BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

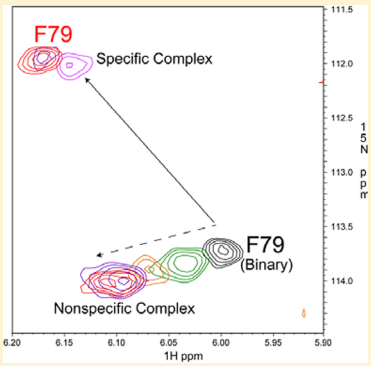
Enzyme-Promoted Base Flipping Controls DNA Methylation Fidelity

Douglas M. Matje,[†] Hongjun Zhou,[†] Darren A. Smith,[‡] Robert K. Neely,[§] David T. F. Dryden,[‡] Anita C. Jones,[‡] Frederick W. Dahlquist,[†] and Norbert O. Reich*[,†]

[†]Department of Chemistry and Biochemistry, University of California, Santa Barbara, California 93106, United States
[‡]EaStCHEM School of Chemistry, The University of Edinburgh, West Mains Road, Edinburgh EH9 3JJ, U.K.
[§]Department of Chemistry, KU Leuven, Celestijnenlaan 200F, 3001 Heverlee, Belgium

Supporting Information

ABSTRACT: A quantitative understanding of how conformational transitions contribute to enzyme catalysis and specificity remains a fundamental challenge. A suite of biophysical approaches was used to reveal several transient states of the enzyme−substrate complexes of the model DNA cytosine methyltransferase M.HhaI. Multidimensional, transverse relaxation-optimized nuclear magnetic resonance (NMR) experiments show that M.HhaI has the same conformation with noncognate and cognate DNA sequences. The high-affinity cognatelike mode requires the formation of a subset of protein−DNA interactions that drive the flipping of the target base from the helix to the active site. Noncognate substrates lacking these interactions undergo slow base flipping, and fluorescence tracking of the catalytic loop corroborates the NMR evidence of a loose, nonspecific binding mode prior to base flipping and subsequent closure of the catalytic loop. This slow flipping transition defines the rate-limiting step for the methylation of noncognate sequences. Additionally, we present spectroscopic evidence of an intermediate along the base flipping pathway that has been predicted but never previously observed. These findings provide important details of how conformational rearrangements are used to balance specificity with catalytic efficiency.



DNA repair and modification require enzymatic access to nucleobases through the rotation of the target base out of the helix and into the enzyme active site, often termed base flipping. This energetically challenging rearrangement provides a unique opportunity to study the role of specific substrate contacts and enzymatic motions in reaching a catalytically poised complex.[1−3] The C5 cytosine methyltransferase M.HhaI provided the first example of an enzyme to exploit base flipping.[1−3] Base flipping has since been recognized as a necessary strategy for nucleotide access converged upon by numerous enzyme classes, including all known DNA methyltransferases, enzymes that recognize misincorporated bases,[4,5] endonucleases,[6,7] RNA-modifying enzymes and non-catalytic transcription factors,[8] and 5-methylcytosine (5mC) recognition domains.[9,10] M.HhaI remains an excellent model for understanding enzymatic base flipping.[11] The high degree of structural homology between M.HhaI and the human DNMT family suggests that insights gained from the bacterial enzyme may contribute to the regulation of these biomedically important enzymes.

M.HhaI transfers a methyl group from *S*-(adenosyl)-L-methionine (AdoMet) to the central cytosine of the palindromic 5′ G<u>C</u>GC 3′ (cognate, $G^1C^2G^3C^4$) sequence. The cognate site is methylated with an up to 19000-fold preference over sequences with single-base pair changes (miscognate sites), and alterations in the methylation kinetics dominate this discrimination.[12] M.HhaI undergoes both minor and dramatic structural rearrangements upon binding to the cognate sequence, including compression of the recognition and catalytic domains, closure of a 20-residue catalytic loop to assemble the active site, and repositioning of side chains into the interhelical space vacated by the flipped base. Discrimination for the cognate sequence is governed by the bases surrounding the target cytosine, as the enzyme is able to flip any nucleoside within the 5′ G_GC 3′ context.[13] A further complication is that any cytosine brought into the active site can undergo methylation, suggesting that base flipping may regulate methylation fidelity. Although base flipping is widespread among diverse enzyme groups, the underlying mechanisms and their contributions to specificity remain unclear.

Systematic examination of the M.HhaI−DNA hydrogen bond network shows that individual hydrogen bonds differentially impact base flipping, catalytic loop closure, and assembly of the active site [ref 14 and the preceding paper (DOI: 10.1021/bi301284f)]. Here we use several complementary biophysical methods to reveal that contacts to the G^1 base pair provide a majority of the energy needed to effect base flipping and that bringing the target base into the active site acts as the energetic barrier to miscognate methylation. Thus, the specificity for modifying particular DNA sequences derives

Received: September 21, 2012
Revised: December 20, 2012
Published: February 14, 2013

largely from the base flipping transition, which most likely involves active participation on the part of the enzyme.

MATERIALS AND METHODS

AdoHcy was purchased from Sigma-Aldrich. Isotopically enriched compounds were ordered from Cambridge Isotopes Laboratories. All other reagents were purchased from Fisher Scientific unless noted.

DNA Preparation. Deoxyoligonucleotides used in this study (Table 1) were purchased in desalted form from IDT

Table 1. Duplex DNA Substrates Used in This Study

substrate	sequence[a]
GCGC	5′ CATGGCGCAGTG 3′
	3′ GTACCGMGTCAC 5′
ACGC	5′ CATGACGCAGTG 3′
	3′ GTACTGMGTCAC 5′
TCGC	5′ CATGTCGCAGTG 3′
	3′ GTACAGMGTCAC 5′
TXGC	5′ CATGTXGCAGTG 3′
	3′ GTACAGMGTCAC 5′

[a]Abbreviations: M, 5-methylcytosine; X, abasic site.

DNA, dissolved in NMR buffer [25 mM potassium phosphate (pH 6.5), 50 mM NaCl, 2 mM EDTA, 5 mM dithiothreitol, and 0.02% sodium azide] lacking DTT or FB, and annealed with the complementary strand by being heated to 65 °C and then slowly cooled to room temperature. DNA concentrations were determined using from the absorbance at 260 nm and molar extinction values calculated using the nearest neighbor method with the algorithm available at http://biophysics.idtdna.com using the Cavaluzzi–Borer correction.

Protein Expression and Purification. The previously described, solubility-enhanced Δ324 M.HhaI enzyme[15,16] was used for study of the wild-type enzyme. Appropriate deoxyoligonucleotides (IDT DNA) were used to introduce the desired mutations into a pET-28 plasmid containing the Δ324 M.HhaI gene by the QuikChange protocol (Stratagene). Mutant plasmids were transformed into ER1727 cells (New England Biolabs) for propagation of the plasmid and the genes sequenced at the University of California, Santa Barbara, Sequencing Facility to confirm the desired mutations were incorporated. Enzymes were expressed in T7Express^{Iq} *Escherichia coli* cells (NEB).

To generate ^{1}H- and ^{15}N-labeled enzyme, the previously described expression protocol,[15,16] based upon that of Marley,[17] was used with minor modifications. When A_{600} reached 0.8, 1 L of cells growing in LB medium was pelleted by centrifugation and suspended in 250 mL of M9 salts containing 0.5 g of [^{15}N]ammonium chloride, 1.5 g of glucose, 2 mM Mg_2SO_4, 0.1 mM $CaCl_2$, and Wolfe's minerals and vitamins in 95% D_2O. The resuspended cells were shaken at 37 °C for 1 h before the addition of IPTG to a final concentration of 1 mM. The cells were grown for ∼16 h at 28 °C before being pelleted and stored at −80 °C.

Enzyme purification was conducted essentially as described above except that the P-11 phosphocellulose column (Whatman) was run prior to the nickel affinity column, which was found to give a more stable enzyme, and DNAase and Triton X-100 were omitted without any consequence with regard to yield or purity. All enzymes were found to be >95% pure via densitometry analysis of sodium dodecyl sulfate–polyacrylamide gels of the purified enzyme fractions. Following dialysis into storage buffer,[15,16] purified enzyme was concentrated before being snap-frozen in LN_2 and stored at −80 °C. It was found that 5–20 mM imidazole in the storage buffer significantly improved enzyme stability in all downstream experiments. The enzyme concentration was determined using the previously calculated molar extinction coefficient of 25500 M^{-1} s^{-1} at 280 nm.[15,16] In preparation for NMR experiments, the concentrated enzyme was exchanged into NMR buffer via P-6 size-exclusion resin (Bio-Rad) equilibrated in NMR buffer. Generally, NMR spectra were collected with 100 μM enzyme, 500 μM AdoHcy, and 0, 100, or 300 μM DNA in NMR buffer.

Observation of Base Flipping. Time-resolved base flipping was monitored on an Applied Photophysics SX.18MV stopped-flow reaction analyzer operating in absorbance mode [refs 14 and 19 and the preceding paper (DOI: 10.1021/bi301284f)]. The source and detector were oriented to give a 10 mm path length for maximal absorbance at 280 nm. An enzyme and an AdoHcy solution in fluorescence buffer (FB) [25 mM potassium phosphate (pH 7.5), 100 mM NaCl, and 1 mM EDTA] were mixed against a solution of DNA in FB to give final concentrations of 2 μM enzyme, 50 μM AdoHcy, and 1.5 μM DNA. Changes in A_{280} were monitored by 1000 data collection points over 1, 10, or 1000 s. WT M.HhaI traces were fit to a three-step reversible model by numerical integration using KinTekSim. C81A data on GCGC were fit to the equation $Y = Y_0 + (Y_{Max} - Y_0)(1 - e^{-kt})$ using Prism 5 (GraphPad).

Tryptophan Reporter Fluorescence. Measurement of the degree of loop closure at equilibrium was conducted using an LB55 fluorimeter as described previously.[14,19,20] In the course of this research, it was discovered that the K91W reporter mutant has changes in the magnitude and rates of observed loop closure in response to cognate and noncognate DNA identical to those of the previously described W41F/K91W enzyme when the enzyme is saturated with AdoHcy. Therefore, the single mutant was used exclusively for the fluorescence experiments described in these studies. The K91W mutant was also observed to have essentially the same binary and ternary NMR spectra as the wild-type enzyme. Time-resolved measurements were taken on an Applied Photophysics SX.18MV apparatus as described previously.[14,20,21]

NMR Spectroscopy. All experiments were conducted at 25 °C either on a Bruker 800 MHz UltraShield Plus II NMR spectrometer equipped with a four-channel (^{1}H/^{13}C/^{15}N/^{2}H) cryoprobe or on a Varian Unity Inova 600 MHz spectrometer equipped with a four-channel (^{1}H/^{13}C/^{15}N/^{2}H) cold probe. The NMR sample contains 100–200 μM ^{15}N-labeled and deuterated M.HhaI in NMR buffer [25 mM potassium phosphate (pH 6.5), 50 mM NaCl, 2 mM EDTA, 5 mM DTT, and 0.02% sodium azide in a 90% H_2O/10% D_2O mixture]. In all DNA binding or titration experiments, two samples were made with identical protein concentrations: one sample without DNA and the other mixed with an excess of DNA. Appropriate volumes of the two samples were swapped to give the desired protein:DNA molar ratio at each titration point. The two-dimensional TROSY-HSQC[21] spectrum was collected, and the data were processed with nmrPipe[22] and ANSIG.[23]

Fluorescence Lifetime. Time-correlated single-photon counting was used to determine fluorescence lifetimes. Samples were measured in UV Quartz cells (Starna) that had a 10 mm path length. Fluorescence decay curves were recorded using an

$$E^{o}S_{M} + D^{h} \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} E^{o}S_{M}D^{d} \underset{k_{stack}}{\overset{k_{flip}}{\rightleftharpoons}} E^{o}S_{M}D^{f} \underset{k_{open}}{\overset{k_{close}}{\rightleftharpoons}} E^{c}S_{M}D^{f} \underset{k_{release}}{\overset{k_{attack}}{\rightleftharpoons}} E^{c}S_{M}D^{a} \xrightarrow{k_{methylation}} E^{c}S_{H}D^{5mC}$$

Figure 1. Precatalytic pathway of M.HhaI. Enzyme in the loop open form (E^o) binds with cofactor AdoMet (S_M) and with DNA where the target base is in the intrahelical state (D^h). The target cytosine is destabilized (D^d) by the development of sequence-specific contacts once the cognate site is located. The cytosine is then fully flipped from the helix (D^f), and the catalytic loop closes (E^c), which brings Cys81 into the proximity to initiate formation of a covalent bond with C5 of cytosine (D^a). Finally, methyl transfer generates 5-methylcytosine (5mC) and AdoHcy (S_H).

Edinburgh Instruments spectrometer (FL920) equipped with TCC900 photon counting electronics. The excitation source was a tunable, mode-locked Ti:Sapphire laser system (Coherent MIRA Ti:Sapphire laser pumped by a Coherent 10 W Verdi CW laser), producing approximately 200 fs pulses at a repetition rate of 76 MHz. A pulse picker (Coherent 9200) was used to reduce the pulse repetition rate to 4.75 MHz. A harmonic generator (Coherent 5−-050) was used to triple the frequency of the source light, to give an excitation wavelength of 295 nm. Fluorescence emission was detected orthogonal to the excitation beam through a polarizer set at the magic angle with respect to the vertically polarized excitation. A band-pass of 18 nm was used in the emission monochromator, and photons were detected using a cooled microchannel plate detector (Hamamatsu R3809 series). The instrument response function was ∼80 ps full width at half-maximum. A 320 nm long-pass filter Schott Glass (Newport) was used to block scattered light during lifetime measurements. Fluorescence decay curves were recorded on a time scale of 50 ns, binned into 4096 channels, to a total of 10000 counts in the peak channel. Decays were measured at emission wavelengths of 360, 380, and 400 nm. All measurements were taken at room temperature (20 °C).

Decay curves were analyzed by using a standard iterative reconvolution method in the F900 (Edinburgh Instruments Ltd.) and FAST (Edinburgh Instruments Ltd.) software packages. A multiexponential decay was assumed (eq 1).

$$I(t) = \sum_{i=1}^{n} A_i \quad \exp\left(\frac{-t}{\tau_i}\right) \tag{1}$$

where A_i and τ_i are the fractional amplitude and fluorescence lifetime, respectively, of the ith decay component and n is the total number of exponentials used in the fit. The fit quality was judged on the basis of the reduced χ^2 statistic and the randomness of the residuals. Typically, a χ^2 value of <1.1 indicated an acceptable fit. During global analysis, a family of decay curves was fitted simultaneously with lifetimes, τ_i, as common parameters. Average lifetimes, $\langle\tau\rangle$, were calculated via eq 2.

$$\langle\tau\rangle = \sum_{i=1}^{n} A_i\tau_i \tag{2}$$

All samples were prepared with the same concentration of enzyme (5 μM), substrate (8−9 μM) ,and AdoHcy (120 μM). These concentrations ensured complete formation of the ternary complex.

RESULTS

NMR Analysis of Enzyme−DNA Interactions. The precatalytic kinetic pathway of M.HhaI is shown in Figure 1. Starting from the transient, nonspecifically associated form of the enzyme with the catalytic loop open (E^o), we find cognate recognition results in several conformational rearrangements in both substrate and enzyme. The predominant form of the M.HhaI−AdoHcy−cognate DNA (ternary) complex observed crystallographically shows the target cytosine to be flipped from the helix and locked into the active site by closure of the catalytic loop (residues 78−98) and covalent addition of cysteine 81 to the target cytosine base ($E^cS_MD^a$ in Figure 1).[3,24,25] At the base of the loop, closure brings the side chain thiol of C81 into position to attack cytosine at the C6 atom to initiate electronic rearrangement and activate position 5 of the cytosine ring. The ternary crystal structures show the electron density of the side chain sulfur to be approximately halfway between a covalent bond and van der Waals distances.[24,25] Previous NMR investigations into the recognition pathway of M.HhaI tracked the slow chemical exchange and large chemical shift changes that occur throughout the enzyme as it undergoes the transition from the enzyme−AdoHcy (binary) complex to the cognate ternary complex.[15,16] In the presence of nonspecific DNA or DNA with a transversion in the recognition sequence (GCGA, miscognate), exchange between the bound and free forms was fast and the chemical shifts remained largely similar to those of the binary enzyme−AdoHcy complex.

At equimolar DNA and enzyme concentrations, the cognate ternary complex has one major set of $^1H-^{15}N$ correlation peaks corresponding to what was expected to represent a complex similar to that observed crystallographically. We used residual dipolar coupling[26] to confirm that the cognate ternary solution state observed in our NMR experiments is analogous to the $E^cS_MD^a$ crystal form. The M.HhaI−AdoHcy−cognate DNA complex was aligned in a mixture of alkyl-polyethylene glycol $C_{12}E_5$ and hexanol,[26] and the structure calculations show good agreement with the cognate ternary structure [Protein Data Bank (PDB) entry 2HRI] and poor correlation with the binary complex (PDB entry 1HMY), confirming the existence of homologous solution and crystal forms of the ternary complex (Figure 1 of the Supporting Information). Attempts to align miscognate complexes and the enzyme−cofactor binary complex in several different formulations of media only resulted in precipitation of the protein.

In cognate complexes with the wild-type (WT) enzyme, two sets of resonances are seen with a ratio of intensities of ∼1:5 for several loop residues. In mutant C81A, only the weaker resonances of the wild-type cognate spectra are observed, suggesting that these resonances correspond to a state in which Cys81 has not added to the cytosine base. Of particular note is Phe79 that sits at the edge of the catalytic loop where the backbone carbonyl forms a hydrogen bond with N4 of the target cytosine when it is flipped into the active site. Two adjacent amide resonances with a roughly 4:1 intensity ratio are observed for Phe79 in the WT ternary complex. In the C81A cognate complex, only one peak corresponding to the minor form is observed at full intensity (Figure 3 of the Supporting Information). Assignment of the minor peak to Phe79 was confirmed by the similarity of the cross-peak patterns in the set of three-dimensional assignment spectra correlating backbone amide and side chain carbon resonances. A similar pattern of

dx.doi.org/10.1021/bi3012912 | Biochemistry 2013, 52, 1677−1685

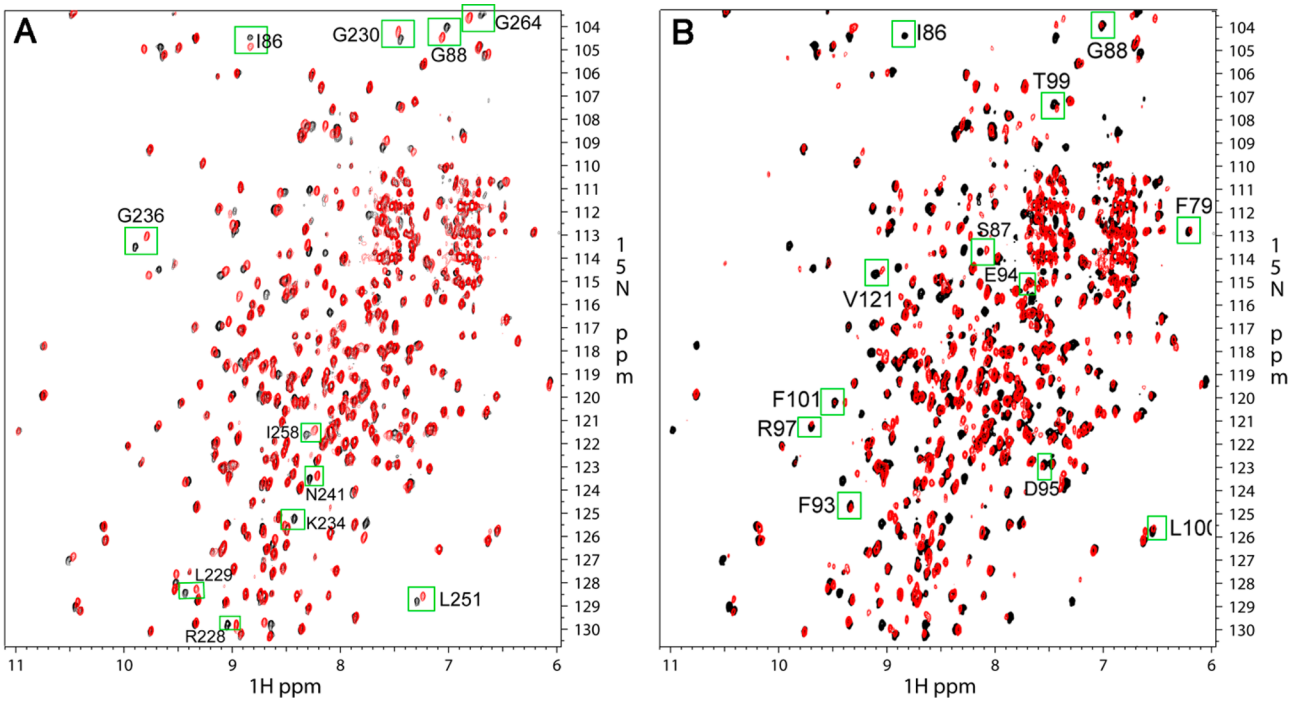


Figure 2. Similarity of the interactions of M.HhaI with miscognate and cognate DNA in ternary complexes involving AdoHcy. (A) With an excess (3-fold) of ACGC DNA (red), superposition of the spectra shows the bound spectrum of M.HhaI closely resembles that of the protein bound with cognate GCGC DNA (black). Because the chemical shift changes from the binary to ternary complex are drastic (up to 1700 Hz in combined ^{1}H and ^{15}N shift at a 600 MHz ^{1}H field), the overall similarity of the two complexes indicates the ACGC ternary complex is largely the same as with GCGC, that is, in a base-flipped, loop-closed form. A number of small differences are indicated with green boxes and residue numbers. Most of the residues come from the protein−DNA contact area in the recognition domain and a few (I86 and G88) from the active site loop region. (B) Comparison of $^{1}H-^{15}N$ correlation spectra of M.HhaI bound with a 3-fold excess of TCGC (red) and GCGC (black) substrates. As with ACGC, most residues from the active site loop (residues 80−100) show a weak resonance with TCGC matching the resonance from the complex with GCGC. The population of the specifically bound species with TCGC cannot be accurately estimated because of the contribution of unknown dynamics and exchange, but it is likely on the order of 10−30%. An excess of the cofactor is present in all experiments.

major and minor peak distribution that changes with the C81A mutant is observed for Phe93 and Arg97, which are involved in DNA recognition. We conclude that the major peaks in the WT complex represent the form with C81 covalently bound to the target cytosine C6, with the minor peaks representing the unattacked form of the closed loop (Figure 4 of the Supporting Information) in rough agreement with previous estimates.[25,26] On the basis of the ^{1}H and ^{15}N chemical shift differences of Phe79 in the two forms and the absence of noticeable exchange broadening of these resonances, we estimate the rate of exchange between the free and covalently modified flipped base is slower than 50 s^{-1}.

We sought to observe if M.HhaI is conformationally altered during the discrimination against miscognate methylation by comparing changes in the $^{1}H-^{15}N$ correlation spectra between the cognate substrate and two substrates where the first base pair of the recognition sequence was changed from G:C to A:T or T:A. The ACGC and TCGC substrates are disfavored 2800- and 19000-fold in k_{cat}/K_M compared to GCGC, respectively.[12] To our surprise, the M.HhaI backbone amide $^{1}H-^{15}N$ correlation spectra for both the ACGC and TCGC substrates showed clear evidence of a cognate ternary-like complex with large chemical shifts and slow chemical exchange. Saturation of the enzyme was ensured using a 3-fold excess of the miscognate substrates. For the M.HhaI−AdoHcy−ACGC complex, this resulted in a single set of peaks that are nearly identical to those observed in the GCGC complex. Subtle differences are seen in the positions of peaks that are close to the G^{1} base pair, likely reflecting minor differences in the environment around this base pair. However, given the magnitude of the chemical shift changes between the binary/nonspecific form and the cognate ternary complex,[15,16] the nearly identical cross correlation peaks demonstrate that the enzyme does not undergo significant structural adaptations to discriminate against miscognate methylation; instead, a kinetic barrier ensures fidelity.

Saturation of M.HhaI with a 3-fold excess of the TCGC substrate also results in a set of peaks representative of the $E^{c}S_{M}D^{a}$ form (Figure 1), but those resonances were significantly weaker and mixed with noncognate/binary-like peaks (Figure 2B). A large excess of the TCGC substrate, up to 10-fold over protein concentration, did not cause significant additional changes in the relative intensity of the peaks. The weaker set of peaks, with approximately 30% of the total population based on peak intensity, closely resemble those from the WT−cognate DNA complex. The stronger set of peaks, slightly shifted from the resonances of the binary complex, are similar to those observed in nonspecific complexes. Similar to the ACGC case, this apparent nonspecific complex undergoes fast exchange on the chemical shift scale with the binary form and the peaks characteristic of the specific complex undergo slow exchange with the other states. A set of weak peaks is observed at positions corresponding to cognate DNA binding for most residues in the active site loop and the enzyme core that interacts with the flipped cytosine. These residues have large chemical shift changes upon binding to

dx.doi.org/10.1021/bi3012912 | *Biochemistry* 2013, 52, 1677−1685

cognate DNA ranging from 300 to 1700 Hz in combined ^{1}H and ^{15}N perturbation measured at a 600 MHz ^{1}H field strength.

We were also able to observe a set of very weak peaks with chemical shifts similar to those of the C81A cognate complex in the TCGC complex. These are in slow exchange and appear to correspond to a base-flipped state but lack the addition of Cys81 to the cytosine. Taken together, these data suggest that the TCGC complex with the WT enzyme and SAH forms three distinct states at saturation. These correspond to an apparent nonspecific complex (70%) and a form very similar to the cognate complex with a mixture of the base-flipped state where Cys81 has added to the target cytosine (20%) and one where cysteine addition has not occurred (10%). These data suggest that covalent attachment of Cys81 to the target base does not contribute inordinately to stabilization of the ternary complex, representing a stabilizing factor of ~4 in the cognate and ACGC cases and a stabilizing factor of ~2 in the TCGC case.

With a 3-fold excess of TCGC DNA, we observed that peaks from several residues in the recognition loops, including Gly236, Leu251, Tyr254, and Phe259, all of which are at the enzyme–DNA interface, are broadened beyond detection. Peak broadening of these recognition residues is consistent with the existence of additional dynamics and chemical exchange due to both specific and nonspecific interactions involving these residues that we are presently investigating.

Our data show that the formation of the specific complex results from the competition between multiple and transient enzyme–DNA species. On the basis of previous and current NMR data, the degree of specificity of M.HhaI against different DNA sequences decreases as follows: GCGC > ACGC > TCGC > GCGA. However, all of the miscognate substrates are bound with approximately the same affinity, 100–500 nM, compared with 10 pM for the cognate substrate.[12] Therefore, we sought to better characterize the nature of the energetic barrier that prevents miscognate methylation.

Base Flipping as the Kinetic Determinant of Specificity. Formation of the cognate ternary complex in a stopped-flow apparatus allows for direct observation of flipping and C6 attack by the active site nucleophile Cys81, observed as a biphasic change in the optical density at 280 nm (Figure 3A, black) [refs 14 and 18 the preceding paper (DOI: 10.1021/bi301284f)]. The first phase corresponds to the hyperchromic increase in absorption at 280 nm when the GC base pair is disrupted by base flipping, and the subsequent decrease in absorption at 280 nm corresponds to the dearomatization associated with the addition of Cys81 to the flipped cytosine base [refs 14 and 19 and the preceding paper (DOI: 10.1021/bi301284f)]. Tracking base flipping in the ACGC and TCGC substrates by the WT enzyme (Figure 3A, red and blue) showed minimal changes in A_{280} over 100 s, in contrast to the 20 milli-absorbance unit (mAU) fluctuation seen with cognate DNA within 10 s. The ACGC and TCGC results are similar to the published results with a TCGA substrate.[19] However, these results do not distinguish between two possible scenarios: the inability of the enzyme to flip and/or maintain a sizable population of extrahelical target base or, alternatively, base flipping and extrahelical stabilization that are so slow that subsequent nucleophilic attack by Cys81 results in little accumulation of the intermediate flipped-but-not-attacked state ($E^{c}S_{M}D^{f}$). To distinguish these possibilities, we used the C81A mutant to prevent nucleophilic attack and dearomatization of the extrahelical cytosine, allowing observation of any

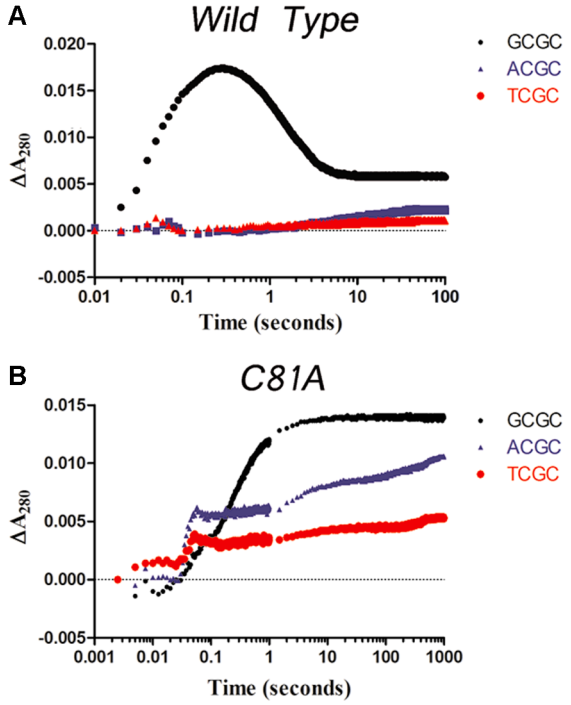


Figure 3. Observation of cytosine flipping by M.HhaI and Cys81Ala M.HhaI. (A) Wild-type M.HhaI with the three substrates used in this study. On cognate DNA (black), base flipping and nucleophilic attack are observed as a biphasic progression over 10 s. ACGC (blue) and TCGC (red) substrates show no accumulation of the flipped base. (B) The Cys81Ala mutant shows no decrease in A_{280} as expected for a mutant that is unable to covalently attack the flipped base. This allows direct observation of any accumulation of extrahelical cytosine in the miscognate substrates. The rapid increase that occurs for miscognate substrates at ~0.03 s is the first observation of a destabilized intermediate.

extrahelical cytosine that accumulates in complex with a miscognate substrate.

Mixing of the GCGC substrate with C81A and AdoHcy results in a monotonic increase in the 280 nm absorbance that does not subsequently decrease (Figure 3B, black traces). Unlike with the WT enzyme, both ACGC and TCGC show a clear accumulation of the base-flipped intermediate in the C81A mutant (Figure 4B, red and blue). With excess ACGC substrate, the increase in optical density through 1000 s is approximately 65% of that seen in the C81A–AdoHcy–GCGC complex, demonstrating a significant accumulation in the population of extrahelical cytosine at equilibrium. Similar to the NMR and fluorescence results (below), the ternary complex with the TCGC substrate results in modest (~33%) accumulation of a flipped target base at equilibrium, further demonstrating a dynamic equilibrium between the bound but unflipped ($E^{o}S_{M}D^{h}$, nonspecific mode) and flipped ($E^{c}S_{M}D^{f}$, cognate mode) states with this substrate.

In the presence of the C81A mutant, both the ACGC and TCGC substrates show polyphasic base flipping traces (Figure 3B). The massive reduction in the base flipping rate with these substrates also allows observation of an apparent intermediate state along the base flipping pathway that is not observed with the GCGC substrate. The first step occurs almost immediately after binding, resulting in a rapid, 5 mAU increase in optical density followed by a slow approach to equilibrium at a rate of 0.003 s^{-1} for ACGC and 0.001 s^{-1} for TCGC (compared to a

dx.doi.org/10.1021/bi3012912 | Biochemistry 2013, 52, 1677–1685

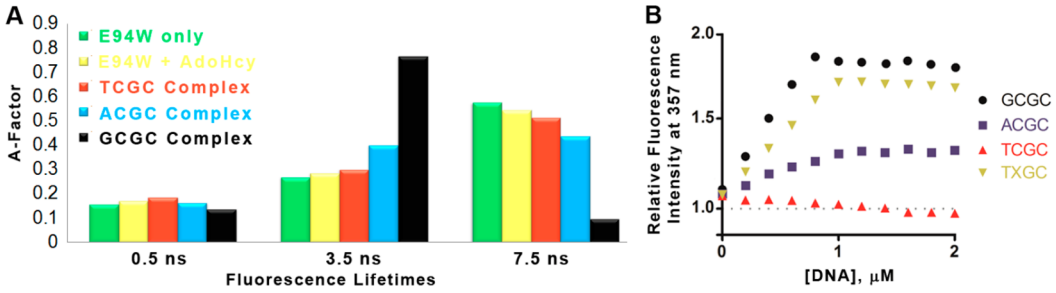


Figure 4. Tracking the position of the catalytic loop with two tryptophan reporter mutants. (A) Fluorescence decay measurements of the W41F/E94W reporter in complex with DNA substrates. The decay components (sTable 1 of the Supporting Information) reveal a distinct increase in the *A* factor of the 3.5 ns lifetime and a corresponding decrease in that of the 7.5 ns lifetime upon loop closure induced by the cognate substrate, GCGC, illustrated here for 380 nm decay data. [The E94W-only reporter gives similar results (sTable 2 of the Supporting Information).] This correlates with the decrease in the fluorescence intensity of the E94W reporter on loop closure. ACGC induces modest loop closure, while TCGC produces essentially no closure of the loop. (B) Equilibrium fluorescence titration of the K91W Trp reporter (1 μM) and AdoHcy (50 μM) with DNA substrates. Closure of the catalytic loop with the cognate substrate, GCGC, results in a large increase in the tryptophan quantum yield.[30,35] The miscognate substrate TCGC (red) causes no fluorescence increase. Binding to ACGC (blue) results in an increase in fluorescence that is approximately $^1/_3$ of that of the cognate substrate, indicating the bound, loop-open form is favored during early equilibrium When the target cytosine of the TCGC sequence is changed to an abasic site, TXGC (yellow), cognatelike closure of the loop is observed, indicating that flipping of the target base is the major energetic barrier to formation of a catalytically competent complex with miscognate substrates. All plotted points represent the average of at least five readings.

rate of 4.6 s^{-1} for the cognate substrate). This suggests at least two sequential steps in the base flipping process with these substrates, unlike the GCGC substrate that is fit well to a single exponential, implying one distinct rate-limiting step. While it is not possible to determine if noncognate and cognate substrates follow identical pathways, we propose the initial rapid increase in absorbance followed by a slow approach to equilibrium is best described by a conformational intermediate along the base flipping pathway where the target base is perturbed from the native helical position before complete rotation into the active site. On the basis of ^{19}F NMR experimental data of the target base as well as computational simulations,[27−29] this intermediate has been proposed to form following the initial development of sequence-specific contacts but has never before been observed for M.HhaI (see Discussion).

The dramatically slower flipping rates for ACGC and TCGC are similar to the corresponding k_{cat} values (0.0021 ± 0.0009 and 0.00082 ± 0.000032 s^{-1}, respectively),[12] suggesting that the rate-limiting step for methylation is the flipping process itself. Similarly, on the basis of the kinetic scheme in Figure 1, the lowered flipping rates account for the observed decreases in affinity of the enzyme for these sequences.[12] Furthermore, because we observe evidence of the flipped state with ACGC and TCGC, this suggests that the equilibrium for the flipped state with the cognate sequence dramatically (>10^4) favors the flipped state. Thus, both contributions to specificity, catalysis and substrate binding, are directly altered by the changes in flipping rates. Stivers and colleagues have argued that another base flipping enzyme acts by taking advantage of the intrinsic opening of U/A, U/G, or T/A base pairs and functions by slowing the rate of restacking of the base pair to stabilize the flipped out form.[30−32] Clearly, the dramatic slowing of flipping observed here cannot be the result of alterations in the restacking rate alone. The closest independent measure of base flipping is the NMR-based imino proton exchange (breathing),[33] although the direct relevance of such measurements to enzymatic base flipping mechanisms remains unclear. For example, it is not clear that base pair breathing[33] and base flipping[30−32] are on the same kinetic path and, if so, how the potentially minor conformational changes leading to breathing relate to the dramatic reorientation occurring during base flipping. Regardless, base pair breathing rates are uniformly 10−100 times faster than the flipping rates we observe for the WT M.HhaI and its cognate sequence.[33] This can still be reconciled by invoking a pre-equilibrium for the M.HhaI base flipping mechanism in which a rapid initial step, comparable to the breathing transition detected by NMR, is followed by a slow transition leading to the stabilization and positioning of the flipped base in the active site. With this mechanism, the second, slow transition may or may not be altered with noncognate sequences.

The Catalytic Loop Serves To Stabilize a Flipped Base and Assemble the Active Site Only. To further evaluate the role of base flipping as a specificity filter, we utilized two well-characterized reporters for loop positioning, the K91W and E94W mutants.[14,19,20] Closing of the loop upon binding to the GCGC substrate results in an increase (K91W) or decrease (E94W) in the quantum yield of the engineered tryptophans that does not occur with nonspecific DNA. Closure of the loop is essential for nucleophilic attack and methyl transfer but is likely impossible because of steric clash if the target cytosine remains in the helical position due to steric occlusion. Therefore, tracking the catalytic loop offers additional means of examining the nature of the M.HhaI−substrate complex.

Measurements of the tryptophan fluorescence decay were used to examine the populations of open and closed loop conformations in complexes with cognate, noncognate, and ACGC substrates. The E94W/W41F mutant was found to give the simplest fluorescence response, allowing quantitative determination of conformational populations (*A* factors). As shown in Figure 4A and sTable 1 of the Supporting Information, in the holo and binary enzyme−AdoHcy forms, the E94W side chain shows three lifetimes, a 7.5 ns component that accounts for 60% of the emitting population and two shorter components of 3.6 ns (27%) and 0.6 ns (13%). Such three-component decays are frequently seen for single-tryptophan proteins and indicate a heterogeneous environment in which the tryptophan side chain is subject to various quenching interactions. (The underlying cause of this complex fluorescence response cannot be readily characterized, and here

dx.doi.org/10.1021/bi3012912 | *Biochemistry* 2013, 52, 1677−1685

we simply employ the decay parameters as an empirical indicator of loop position.) Complexation with the cognate GCGC substrate leaves the values of the lifetimes essentially unchanged but results in a large decrease in the population of the 7.5 ns component, to 11%, with a concomitant increase in the population of the 3.6 ns component to 76%; the *A* factor of the shortest component is unchanged. As there is no evidence of an interconversion of the loop states in the absence of DNA, the decay parameters for the holo and cognate ternary states are expected to represent the extremes of completely loop-open and completely loop-closed, respectively. We interpret the large increase in A_2 and decrease in A_3 (Figure 4A) to be a signature of loop closure. Complexation with the noncognate TCGC substrate results in a negligible change in the *A* factors (or lifetimes) compared with those of the holo/binary states, indicating little stable closure of the catalytic loop at equilibrium with this substrate. The decay parameters for the ACGC complex show a significant change in A_2 and A_3, but much less than that seen for complete loop closure, indicating only modest closure of the catalytic loop with this miscognate substrate. The decay curve of this complex can be reproduced exactly by a linear combination of the decay functions of the loop-open and loop-closed states, indicating the presence of 80% loop-open and 20% loop-closed states. This suggests that the loop conformation can be described by a simple equilibrium between open and closed states and that sequences other than the cognate form do not cause the loop to take on a conformation distinct from closed and open.

The decay parameters reported here are quantitatively consistent with previously reported fluorescence intensity measurements.[14,19,20] Moreover, the E94W and E94W/W41F mutants are observed to have essentially identical decay signals when the enzyme is saturated with cofactor (see sTable 2 of the Supporting Information). This demonstrates that the fluorescence of the native tryptophan is efficiently quenched by cofactor binding and signal changes in the ternary complexes of the E94W mutant are solely from the tryptophan engineered into the loop.

To test the hypothesis that maintenance of the cytosine in the extrahelical state is the major specificity filter for M.HhaI, the base of the target cytosine was removed from the TCGC substrate to leave only the phosphate backbone and ribose ring (an abasic site, X). Substrates containing cognate sequences with an abasic site at the target cytosine are bound more tightly by M.HhaI[34] and cause significantly faster closure of the loop, which was previously interpreted to demonstrate that base flipping is the major energetic barrier prior to chemistry.[34] We monitored the position of the catalytic loop with the TXGC substrate and the K91W reporter so that flipping or stabilization of the target base by the enzyme was not necessary for loop closure. As predicted, near-complete closure of the loop was observed when the TXGC substrate was complexed with the K91W reporter, demonstrating for the first time complete closure of the catalytic loop on a substrate containing a transversion in the recognition sequence (Figure 4B). The results also suggest that flipping of the target base into the active site is the major specificity filter against noncognate methylation by the methyltransferase.

DISCUSSION

Single-base pair changes to the recognition sequence of M.HhaI are observed to cause position-dependent variations in M.HhaI–substrate interactions. Among the miscognate DNA sequences examined (ACGC, TCGC, and, previously, GCGA), ACGC is found to have the highest affinity and strongest ability to be brought into the specific binding mode. Such interactions are greatly weakened with TCGC and essentially nonexistent with GCGA or a random DNA sequence. With GCGA and nonspecific DNA, all interactions with the enzyme are weakened, and only weak interactions at the enzyme–DNA binding interface could be detected as representing the nonspecific binding mode. The results presented here indicate a graded, rather than binary, transition from nonspecific to specific interaction modes, perhaps similar to the structural snapshots observed in crystal structures of T4DAM.[35] There it was observed that recognition interactions with one-fourth of the cognate sequence that did not promote base flipping preceded the formation of contacts to the base pairs that promoted base flipping. The *E. coli* DAM has also been suggested to actively promote base flipping, with individual side chain–nucleobase interactions contributing differentially to base flipping and binding.[36] Active participation by flipping enzymes in overcoming the energetically costly base flipping step provides a formidable kinetic barrier that prevents miscognate methylation.

Our base flipping results with the miscognate substrates reveal glimpses of a long-postulated intermediate along the base flipping pathway. Previous molecular dynamics simulations of the base flipping pathway and ^{19}F monitoring of the target cytosine in M.HhaI complexes have suggested the existence of a preflipped state in which the target base is destabilized at ~15° from the native helical position prior to full rotation into the active site.[27−29] The rapid but modest increase in absorbance between 10 and 100 ms seen in Figure 3B is suggestive of the reduction in hypochromicity expected by slight displacement of a base. The intermediate also accumulates within the time regime of miscognate association. This association is sufficient to result in the subtle distortion of the DNA helix from the native B-form state that is observed in the crystal structure and driven by enzyme–DNA complementarity. Juxtaposition of the catalytic and recognition domains serves to subtly separate the strands to reduce the energetic barrier of breaking the interbase hydrogen bonds. The further development of enzyme–substrate complementarity, particularly along the phosphate backbone and the two interactions between the two N^{ζ} atoms of Arg240 and the O6 and N7 atoms of G^1, serves to further lower the barrier to flipping. These findings also explain the loss of catalytic burst previously observed with miscognate substrates,[12] as base flipping supplants product release as the rate-limiting step during multiple turnovers. Recently, human UDG was crystallized in a form that trapped a uracil analogue in a position ~30° from the native state, described as an "early intermediate" along the pathway that results from enzymatic destabilization of the native state.[32] Similar to M.HhaI, UDG utilizes a mobile loop to intercalate into the space vacated by the target base to ensure maintenance of the flipped base without providing significant binding energy from substrate-specific interactions.[30−32] However, UDG does not utilize contacts to the surrounding bases for binding or flipping energy, instead only discriminating against nonsubstrates such as thymine in the active site after flipping.

Base flipping by M.HhaI is a clear example in which binding energy provided by contacts separate from the chemical target is directly translated into substrate reorganization needed to overcome the major energetic barrier and control fidelity. Our results show that M.HhaI is able to destabilize the target base

via contacts at the $C^2G^3C^4$ base pairs, but even after prolonged incubation, the equilibrium lies in favor of the nonflipped form with the TCGC substrate. Thus, the reduction in ground state binding energy and the inability to overcome the flipping barrier favor dissociation from miscognate sites with a variation at position G^1. In contrast, contacts to the G^3C^4 base pairs serve to prevent association with miscognate sites in the first place. These results agree with the conservation of elements involved in recognizing the G^1C^2 pairs, including Arg240 and Thr250, while there is little conservation among the G^3C^4 recognition residues. This modularity in the role of substrate contacts has allowed enormous diversification of DNA methyltransferases in accepting thousands of substrates via a common mechanism, with conservation of the base flipping elements allowing drift of recognition elements toward new cognate sequences.

■ ASSOCIATED CONTENT

Supporting Information

Additional NMR spectra, further explanation of the fluorescence lifetime results, and complete materials and methods. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: reich@chem.ucsb.edu. Phone: (805) 893-8368. Fax: (805) 893-4120.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Cheng, X., and Blumenthal, R. M. (1996) Finding a basis for flipping bases. *Structure 4* (6), 639−645.
(2) Roberts, R. J., and Cheng, X. D. (1998) Base flipping. *Annu. Rev. Biochem. 67*, 181−189.
(3) Klimasauskas, S., et al. (1994) Hhal methyltransferase flips its target base out of the DNA helix. *Cell 76* (2), 357−369.
(4) Parikh, S. S., et al. (1998) Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. *EMBO J. 17* (17), 5214−5226.
(5) Guan, Y., et al. (1998) MutY catalytic core, mutant and bound adenine structures define specificity for DNA repair enzyme superfamily. *Nat. Struct. Biol. 5* (12), 1058−1064.
(6) Horton, J. R., et al. (2006) DNA nicking by HinP1I endonuclease: Bending, base flipping and minor groove expansion. *Nucleic Acids Res. 34* (3), 939−948.
(7) Tsutakawa, S. E., et al. (2011) Human flap endonuclease structures, DNA double-base flipping, and a unified understanding of the FEN1 superfamily. *Cell 145* (2), 198−211.
(8) Schubot, F. D., et al. (2001) Crystal structure of the transcription factor sc-mtTFB offers insights into mitochondrial transcription. *Protein Sci. 10* (10), 1980−1988.
(9) Hashimoto, H., et al. (2008) The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. *Nature 455* (7214), 826−829.
(10) Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y., and Shirakawa, M. (2008) Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature 455* (7214), 818−821.
(11) Cheng, X. D. (1995) Structure and Function of DNA Methyltransferases. *Annu. Rev. Biophys. Biomol. Struct. 24*, 293−318.
(12) Youngblood, B., Buller, F., and Reich, N. O. (2006) Determinants of Sequence-Specific DNA Methylation: Target Recognition and Catalysis Are Coupled in M.HhaI. *Biochemistry 45* (51), 15563−15572.
(13) Vilkaitis, G., Dong, A., Weinhold, E., Cheng, X., and Klimašauskas, S. (2000) Functional Roles of the Conserved Threonine 250 in the Target Recognition Domain of HhaI DNA Methyltransferase. *J. Biol. Chem. 275* (49), 38722−38730.
(14) Matje, D. M., and Reich, N. O. (2012) Molecular drivers of base flipping during sequence-specific DNA methylation. *ChemBioChem 13* (11), 1574−1577.
(15) Zhou, H., Purdy, M. M., Dahlquist, F. W., and Reich, N. O. (2009) The recognition pathway for the DNA cytosine methyltransferase M.HhaI. *Biochemistry 48* (33), 7807−7816.
(16) Zhou, H., Shatz, W., Purdy, M. M., Fera, N., Dahlquist, F. W., and Reich, N. O. (2007) Long-range structural and dynamical changes induced by cofactor binding in DNA methyltransferase M.HhaI. *Biochemistry 46* (24), 7261−7268.
(17) Marley, J., Lu, M., and Bracken, C. (2001) A method for efficient isotopic labeling of recombinant proteins. *J. Biomol. NMR 20* (1), 71−75.
(18) Gerasimaite, R., et al. (2011) Direct observation of cytosine flipping and covalent catalysis in a DNA methyltransferase. *Nucleic Acids Res. 39* (9), 3771−3780.
(19) Estabrook, R. A., and Reich, N. (2006) Observing an Induced-fit Mechanism during Sequence-specific DNA Methylation. *J. Biol. Chem. 281* (48), 37205−37214.
(20) Estabrook, R. A., Nguyen, T. T., Fera, N., and Reich, N. O. (2009) Coupling Sequence-specific Recognition to DNA Modification. *J. Biol. Chem. 284* (34), 22690−22696.
(21) Pervushin, K., Riek, R., Wider, G., and Wuthrich, K. (1997) Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. U.S.A. 94*, 12366−12371.
(22) Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) Nmrpipe: A Multidimensional Spectral Processing System Based on Unix Pipes. *J. Biomol. NMR 6* (3), 277−293.
(23) Kraulis, P. J. (1989) Ansig: A Program for the Assignment of Protein H-1 2d-Nmr Spectra by Interactive Computer-Graphics. *J. Magn. Reson. 84* (3), 627−633.
(24) Horton, J. R., Ratner, G., Banavali, N. K., Huang, N., Choi, Y., Maier, M. A., Marquez, V. E., MacKerell, A. D., Jr., and Cheng, X. (2004) Caught in the act: Visualization of an intermediate in the DNA base-flipping pathway induced by HhaI methyltransferase. *Nucleic Acids Res. 32*, 3877−3886.
(25) O'Gara, M., Klimasauskas, S., Roberts, R. J., and Cheng, X. (1996) Enzymatic C5-cytosine methylation of DNA: Mechanistic implications of new crystal structures for HhaI methyltransferase-DNA-AdoHcy complexes. *J. Mol. Biol. 261*, 634−645.
(26) Lipsitz, R. S., and Tjandra, N. (2004) Residual dipolar couplings in NMR structure analysis. *Annu. Rev. Biophys. Biomol. Struct. 33*, 387−413.
(27) Klimasauskas, S., Szyperski, T., Serva, S., and Wuthrich, K. (1998) Dynamic modes of the flipped-out cytosine during HhaI methyltransferase-DNA interactions in solution. *EMBO J. 17* (1), 317−324.
(28) Huang, N., and MacKerell, A. D., Jr. (2004) Atomistic View of Base Flipping in DNA. *Philos. Trans. R. Soc., A 362*, 1439−1460.
(29) Huang, N., Banavali, N. K., and MacKerell, A. D., Jr. (2003) Protein-facilitated base flipping in DNA by cytosine-5-methyltransferase. *Proc. Natl. Acad. Sci. U.S.A. 100* (1), 68−73.
(30) Stivers, J. T., and Kivie, M. (2004) *Site-Specific DNA Damage Recognition by Enzyme-Induced Base Flipping*, Vol. 77, pp 37−65, Academic Press, San Diego.
(31) Jiang, Y. L., and Stivers, J. T. (2002) Mutational analysis of the base-flipping mechanism of uracil DNA glycosylase. *Biochemistry 41* (37), 11236−11247.
(32) Parker, J. B., Bianchet, M. A., Krosky, D. J., Friedman, J. I., Amzel, L. M., and Stivers, J. T. (2007) Enzymatic capture of an extrahelical thymine in the search for uracil in DNA. *Nature 449*, 433−437.
(33) Coman, D., and Russu, I. M. (2005) A nuclear magnetic resonance investigation of the energetics of basepair opening pathways in DNA. *Biophys. J. 89*, 3285−3292.

(34) O'Gara, M., Horton, J. R., Roberts, R. J., and Cheng, X. (1998) Structures of HhaI methyltransferase complexed with substrates containing mismatches at the target base. *Nat. Struct. Biol.* 5 (10), 872−877.
(35) Yang, Z., Horton, J., Zhou, L., Zhang, X., Dong, A., Zhang, X., Schlagman, S., Kossykh, V., Hattman, S., and Cheng, X. (2003) Structure of bacteriophage T4 adenine methyltransferase. *Nat. Struct. Biol. 10*, 849−855.
(36) Horton, J. R., Liebert, K., Bekes, M., Jeltsch, A., and Cheng, X. (2006) Structure and Substrate Recognition of the *Escherichia coli* DNA Adenine Methyltransferase. *J. Mol. Biol. 358*, 559−570.