Base eversion and shuffling by DNA methyltransferases

Hillary CM Nelson[1] and Timothy H Bestor[2]

The structures of two DNA cytosine methyltransferases reveal two novel methods of gaining access to the substrate cytosine residue, both of which involve eversion of the cytosine in a process that may require DNA bending. In one instance there is also widespread base shuffling and distortion of the DNA.

Addresses: [1]Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA 94720, USA and [2]Department of Genetics and Development, College of Physicians and Surgeons of Columbia University, New York, NY 10032, USA.

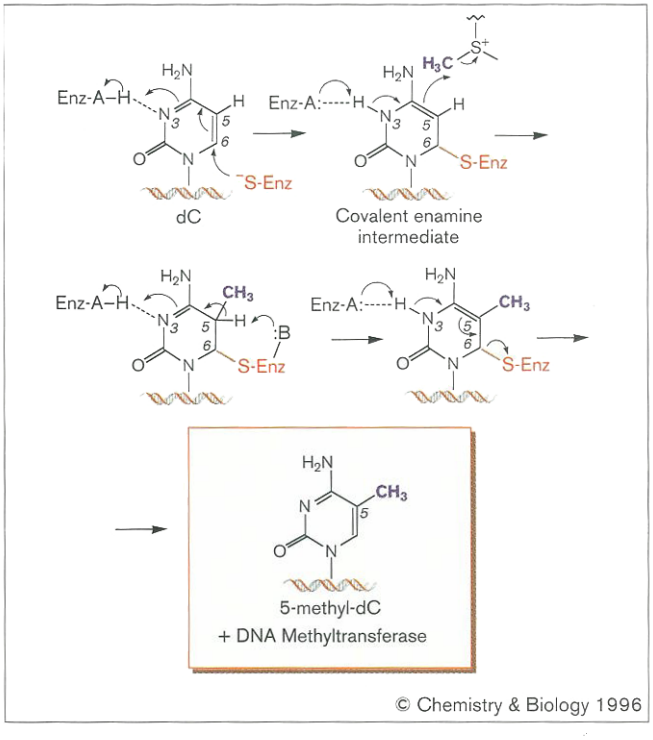
Chemistry & Biology June 1996, **3**:419–423

Most organisms modify their DNA by postreplicative methylation of the C5 atom of specific cytosine residues. Cytosine methylation functions in host defense in bacteria (unmethylated DNA is degraded by specific restriction endonucleases), and is essential for mammalian and flowering plant development [1,2]. Recent progress in elucidating the biology of cytosine methylation has been matched by discoveries of the remarkable structural and catalytic features of the enzymatic transmethylation reaction.

Cytosine-5 methylation has been termed a 'chemically improbable reaction' [3], as it must overcome formidable energetic and stereochemical barriers. Santi *et al.* [4] addressed the energy problem by proposing that DNA cytosine methyltransferases might use a reaction mechanism analogous to that of thymidylate synthetase, and suggested that the C5 position of cytosine is activated by covalent addition of an enzyme cysteine thiolate to the neighboring C6 position to create the C5 carbanion, which would then attack the sulfonium-linked methyl group of S-adenosyl L-methionine (AdoMet). Following methyl transfer, abstraction of the C5 proton would restore the C5–C6 double bond and allow release of free enzyme by β elimination. It was later noted that the C5 carbanion is unlikely to exist under physiological conditions [5], and that the C4–C5 enamine was the more likely nucleophilic intermediate [6]. The proposed reaction scheme, which is probably shared by all enzymes that modify the 5 position of pyrimidines, is shown in Figure 1.

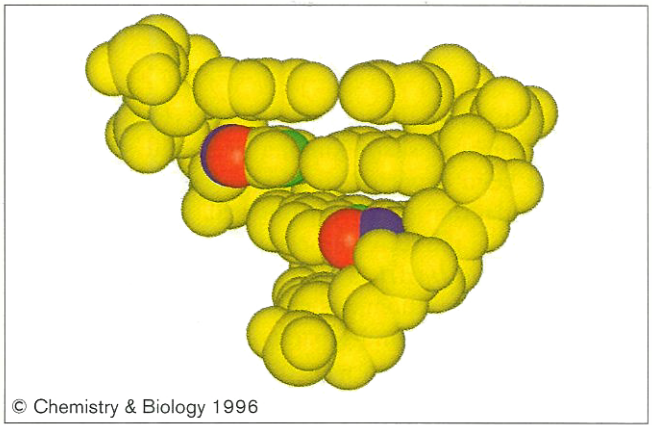
The above reaction scheme helps to resolve the energy issue but introduces severe steric embarrassment; the N3 position of cytosine is hydrogen bonded to a guanosine on the complementary strand and is therefore deep within the structure of B form DNA and inaccessible to the enzyme acid, and the attack trajectories of the enzyme cysteinyl thiolate and AdoMet methyl groups are almost entirely occluded by flanking nucleotides (Fig. 2). Heavy distortions (such as sharp kinking or strand separation) might have been expected to allow access of enzyme to substrate. Erlanson *et al.* [7] showed that interstrand crosslinks within the recognition sequence (GGCC) of the methyltransferase *Hae*III (M.*Hae*III) failed to prevent the transmethylation reaction; this result eliminated the possibility of a global strand-separation pathway, although single base pair separation was still a possibility. An extrahelical cytosine was therefore proposed as a reaction intermediate. When the first co-crystal structure of a DNA methyltransferase (M.*Hha*I)–DNA covalent transition state intermediate became available, it was still a surprise to see that the substrate cytosine is neatly everted from the helix (Fig. 3) [8]. The structure of a second DNA–DNA methyltransferase complex (which involved M.*Hae*III) also contained the everted base, but had several significant structural differences [9]. In the case of M.*Hha*I, the space vacated by the everted base is occupied

Figure 1



Chemistry of DNA (cytosine-5)-methyltransferases. The reaction mechanism was proposed by Chen *et al.* [6] for enzymes that modify the 5 position of pyrimidines, and is a modification of a mechanism originally proposed by Santi *et al.* [4]. Activation is by attack of an enzyme cysteine thiolate to form an enamine intermediate, which is resolved by β elimination after methyl transfer.

Figure 2



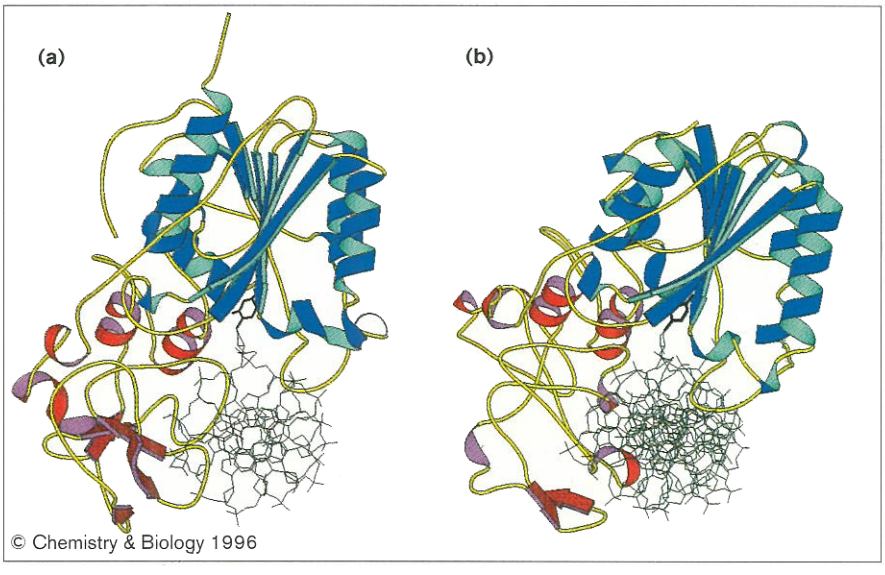
Space-filling model of the M.*Hha*I recognition sequence (5'-GCGC-3') in which C6 is shaded purple, C5 red and N3 green. Note that the attack trajectories of the thiolate on C6 and a methyl group on C5 are occluded, and that the N3 position is buried within the helix; the reaction shown in Figure 1 thus cannot occur in DNA of pure B form. These stereochemical considerations, in addition to the need for protons to be transferred to the substrate cytosine, led Verdine and colleagues [6] to propose that the target cytosine was extrahelical at the time of methyl transfer.

by a glutamine residue, present on a protein loop that enters the site from the major groove, and by a serine residue introduced from the minor groove. These residues hydrogen bond to each other, and the glutamine side chain bonds to the orphan guanosine and also makes hydrophobic contacts with bases formerly stacked against the everted cytosine (reviewed in [10]). The geometry of the remainder of the oligonucleotide is largely B-form in the crystal analyzed. Views of the M.*Hha*I–DNA and M.*Hae*III–DNA complexes are shown in Figure 3.

Although many features of the structures were unexpected, there were precedents for extrahelical bases (different in conformation from the extrahelical cytosine present in the methyltransferase structures) in existing structures. The crystal structure of d(CGCAGAATTCGCG)$_2$ shows two 'bulged' adenines that have no pairing partners [11]; one of these adenines stacks into the helix, while the other loops out and is stabilized by crystal packing forces. Solution NMR structures of DNA with abasic sites have shown that the orphaned base is predominantly extrahelical [12,13], although the structure is dependent on the type of abasic site [14] and the local sequence [15].

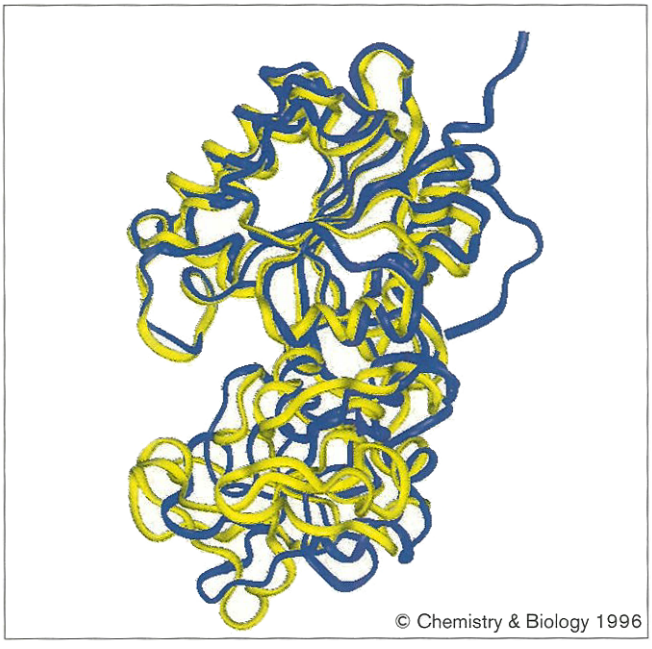
Given the conservation of sequence among DNA cytosine methyltransferases [16–18], it might have been expected that other cytosine methyltransferases would have similar structures and would produce very similar structural distortions in DNA around the target site. This was not, however, borne out by the structure of M.*Hae*III, a DNA cytosine methyltransferase that is found in bacteria of the same genus as those that contain M.*Hha*I and that has a related recognition sequence (GGCC for M.*Hae*III and GCGC for M.*Hha*I). Instead of a common structural framework with substitutions at key contact residues, both the sequences and the structures of the target recognition domains are entirely different, as are the pattern and types of DNA–protein contacts; only the structure of the large domains responsible for cofactor binding and key catalytic steps is strongly conserved. Quite unexpected were the extensive DNA rearrangements caused by M.*Hae*III that occurred in addition to base eversion [9].

Figure 3



Ribbon diagrams of **(a)** M.*Hha*I–DNA and **(b)** M.*Hae*III–DNA complexes. The conserved large domain, which is involved in catalysis, is shaded blue; the small domain, which is responsible for sequence recognition, is shaded red. The everted base is extended into the catalytic pocket of the large domain. Protein loops from the small domain make contacts with base edges in the major groove, and multiple backbone contacts are made by the large and small domains [8,9].

Figure 4



Superposition of ribbon traces of M.*Hha*I (blue) and M.*Hae*III (yellow). DNA has been removed from the structures depicted in Figure 3, and the proteins rotated clockwise about the vertical axis by ~120°. Notice the close congruence within the large domain (top) and the lack of structural similarity within the small domain (bottom).

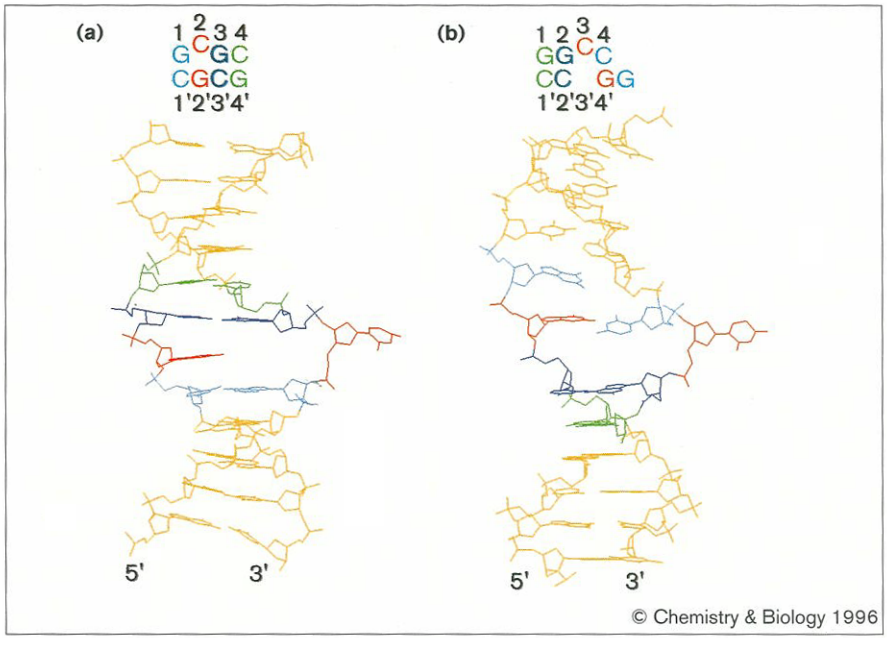
As with M. *Hha*I, M.*Hae*III in a complex with DNA is divided into two major domains. The small domain has been shown to confer sequence specificity to the reaction [17,19], and makes base-specific contacts in the major groove. It has little repeating secondary structure, and most of the specific contacts occur via side chains in loops or turns; this is also true for M.*Hha*I, except that the contacts are largely with backbone atoms or bridging water molecules. This type of binding (using loops instead of helices or sheets) has been recently seen in sequence-specific DNA binding proteins such as p53 [20] and NF-κB [21,22]. The small domains of M.*Hae*III and M.*Hha*I have virtually no sequence similarity to each other, and their structures are not superimposible (Fig. 4).

The sequence and structure of the large domain of the two enzymes are highly conserved. Of the 10 conserved sequence motifs between the two enzymes, 9 are located in the large domain, which contains a central seven-stranded β-sheet with two parallel α-helices on one side and three on the other (all shaded blue in Fig. 3). This folding pattern is reminiscent of the nucleotide binding fold present in many types of enzymes and is also seen in three other methyltransferases: M.*Taq*I [23], vaccinia virus RNA 2' O-methyltransferase [24], and catechol O-methyltransferase [25], all of which catalyze very different reactions from the cytosine methyltransferases. Part of the AdoMet binding site is even more widely conserved, and is discernible in most proteins that bind AdoMet or S-adenosyl L-homocysteine [26]. These observations have provoked suggestions that a single ancestral methyltransferase may have existed [27,28].

The most dramatic aspect of the M.*Hae*III complex is the massively rearranged base pairing of the DNA around the target cytosine (Fig. 5). The substrate cytosine is rotated nearly 180° with respect to the sugar-phosphate backbone

Figure 5

DNA structures in **(a)** M.*Hha*I and **(b)** M.*Hae*III DNA methyltransferase–DNA co-crystals. Bases that are normally paired are illustrated in the same color, so the shuffled base pairing in the M.*Hae*III structure is conspicuous as a red and light blue pair.



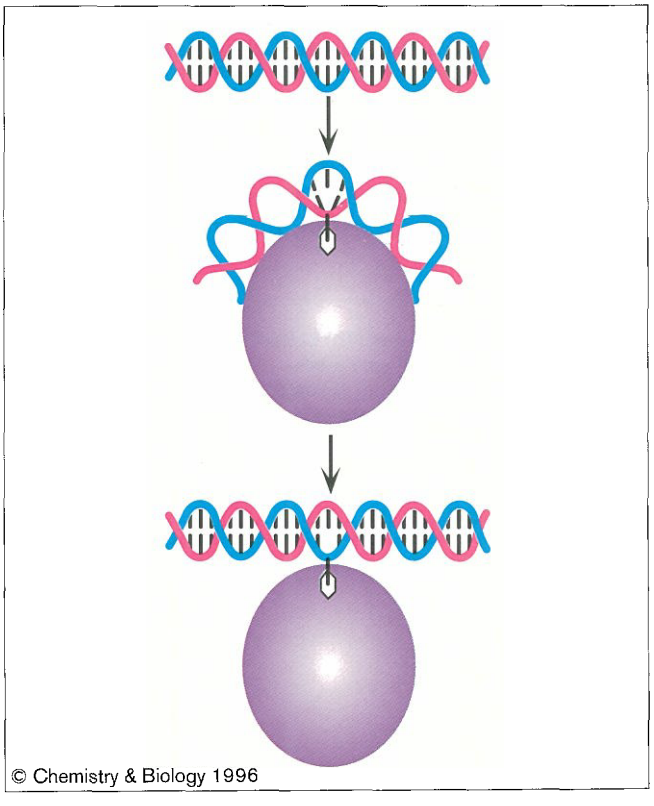
(as in the M.*Hha*I structure), but the resulting void is not occupied by protein loops; the orphan G3' guanine shuffles to pair with the C4 cytosine, and the G4' guanine is orphaned. This shuffled base pairing severely disrupts local DNA structure, which seems to be stabilized by contact of an arginyl side chain with G4' guanine in the major groove and by intercalation of an isoleucyl side chain from the major-groove side between G2' and C3'. Intercalated hydrophobic residues have also been found in the minor grooves of protein–DNA complexes with unusual DNA structures such as PurR [29] and the TATA-binding protein [30,31].

Although the co-crystal structures of M.*Hae*III and M.*Hha*I have given important insights into the mechanism of the transmethylation reaction (and have suggested a widespread role for extrahelical bases [32]), several important questions cannot be answered by studying the trapped transition-state intermediates. The kinetic pathway is unclear, especially with regard to eversion mechanism. Do the enzymes merely capture an extreme thermal excursion of the substrate cytosine, or do they actively participate in the eversion reaction? Is base shuffling spontaneous, or is it induced by interaction with the enzyme? Capture of spontaneously everted and shuffled bases might seem unlikely were it not for the extremely low k_{cat} values characteristic of cytosine methyltransferases (0.03 for M.*Hha*I; the value for M.*Hae*III is probably similar). Thermal motion may present an extrahelical base to the enzyme if given enough time (i.e., many seconds). It is also possible, however, that eversion and base rearrangement occur after binding. For example, M. *Hae*III could bind to half of the binding site, as well as to the undistorted part of the phosphate backbone. Binding-site rearrangement might be aided by Ile221, which moves into the cleft formed after DNA rearrangement. Infiltration of protein loops and amino acid side chains into the DNA, however, requires substantial prior structural rearrangements, and the structures of the particular transition-state intermediates present in the crystals provide no evidence that the protein elements exert force on the substrate DNA.

Recent footprinting data do, however, suggest that a strained intermediate (present prior to the state observed in the crystals) may exert a force that favors base eversion, and raise some interesting questions about the relationship of the intermediate present in the crystal to those that exist in solution. Renbaum and Razin [33] footprinted M.*Hha*I and M.*Sss*I (a *Spiroplasma* cytosine methyltransferase that recognizes the sequence 5'-CG-3') using DNase I, dimethylsulfate or hydroxyl radicals. The footprints of the two enzymes were nearly identical and were unexpectedly large; DNase I footprints spanned 18–21 nucleotides, whereas hydroxy radical footprints were 16 nucleotides long and indicated exposure of the major groove. The oligodeoxynucleotide present in the M.*Hha*I complex was 13 nucleotides long; crystallographic contacts spanned 10 nucleotides on the non-substrate strand and 6 on the substrate strand. The M.*Hae*III oligodeoxynucleotide was 19 nucleotides long, and only 7 bases are contacted on the non-substrate strand, and 6 on the substrate strand. Crystal packing forces may favor the linear conformation, or the short oligonucleotides used in the co-crystallization may not form a sufficient number of contacts with the protein surface to produce the bend that is expected based on the longer footprint seen in solution.

It is thus not unlikely that M.*Hae*III and M.*Hha*I contact larger regions of DNA than were present in the co-crystals, and that these contacts might lead to bending of the DNA towards the minor groove, compressing the target cytosine between neighboring bases and favoring its eversion (Fig. 6). This suggestion is supported by the observation that the 13-base oligonucleotide used in the co-crystallization with M.*Hha*I binds to the enzyme only very weakly in mobility shift assays compared to longer oligonucleotides [33]. The high reactant concentrations

Figure 6



Bend-aided base eversion. High resolution footprinting studies suggest that bacterial cytosine methyltransferases make contacts over a larger region than was represented in the oligonucleotides used in the crystallization studies [33]. These contacts may bend the DNA in solution to compress the target cytosine between neighboring bases, favoring its eversion into the active site pocket of the enzyme.

and long incubation times involved in preparation of the crystals may have compensated for the slow reaction rate with the short substrates. This model makes two predictions: DNA substrates above a certain length will be bent in solution, and the rate of base eversion with short oligonucleotides will be much reduced compared to that seen with longer substrates.

The M.*Hha*I and M.*Hae*III structures illuminate several remarkable properties of these enzymes: nearly complete separation of sequence recognition and catalysis into two domains, nearly complete divergence of sequence and structure of the small (specificity) domain despite close relatedness in species of origin and in recognition sequence between M.*Hha*I and M.*Hae*III, and novel solutions to difficult energetic and stereochemical problems. Several important problems remain to be addressed; the kinetic path of base eversion and the pathway by which the covalent enzyme–DNA complex is resolved are particularly interesting questions.

Acknowledgements

Supported by N.I.H. grants GM00616 and CA60610 (to T.H.B.) and GM44086 (to H.C.M.N.).

References

1. Bestor, T. & Tycko, B. (1996). Creation of genomic methylation patterns. *Nat. Genet.* **12**, 363–367.
2. Li, E., Bestor, T.H. & Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926.
3. Chen, L., MacMillan, A., Chang, W., Ezaz-Nipkay, K., Lane, W. & Verdine, G. (1991). Direct identification of the active-site nucleophile in a DNA (cytosine-5)-methyltransferase. *Biochemistry* **30**, 11018–11025.
4. Santi, D., Garrett C. & Barr P. (1983). On the mechanism of inhibition of DNA-cytosine methyltransferases by cytosine analogs. *Cell* **33**, 9–10.
5. Gerlt, J. & Gassman, P. (1992). Understanding enzyme-catalyzed proton abstraction from carbon acids: details of stepwise mechanisms for β-elimination. *J. Am. Chem. Soc.* **114**, 5928–5934.
6. Chen, L., MacMillan, A. & Verdine, G. (1993). Mutational separation of DNA binding from catalysis in a DNA cytosine methyltransferase. *J. Am. Chem. Soc.* **115**, 5318–5319.
7. Erlanson, D., Chen, L. & Verdine, G. (1993). DNA methylation through a locally unpaired intermediate. *J. Am. Chem. Soc.* **115**, 12583–12584.
8. Klimasauskas, S., Kumar, S., Roberts, R. & Cheng, X. (1994). *Hha*I methyltransferase flips its target base out of the DNA helix. *Cell* **76**, 357–369.
9. Reinisch, K., Chen, L., Verdine, G. & Lipscomb, W. (1995). The crystal structure of *Hae*III methyltransferase covalently complexed to DNA: an extrahelical cytosine and rearranged base pairing. *Cell* **82**, 143–153.
10. Bestor, T. & Verdine, G. (1994). DNA methyltransferases. *Curr. Opin. Cell Biol.* **6**, 380–389.
11. Joshua-Tor, L., Frolow, F., Appella, E., Hope, H., Rabinovich, D. & Sussman, J. (1992). Three-dimensional structures of bulge-containing DNA fragments. *J. Mol. Biol.* **225**, 397–431.
12. Cuniasse, P., Fazakerley, G.V., Guschlbauer, W., Kaplan, B.E. & Sowers, L. (1990). The abasic site as a challenge to DNA polymerase: a nuclear magnetic resonance study of G, C, and T opposite a model abasic site. *J. Mol. Biol.* **213**, 303–314.
13. Kouchakdjian, M., Eisenberg, M., Johnson, F., Grollman, A.P. & Patel, D.J. (1991). Structural features of an exocyclic adduct positioned opposite an abasic site in a DNA duplex. *Biochemistry* **30**, 3262–3270.
14. Withka, J.M., Wilde, J.A., Bolton, P.H., Mazumder, A. & Gerlt, J.A. (1991). Characterization of conformational features of DNA heteroduplexes containing aldehydic abasic sites. *Biochemistry* **30**, 9931–9940.
15. Singh, M.P., Hill, G.C., Péoc'h, D., Rayner, B., Imbach, J.L. & Lown, J.W. (1994). High-field NMR and restrained molecular modeling studies on a DNA heteroduplex containing a modified apurinic abasic site in the form of covalently linked 9-aminoellipticine. *Biochemistry* **33**, 10271–10285.
16. Posfai, J., Bhagwat, A.S., Posfai, G. & Roberts, R.J. (1989). Predictive motifs derived from cytosine methyltransferases. *Nucleic Acids Res.* **17**, 2421–2435.
17. Lauster, R., Trautner, T. & Noyer-Weidner, M. (1989). Cytosine-specific type II DNA methyltransferases. A conserved enzyme core with variable target-recognizing domains. *J. Mol. Biol.* **206**, 305–312.
18. Bestor, T.H., Laudano, A., Mattaliano, R. & Ingram, V. (1988). Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzyme is related to bacterial restriction methyltransferases. *J. Mol. Biol.* **203**, 971–983.
19. Klimasauskas, S., Nelson, J. & Roberts, R. (1991). The sequence specificity domain of cytosine-C5 methyltransferases. *Nucleic Acids Res.* **19**, 6183–6190.
20. Cho, Y., Gorina, S., Jeffrey, P.D. & Pavletich, N.P. (1994). Crystal structure of a p53 tumor suppressor–DNA complex: understanding tumorigenic mutations. *Science* **265**, 346–355.
21. Ghosh, G., van Duyne, G., Ghosh, S. & Sigler, P.B. (1995). Structure of NFκB p50 homodimer bound to a κB site. *Nature* **373**, 303–310.
22. Muller, C.W., Rey, F.A., Sodeoka, M., Verdine, G.L. & Harrison, S.C. (1995). Structure of the NFκB p50 homodimer bound to DNA. *Nature* **373**, 311–317.
23. Labahn, J., *et al.*, & Saenger, W. (1994). Three dimensional structure of the adenine-specific DNA methyltransferase M.*Taq*I in complex with the cofactor S-adenosylmethionine. *Proc. Natl. Acad. Sci. USA* **91**, 10957–10961.
24. Hodel, A.E., Gershon, P.D., Shi, X.N. & Quiocho, F.A. (1996). The 1.85 Å structure of vaccinia protein VP39: a bifunctional enzyme that participates in the modification of both mRNA ends. *Cell* **85**, 247–256.
25. Vidgren, J., Svensson, L. & Liljas, A. (1994). Crystal structure of catechol O-methyltransferase. *Nature* **368**, 354–358.
26. Ingrosso, D., Fowler, A., Bleibaum, J. & Clarke, S. (1989). Sequence of the D-aspartyl/L-isoaspartyl protein methyltransferase from human erythrocytes: common sequence motifs for protein, DNA, RNA, and small molecule S-adenosyl L-methionine-dependent methyltransferases. *J. Biol. Chem.* **264**, 20131–20139.
27. Cheng, X. (1995). DNA modification by methyltransferases. *Curr. Opin. Struct. Biol.* **5**, 4–10.
28. Cheng, X. (1995). Structure and function of DNA methyltransferases. *Annu. Rev. Biophys. Biomol. Struct.* **24**, 293–318.
29. Schumacher, M.A., Choi, K.Y., Zalkin, H. & Brennan, R.G. (1994). Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by alpha helices. *Science* **266**, 763–770.
30. Kim, J.L., Nikolov, D.B. & Burley, S.K. (1993). Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* **365**, 520–527.
31. Kim, Y., Geiger, J.H., Hahn, S. & Sigler, P.B. (1993). Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**, 512–520.
32. Roberts, R. (1995). On base flipping. *Cell* **82**, 9–13.
33. Renbaum, P. & Razin, A. (1995). Footprint analysis of M.*Sss*I and M.*Hha*I methyltransferases reveals extensive interactions with the substrate DNA backbone. *J. Mol. Biol.* **248**, 19–26.