benkovic, S.J. and Modrich, P. (1989) *J. Biol. Chem.* **264**, 11807–11815.
- Wright, D.J., King, K. and Modrich, P. (1989) *J. Biol. Chem.* **264**, 11816–11821.
- Aiken, C.R., Fisher, E.W., and Gumpert, R.I. (1991) *J. Biol. Chem.* **266**, 19063–19069.
- Vovis, G.F., Horiuchi, K., and Zinder, N.D. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 3810–3813.
- Gunthert, U., Jentsch, S. and Freund, M. (1981) *J. Biol. Chem.* **256**, 9346–9351.
- Renbaum, P., Abrahamove, D., Fainsod, A., Wilson, G.G., Rottem, S., and Razin, A. (1990) *Nucl. Acids Res.* **18**, 1145–1152.
- Rubin, R.A. and Modrich, P. (1977) *J. Biol. Chem.* **252**, 7265–7272.
- Fox, K.R. (1988) *Biochem. Biophys. Res. Comm.* **155**, 779–785.