Supplemental Data

Functional validation of the M.HhaI mutants in this study was provided by several studies. The introduction of a Trp residue into the catalytic loop has almost no impact on methylation of the cognate DNA sequence when compared to WT enzyme. Many parameters were measured and reported (supplemental Table 1) for W41F, W41F/K91W, and W41F/E94W with various DNA substrates and cofactors. The K_D values for various DNA substrates confirm that all enzymes are bound to these substrates under the conditions of the stopped-flow experiments. Also, K_M values for AdoMet confirm mutant enzymes are bound to cofactor during stopped-flow experiments, and that the AdoMet binding pocket is not disturbed. K_M values were measured for W41F to confirm our results were consistent with previous reported values of K_M (19) and to further validate our observed K_M values for W41F/K91W and W41F/E94W. Although a ~10-fold reduction in affinity from WT for both DNA and cofactor substrates was observed for W41F/K91W and W41F/E94W, it was i) predominantly due to the W41F mutation (19) and ii) was accounted for during stopped-flow experiments by using a large excess of DNA substrates (750 nM) and cofactors (500 μM). Stopped-flow data collected during catalysis was also measured with increased concentrations to further ensure that enzymes were fully bound and saturated. Order of addition stopped-flow experiments showed no difference between enzyme mixed with DNA and cofactor, and enzyme/cofactor mixed with DNA. Enzyme/DNA mixed with cofactor looked nearly identical but was missing the very fast initial loop closure previously reported and could only be seen at lower temperatures (~4 ºC) (12) (data not shown).

In Figure 2 we show fluorescence changes monitored using a stopped-flow apparatus during various substrate and cofactor additions to the two M.HhaI mutants W41F/K91W and W41F/E94W. Previous measurements with just enzyme and DNA (binary complex) revealed a coupling between base flipping and loop motion (12). Loop motion in the binary state is significantly faster and only minimally populates the closed conformer when compared to the ternary complex (data not shown). Rate constants for binary loop motion had to be measured at 4 ºC over 2 seconds and showed a dependence on DNA concentration (12). Here, we monitored ternary loop motion at 22 ºC over 50 seconds and no concentration dependence was observed. Thus, the enzyme immediately recognizes cognate DNA but must make slower rearrangements for cofactor binding and catalysis since no concentration dependence was observed. Although slower, cofactor additions greatly increase the population of the closed conformer as determined by the maximum of observed fluorescence intensity (0.03 binary, 0.19 ternary, data not shown). This result is consistent with NMR studies on M.HhaI (22). While cofactor itself does not mediate loop motion, in the presence of a DNA substrate, it greatly increases the population of the closed conformer.

Loop motion was observed with NC-2AP DNA when mixed with AdoMet where it was predominantly absent when mixed with AdoHcy (Fig. 2*B* and *E*-*G*). Surprisingly, the change in signal with NC-2AP DNA and AdoMet was in the same direction for both W41F/K91W and W41F/E94W, contrary to all other traces (Fig. 2*E* and *F*; data not shown for W41F/K91W). Since 2AP is a fluorescent base analog with spectral properties ideal for energy transfer with Trp (23), it is most likely absorbing the Trp fluorescent signal as the two molecules are brought into proximity upon loop closure. This would account for the decrease in signal seen with both mutants. Also, since almost no change in fluorescent signal was observed with NC-2AP DNA and AdoHcy for both mutants, it suggests that the loop did not change position and was not in proximity to transfer energy to 2AP (Fig. 2*B* and *G*). Thus, the loop partially closes with NC-2AP DNA and AdoMet but not with AdoHcy and this partial closing allows for the retention of the catalytic function, since k_{chem} for NC-2AP DNA is only reduced 4-fold from COG DNA.

That DNA-dependent loop closure was not mediated through direct contacts from the loop to the DNA substrate (I86 to N^2) was very surprising; alternative paths of communication between the catalytic loop and the recognition domain are extremely difficult to identify and characterize. M.HaeIII, the only other enzyme from the cytosine C^5 DNA methyltransferase family with a co-crystal structure (24), has R227 contacting the guanosine directly above the target

cytosine in a very similar fashion to M.HhaI. While this similarity supports its importance, the lack of a detailed functional analysis of M.HaeIII precludes definitive comparisons.

For M.HhaI, two possible mechanisms seem reasonable for the major groove R240/guanosine interaction to control loop positioning. Either, i) R240 interacts with Q237, which contacts a loop residue, S87, and information passes directly through the DNA or, ii) it passes through the overall protein scaffold from the small recognition domain, across the active site, and towards the loop. In support of the first hypothesis, Q237 lies within the small DNA recognition domain and was previously implicated in modulating base flipping and loop closure (25). Q237 makes the only other hydrogen bond from the DNA to the loop (other than I86 to guanosine N^2) by contacting S87 which is positioned on the opposite side of the DNA in the catalytic loop. The interaction between R240 and the guanosine at position one, in addition to the proximity of R240 to Q237 and G236, provides a plausible connectivity to the catalytic loop (supplemental Figure 2). A similar structural arrangement was found in a DNA repair enzyme (MutY adenine glycosylase, 1VRL.pdb) (26). However, if Q237 mediates loop positioning, then unmethylated DNA would be expected to have altered kinetic rate constants from hemi-methylated DNA since the C^5 methyl group sterically clashes with the backbone carbonyl of Q237 and alters its orientation (27); no significant kinetic differences were observed (supplemental Table 2). Further, if the direct contact to the loop, (I86 to guanosine N^2) did not modulate loop positioning, it seems unlikely that it would instead go through Q237. Thus, a conclusive path between R240 and the catalytic loop which passes through the DNA could not be identified.

An alternative mechanism that could account for the communication between the small, recognition domain and the catalytic loop is through the entire protein scaffold (Fig. 1*A*). Evidence in support of this mechanism comes from SCA on the cytosine C^5 methyltransferase family which shows an allosteric network throughout the loop which traverses the active site and leads towards the recognition domain (supplemental Movie 1). SCA has been used to identify distal elements which must work in concert for binding and catalysis (28). Unfortunately, a complete path could not be identified because SCA is unable to examine the recognition domain due to the lack of sequence conservation within the cytosine C^5 methyltransferase family. This family of enzymes performs the same catalytic activity but on a wide ensemble of DNA sequences leading to high sequence conservation in the catalytic domain and low conservation in the recognition domain (29).

Also, the lack of loop reorganization with the NC 4-6 DNA substrates (data not shown) which have modifications at position four of the recognition sequence, further support this path of communication between the recognition domain and the catalytic loop (15). These non-cognate DNA substrates only make direct contacts to protein backbones which will most likely propagate structural changes over greater distances; side chains can absorb subtle changes in conformation leading to almost no impact on adjacent residues. Also, there is no obvious structural basis for how interactions involving position four of the recognition sequence might be propagated through Q237.

Supplemental References

22. Klimasauskas, S., Szyperski, T., Serva, S., and Wuthrich, K. (1998) *Embo Journal* **17,** 317-324
23. Lakowicz, J. R. (1999) *Principles of Fluorescence Spectropscopy*, 2nd Ed., Kluwer Academic/Plenum Publishers, New York
24. Reinisch, K. M., Chen, L., Verdine, G. L., and Lipscomb, W. N. (1995) *Cell* **82,** 143-153
25. Daujotyte, D., Serva, S., Vilkaitis, G., Merkiene, E., Venclovas, D., and Klimasauskas, S. (2004) *Structure* **12,** 1047-1055
26. Fromme, J. C., Banerjee, A., and Verdine, G. L. (2004) *Current Opinion in Structural Biology* **14,** 43-49
27. Ogara, M., Klimasauskas, S., Roberts, R. J., and Cheng, X. D. (1996) *Journal of Molecular Biology* **261,** 634-645
28. Lockless, S. W. and Ranganathan, R. (1999) *Science* **286,** 295-299
29. Kumar, S., Cheng, X. D., Klimasauskas, S., Mi, S., Posfai, J., Roberts, R. J., and Wilson, G. G. (1994) *Nucleic Acids Research* **22,** 1-10

Supplemental Figure and Movie Legends

SUPPLEMENTAL FIGURE 1. **Mechanism of M.HhaI methylation.** *A*, Nucleophillic attack of C81 onto the target cytosine to activate the C5 attack onto AdoMet. *B*, β-elimination regenerates the active site. *C*, The product, 5-methyl cytosine.

SUPPLEMENTAL FIGURE 2. **Kinetic scheme for the simulation.** Enzyme conformers (C1-C7), each having different fluorescence intensities (I1-I7), and forward and reverse rate constants (k1-k11) were minimized to fit the observed data.

SUPPLEMENTAL FIGURE 3. **Residues near R240.** a, Q237, G236, and R240 are shown in the binary complex (green), the nonspecific DNA complex (cyan), and the cognate DNA complex (pink). In the binary complex R240 only interacts with the carbon side chain of Q237 while in the cognate DNA complex it is interacting with the backbone carbonyl of G236. Q237 positioning sterically clashes with DNA a double helix and is skewed in the nonspecific DNA complex to allow room for the DNA. Q237 positioning appears dependent on R240; the backbone interaction to Q236 could easily modulate Q237 positioning which then modulates loop repositioning via the Q237/S87 interaction. b, A close-up of the interactions between I86, S87, R240, Q237, and the guanosine at position 1 of the recognition sequence. The flipped cytosine is in green. Recognition interactions between R240 and the guanosine could control loop positioning through Q237 and S87.

SUPPLEMENTAL MOVIE 1. **SCA network in M.HhaI.** The network is highlighted in orange and shown in sticks over an enzyme backbone in white, DNA backbone, cofactor, and flipped base in purple, and I86, S87, Q237, R240, and the guanosine at position 1 in cyan. The low sequence homology of the recognition domain and the dominant backbone interactions of the ß–sheets are absent in the SCA preventing complete identification of the network. However, the network travels all throughout the catalytic loop, across the active site, and towards the small recognition domain.

Supplemental Tables

SUPPLEMENTAL TABLE 1. **Kinetic and thermodynamic properties of enzymes studied with a variety of DNA substrates at 22 ºC.** N.D. is not determined.

M.HhaI	k_{cat} (ms^{-1})	K_M AdoMet (nM)	K_D MCOG DNA AdoHcy (nM)	K_D MCOG DNA AdoMet (nM)	K_D NS DNA (nM)	K_D NC-1 DNA (nM)	K_D NC-PUR DNA (nM)	K_D NC-2AP DNA (nM)	K_D NC-INO DNA (nM)
WT	55 +/- 4	138 +/- 18	47 +/- 29	77 +/- 7	776 +/- 99	268 +/- 27	51 +/- 8	18 +/- 3	11 +/- 3
W41F/K91W	42 +/- 5	1861 +/- 255	132 +/- 29	201 +/- 31	487 +/- 110	193 +/- 38	82 +/- 37	118 +/- 28	14 +/- 3
W41F/E94W	46 +/- 5	1266 +/- 113	64 +/- 30	420 +/- 47	451 +/- 56	339 +/- 128	180 +/- 65	29 +/- 5	13 +/- 2
W41F	50 +/- 6	1099 +/- 161	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.

SUPPLEMENTAL TABLE 2. **Rate constants of observed loop motion.** Magnitudes of each phase for biphasic rate constants are shown in parenthesis.
[a]: Rates are measured for increasing and decreasing signal. The first two values are biphasic initial loop closure with percent magnitude in parenthesis and the third value is loop re-opening.
[b]: Data not shown.
[c]: Initial rates (< 5 sec) gave a decrease in signal for both enzymes, convoluting a full analysis.

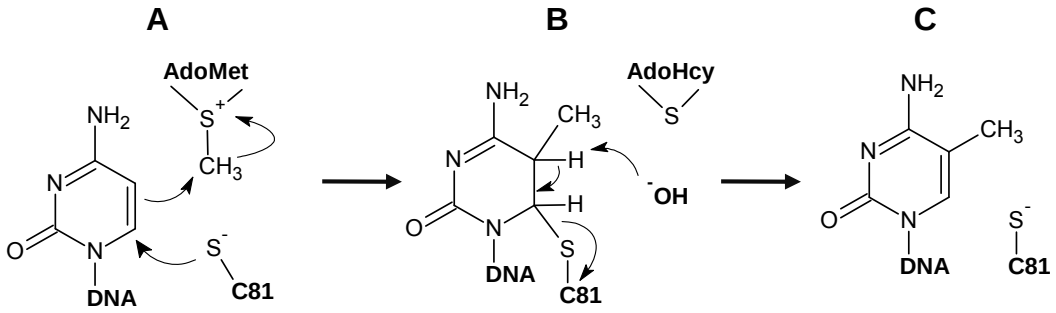
DNA / Cofactor	Loop Kinetics (s^{-1}) W41F/K91W	Loop Kinetics (s^{-1}) W41F/E94W
COG / AdoHcy	0.40+/- 0.07	1.17 +/- 0.09
COG / AdoMet[a]	37.6 +/- 3.1 (65)	42.4 +/- 3.5 (80)
	6.36 +/- 0.24 (35)	3.35 +/- 0.06 (20)
	0.092 +/- 0.005	0.084 +/- 0.004
UCOG / AdoHcy[b]	0.40 +/- 0.01	0.53 +/- 0.02
UCOG / AdoMet[ab]	19.7 +/- 0.7 (94)	33.5 +/- 0.97 (84)
	25.5 +/- 1.6 (06)	7.1 +/- 0.4 (16)
	0.064 +/- 0.002	0.062 +/- 0.012
MCOG / AdoHcy	0.98 +/- 0.15	1.35 +/- 0.05
MCOG / AdoMet	0.54 +/- 0.11	0.84 +/- 0.08
NC-PUR / AdoHcy	0.26 +/- 0.05	0.31 +/- 0.02
NC-PUR / AdoMet	1.97 +/- 0.16[b]	2.33 +/- 0.03
NC-INO / AdoHcy	0.55 +/- 0.13	0.82 +/- 0.14
NC-INO / AdoMet[a]	27.9 +/- 4.5 (52)[b]	31.3 +/- 1.2 (93)
	15.6 +/- 2.5 (48)[b]	1.5 +/- 0.3 (07)
	0.065 +/- 0.006[b]	0.085 +/- 0.001
NC-2AP / AdoHcy	N.D.	0.15 +/- 0.01
NC-2AP / AdoMet[c]	0.77 +/- 0.20[b]	1.09 +/- 0.15

SUPPLEMENTAL TABLE 3. **Kinetic and thermodynamic properties of WT and R240A M.HhaI at 22 ºC.**

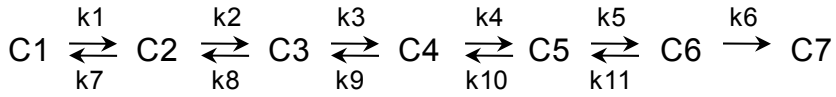
Parameter	WT	R240A
k_{chem} COG DNA (ms^{-1})	481 +/- 88	60 +/- 5
k_{chem} NC-PUR DNA (ms^{-1})	44 +/- 5	65 +/- 5
k_{chem} NC-2AP DNA (ms^{-1})	74 +/- 11	73 +/- 14
k_{chem} NC-INO DNA (ms^{-1})	88 +/- 20	32+/- 2
K_D COG DNA	17 +/- 11 pM	394 +/- 75 nM
K_D AB DNA	247 +/- 48 pM	40 +/- 22 pM

Supplemental Figures

SUPPLEMENTAL FIGURE 1



SUPPLEMENTAL FIGURE 2



SUPPLEMENTAL FIGURE 3