SURVEY AND SUMMARY

AdoMet-dependent methylation, DNA methyltransferases and base flipping

Xiaodong Cheng* and Richard J. Roberts1

Emory University School of Medicine, Department of Biochemistry, 1510 Clifton Road, Atlanta, GA 30322, USA and
1New England Biolabs, 32 Tozer Road, Beverly, MA 01915, USA

Received February 5, 2001; Revised and Accepted March 19, 2001

ABSTRACT

Twenty AdoMet-dependent methyltransferases (MTases) have been characterized structurally by X-ray crystallography and NMR. These include seven DNA MTases, five RNA MTases, four protein MTases and four small molecule MTases acting on the carbon, oxygen or nitrogen atoms of their substrates. The MTases share a common core structure of a mixed seven-stranded β-sheet (6↓7↑5↓4↓1↓2↓3↓) referred to as an ‘AdoMet-dependent MTase fold’, with the exception of a protein arginine MTase which contains a compact consensus fold lacking the antiparallel hairpin strands (6↓7↑). The consensus fold is useful to identify hypothetical MTases during structural proteomics efforts on unannotated proteins. The same core structure works for very different classes of MTase including those that act on substrates differing in size from small molecules (catechol or glycerol) to macromolecules (DNA, RNA and protein). DNA MTases use a ‘base flipping’ mechanism to deliver a specific base within a DNA molecule into a typically concave catalytic pocket. Base flipping involves rotation of backbone bonds in double-stranded DNA to expose an out-of-stack nucleotide, which can then be a substrate for an enzyme-catalyzed chemical reaction. The phenomenon is fully established for DNA MTases and for DNA base excision repair enzymes, and is likely to prove general for enzymes that require access to unpaired, mismatched or damaged nucleotides within base-paired regions in DNA and RNA. Several newly discovered MTase families in eukaryotes (DNA 5mC MTases and protein arginine and lysine MTases) offer new challenges in the MTase field.

INTRODUCTION

Methyl transfers are alkylation reactions central to cellular biochemistry, and S-adenosyl-L-methionine (AdoMet) is by far the most commonly used methyl donor molecule. The AdoMet methyl group is bound to a charged sulfur atom, which thermodynamically destabilizes the molecule and makes the relatively inert methylthiol of the methionine moiety very reactive (1) toward polarizable nucleophiles (N, O and S) and activated C atoms (carbanions). The AdoMet-dependent methyltransferases (MTases) act on a wide variety of target molecules, including DNA, RNA, proteins, polysaccharides, lipids and a range of small molecules. Since 1993, 20 of these enzymes have been characterized structurally, almost all with bound AdoMet or the reaction product S-adenosyl-l-homocysteine (AdoHcy) and some with bound substrates.

In prokaryotes, over 3000 restriction–modification systems have been discovered so far, and they have been found across the full spectrum of known bacterial species, including both eubacteria and archaea (see http://rebase.neb.com/) (2). While the functions and various types of restriction–modification systems are discussed in this issue and elsewhere (for example, reviewed in 3), structures for six type II MTases are currently available: two 5-methylcytosine (5mC), one N4-methylcytosine (N4mC) and three N6-methyladenine (N6mA) MTases (Table 1).

In mammals, recent publications clearly define, in addition to Dnmt1 (4), two additional distinct phylogenetic 5mC MTase families, Dnmt2 (5–7) and Dnmt3a and Dnmt3b (8). All of the enzymes preserve the same basic organization of the conserved MTase motifs in a C-terminal catalytic domain that resembles the prokaryotic 5mC MTases (reviewed in 9–11). Less is known about the enzymology and structure of the recently described enzymes, but considering that methylation in eukaryotes is found predominantly in the sequence CpG or CpNpG, it is likely that the eukaryotic enzymes have similar sequence specificities. However, embryonic stem cells have significant 5mC residues at CpA and, to a lesser extent, at CpT (12). The unique features of the individual 5mC MTase families are found in the large N-terminal domain that varies in size between the different MTase families and is likely responsible for the regulation of their diverse biological functions—in the normal development of animal (13,14) and plant species (15–17), gene repression (18), X chromosome inactivation (19), genome imprinting (20) and replication timing (21). The tasks of dissecting the functional roles of the N-terminal region of the different 5mC MTase families are just beginning. How genome methylation patterns are established during development and how the altered function of these enzymes contributes to the

*To whom correspondence should be addressed. Tel: +1 404 727 8491; Fax: +1 404 727 3746; Email: xcheng@emory.edu
aberrant methylation changes that accompany embryonic development (13,14) and disease (14,22,23) is presently unknown.

While some proteins exert their effects simply through binding interactions, other proteins both bind to DNA and catalyze chemical reactions. These include polymerases, nucleases, glycosylases, MTases and various integrases and recombinases that rearrange DNA segments. DNA binding frequently deforms the usual B-helix and bending and kinking of DNA is common. Proteins that perform chemistry on the DNA bases have a difficult accessibility problem. This problem was resolved in 1994 by the discovery of ‘base flipping’—when a structure was reported for the ternary complex of the 5mC DNA MTase, M. HhaI, its DNA substrate and the methyl donor AdoMet (24). It was proposed that other classes of DNA MTases and some DNA glycosylases might also use base flipping to gain access to DNA bases (24). Both predictions proved accurate.

The purpose of this review is to provide a brief overview of the known structures of AdoMet-dependent MTases, particularly DNA MTases, and summarize current knowledge about base flipping, including all systems in which it is proven to occur. Finally, we will focus on the available structures of the HhaI MTase that has served as a structural paradigm for DNA MTases and DNA base flipping enzymes.

**AN AdoMet-DEPENDENT MTase FOLD**

At this time, the structural characterization of 20 AdoMet-dependent MTases has been reported (Table 1). These include DNA MTases (two generating 5mC, one generating N4mC and three generating N6mA), RNA MTases (two closely related enzymes generating N6mA in rRNA, one generating mRNA cap-specific 2′-O-methylribose), protein MTases generating glutamyl carboxymethyl ester, L-isoaspartate methyl ester and methylarginine, and small molecule MTases that act on catechol, glycine, chalcone and isoflavone. Of the known MTase structures, two transfer the methyl group to a carbon atom, six to oxygen atoms and nine to nitrogen atoms (Table 1). The substrates for human DNMT2, Methanococcus jannaschii fibrillarin homolog and Escherichia coli FtsJ are unknown.

A striking feature of all the structures is that they share a common core structure referred to as an ‘AdoMet-dependent MTase fold’ (Figs 1–4). Many of these proteins have domains
outside the core structure that play a role in substrate recognition or in separate functions. Central to this shared core structure is a mixed seven-stranded $\beta$-sheet (green in Figs 1–4). Strand 7 is antiparallel to the other six strands, and is inserted into the sheet between strands 5 and 6 ($\downarrow\downarrow\uparrow\downarrow\downarrow\downarrow\downarrow\downarrow\downarrow\downarrow\downarrow$); this feature is thus far characteristic of the AdoMet-dependent MTase fold.

Figure 1. Examples of DNA MTases (see Table 1). M. HhaI (5mC), M. HaeIII (5mC), human DNMT2, M. DpnII (N6mA, $\alpha$), M. PvuII (N4mC, $\beta$), M. RsrI (N6mA, $\beta$) and M. TaqI (N6mA, $\gamma$). The AdoMet-dependent MTase fold is colored in green ($\beta$ strands), cyan ($\alpha$ helices) and red (the loops after the carboxyl ends of $\beta$ strands). The region(s) outside the MTase fold is colored in gray.

Figure 2. Examples of RNA MTases. VP39 (mRNA nucleoside-2$'$-O), ErmC$'$ or closely related ErmAM (rRNA N6mA), FtsJ and fibrillarin homolog.

Figure 3. Examples of protein MTases. CheR (glutamate-O), PRMT3 (arginine-N), Hmt1 (arginine-N) and PIMT (isoaspartate-O).
MTases, with one exception. The protein arginine (R) MTase (PRMT) lacks the anti-parallel β-hairpin (Δ4–7) (Fig. 3) (25). In addition, the protein isoaspartyl MTase (PIMT) has the order of strands 6 and 7 reversed in the primary sequence (26). It is possible that the insertion of the antiparallel strand 7 between strands 5 and 6 plays an important functional role in some MTases. In HhaI DNA MTase the loop between strands 6 and 7 stabilizes both the substrate (cytosine) and the DNA-recognition domain. In a simplified version, strands 1–3 form the part of the MTase that interacts with the AdoMet, while strands 4–7, particularly the loop after strand 4, which usually contains the catalytic amino acids, are in the region binding the diverse substrates methylated by these enzymes.

The high degree of structural similarity among AdoMet-dependent MTases is not reflected by a corresponding degree of sequence conservation (27). In fact, only three positions are highly conserved among the structurally-characterized enzymes (27): a Gly (motif I) in the loop after strand 1, a negatively-charged Asp/Glu (motif II) in the carboxyl end of strand 2 and a hydrophobic Val/Ile/Leu (motif IV) within strand 4. Only one of these (motif I) is obvious from the sequence alone without structural guidance (28), though substantial conservation can exist within a particular subfamily of MTase (defined according to their reactions and substrates, such as the DNA 5mC MTases).

Structural proteomics
While the lack of sequence-level conservation has made it difficult to determine relationships among the various AdoMet-dependent MTases, the high degree of conservation of the MTase fold has served as a prototype to identify hypothetical MTases of unannotated proteins. In one of the first structural proteomics efforts (http://www-kimgrp/lbl.gov/genomics/proteinlist.html), the crystal structure of the fibrillarin homolog from M. jannaschii (Mj0697) revealed a MTase-like C-terminal domain (Fig. 2) (29). Fibrillarin is a phylogenetically conserved protein essential for efficient processing of pre-rRNA through its association with a class of small nucleolar RNAs (snoRNAs) during ribosomal RNA biogenesis (30). A majority of fibrillarin-associated snoRNAs function in rRNA 2′-O-methylation within a base-paired duplex region. In addition, a structure of Mj0882 deposited in PDB (1 dus) has the MTase fold (31).

Other examples are the two new structures of proteins (HI0319 and HI0766) from Haemophilus influenzae (http://s2f.umbi.umd.edu). HI0319 is homologous to some MTases acting on small molecules and has a core structure with a MTase fold and a bound AdoHcy (32). HI0766 is a truncated version of SpoU, a MTase acting on tRNA (33). It is worth noting that AdoMet concentrations in E.coli were reported to be in the 300–500 μM range (34) and copurification of AdoMet has been noted for many DNA MTases expressed in E.coli (35–37). Bound AdoMet or AdoHcy has been found in almost all (except yeast Hmt1; 38) structurally characterized known MTases, no matter whether extra cofactor was added or not during crystallization.

Circular permutation of DNA MTases
An analysis of the family of DNA 5mC MTases revealed 10 conserved amino acid sequence motifs (39) and a target-recognition domain(s) (TRD) (40). Nine of these motifs were also found to occur in the three different linear orders (families α, β and γ) in the DNA amino MTases (41). Currently, we have examples for each family (Table 2). The 5mC (M.HhaI and M.HaeIII) and the γ (M.TaqI) MTases differ only in the placement of one helix and its associated conserved motif X (respectively at the C- and N-termini of the protein) (42). In the β MTases M.PvuII and M.RsrI, the N-terminus is just upstream of strand 3 (43,44), which means that the conserved motifs are in a simple permuted order relative to the γ MTases (45). The α MTase M.DpnII has a large insertion (TRD) between strands 2 and 3 (46). Despite the altered linear order of the sequence motifs and secondary structure elements, they are still able to yield the same tertiary structure and maintain the core MTase fold.

DNA BASE FLIPPING
It might seem surprising that the same core structure works for a very different class of MTases that act on substrates differing in size from small molecules (catechol or glycine) to protein and DNA. The solution to this apparent puzzle is that macro-molecule MTases do not methylate protein or DNA per se, but methylate a specific base within a DNA molecule or a specific amino acid within a protein. This raises the problem of

Figure 4. Examples of small molecule MTases: COMT (catechol-O), GNMT (glycine-N), and IOMT (isoflavone-O) or closely related ChOMT (chalcone-O).
substrate accessibility. How can the MTase deliver the target to its catalytic center? One elegant solution found by the DNA MTases is shown in Figure 5A, and was originally discovered for HhaI MTase (24). In the protein–DNA complex, the target cytosine is no longer buried within the double helix, but has been rotated on its flanking sugar–phosphate bonds so that it projects out into a typically concave catalytic pocket. No covalent bonds were broken to carry out the base flipping process, however, the base pairing hydrogen bonds were broken and the stacking \( \pi \) interactions with the adjacent base pairs were lost.

There are still a number of things we do not understand about base flipping, including how it is initiated and how (or if) it is related to the recognition of the substrate sequence. Two interesting features are known, however. First, in the structure of M.HhaI with a DNA substrate having an abasic (apurinic/apyrimidinic or AP) site at the position of the target cytosine, the enzyme still moves the sugar–phosphate backbone to the ‘flipped out’ position (Fig. 5A) (47). A similar conformation is also observed for the flipped-out abasic nucleotide in four glycosylase–DNA complexes: uracil DNA glycosylase (UDG) (48) (Fig. 5B), mismatch-specific uracil glycosylase (MUG) (49,50), alkyladenine glycosylase (AAG) (51) (Fig. 5C) and alkylation glycosylase (AlkA) (52) (Fig. 5E). It thus appears that the base \( \text{per se} \) is not the target for the structural change in the DNA. Since an abasic sugar is flipped by M.HhaI, AAG and AlkA, UDG and MUG, there is clearly nothing special about the cytosine, 3-methyladenine, uracil, thymine or any of the bases that is required for flipping. Thus, we conclude that it is the backbone that is targeted for rotation by the enzyme and the base is merely carried along with it (53).

Secondly, base flipping by M.HhaI will work with guanine and uracil (47) or probably any base at the target position (54,55). The methylation reaction is sensitive to the base at the target position, but the base flipping step is not. This property has been used to measure base flipping because 2-aminopurine fluorescence increases dramatically when it is removed from the stacking environment in double helical DNA (56). Examples include M.EcoRI (57,58), M.TaqI and M.HhaI (59).

<table>
<thead>
<tr>
<th>Family</th>
<th>Motif order (linear)</th>
<th>Strand order (linear)</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>1 (AdoMet)-TRD-IV (DPPY)</td>
<td>1-2-TRD-3-4-5-6-7</td>
<td>M.DpnII (N6mA)</td>
</tr>
<tr>
<td>( \beta )</td>
<td>IV (DPPY)-TRD-1 (AdoMet)</td>
<td>3-4-5-6-7-TRD-1-2</td>
<td>M.RsrII (N6mA) and M.PvuII (N4mC)</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>1 (AdoMet)-IV (NPPY)-TRD</td>
<td>1-2-3-4-5-6-7-TRD</td>
<td>M.TaqI (N6mA)</td>
</tr>
<tr>
<td>5mC</td>
<td>1 (AdoMet)-IV (PC)-TRD</td>
<td>1-2-3-4-5-6-7-TRD</td>
<td>M.HhaI and M.HaeIII (5mC)</td>
</tr>
</tbody>
</table>

Figure 5. Examples of DNA base flipping proteins (see Table 3) in complex with oligonucleotide containing an abasic site. (A) M.HhaI (a DNA 5mC MTase), (B) human UDG, (C) human AAG, (D) E.coli endonuclease IV, (E) E.coli AlkA and (F) human HAP1. The protein is colored gray, the DNA is represented as a magenta stick model with the flipped abasic site in green, usually buried in a surface pocket in the protein.
However, the practice of this technique needs some precautions. Large changes in 2-aminopurine fluorescence have been observed in *M. Eco* P15I (60) and *M. Eco* RV (62) for substrates which do not carry the substitution at the site of methylation/flipping. On the other hand, small or no changes have been observed in *M. Eco* RV (62) and *M. Rsr* I (63) for substrates which do carry the substitution at the site of flipping/methylation. Therefore, changes in 2-aminopurine fluorescence should not be regarded as definitive proof of base flipping but as evidence for conformational or environmental changes of the DNA which lead to unstacking of the 2-aminopurine base. Base flipping is merely one example of such a change.

Several other examples of base flipping have now been observed in structures of protein–DNA complexes (Table 3).

### Table 3. Known base-flipping systems

<table>
<thead>
<tr>
<th>Specific protein</th>
<th>Catalytic reaction</th>
<th>PDB</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DNA MTases</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>M. Hha</em>I</td>
<td>Forms 5-methylcytosine in DNA</td>
<td>9mht</td>
<td>24,47</td>
</tr>
<tr>
<td><em>M. Hae</em>III</td>
<td>Forms 5-methylcytosine in DNA</td>
<td>1dct</td>
<td>64</td>
</tr>
<tr>
<td><em>M. Taq</em>I</td>
<td>Forms N6-methyladenine in DNA</td>
<td>1G38</td>
<td>65</td>
</tr>
<tr>
<td><strong>DNA glycosylases</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T4 endonuclease V</td>
<td>Removes pyrimidine dimers from DNA</td>
<td>1vas</td>
<td>78</td>
</tr>
<tr>
<td>Human UDG</td>
<td>Removes uracil from DNA</td>
<td>2ssp</td>
<td>48,71,135</td>
</tr>
<tr>
<td><em>E. coli</em> MUG</td>
<td>Removes uracil or thymine from DNA containing G:T or G:U</td>
<td>49,50</td>
<td></td>
</tr>
<tr>
<td>Human AAG</td>
<td>Removes 3-methyladenine from DNA</td>
<td>1bnk</td>
<td>51</td>
</tr>
<tr>
<td><em>E. coli</em> AlkA</td>
<td>Removes 3-methyladenine from DNA</td>
<td>1diz</td>
<td>52</td>
</tr>
<tr>
<td>hOGG1</td>
<td>Remove 8-oxoguanine from DNA</td>
<td>1ebm</td>
<td>74</td>
</tr>
<tr>
<td><strong>AP endonucleases</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>E. coli</em> endonuclease IV</td>
<td>Cleaves the DNA backbone 5’ of AP sites</td>
<td>1qum</td>
<td>79</td>
</tr>
<tr>
<td>Human AP endonuclease (HAP1 or APE1)</td>
<td>Cleaves 5’ to AP sites</td>
<td>1dew</td>
<td>80</td>
</tr>
</tbody>
</table>

*The coordinates of MUG–DNA complexes are currently not available in PDB.*

![Figure 6](image-url). 30S ribosomal subunit flipped-out A1492 and A1493 from helix 44 of 16S RNA by binding of (left) paromomycin (seen in the difference electron density) and (right) initiation factor IF1 (in purple). The protein S12 is in orange, helix H44 in cyan.

*M.Eco*P15I (60) and *E. coli* UDG (61). However, the practice of this technique needs some precautions. Large changes in 2-aminopurine fluorescence have been observed in *M. Eco*P15I (60) and *M. Eco*RV (62) for substrates which do not carry the substitution at the site of methylation/flipping. On the other hand, small or no changes have been observed in *M. Eco*RV (62) and *M. Rsrl* I (63) for substrates which do carry the substitution at the site of flipping/methylation. Therefore, changes in 2-aminopurine fluorescence should not be regarded as definitive proof of base flipping but as evidence for conformational or environmental changes of the DNA which lead to unstacking of the 2-aminopurine base. Base flipping is merely one example of such a change.

Several other examples of base flipping have now been observed in structures of protein–DNA complexes (Table 3). In the structure of *M. Hae*III bound to DNA, the substrate cytosine is flipped out from the DNA helix (64), as observed for *M. Hha*I. The *M. Hae*III structure is, so far, unique in having some rearrangement of the bases adjacent to the flipped cytosine. Recently, the structure of a ternary complex between *M. Taq*I, an adenine-N6-specific DNA MTase, and its two substrates, the specific DNA and a non-reactive AdoMet analog, revealed a flipped adenine (65). However, the three available MTase–DNA complexes use three different means of stabilizing DNA structure with a flipped base (66). The latest and most amazing observations on base flipping came from structure of the 30S ribosomal subunit complexed with the antibiotics paromomycin (67) or initiation factor IF1 (68). Paromomycin binds in the major groove of helix H44 of 16S RNA and flips out the functionally important bases A1492 and A1493; the flipped-out bases point directly into the ribosomal A site and interact with the minor groove of the codon–anticodon duplex (Fig. 6, left). Binding of IF1 occludes the A site and flips out A1492 and A1493 and buries them in pockets in IF1 (Fig. 6, right).

### DNA repair enzymes

Besides DNA MTases, base flipping has since been found in a variety of DNA repair enzymes (53,69) that function through either the base excision repair pathway (see references in
Structures of sylases that remove the damaged base by cleaving the glyco-
3-methyladenine DNA glycosylase II (AlkA) induces a hairpin–helix DNA glycosylase–DNA complexes have revealed the base-binding pocket. The enzyme’s active site. In the structure of the modified DNA complex, the non-hydrolyzable substrate analog is flipped out of the DNA helix. Further work is needed to settle the issue.

Escherichia coli MUG. MUG removes pyrimidines from mismatches opposite guanine arising from deamination of cytosine (to uracil) or 5-methylcytosine (to thymine). Structures of E.coli MUG have been obtained for an abasic–DNA product complex (49) and a non-hydrolysable deoxyuridine analog (50). The crystal structures show great similarity to UDG, especially around the active site. In the structure of the native DNA complex, the glycosidic bond of the uracil base has been hydrolyzed and a flipped-out abasic site is left in the enzyme’s active site. In the structure of the modified DNA complex, the non-hydrolyzable substrate analog is flipped out and bound in the base-binding pocket.

HhaI DNA glycosylase. The crystal structures of two helix–hairpin–helix DNA glycosylase–DNA complexes have revealed a common structural feature of base flipping. Escherichia coli 3-methyladenine DNA glycosylase II (AlkA) induces a 66° bend in the DNA with a marked widening of the minor groove and flips a 1-azaribose abasic nucleotide out of the DNA. The crystal structures of a helix–hairpin–helix DNA glycosylase–DNA complexes have revealed a common structural feature of base flipping. Escherichia coli 3-methyladenine DNA glycosylase II (AlkA) induces a 66° bend in the DNA with a marked widening of the minor groove and flips a 1-azaribose abasic nucleotide out of the DNA. The crystal structures of a helix–hairpin–helix DNA glycosylase–DNA complexes have revealed a common structural feature of base flipping. Escherichia coli 3-methyladenine DNA glycosylase II (AlkA) induces a 66° bend in the DNA with a marked widening of the minor groove and flips a 1-azaribose abasic nucleotide out of the DNA. The crystal structures of a helix–hairpin–helix DNA glycosylase–DNA complexes have revealed a common structural feature of base flipping. Escherichia coli 3-methyladenine DNA glycosylase II (AlkA) induces a 66° bend in the DNA with a marked widening of the minor groove and flips a 1-azaribose abasic nucleotide out of the DNA. The crystal structures of a helix–hairpin–helix DNA glycosylase–DNA complexes have revealed a common structural feature of base flipping. Escherichia coli 3-methyladenine DNA glycosylase II (AlkA) induces a 66° bend in the DNA with a marked widening of the minor groove and flips a 1-azaribose abasic nucleotide out of the DNA. The crystal structures of a helix–hairpin–helix DNA glycosylase–DNA complexes have revealed a common structural feature of base flipping. Escherichia coli 3-methyladenine DNA glycosylase II (AlkA) induces a 66° bend in the DNA with a marked widening of the minor groove and flips a 1-azaribose abasic nucleotide out of the DNA. The crystal structures of a helix–hairpin–helix DNA glycosylase–DNA complexes have revealed a common structural feature of base flipping. Escherichia coli 3-methyladenine DNA glycosylase II (AlkA) induces a 66° bend in the DNA with a marked widening of the minor groove and flips a 1-azaribose abasic nucleotide out of the DNA. The crystal structures of a helix–hairpin–helix DNA glycosylase–DNA complexes have revealed a common structural feature of base flipping. Escherichia coli 3-methyladenine DNA glycosylase II (AlkA) induces a 66° bend in the DNA with a marked widening of the minor groove and flips a 1-azaribose abasic nucleotide out of the DNA. The crystal structures of a helix–hairpin–helix DNA glycosylase–DNA complexes have revealed a common structural feature of base flipping. Escherichia coli 3-methyladenine DNA glycosylase II (AlkA) induces a 66° bend in the DNA with a marked widening of the minor groove and flips a 1-azaribose abasic nucleotide out of the DNA. The crystal structures of a helix–hairpin–helix DNA glycosylase–DNA complexes have revealed a common structural feature of base flipping. Escherichia coli 3-methyladenine DNA glycosylase II (AlkA) induces a 66° bend in the DNA with a marked widening of the minor groove and flips a 1-azaribose abasic nucleotide out of the DNA. The crystal structures of a helix–hairpin–helix DNA glycosylase–DNA complexes have revealed a common structural feature of base flipping. Escherichia coli 3-methyladenine DNA glycosylase II (AlkA) induces a 66° bend in the DNA with a marked widening of the minor groove and flips a 1-azaribose abasic nucleotide out of the DNA. The crystal structures of a helix–hairpin–helix DNA glycosylase–DNA complexes have revealed a common structural feature of base flipping. Escherichia coli 3-methyladenine DNA glycosylase II (AlkA) induces a 66° bend in the DNA with a marked widening of the minor groove and flips a 1-azaribose abasic nucleotide out of the DNA. The crystal structures of a helix–hairpin–helix DNA glycosylase–DNA complexes have revealed a common structural feature of base flipping. Escherichia coli 3-methyladenine DNA glycosylase II (AlkA) induces a 66° bend in the DNA with a marked widening of the minor groove and flips a 1-azaribose abasic nucleotide out of the DNA. The crystal structures of a helix–hairpin–helix DNA glycosylase–DNA complexes have revealed a common structural feature of base flipping. Escherichia coli 3-methyladenine DNA glycosylase II (AlkA) induces a 66° bend in the DNA with a marked widening of the minor groove and flips a 1-azaribose abasic nucleotide out of the DNA. The crystal structures of a helix–hairpin–helix DNA glycosylase–DNA complexes have revealed a common structural feature of base flipping. Escherichia coli 3-methyladenine DNA glycosylase II (AlkA) induces a 66° bend in the DNA with a marked widening of the minor groove and flips a 1-azaribose abasic nucleotide out of the DNA. The crystal structures of a helix–hairpin–helix DNA glycosylase–DNA complexes have revealed a common structural feature of base flipping. Escherichia coli 3-methyladenine DNA glycosylase II (AlkA) induces a 66° bend in the DNA with a marked widening of the minor groove and flips a 1-azaribose abasic nucleotide out of the DNA. The crystal structures of a helix–hairpin–helix DNA glycosylase–DNA complexes have revealed a common structural feature of base flipping. Escherichia coli 3-methyladenine DNA glycosylase II (AlkA) induces a 66° bend in the DNA with a marked widening of the minor groove and flips a 1-azaribose abasic nucleotide out of the DNA. The crystal structures of a helix–hairpin–helix DNA glycosylase–DNA complexes have revealed a common structural feature of base flipping. Escherichia coli 3-methyladenine DNA glycosylase II (AlkA) induces a 66° bend in the DNA with a marked widening of the minor groove and flips a 1-azaribose abasic nucleotide out of the DNA. The crystal structures of a helix–hairpin–helix DNA glycosylase–DNA complexes have revealed a common structural feature of base flipping. Escherichia coli 3-methyl adenine DNA glycosylase (AAG) that removes 3-methyladenine and a wide variety of other damaged bases from DNA. Lau et al. (51) reported a structure for AAG complexed to DNA containing a pyrrolidine abasic nucleotide, a potent inhibitor of excision repair glycosylases, which is flipped into the enzyme’s active site (Fig. 5C).

T4 endonuclease V. T4 endonuclease V is a DNA glycosylase/AP lyase that can initiate repair of cis–syn cyclobutane pyrimidine dimers in DNA by cleaving the glycosidic bond of the 5′ pyrimidine and then cleaving the phosphodiester backbone. T4 endonuclease V does not flip out the damaged bases, but rather it kinks the dimer-containing DNA, at an angle of ~60°, and moves the nucleotide opposite the 5′ pyrimidine of the dimer into a binding pocket on the surface of the enzyme (78). The key feature associated with endonuclease V base flipping is that the hole in DNA, created by movement of the nucleotide, is filled by the enzyme inserting its active site amino acids into that hole. Thus, through a change in the structure of the DNA the enzyme is correctly positioned to carry out a nucleophilic attack on the 5′ pyrimidine of the dimer.

AP endonucleases

Escherichia coli endonuclease IV is the hydrolytic AP endonuclease. The structure of the enzyme complexed with AP-containing DNA revealed that the protein loops intercalate side chains at the abasic site, compress the DNA backbone, bend the DNA, and promote base flipping of both the AP site and its mismatched guanine (79). The extrahelical AP site is bound in an enzyme pocket and the target scissile phosphate bond is cleaved (Fig. 5D). A flipped abasic nucleotide also occurs in another structurally unrelated AP endonuclease, human APE1 (80) (Fig. 5F). In addition, E.coli exonuclease III (81) has a structure similar to human APE1 and will very likely bind to DNA and flip the abasic nucleotide in a manner similar to APE1.

HHAI DNA MTase

HhaI MTase, a 327-residue protein, methylates the internal cystine of its recognition sequence 5′-GCGC-3′,CGCG-5′ (82,83). Crystal structures of M.HhaI are available in various combinations with cofactor and DNA (Table 4), and represent a rich structural paradigm for AdoMet-dependent MTases and DNA base flipping enzymes. The structural information also provides the basis for studies of this enzyme using other techniques: NMR (84), 2-aminoaridine fluorescence (59), a chemical probe for thymine in DNA (85), molecular dynamics (86), DNA π-electron transfer (87), biochemistry and kinetics (88), and mutagenesis (89,90).

Structures are available for M.HhaI bound to AdoMet (91,92), non-covalently bound to all three states of DNA methylation (unmethylated, hemimethylated and fully-methylated) (93,94), base pair mismatch-containing DNA (47) and nucleoside analog-containing DNA (24,95,96). The protein has two lobes, a conserved AdoMet-dependent catalytic core domain and a DNA-recognition domain (91), which is common to all structurally-characterized DNA MTases, M.HaeIII (64), M.TaqI (97), M.PvuII (43), M.DpnII (46) and M.RsrI (44).
Three states of DNA methylation

An important aspect of DNA methylation is the ability of DNA MTases to distinguish DNA substrates with methyl groups on one strand (hemimethylated DNA) from those which carry no methyl groups. In bacteria, many of the type II enzymes such as M.HhaI are equally active on unmethylated and hemimethylated DNA, but with an increased affinity for asymmetrically methylated DNA [M.EcoRI (98), M.MspI (99), M.HaeIII (100), M.EcoRV (37), M.RsrI (63)]. In mammals, human ‘maintenance’ MTase Dnmt1 shows a preference for hemimethylated DNA in vitro (101). Although mouse Dnmt3 may represent ‘the long sought after de novo MTase’ (8,14), recombinant Dnmt3 in vitro methylates unmethylated and hemimethylated DNA at equal rates (8).

The three M.HhaI structures (3mht contains unmodified C, 4mht contains symmetrically methylated C and 5mht contains a hemimethylated substrate) are rather similar, with one of the target nucleotides (only the unmethylated target C in 5mht) flipped out of the DNA helix and fitting snugly into the active site of the enzyme. On the complementary strand, the target C (methylated or unmethylated) remains stacked in the DNA helix. Both the N4 (NH2) group and the methyl group of 5mC (methylated or unmethylated) remains stacked in the DNA helix. Both the N4 (NH2) group and the methyl group of 5mC (methylated or unmethylated) remains stacked in the DNA helix. Both the N4 (NH2) group and the methyl group of 5mC (methylated or unmethylated) remains stacked in the DNA helix.

In 4mht the 5mC residue in the fully methylated oligonucleotide, which is the reaction product, is flipped out of the DNA helix in the same manner as with the unmodified target C in 3mht and 5mht. This is surprising but consistent with biochemical data, which suggest that the binding specificity for M.HhaI is asymmetric 5′-GXC-G-3′/3′-CGC-G-5′ and determined by the nucleotides neighboring the target (X) nucleotide (54). In other words, the MTase does not depend on the flippable base for its binding specificity.

Nucleoside analogs

The fact that HhaI MTase does not show much binding specificity for the flippable base may reflect a need to leave that base unencumbered by recognition contacts. This provides an opportunity to probe the structural and chemical interactions involved in sequence-specific recognition and catalysis using nucleotide analogs incorporated into synthetic oligonucleotides in the position of the target nucleotide. One prominent analog is 2-aminopurine, which has been used extensively to probe base flipping (see above).

There is also potential clinical significance in designing novel DNA MTase inhibitors, which may be used to reverse the effects of DNA methylation (102). The effects include mutagenicity caused by the spontaneous deamination of 5mC to T and promoter hypermethylation which causes gene silencing and is believed to be a problem in many cancers. Three nucleotide analogs/inhibitors have been used in structural studies including a 4′-thionucleoside (95) and two mechanism-based inhibitors of 5mC MTases, namely 5-fluoro-2′-deoxyctydine (103,104) and 5,6-dihydro-5-azacytosine (a hydrolytically stable replacement for 5-azacytosine) (105).

A complex containing 5-fluoro-2′-deoxyctydine as the target base was first crystallized in the presence of AdoMet, which resulted in the formation of a covalent linkage between the S atom of the enzyme nucleophile Cys81 and the C6 atom of cytosine generating 5-methyl-5-fluoro-2′-deoxydihydroctydine and AdoHcy (24). Interestingly, 5mC RNA MTases contain sequence motifs that are very similar to those found in the DNA 5mC MTases, including the Pro–Cys, but use a different Cys as a catalytic nucleophile (106).

When 4′-thio-2′-deoxyctydine is incorporated as the target base in the recognition sequence for M.HhaI, binding to the 4′-thio-methylated DNA is almost identical to that of the unmodified DNA under equilibrium conditions (95). In contrast, methyl transfer was strongly inhibited in solution. Surprisingly, the flipped 4′-thio-2′-deoxyctydine in the 6mht crystal structure was partially methylated, the result of a much slowed reaction that allowed direct visualization of the methyl transfer between the donor atom (the AdoMet S) and the acceptor atom (the activated cytosine C5) (see figure 9 of ref. 95).

5,6-Dihydro-5-azacytosine (DHAC) contains a cytosine-like ring lacking aromatic character with an sp3-hybridized carbon (CH3 group) at position 6 and an NH group at position 5, thus resembling the transition state of the dihydroctydine intermediate in the reaction pathway of the 5mC DNA MTases (see...
figure 2 of ref. 107). The 10mH structure, containing DHAC as the target, showed that DHAC occupies the active site of M.HhaI but with no covalent bond formed between the sulfur atom of nucleophile Cys81 and the pyrimidine C6 carbon (96). This result indicates that the DHAC-containing DNA, behaving as a transition-state mimic, is sufficient to produce strong inhibition of DNA MTases.

Mismatched bases

Following the discovery of base flipping by M.HhaI, the effects of replacing the target cytosine by mismatched bases, including adenine, guanine, thymine and uracil, were investigated (54,55). Many DNA MTases were found to bind even more tightly to such mismatched substrates, including human Dnmt1 (108), M.HpaII (55), M.EcoRV (109), M.EcoRII (110), T4 Dam (111), M.EcorI (58), M.EcorI241 (112) and M.EcoP15I (60). Furthermore, the uracil can be enzymatically methylated and converted to thymine at low efficiencies (54,55); and the binding of these MTases at the G:U mismatch prevented its repair by uracil DNA glycosylase binding of these MTases at the G:U mismatch prevented its repair by uracil DNA glycosylase in vitro (55). The highest affinity of M.HhaI was for a gap formed by removal of the target nucleotide and both phosphodiester linkages (54).

Three ternary structures (7mht, 8mht and 9mht) of M.HhaI complexed with AdoHcy and a non-palindromic oligonucleotide containing a G:A, G:U or G:AP mismatch at the target base pair, respectively, have been determined (47). The mismatched adenine, uracil and abasic site are flipped out and located in the enzyme’s active site, respectively. It seems likely that this active site pocket is non-specific for binding, but specific for methylation. In the light of the non-specific binding pocket, the DNA MTase may be more related to the repair enzymes such as AAG that have broad substrate specificity. On the other hand, the methylation reaction is specific in that catalysis occurs only when the flipped base is cytosine or uracil (at low efficiency), more closely resembling the behavior of MUG.

M.HhaI–AdoMet binary complexes

Two distinct binding orientations of AdoMet have been observed in the binary M.HhaI–AdoMet (91,92), suggesting that the enzyme–AdoMet complex does form but lacks catalytic competence in the absence of DNA. These findings were supported by pre-steady-state partitioning analysis of M.HhaI (88). This reconciles the proposed ordered mechanism of binding in which the DNA is bound first (107) and explains the stable association between AdoMet and M.HhaI (36). It is interesting to note the differences in the order of substrate binding between the 5mC MTases and the amino MTases examined so far. 5mC MTases M.HhaI (88,107), M.MspI (113) and mouse Dnmt1 (114) follow a sequential mechanism with the DNA binding first. In Wu and Santi’s ordered mechanism (107), after methyl transfer, AdoHcy dissociates followed by methylated DNA. In support of this ordered mechanism, the structure of M.HaeIII (another 5mC MTase) does not have AdoMet or AdoHcy bound, but the C5 position of the target 5-fluoro-2’-deoxycytosine carries both the fluorine and the methyl group (64), indicating that methyl transfer has taken place and M.HaeIII had diffused out of the enzyme–DNA complex. The amino MTases have all three possibilities: M.EcoRV (α), M.RsrI (β) and M.EcorI (γ) all bind AdoMet first (62,63,98,115), CcrM (an orphan MTase) binds DNA first (116), and M.EcoP15I (a type III MTase) and M.EcoI (β) both have a random equilibrium mechanism (117,118).

M.HhaI–DNA binary complex

The binary M.HhaI–DNA complex has been studied by NMR (84). The solution study showed that the initial product of binding is a complex containing normal B-DNA. Only later does a conformational change take place that results in the flipped base in an intermediate position (or a series of intermediates). The intermediate position is neither in its original stacked position in the DNA helix nor in its final position in the catalytic pocket of the enzyme. The addition of AdoHcy greatly enhances the locking of the flipped-out base into the enzyme’s active site pocket (the methyl transfer turnover would take place with the addition of AdoMet). The two-state model of flipping has been used to fit the kinetic data for M.EcorI (N6mA), M.HhaI (5mC) and M.EcoRV (N6mA) (58,62,90).

PROSPECTIVES

As we move into the new millenium, we are in a fast moving world of genomics and proteomics. There is a tremendous amount of work to be done to understand the binding, catalytic and kinetic mechanisms of both existing and newly identified families of AdoMet-dependent MTases. Computer searches of newly completed genomes show an abundance of new, putative MTases including many for which substrates cannot even be guessed. There is a dire need for more biochemistry to characterize this large family of enzymes. Even in the cases of MTases for which structures are known, only a few have the target of methylation bound. The phenomenon of base flipping is appearing in a number of DNA repair enzymes following its initial discovery in the DNA MTases, and many additional enzymes are likely candidates for use of this novel mechanism. It remains to be proven whether base flipping is an active process in which the protein rotates the sugar–phosphate backbone and brings the base out of the helix or a passive one in which the protein binds to a transiently flipped base.

ACKNOWLEDGEMENTS

We thank Dr Alan M. Friedman for providing PMT coordinates, Drs Mark A. Saper and Ursula Jakob for FTSJ coordinates, Drs Dick Gumport and Mair Churchill for M.RsrI coordinates, Dr Osnat Herzberg for communicating results before publication, and Drs Venki Ramakrishnan and Andrew Carter for providing images for Figure 6. Work in our laboratories is supported by grants from the National Institutes of Health (GM46127 to R.J.R. and GM61355 to X.C.).

REFERENCES


