M.*Hha*I binds tightly to substrates containing mismatches at the target base

Saulius Klimašauskas+ and Richard J. Roberts[1*]

Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA and [1]New England Biolabs, Beverly, MA 01915, USA

ABSTRACT

The (cytosine-5) DNA methyltransferase M.*Hha*I causes its target cytosine base to be flipped completely out of the DNA helix upon binding. We have investigated the effects of replacing the target cytosine by other, mismatched bases, including adenine, guanine, thymine and uracil. We find that M.*Hha*I binds more tightly to such mismatched substrates and can even transfer a methyl group to uracil if a G:U mismatch is present. Other mismatched substrates in which the orphan guanine is changed exhibit similar behavior. Overall, the affinity of DNA binding correlates inversely with the stability of the target base pair, while the nature of the target base appears irrelevant for complex formation. The presence of a cofactor analog, *S*-adenosyl-L-homocysteine, greatly enhances the selectivity of the methyltransferase for cytosine at the target site. We propose that the DNA methyltransferases have evolved from mismatch binding proteins and that base flipping was, and still is, a key element in many DNA–enzyme interactions.

INTRODUCTION

The (cytosine-5) DNA methyltransferases catalyze the transfer of a methyl group from *S*-adenosylmethionine (AdoMet) to the 5-position of cytosine in DNA (reviewed in 1). The key step involves the formation of a covalent bond between a cysteine residue in the catalytic site of the enzyme and the 6-position of cytosine (2,3). This generates a delocalized carbanion that can accept the methyl group from the sulfonium center of AdoMet, forming a methylated dihydrocytosine intermediate. Loss of the hydrogen from the 5-position and release of the enzyme leads to the methylated product. If the proton at position 5 is replaced by a fluorine atom, then this final step is blocked and covalent complexes between the methyltransferase and DNA can be trapped and have been isolated and characterized (3,4).

Recently we have described crystal structures for a binary complex containing M.*Hha*I and AdoMet (5) and for a ternary complex containing M.*Hha*I covalently bound to an oligonucleotide containing 5-fluorocytosine and the end product of the reaction *S*-adenosylhomocysteine (AdoHcy) (6). The 2.8 Å structure of another ternary complex, involving a non-methylated

13mer duplex (7) is very similar to that of the covalently-trapped DNA. The surprising finding was that in the co-crystal structures containing DNA the target cytosine is flipped 180° out of the helix and into a pocket in the enzyme where catalysis takes place. Although the enzyme makes clear contacts with the orphan guanine residue, nothing is known about the contacts that might be made with the target cytosine during the initial binding reaction. Methyltransferases have provided the first example of base flipping during protein–DNA interaction and it is clearly of interest to understand how such flipping is achieved and whether it is likely to be found more generally.

We now report studies of the interactions between M.*Hha*I and oligonucleotides that contain mismatches at the target base pair. The mismatches studied include modifications at the target cytosine, as well as modifications of its partner, the orphan guanine. Both DNA binding and methyltransferase activity have been tested.

MATERIALS AND METHODS

Purified oligonucleotides and T4 polynucleotide kinase were obtained from New England BioLabs. [γ-^{32}P]ATP and [methyl-^{3}H]-*S*-Adenosyl-L-methionine ([^{3}H]AdoMet) were purchased from Amersham.

Concentrations of oligonucleotides were determined spectrophotometrically in 10 mM Tris–HCl (pH 8.0), 1 mM Na$_2$EDTA (TE buffer) based on their A$_{260}$–A$_{320}$ values. Molar extinction coefficients ε_{260} for strands were calculated as sums of individual contributions from deoxynucleotides. Duplexes were prepared by annealing individual strands in TE buffer from 80 to 20°C over 5–7 h.

Oligonucleotides were 5'-labeled with polynucleotide kinase and [γ-^{32}P]ATP, followed by gel filtration through Quick-Spin G-25 columns (Boehringer Mannheim). The following series of oligonucleotides were used (shown 5'→3', M = 5-methylcytosine, italics indicates the four bases of the recognition sequence):

GACTGGTACAGTATCAG*GCGC*TGACCCACAACATCCG
GACTGGTACAGTATCAG*GMGC*TGACCCACAACATCCG
GACTGGTACAGTATCAG*GUGC*TGACCCACAACATCCG
GACTGGTACAGTATCAG*GTGC*TGACCCACAACATCCG
GACTGGTACAGTATCAG*GAGC*TGACCCACAACATCCG
GACTGGTACAGTATCAG*GGGC*TGACCCACAACATCCG
GACTGGTACAGTATCAG

---

* To whom correspondence should be addressed

+Present address: Institute of Biotechnology FERMENTAS, Vilnius, Lithuania

```
GACTGGTACAGTATCAGG
GACTGGTACAGTATCAGGC
GACTGGTACAGTATCAGGCG
            CGCTGACCCACAACATCCG
             GCTGACCCACAACATCCG
              CTGACCCACAACATCCG
               TGACCCACAACATCCG
TCGGATGTTGTGGGTCAGCGCCTGATACTGTACCAGT
TCGGATGTTGTGGGTCAGMGCCTGATACTGTACCAGT
TCGGATGTTGTGGGTCAGMICCTGATACTGTACCAGT
TCGGATGTTGTGGGTCAGMACCTGATACTGTACCAGT
TCGGATGTTGTGGGTCAGMTCCTGATACTGTACCAGT
TCGGATGTTGTGGGTCAGMCCCTGATACTGTACCAGT
```

M.*Hha*I was essentially purified as earlier described (8); exhaustive dialysis prior to FPLC yielded homogeneous AdoMet-free enzyme (7). M.*Hha*I concentrations were estimated using the BioRad Coomassie G-250 assay and BSA as a standard.

Electrophoretic mobility shift analysis

Binding reactions were performed in 50 mM NaCl, 10 mM Tris–HCl (pH 7.4), 0.5 mM Na$_2$EDTA, 2–5 µg/ml poly(dA–dT), 200 µg/ml bovine serum albumin (BSA) and 6–8% glycerol. The poly(dA–dT) was added to eliminate non-specific methyltrans-ferase–substrate complexes that otherwise appear as lower mobility or smeared bands. Typical reactions of 10–15 µl were incubated for 20–40 min at 20°C and loaded onto a running 7–12.5% polyacrylamide gel. Electrophoresis was performed in 90 mM Tris–borate (pH 8.3), 2 mM Na$_2$EDTA for 1–1.5 h at 7–10 V/cm. Prior to loading, gels were prerun for 1.5–2 h. The gels were dried on Whatman 3MM paper and autoradiographed to Kodak XAR film or quantitated with a Fuji BAS-2000 phosphor-imager.

To normalize data from different experimental datasets the following procedure was employed. One dataset was chosen as a reference. A second dataset was then scaled to the reference dataset using a least squares procedure. For instance, for an oligonucleotide containing the mismatch GU, the ratio R = c.p.m. bound/c.p.m. free was determined in both the reference dataset and a different experimental dataset. These ratios were then plotted as a single data point x,y where x = log$R$ for the reference and y = log$R$ for the different experimental set. This was continued over most matching pairs of data points between the two sets and a straight line drawn through the points using a least squares approximation. The resulting straight line is defined by an equation $y = a + bx$ and is characterized by the slope b and the intercept a. The values of a and b were then used as scaling coefficients to recalculate a scaled value, logR'. Mean values were then calculated for each mismatch from the appropriate datasets and used to produce Figure 7.

Enzymatic methylation reactions were carried out at 20 or 37°C in 50 mM Tris–HCl (pH 7.4), 10 mM Na$_2$EDTA, 6 mM 2-mercaptoethanol, 200 µg/ml BSA in the presence of 130–350 nM [³H]AdoMet (15 or 78 Ci/mmol). Reactions were terminated by addition of AdoHcy to 10–100 µM concentration (2). Samples were spotted onto DE-81 filters (Whatman), washed with 0.2 M ammonium bicarbonate (5 × 2 ml), water (2 × 1 ml) and ethanol (2 × 1 ml), dried and counted in a liquid scintillation spectrometer. Initial velocities were determined under conditions of <5% substrate consumption.
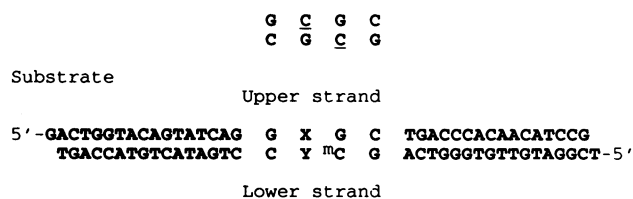

Figure 1. The M.*Hha*I recognition sequence and the design of oligonucleotides containing mismatches at the target site.

Analysis of the methylated nucleotide

Duplex oligonucleotides (13 µM) were subjected to methylation with M.*Hha*I (3 µM) in the presence of [³H]AdoMet (15 Ci/mmol) for 6 h. The samples were treated with proteinase K for 1 h at 56°C, chloroform-extracted and purified by passing through a G-25 spin column. The oligonucleotides were digested with nuclease P1 at 37°C for 2 h and concentrated *in vacuo*. The samples were applied to thin layer cellulose plates and chromato-graphed in one dimension, along with standard pThd and pdm5Cyd, as described earlier (9). Following inspection under UV light, the plates were cut into several sections. The cellulose layer was then scraped off into scintillation vials with 10 ml Cytoscint and analyzed for ³H radioactivity. Controls were performed with no DNA present in the methylation reaction or nuclease P1 omitted from the hydrolysis reaction.

RESULTS

Duplex oligonucleotides (37mer) containing modifications at the target G:C base pair, indicated by Y:X in Figure 1, were used to study the binding and catalytic properties of M.*Hha*I. All substrates contained 5-methylcytosine (m5C) on the lower strand, so that an asymmetric, hemimethylated site was present that should favor binding in a single orientation. Such preferential binding is common among methyltransferases (10,11) and probably reflects the fact that hemimethylated sites are the predominant substrate *in vivo*.

Changes in the target cytosine

In the first series of experiments we changed the target cytosine (X) to a variety of other bases. It was replaced by each of the other three natural bases and also with m5C and uracil. As can be seen in Figure 2, introduction of a mismatch opposite the orphan guanine leads to the formation of very stable protein–DNA complexes. Yet another substrate, termed G:Δ, contains the orphan guanine opposite a gap formed by removal of the target cytosine and both phosphodiester linkages. Remarkably, this substrate shows the tightest binding in the series. A duplex containing a C→m5C substitution, G:M, is bound with substantially lower affinity than the native substrate. Since this is the end product of the reaction this is not surprising. Overall, the stability of the binary complexes involving G:X base pairs increases in the order m5C < C << G, A, T, U < gap. The strength of the G:X base pairs as judged from the corresponding duplex T_m values is exactly the opposite: m5C > C >> G, A, T (12–14). Thus there is an inverse correlation between the stability of the methylase–DNA complexes and the strength of the base pair in which the target cytosine is located.
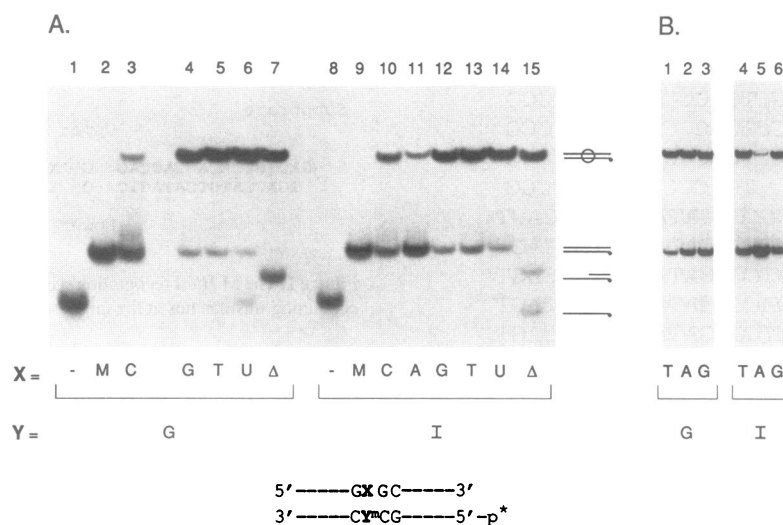
A.   B.


```
5'-----GX GC-----3'
3'-----CYᵐCG-----5'-p*
```

Figure 2. Equilibrium binding between M.*Hha*I and DNA duplexes containing G:X or I:X base pairs. 5'-^{32}P-labeled 37mer duplexes (4.5 nM) were incubated at 20°C for 20 min in the presence of 16 nM M.*Hha*I and 2 μg/ml poly(dA–dT)·poly(dA–dT). Samples were resolved by 10% PAGE and autoradiographed as described in Materials and Methods. – (lanes 1 and 8), no upper strand; M (lanes 2 and 9), m5C; Δ (lanes 7 and 15), no nucleotide (see Figs 3 and 5); p*, labeled strand. The schematic between (A) and (B) shows the structures of the labeled bands detected.

If the orphan guanine in the lower strand is substituted by its close analog inosine, a similar pattern is observed. The binding of I:M and I:C duplexes is slightly increased, while the binding of I:X mismatches diminishes, as compared with the G-containing DNAs. This again correlates inversely with the known effects of inosine on the thermal stability of the corresponding duplexes. For instance, the I:C and I:M base pairs are destabilized (due to the lack of one hydrogen bond), whereas the I:U, I:A and I:G mismatches are more stable than the analogous G:X substrates (15–17). A large decrease in binding is observed when G:A is replaced by I:A (Fig. 2B). This is consistent with the inverse correlation above, since the stability of an I:A base pair is usually higher than that of G:A, although the extent of such stabilization is very much dependent on surrounding sequences (18).

Changes in the orphan guanine

Another series of substrates contained permutations of the orphan guanine in the lower strand. The strength of binary interactions with Y:C base pairs (Fig. 3, left panel) increases in the order (Y=): G < I < A < T, C. Here again there is an inverse correlation, as the weakest base pairs, C:C and C:T (12,19), are bound with the highest affinity. It should be noted that in the ternary covalent complex (6) the orphan guanine is involved in multiple contacts with the protein and the above replacements should cause some loss of interaction energy. Indeed, the binding experiment where the orphan guanine is exchanged and no target base is present on the upper strand (Fig. 3, right panel) reveals a substantial decline in the affinity of binary complexes (Y=): G > I > T > A > C. In this case no base pair interactions are possible in unbound DNA and the observed drop in affinities mostly reflect the binding energy penalty associated with the orphan base replacements.

Changes in both the orphan guanine and the target cytosine

Figure 4 shows an experiment where both the target cytosine and the orphan guanine are changed. Only when A:T or T:A base pairs were present was binding almost completely abolished, even if
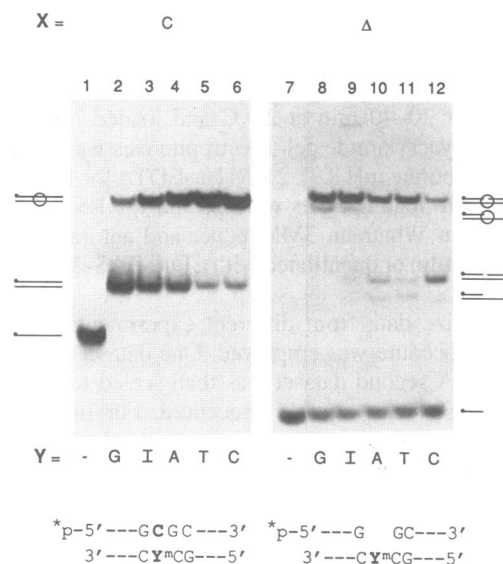


```
*p-5'---GCGC---3'   *p-5'---G  GC---3'
  3'---CYᵐCG---5'     3'---CYᵐCG---5'
```

Figure 3. Equilibrium binding between M.*Hha*I and DNA duplexes containing a Y:C base pair (left) or an orphan base Y opposite a deletion (right). Experimental conditions and symbols as in Figure 2.

AdoHcy is present. Since these two substrates represent non-specific sequences (GAGC and GTGC) this is expected, due to the specificity of M.*Hha*I. Nevertheless, it may be considered surprising in view of the proclivity of other combinations to show such strong binding. All other permutations tested permit binding, although when both X and Y are exchanged the mismatches are bound less efficiently than most others, with modest increases observed when AdoHcy is present.

Effect of position within the recognition sequence

All of the above modifications were located at the target G:C base pair within the specific recognition sequence GCGC for M.*Hha*I.
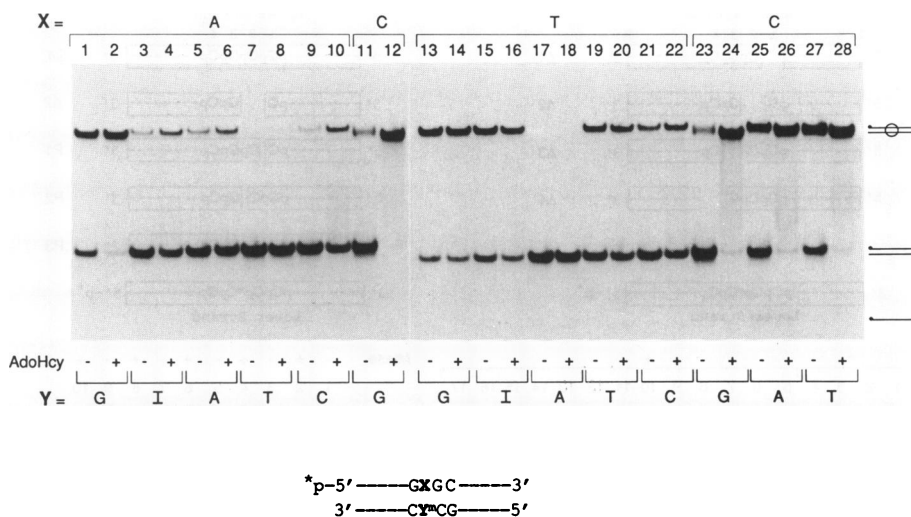
Figure 4. Equilibrium binding between M.*Hha*I and DNA duplexes containing a Y:X base pair in the presence (+) or absence (−) of AdoHcy. 5′-^{32}P-labeled 37mer duplexes (4 nM) were incubated in the presence of 15 nM M.*Hha*I and 50 μM (if any) AdoHcy. Other experimental conditions and symbols as in Figure 2.

To show that the position of the mismatch, as well as its sequence context, was important we varied the location of the lesion along the recognition sequence. Since substrate oligonucleotides containing a gap at the target cytosine show the maximum effect on binding (Fig. 2), we chose to use substrates with gaps at varying positions on the upper strand. Substrates with nucleotide deletions at all four positions (termed ΔN, where N = 1, 2, 3 or 4) were prepared by annealing a lower strand 37mer oligonucleotide with complementary pairs of upper strand oligonucleotides (Fig. 5, left). The controls for each substrate included both partial duplexes involving the labeled lower strand and either the left or right component of the detached upper strand. The only substrates interacting with the methylase were the Δ2 duplex (identical with G:Δ in Fig. 2) and its relative lacking the 3′ component on the upper strand (lanes 8 and 6). Almost identical patterns were observed in the presence or absence of AdoHcy (not shown). This experiment clearly demonstrates that binding is specific for substrates in which the target cytosine is removed from the GCGC site. Duplexes lacking other nucleotides were poor substrates, even though they retain the target cytosine. Apparently, both the presence of the recognition sequence and the exact positioning of the deletion site are necessary for the formation of high affinity complexes, since the other substrates (N = 1, 3 or 4) can be also regarded as specific C or G deletions within non-specific sequences. The only surprise is that a partial recognition sequence forms a binding complex (lane 6). The nature and significance of this complex will require further investigation.

Effect of nicked substrates

Another series of experiments examined duplexes containing dephosphorylated nicks in the upper strand (Fig. 5, right panel). Here all four specific base pairs are preserved. Remarkably, removal of either 5′ or 3′ phosphate around the target cytosine (duplexes P1 and P2) leads to a 2.6- and 26-fold improvement in binding over the cognate substrate. The previously described Δ2 duplex, which lacks both phosphates and the target nucleoside, is bound most efficiently (45-fold better than G:C). Although both of the target cytosine phosphodiesters are involved in multiple

interactions with the methyltransferase in the ternary complex, they also suffer conformational strain associated with the flipped out cytosine. It is these phosphodiesters that show maximal deviations from the optimal range of torsion angles (6). The loss of the interaction energy is apparently offset by alleviating the conformational strain in the upper strand of the substrate DNA. In contrast, the loss of a single phosphate contact in P3 is not compensated for by other factors and so leads to a lower binding affinity.

Effects of cofactor on the methyltransferase–DNA interactions

Addition of the cofactor AdoMet or the reaction end product AdoHcy leads to the formation of ternary complexes. Both compounds are known to facilitate the formation of specific complexes (10,11,20), although their effects on the methyltransferase–DNA interactions are not identical. AdoHcy is a competitive inhibitor with respect to AdoMet, as it lacks the mobile methyl group and cannot support enzymatic turnover. These differences are illustrated in the I:X series of duplexes shown in Figure 6 (left panel). In the presence of AdoHcy the I:C and G:C substrates (lanes 12 and 16) are bound with very high affinity. These non-productive ternary complexes represent a dead-end branch in the enzymatic reaction (2). Lanes 6 and 7 in Figure 6 appear identical, because in the presence of AdoMet enzymatic conversion of C to m5C has taken place. The reaction product, a fully methylated duplex, is poorly bound by M.*Hha*I in the presence of AdoMet. The behavior of the fully methylated duplex (lane 11) in the presence of AdoHcy is different. A substantial portion of radioactivity in lanes 11 and 17 appears as a smear between the bound and free DNA bands, indicating that most of the initial complex has fallen apart during electrophoresis. This suggests that this ternary complex dissociates more rapidly than the others. The free DNA band in lane 17 constitutes half of the total counts, indicating that the ratio of complex:free DNA is 1:1 at equilibrium, much lower than that observed for the canonical site (lane 16). The apparent kinetic and thermodynamic instability of the methyltransferase–AdoHcy–G:M complex is important mechanistically, since dissociation of the final reaction complex contributes to the turnover velocity and in fact may be rate limiting.
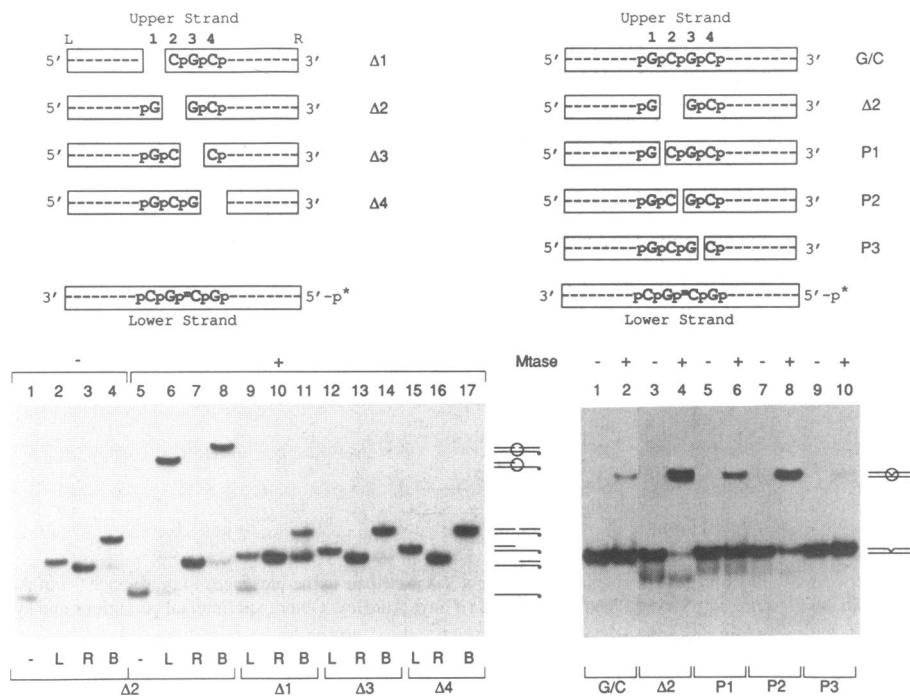
Figure 5. Equilibrium binding between M.*Hha*I and DNA duplexes containing gaps in the upper strand of the recognition site. The gap was formed by loss of the nucleoside and both 3′ and 5′ phosphates. (**Top**) Structures of the substrate duplexes. Numbers indicate nucleotides within the recognition site. (**Bottom**) Autoradiographs of mobility shift experiments. (**Left**) 5′-^{32}P-labeled gapped DNA substrates (1.6 nM) as shown above were incubated with 9 nM M.*Hha*I and 200 μM AdoHcy and resolved by 12.5% PAGE. L, R and B indicate that either right, left or both upper strand components were included in the annealing reaction. (**Right**) Labeled duplexes (3.2 nM) as shown above were incubated with 19 nM M.*Hha*I and analyzed by 12.5% PAGE.
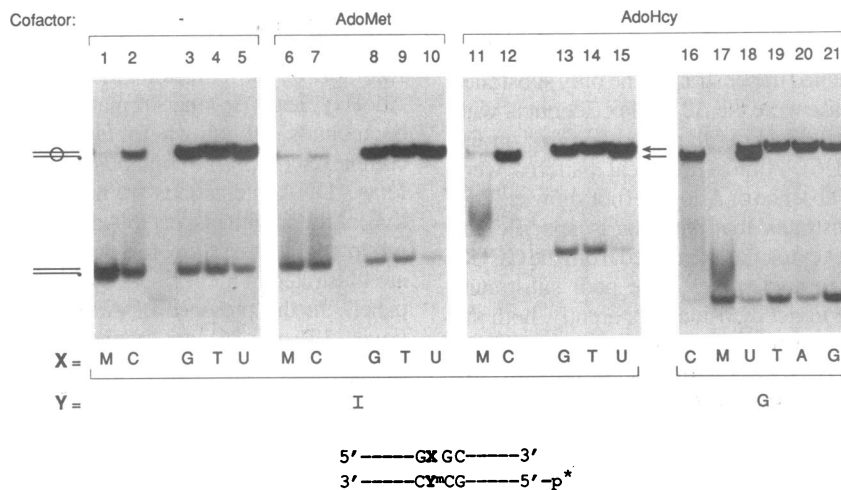

Figure 6. Equilibrium binding between M.*Hha*I and DNA duplexes containing I:X or G:X base pairs in the presence of 200 μM AdoMet, AdoHcy or no cofactor (–). Other experimental conditions and symbols as in Figure 2.

Another distinctive feature of the AdoHcy complexes is the heterogeneity of their electrophoretic mobility. There is hardly any discernible variation in mobility within binary complexes (Fig. 6, left panel) or AdoMet complexes (second panel). In contrast, the ternary complexes involving AdoHcy show clear and reproducible differences. The I:M and I:C complexes (lanes 11 and 12) move faster through the gel than do I:G or I:T complexes. Curiously, the I:U duplex is partitioned almost evenly between the two species. A similar regularity is observed with the G:X substrates that were analyzed under conditions allowing a clearer separation of the electrophoretic bands (see right panel). Here again, the U complex appears as a doublet (lane 18), although the ratio of the bands is consistently different. The two bands are stable enough to survive prolonged electrophoresis with no signs of interconversion. We do not have an explanation as to why G:U and I:U mismatches form two kinds of stable
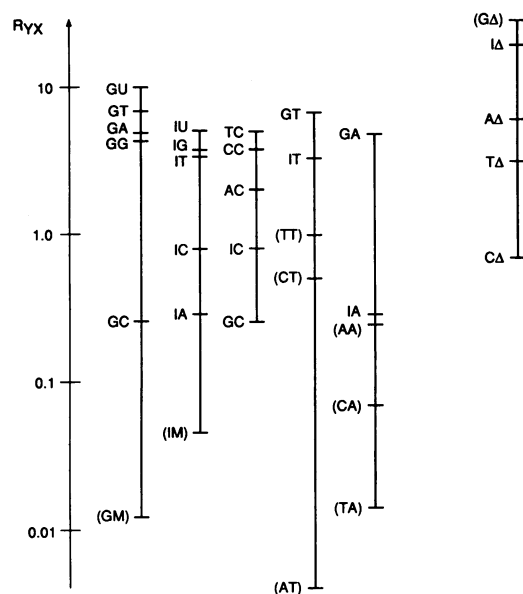
Figure 7. Schematic representation of the relative affinities (R_{YX}) of the binding between M.*Hha*I and the DNA duplexes containing Y:X combinations as shown in Figure 1. Six vertical lines show data for illustrative combinations (from left to right): G:X, I:X, Y:C, Y:T, Y:A and Y:Δ. Note that data from several different experimental datasets have been merged for this comparison and mean values from several determinations are presented. In all cases the variations found in multiple (2–5) measurements are less than 0.08 log units (20%), except for I:M, where the variation was 0.27 log units (80%). Values for other DNAs in parentheses were determined from single measurements.

conformers upon interaction with M.*Hha*I and AdoHcy. However, it is interesting to note that the G:C complex also runs faster than most complexes. It may indicate that the faster running substrates are all bound in a similar compact structure in which the catalytic loop has undergone the full conformational change required for catalysis (6).

Quantitation of DNA–M.*Hha*I interactions

We attempted to determine dissociation constants for the binary interaction between M.*Hha*I and its substrates using Scatchard analysis. K_d values derived from the reverse titration experiments for duplexes G:T and G:U were 2.2 and 0.5 nM respectively. Unfortunately, the same approach gave unsatisfactory results for other complexes of lower affinity (G:C, I:C, G:M and I:M). At higher concentrations of M.*Hha*I these complexes exhibited non-linear behavior, precluding a reliable assessment of K_d values. The problem could not be alleviated by changing either the conditions of the binding reaction or the gel electrophoretic step.

As an alternative approach we compared affinities by measuring the binding ratios, defined as R = bound oligonucleotide/free oligonucleotide. In a linear binary system R is proportional to $1/K_d$ if the concentrations of all reactants and other conditions are identical. However, there was considerable variation in the absolute values observed in our experiments, because different reaction conditions were often employed. Nevertheless, the relative numbers proved quite reproducible under any given set of conditions. Datasets from four experiments were normalized to a fifth reference dataset using a least squares procedure. The data were then merged by averaging the normalized R values

from the five experiments. The mean values obtained are presented on a logarithmic scale (Fig. 7), where relative binding ratios are organized into vertical 'stacks' that correspond to the individual scans shown in Figures 2, 3 and 4. Note that the values in any one vertical stack may come from more than one experimental dataset. The schematic in Figure 7 provides a consolidated view of how 37mer duplexes containing various base pairs interact with M.*Hha*I in the absence of cofactor.

Enzymatic modification of mismatch duplexes

In the light of the unexpected ability of M.*Hha*I to bind tightly to mismatch duplexes it was of interest to see if any could behave as substrates in the catalytic reaction. First, the G:X duplexes were incubated with M.*Hha*I in the presence of [^{3}H]AdoMet. The expected levels of radioactivity were incorporated into the canonical DNA containing G:C at the target base pair. Of the other duplexes only G:U gave significant counts above background (Table 1). Analysis of the modified nucleotide confirmed that the product of this reaction was thymine (Fig. 8). This means that the cytosine methyltransferase M.*Hha*I catalyzes the transfer of a methyl group from AdoMet onto the C5-position of a uracil residue when the latter replaces the target cytosine residue within the specific recognition sequence. Remarkably, the G:U duplex is the only one in the G:X series that in the presence of AdoHcy is able to form the same type of ternary complex as does the canonical 37mer (Fig. 6, right). Chemically, cytosine and uracil have much in common, as they share the general mechanism of enzymatic transfer of 1-carbon units onto the C5 of pyrimidines (1). The M.*Hha*I turnover rate in the methylation reaction (at 42 nM duplex and 1.1 μM AdoMet, 37°C) for the G:U duplex is 5 × 10^{-3}/min, which is almost three orders of magnitude lower than the value of 1.4/min observed for the G:C hemimethylated duplex (this study) or 1.3/min for poly(dG–dC) (2).

Table 1. Enzymatic methylation of G:X mismatches with M.*Hha*I

DNA[a]	^{3}H incorporated (c.p.m.)[b]	
	20 min	40 min
..GMGC..	240	470
..GCGC..	35 900	41 000
..GUGC..	760	1130
..GUGC..[c]	210	560
..GMGC..[c]	240	480

[a]Only the detail of the upper strand of the recognition sequence is shown. In each case the lower strand was premethylated (5'..GMGC..3').
[b]Reaction conditions were 1.2 μM DNA, 1.2 μM M.*Hha*I, 1.3 μM [^{3}H]AdoMet (37°C).
[c]These are controls in which only the single upper strand was present.

The Y:C series of duplexes were also tested for their ability to function as substrates in the methylation reaction. All of them were good substrates of the methyltransferase, with a slight spread in their methylation efficiencies (Table 2). Initial reaction velocities at 20–80 nM duplex concentration at both 20 and 37°C decrease in the order: Y = A > I, T > G > C. At 20°C the velocity ratios from the experiment of Table 2 are 2.0 (A):1.3 (I):1.2 (T):1.0 (G):0.9 (C) (data not shown). This contrasts with the relative measured stabilities (T > C > A > I > G) of these duplexes in binary complexes with M.*Hha*I, where there is an inverse correlation with the strength of the Y:C base pair (Fig. 3). Nor is
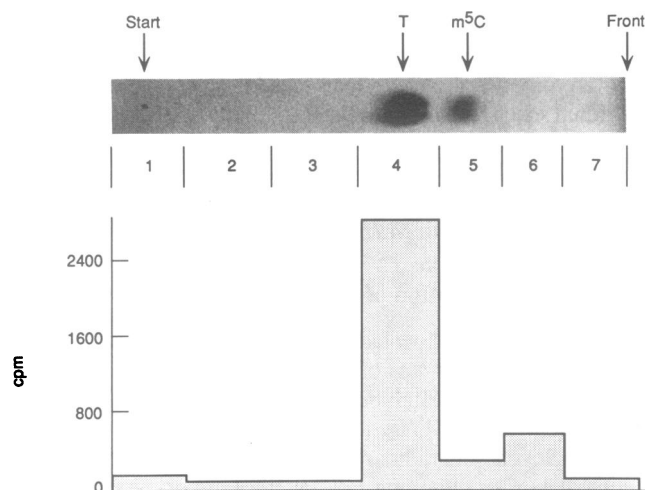
Figure 8. Analysis of the methylated nucleotide produced by M.*HhaI* after incubation with a duplex substrate containing a G:U mismatch at the target. The G:U duplex was methylated in the presence of M.*HhaI* and [^{3}H]AdoMet as described in the Materials and Methods. DNA was hydrolyzed with nuclease P1 and the resulting mononucleotides were mixed with thymidylic acid and 5-methyldeoxycytidylic acid as standards. The mixture was separated by TLC on cellulose and the plate was photographed under UV-illumination (top) to locate the standards. It was then cut into sections as indicated and analyzed for ^{3}H radioactivity (bottom). The small peak present in section 6 was also present in the undigested control.

there a discernible correlation with the stability of ternary complexes involving AdoHcy (Fig. 4, lanes 23–28).

Table 2. Enzymatic methylation of Y:C mismatches with M.*HhaI*

DNA[a]	^{3}H incorporated (c.p.m.)[b]	
	20 min	40 min
3′..CGMG..	8400	17 400
3′..CIMG..	11 700	22 500
3′..CAMG..	17 200	35 800
3′..CTMG..	12 100	20 800
3′..CCMG..	8600	15 800
3′..CGCG..[c]	4200	

[a]Only the detail of the lower strand of the recognition sequence is shown. In each case the upper strand was unmethylated (5′..GCGC..3′).

[b]Reaction conditions were 80 nM DNA, 0.4 nM M.*HhaI*, 260 nM [^{3}H]AdoMet (20°C).

[c]Typical results for an unmethylated substrate, where the initial velocity is approximately half that of a hemimethylated substrate.

DISCUSSION

Previously it has been shown that DNA binding in a ternary complex containing M.*HhaI*, DNA and AdoHcy results in the target cytosine residue being flipped out of the helix and into a pocket in the enzyme (6). In the present study we have mainly focused on binary complexes between M.*HhaI* and DNA with no cofactor present.

While the stable binding of M.*HhaI* to substrates containing a mismatch may appear unexpected at first sight, it can be partially rationalized on thermodynamic grounds. The energy needed to disrupt the target G:C base pair is offset from the total energy

gained upon formation of multiple protein–DNA contacts. The weaker the base pair, the less energy is required to disrupt it and the higher the stability of the resulting complex. Similarly, for the substrates containing a gap, where there is no base pair to disrupt and no base to flip out, tight binding might be expected. However, it should be noted that substitution of either the target cytosine or the orphan guanine will also have energetic costs, as a result of lost or altered interactions in the complex. In the light of the flipping mechanism, the inverse correlation observed between base pair strength and binding affinity seems intuitively clear, but must be viewed with caution until the additional effects of lost contacts are better understood.

Although we have not directly demonstrated that the complexes contain a flipped base, it seems likely on several grounds. In each of these mismatch interactions the complexes formed all display greater stabilities than that of the cognate G:C base pair and they show similar gel shift mobilities. Furthermore, whenever cytosine or uracil is at the appropriate location, the complex is catalytically competent and the target base can be methylated. Some additional support for this idea is also available. Earlier biochemical studies of M.*HhaI* showed that in the absence of cofactor, M.*HhaI* catalyzes the exchange of the proton at the 5-position of the substrate cytosine for a proton from water (2). Such catalysis requires the formation of a transient covalent bond between the active site cysteine residue and the target cytosine, which only seems possible if the target cytosine is flipped out of the helix.

Recent experimental evidence from other laboratories suggests that this tight binding of a (cytosine-5) methyltranferase to mismatches may also be true for M.*HpaII* (21) and M.*MspI* (22). In the latter study it was found that elimination of the 4-amino group of the target cytosine results in unusually stable complexes between M.*MspI* and DNA. This chemical modification would abolish one of the three hydrogen bonds to the orphan guanine. It thus becomes unnecessary to postulate covalent binding to this analog to explain the enhanced binding. Enhanced binding, without covalent bond formation, has also been observed when the active site cysteine residue is changed to glycine or alanine (3,20,23).

Addition of the cofactor analog AdoHcy (or AdoMet) leads to further stabilization of the methyltransferase–DNA complexes, which appears to be a common effect described for the majority of methyltransferases examined (10,11). Our data show that the stabilization is substantial only if the target position is occupied by cytosine and, to a lesser extent, m5C or uracil. One of the effects of AdoHcy binding is apparently stabilization of the active site loop in its closed conformation, where the active site is assembled around the target residue (6). Additional favorable contacts are generated between the target base and the active site residues, with cytosine providing the best match, followed by its closest analogs uracil and m5C. This is the first point in the reaction mechanism where the nature of the target base becomes crucial. It also highlights the molecular basis for an active role for the cofactor in sequence discrimination, which had been earlier documented for a number of bacterial methyltransferases (10,11). Since complexes containing either thymine or purine bases in place of the normal target cytosine lead to complexes of low electrophoretic mobility (Fig. 6), it seems that these bases cannot be accommodated properly in the active site. Thus the methyltransferase is able to discriminate effectively between the correctly oriented G:C base pair and the other three base pairs naturally found in DNA.

Earlier, increased methylation velocities had been observed for human DNA methylase upon interaction with DNA substrates containing mismatches in the recognition site (24–26). These effects were to some extent attributed to the same mechanistic feature, although no clear cut correlations could be established. Our methylation velocity studies show a rather small effect of mismatches on the overall reaction efficiency, as compared with the human enzyme, suggesting that the rate limiting steps in these two cases are different. Provided base separation occurs upon formation of a protein–DNA complex, the affinity of binding may be a better measure of the energetic change in the substrate DNA than is the overall reaction velocity. Our findings may be especially useful for studies of enzymatic reactions that employ local opening of the DNA helix.

Evolution of methyltransferases

It was unexpected that M.*Hha*I would form such stable complexes with oligonucleotides containing a wide variety of mismatches in the target G:C base pair. It suggests that DNA recognition by M.*Hha*I consists of two components. One is the specific interaction with the outer three base pairs of the recognition sequence. The other is the location of a correctly-positioned cytosine or other base that can be flipped out of the helix. Since both partners in the G:C base pair that is the target for flipping can be substituted, it appears that the key feature of the latter recognition is the ability to be flipped or, in the case of a deletion, not interfering with flipping. We conclude that M.*Hha*I initially recognizes and binds 5'-G–GC/CGCG-3' sites (where – represents 'no base in the major groove'). The canonical GCGC sequences are then only bound if/when the cytosine at position 2 is flipped out of the helix. This could happen by the enzyme catching an intermediate during normal base pair breathing, since the reported half-life for this process is 1–100 ms (13), while the reaction turnover is ~1 per min (2). However, flipping could also be an enzyme-assisted process. It is reasonable to ask how such a mechanism of recognition and binding could have arisen. The present results suggest that perhaps the methyltransferases have evolved from DNA mismatch binding proteins.

Early in evolutionary history, when DNA was first being used as the genetic material, one can imagine that the primitive DNA polymerases were much less faithful than they are today. It is likely that mismatches were commonly present in such early DNA molecules. However, because of the need to preserve sequence integrity in the newly emerging genetic material, it would have been essential to correct those mismatches. Since proteins or, more likely, shorter polypeptides were similarly primitive, such correction would probably involve the interaction of several different discrete polypeptides. What better way to achieve this than to select polypeptides able to bind mismatches and flip one of the bases out of the helix? In this way both the base and its surrounding phosphodiester bonds would be available for the action of other polypeptides that might effect its removal and subsequent replacement. In such a model, base flipping would have been an early discovery during the evolution of DNA as the genetic material. The principal advantage is that it allows the cooperation of two or more polypeptides to effect the correction or other manipulation of the DNA bases.

We suggest that the m5C-methyltransferases have arisen by the aggregation of protein domains with the ability: (i) to recognize mismatches in DNA; (ii) to perform sequence-specific recognition;

(iii) to carry out the methylation reaction. In this model the key original module was the mismatch recognition system, which could accomplish base flipping. If such an early evolutionary origin for base flipping is correct, then it is to be expected that it will occur more promiscuously than has thus far been reported. This idea is supported by the recent crystal structure for a uracil-DNA glycosylase, which also uses a base flipping mechanism (27).

REFERENCES

1 Kumar,S., Cheng,X., Klimasauskas,S., Mi,S., Posfai,J., Roberts,R.J. and Wilson,G.G. (1994) *Nucleic Acids Res.*, **22**, 1–10.
2 Wu,J.C. and Santi,D.V. (1987) *J. Biol. Chem.*, **262**, 4778–4786.
3 Chen,L., MacMillan,A.M., Chang,W., Ezaz-Nikpay,K., Lane,W.S. and Verdine,G.L. (1991) *Biochemistry*, **30**, 11018–11025.
4 Friedman,S. and Ansari,N. (1992) *Nucleic Acids Res.*, **20**, 3241–3248.
5 Cheng,X., Kumar,S., Posfai,J., Pflugrath,J.W. and Roberts,R.J. (1993) *Cell*, **74**, 299–307.
6 Klimasauskas,S., Kumar,S., Roberts,R.J. and Cheng,X. (1994) *Cell*, **76**, 357–369.
7 Cheng,X., Kumar,S., Klimasauskas,S. and Roberts,R.J. (1993) *Cold Spring Harbor Symp. Quant. Biol.*, **58**, 331–338.
8 Kumar,S., Cheng,X., Pflugrath,J.W. and Roberts,R.J. (1992) *Biochemistry*, **31**, 8648–8653.
9 Klimasauskas,S., Nelson,J.L. and Roberts,R.J. (1991) *Nucleic Acids Res.*, **19**, 6183–6190.
10 Dubey,A.K. and Roberts,R.J. (1992) *Nucleic Acids Res.*, **20**, 3167–3173.
11 Chen,L., MacMillan,A.M. and Verdine,G.L. (1993) *J. Am. Chem Soc.*, **115**, 5318–5319.
12 Aboul-ela,F., Koh,D., Tinoco,I. and Martin,F.H. (1985) *Nucleic Acids Res.*, **13**, 4811–4824.
13 Moe,J.G. and Russu,I.M. (1992) *Biochemistry*, **31**, 8421–8428.
14 Jurgaitis,A.P., Butkus,V., Klimasauskas,S. and Janulaitis,A. (1988) *Bioorg. Khim.*, **14**, 158–165.
15 Martin,F.H., Castro,M.M., Aboul-ela,F. and Tinoco,I. (1985) *Nucleic Acids Res.*, **13**, 8927–8938.
16 Kawase,Y., Iwai,S., Miura,K. and Ohtsuka,E. (1986) *Nucleic Acids Res.*, **14**, 7727–7736.
17 Oda,Y., Uesugi,S., Ikehara,Y. and Ohtsuka,E. (1991) *Nucleic Acids Res.*, **19**, 5263–5267.
18 Leonard,G.A., Booth,E.D., Hunter,W.N. and Brown,T. (1992) *Nucleic Acids Res.*, **20**, 4753–4759.
19 Werntges,H., Steger,G., Riesner,D. and Fritz,H.-J. (1986) *Nucleic Acids Res.*, **14**, 3773–3790.
20 Wyszynski,M.W., Gabbara,S., Kubareva,E.A., Romanova,E.A., Oretskaya,T.S., Gromova,E.S., Shabarova,Z.A. and Bhagwat,A.S. (1993) *Nucleic Acids Res.*, **21**, 295–301.
21 Yang,A.S., Shen,J.-C., Zingg,J.-M., Mi,S. and Jones,P.A. (1995) *Nucleic Acids Res.*, **23**, 1380–1387.
22 Ford,K., Taylor,C., Connolly,B. and Hornby,D.P. (1993) *J. Mol. Biol.*, **230**, 779–786.
23 Mi,S. and Roberts,R.J. (1993) *Nucleic Acids Res.*, **21**, 2459–2464.
24 Smith,S.S., Hardy,T.A. and Baker,D.J. (1987) *Nucleic Acids Res.*, **15**, 6899–6916.
25 Smith,S.S., Kan,J.L.C., Baker,D.J., Kaplan,B.E. and Dembek,P. (1991) *J. Mol. Biol.*, **217**, 39–51.
26 Tan,N.-W. and Li,B.F.L. (1990) *Biochemistry*, **29**, 9234–9240.
27 Savva,R., McAuley-Hecht,K., Brown,T. and Pearl,L. (1995) *Nature*, **373**, 487–493.