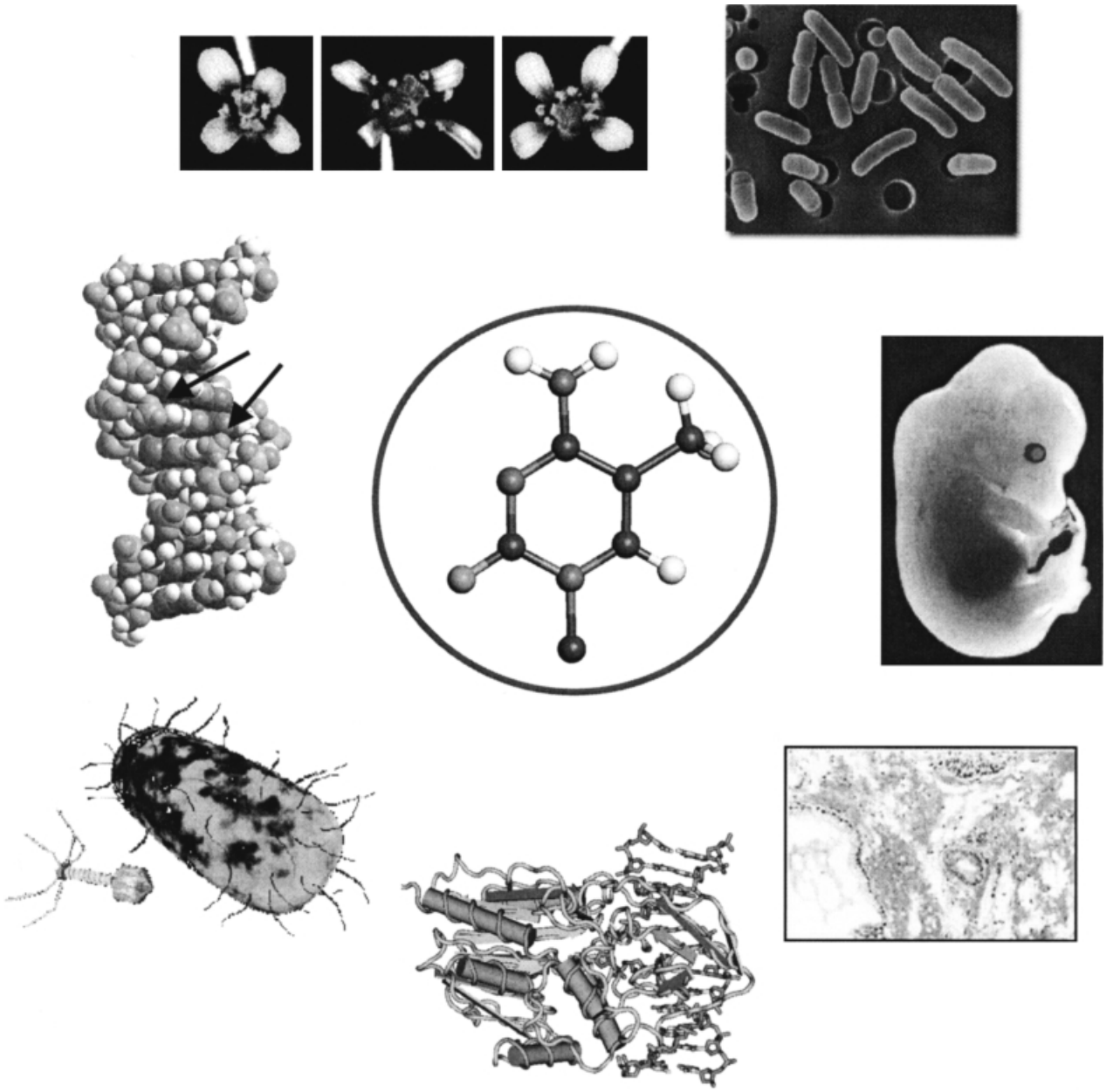
The Chemistry and Biology of DNA Methylation



The central picture shows the structure of 5-methylcytosine, the major methylated base in metazoa. The other pictures (clockwise from the bottom left) show: 1) An *E. coli* cell together with a bacteriophage. Protection of bacteria from phage infection is a central function of DNA methylation in prokaryotic restriction/modification systems. 2) A view into the major groove of a methylated CG site in B-DNA. The methyl groups of 5-methylcytosine are labeled with arrows. 3) Changes of flower phenotype caused by methylation of the Superman gene in *A. thaliania*. 4) Dividing *E. coli* cells. Control of DNA replication and cell division is a second important function of prokaryotic DNA methyltransferases. 5) A mouse embryo. DNA methylation is essential in mammals; animals in which one of the three known active methyltransferases is inactivated die during embryogenesis or after birth. 6) A prostatic adenocarcinoma as an example of a human tumor. Alterations in the pattern of DNA methylation are often found in tumor cells. 7) The structure of the *M.Hha*I DNA methyltransferase complexed with DNA.

  1439-4227/02/03/04 $ 20.00+.50/0 ChemBioChem **2002**, 3, 274 – 293

Beyond Watson and Crick: DNA Methylation and Molecular Enzymology of DNA Methyltransferases

Albert Jeltsch*[a]

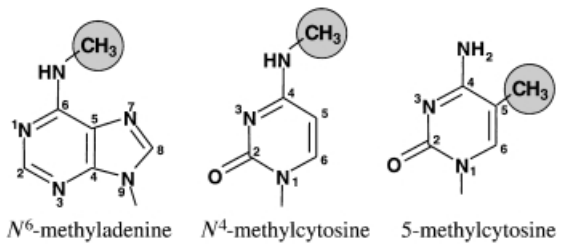
DNA methyltransferases catalyze the transfer of a methyl group from S-adenosyl-L-methionine to cytosine or adenine bases in DNA. These enzymes challenge the Watson/Crick dogma in two instances: 1) They attach inheritable information to the DNA that is not encoded in the nucleotide sequence. This so-called epigenetic information has many important biological functions. In prokaryotes, DNA methylation is used to coordinate DNA replication and the cell cycle, to direct postreplicative mismatch repair, and to distinguish self and nonself DNA. In eukaryotes, DNA methylation contributes to the control of gene expression, the protection of the genome against selfish DNA, maintenance of genome integrity, parental imprinting, X-chromosome inactivation in mammals, and regulation of development. 2) The enzymatic mechanism of DNA methyltransferases is unusual, because these enzymes flip their target base out of the DNA helix and, thereby, locally disrupt the B-DNA helix. This review describes the biological functions of DNA methylation in bacteria, fungi, plants, and mammals. In addition, the structures and mechanisms of the DNA methyltransferases, which enable them to specifically recognize their DNA targets and to induce such large conformational changes of the DNA, are discussed.

KEYWORDS:

DNA methylation · enzyme catalysis · epigenetics · gene expression · protein–DNA interactions

1. Introduction: DNA Methylation as an Extension of the Genetic Code

It has been known for decades that DNA from various sources contains the methylated bases N^6-methyladenine, 5-methylcytosine, and N^4-methylcytosine in addition to the four standard nucleobases (Scheme 1). It should be noted that these methylated bases are natural components of DNA; this distinguishes them from a large variety of chemically modified bases that can be formed by alkylation or oxidative damage of the DNA. DNA methylation is introduced enzymatically by DNA methyltransferases (MTases) after DNA replication. These enzymes use *S*-adenosyl-L-methionine (AdoMet) as the donor of an activated methyl group and modify the DNA in a sequence-specific manner, usually at palindromic sites (Table 1). The methylation does not interfere with the Watson/Crick pairing properties of adenine and cytosine but the methyl group is positioned in the major groove of the DNA, where it can easily be detected by proteins interacting with the DNA. Thereby, methylation adds extra information to the DNA that is not encoded in the sequence, and the methylated bases can be considered the 5th, 6th, and 7th letters of the genetic alphabet.



Scheme 1. *Structures of methylated bases occurring in DNA.*

The process of DNA methylation is intimately interwoven with DNA replication, which inherently destroys DNA methylation because the newly synthesized DNA strand does not carry any methylation. Since it is usually palindromic sequences that are modified, a methylation mark is present on both strands of the DNA (for example: G^mATC/G^mATC = "fully methylated"). During semiconservative DNA replication, these sites are converted into hemimethylated ones ($G^mATC/GATC$), which are re-transformed by a DNA methyltransferase into the fully methylated state.

The focus of this review is the chemistry, enzymology, and biological function of DNA methylation. There exist several other excellent reviews on DNA methylation to whom the reader is referred for additional details on structures and mechanisms of DNA MTases,[1–5] DNA methylation in prokaryotes,[6–8] and eukaryotic DNA methylation.[9–15]

[a] *Priv.-Doz. Dr. A. Jeltsch*
Institut für Biochemie, FB 8
Justus-Liebig-Universität
Heinrich-Buff-Ring 58
35392 Giessen (Germany)
Fax: (+49) 641-99-35409
E-mail: Albert.Jeltsch@chemie.bio.uni-giessen.de

Table 1. *Properties of typical DNA MTases.*

	Recognition sequence	Size[a]	Classification	Comment
prokaryotic enzymes:				
M.*Hha*I	G<u>C</u>GC	327	cytosine-C5	part of an RM system
M.*Hae*III	GG<u>C</u>C	330	cytosine-C5	part of an RM system
M.*Eco*RV	G<u>A</u>TAC	298	adenine-N^6 (α type)	part of an RM system
E. coli dam	G<u>A</u>TC	278	adenine-N^6 (α type)	
M.*Pvu*II	CAG<u>C</u>TG	336	cytosine-N^4 (β type)	part of an RM system
M.*Taq*I	TCG<u>A</u>	421	adenine-N^6 (γ type)	part of an RM system
eukaryotic enzymes:				
Dnmt1 (mouse)	<u>C</u>G	1620	cytosine-C5	high preference for hemimethylated CG/^{m}CG sites[228]
Dnmt2 (mouse)	?	415	?	catalytic activity has not yet been shown
Dnmt3a (mouse)	<u>C</u>G	908	cytosine-C5	also methylates non-CG sites with high activity[136, 206]
Dnmt3b (mouse)	<u>C</u>G	859	cytosine-C5	specificity has not yet been investigated in detail
Met1 (*A. thaliana*)	<u>C</u>G	1537	cytosine-C5	Dnmt1-type
CMT3 (*A. thaliana*)	<u>C</u>NG	839	cytosine-C5	chromomethylase
DRM2 (*A. thaliana*)	?	626	cytosine-C5 ?	Dnmt3-type
Masc1 (*A. immersus*)	?	537	cytosine-C5	de novo DNA MTase in vivo, inactive in vitro
Masc2 (*A. immersus*)	?	1356	cytosine-C5	active in vitro, function in vivo unknown

[a] Size is given as the number of amino acid residues.

2. The Role of DNA Methylation in Prokaryotes

In prokaryotes, all three types of DNA methylation described above are observed. Here, DNA methylation has three major biological roles: 1) distinction of self and nonself DNA, 2) direction of postreplicative mismatch repair, and 3) control of DNA replication and cell cycle. The first of these issues is associated with restriction/modification systems (RM systems)[16, 17] which function as a defense against infection of bacteria by bacteriophages and are the source of the overwhelming majority of DNA methyltransferases found in prokaryotes. In addition to a methyltransferase, RM systems also contain a restriction endonuclease that specifically cleaves DNA invading the cell at defined recognition sites.[18–20] The cellular DNA is protected against this attack by methylation within the recognition site. If a phage DNA escapes from cleavage, the phage can infect the cell and its progeny carries the same methylation pattern as the host cell. Then, other bacteria containing the same RM system are no longer protected against infection. Therefore, it is advantageous that different bacteria carry RM systems with different recognition sequences and, as a result, a large diversity of RM systems with different recognition sequences has been developed through the billions of years of bacteria/bacteriophage coevolution. So far, over 2000 different RM systems are known, 700 different DNA MTases have been sequenced which recognize and methylate almost 300 different DNA sequences (see: http://www.neb.com/rebase[21]). In these systems the sequence context of the DNA methylation carries the information that allows the cell to discriminate between self and nonself DNA.

The role of DNA methylation in DNA repair and control of DNA replication is best understood in the *Escherichia coli* dam system, where DNA is modified at adenine residues in GATC sequences.[22, 23] Similar systems are present in other γ-proteobacteria. In these bacteria the methylation status of GATC sites cycles between hemimethylation (GATC/G^{m}ATC) immediately after DNA replication and full methylation (G^{m}ATC/G^{m}ATC) after remethylation by the dam MTase; at most sites this occurs 2–4 s after replication.[24] During the short time span between DNA replication and dam methylation, a directed repair of replication errors is possible, because the methylation mark allows the unmethylated daughter strand which must be repaired and the methylated original template strand whose nucleotide sequence is correct to be distinguished. Moreover, dam methylation is used to couple the bacterial cell cycle to DNA replication, because a number of gene promotors are induced in the hemimethylated state.[23, 25] The only DNA region that remains hemimethylated for a longer period ($\approx$20 min) is the origin of DNA replication,[26, 27] because the SeqA protein binds to it and, thereby, prevents dam methylation. This is used for the control of DNA replication, because hemimethylated origins of replication are not active. Taken together, hemimethylation of dam sites is used to encode two kinds of information: 1) it indicates that a DNA molecule already has been replicated, and 2) it labels the parental strand after DNA replication. Moreover, dam

Albert Jeltsch was born in Kassel (Germany) in 1966. He studied Biochemistry at the University of Hannover and received his doctoral degree in 1994 in the group of Alfred Pingoud for work dealing with the molecular enzymology of restriction endonucleases. He finished his Habilitation on the molecular enzymology of DNA methyltransferases in 1999. Today, he is assistant professor at the Institute for Biochemistry of the Justus-Liebig University Giessen, where he heads a group working on DNA methylation. He received the Gerhard-Hess award of the Deutsche Forschungsgemeinschaft in 1999 and the BioFuture award of the Bundesministerium für Bildung und Forschung in 2001.

methylation plays a role in the segregation of the *E. coli* chromosome. Despite these important functions, dam methylation is not essential in *E. coli*, but it is involved in the pathogenicity of different bacteria like *Bordetella pertussis*,[28] *E. coli*,[29] *Salmonella thyphimurium*,[30–33] and *Neisseria meningitidis*,[34] a fact that suggests that inhibitors of dam methylation might be effective as antibacterial drugs.

A similar system of DNA methylation is observed in *Caulobacter cresentus* and other α-proteobacteria.[35] The *Ccr*M MTase modifies adenine residues within GANTC sequences. Its biology is less clear understood than that of the dam system, but it is known that the expression of the *Ccr*M MTase is correlated with the cell cycle of *C. cresentus*; this suggests that DNA methylation is involved in cell cycle control in these organisms. In contrast to the dam enzyme, the *Ccr*M MTase is an essential protein in α-proteobacteria.[36] Since the group of α-proteobacteria contains important human pathogens and adenine-MTases appear not to exist in higher eukaryotes, the *Ccr*M MTases are attractive drug targets—promising inhibitors that are specific for adenine MTases have just been described.[37]

2.1. The structure of prokaryotic DNA methyltransferases

According to the chemistry of methylation, C-MTases which form a C—C bond (cytosine-C5 MTases) can be distinguished from N-MTases which form a C—N bond (adenine-N^6 and cytosine-N^4 MTases). In general, both types of MTases are two-domain proteins comprising one large and one small domain with the DNA binding cleft being located at the domain interface. The large domain contains a set of up to ten conserved amino acid motifs, which differ between C- and N-MTases.[38–40] N-MTases can be further subdivided into three classes (α, β, and γ) that differ from each other with respect to the position of insertion of the small domain into the framework of the large domain as well as by a circular permutation of the amino acid sequence of the large domain.[4, 38, 40, 41] These differences result in a different arrangement of the most conserved amino acid sequence motifs that is characteristic for each class. The C-MTases are most similar to the γ class of N-MTases. Most cytosine-N^4 MTases are members of the β group of N-MTases, but some examples of cytosine-N^4 MTases are also found in the α and γ groups.

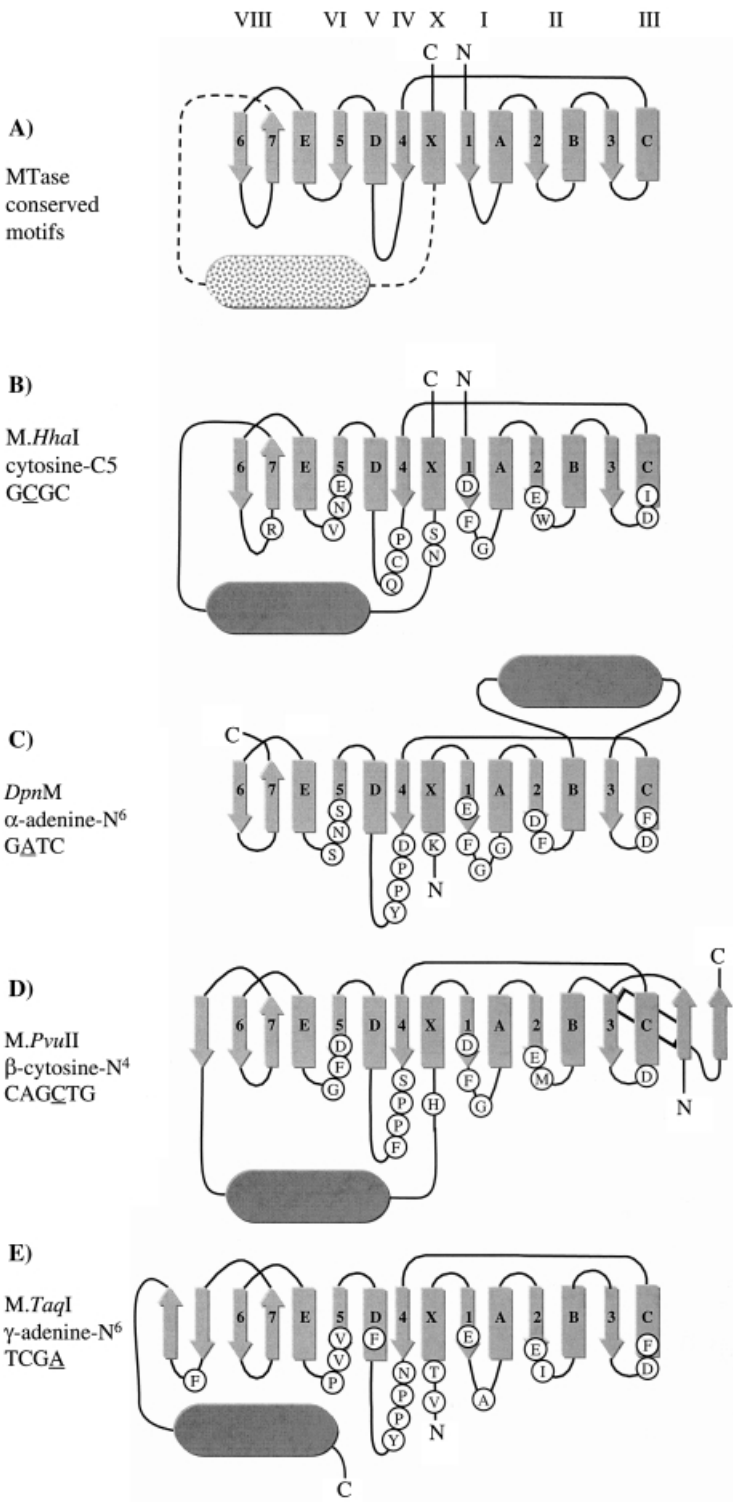


Figure 1. *Comparison of the topologies and location of conserved motifs in DNA MTases of different families. A) Locations of the conserved motifs displayed on the framework of the structure of the large domain of cytosine-C5 MTases. B) Topology and locations of important amino acid residues in cytosine-C5 MTases. C) Topology and locations of important amino acid residues in α-N MTases. D) Topology and locations of important amino acid residues in β-N MTases. E) Topology and locations of important amino acid residues in γ-N MTases.*

2.1.1. The structure of the large domain

The large domain of DNA MTases forms the binding site for the AdoMet and the catalytic center of the enzyme. The large domains of all DNA MTases share a common structural core which consists of a six-stranded parallel β sheet with a seventh strand inserted in an antiparallel fashion between the fifth and sixth strands (Figure 1 A). The β sheet begins in the middle with strand 1 (6↑ 7↓ 5↑ 4↓ 1↑ 2↑ 3↑). This center to right, center to left architecture generates a topological switch-point in the middle which separates two subdomains from each

other. The structure of both the subdomains is based on a β sheet that is flanked by 1 – 2 α helices on either side with a doubly wound open α/β/α sandwich as the core structure. The right subdomain (β1 – β3) creates the AdoMet binding site, the left one (β4 – β7) the binding site for the extrahelical target base (see below). Both binding sites are hydrophobic pockets that are located at equivalent positions within the subdomains, a fact which suggests that the catalytic domain might have arisen by a gene duplication.[40] This type of structure is also observed in other AdoMet-dependent MTases like RNA MTases, protein MTases, and small-molecule MTases.[4, 5] The large subdomains of C- and N-MTases contain up to ten characteristic amino acid motifs, where the motifs I (DXFXGXG), IV (GFPCQ), and VI (ENV) are most conserved in C-MTases[42] and the motifs I (DXFXGXG) and IV ((D/N/S)PP(Y/F)) in N-MTases.[40] Although only motif I is similar between both classes of enzymes, it has been found that corresponding motifs are located at structurally equivalent positions in the proteins and form similar interactions to the cofactor and other parts of the protein[4, 5]—a clear example of proteins with almost no similarity at the amino acid sequence level that share a remarkably conserved structure. For example, in both types of enzymes residues located within the motifs IV and VI interact with the flipped target base and play the most important roles in catalysis (see below).

2.1.2. The structure of the small domain

The small domains of different DNA MTases are dissimilar in amino acid sequence, size, and structure. So far structures of two bacterial C-MTases (M.*Hha*I[43] and M.*Hae*III[44]) are known. Although the catalytic domains of both proteins can be easily superimposed on each other, the structures of the small domains are dissimilar. The small domain of M.*Hha*I comprises 81 amino acid residues and consists of 7 β strands whereas that of M.*Hae*III (92 amino acid residues) has no extensive secondary structure. The heterogeneity is even larger in the family of N-MTases, where, so far, four structures are known: M.*Taq*I,[45] M.*Pvu*II,[46] *Dpn*M,[47] and M.*Rsr*I.[48] The size of the "small" domain of the M.*Taq*I MTase is 177 amino acid residues. It mainly consists of β strands with three α helices forming a three helical bundle on one end. In contrast the small domain of M.*Pvu*II just comprises one α helix with the two adjacent loops (26 amino acid residues in total). In *Dpn*M the small domain (91 amino acid residues) is composed of an α-helical cluster comprising four α helices, and in M.*Rsr*I approximately 40 amino acid residues are folded into a short 3_{10}-helix and two α helices arranged in a helical bundle. This divergence in structure goes in parallel with a difference in function, because the small domains of DNA MTases form many but not all of the sequence-specific contacts between the DNA and the MTase (see Section 2.2.5). These contacts mediate the recognition of the DNA sequence of the target site that is characteristic for each enzyme.

2.1.3. Target-base flipping

The most interesting structural and mechanistic feature of DNA MTases became apparent with the first crystal structure of a DNA MTase in complex with substrate DNA, because these enzymes completely flip the target base out of the DNA helix (Figure 2) and bind it into a hydrophobic pocket in the large domain of the enzyme that creates the active site for methyl group transfer.[44, 49, 50] The target-base binding pocket is formed by residues from motifs IV, VI, and VIII which are located in the large domain of the MTase at the ends of β4 and β5 and at the beginning of β7, respectively. The reason for this unusual mechanism might be in the catalytic mechanism of C- and N-MTases (see below) which requires an intimate contact of the enzymes to the aromatic ring system of the target base; this contact would not be possible if the base were located inside the DNA double helix. Later, other DNA-interacting enzymes were identified that also make use of base flipping, perhaps because they also require a close contact to the bases of the DNA or they must open up space for the catalytic machinery working at the backbone of the DNA. These include many DNA repair enzymes like uracil-DNA glycosylase and T4 endonuclease V.[51–54]

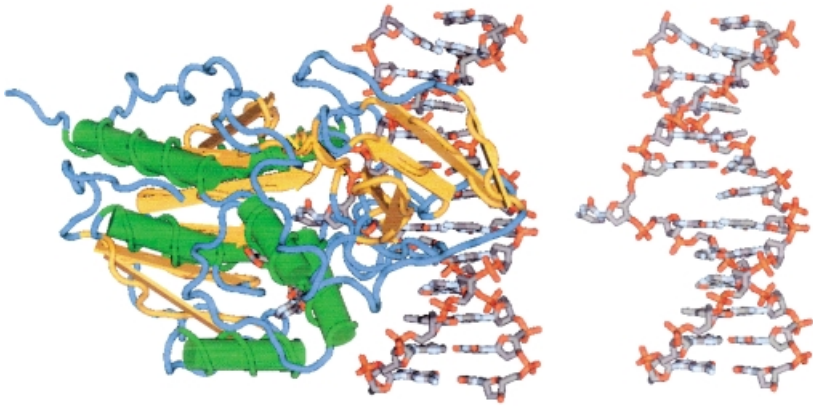


Figure 2. *Structure of the M.*Hha*I – DNA complex. On the right-hand side, the DNA is displayed alone to allow the target base, rotated out of the DNA helix, to be easily seen.*

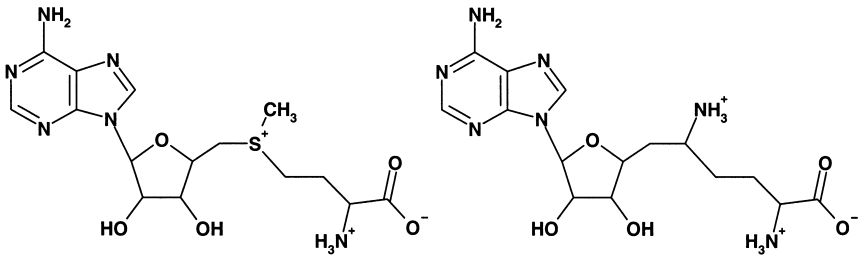
So far, structures of three different DNA MTases in complex with target DNA are available. Base flipping is observed in all of them, however, the structural adaptations of the DNA after base flipping are dissimilar in all three cases. In M.*Hha*I (G<u>C</u>GC, the modified base is underlined), the DNA retains an almost B-DNA-like structure with the exception of the flipped target base. The N1, N^2, and O^6 positions of the orphan guanine base are contacted by a glutamine residue from the small domain of the enzyme that is inserted into the DNA and occupies the space of the flipped cytosine.[49] In the M.*Hae*III – DNA structure, the bases of the recognition sequence (GG<u>C</u>C) undergo extensive rearrangements. The orphan guanosine is tilted and pairs with the cytosine in the 3′ direction from the target cytosine. Thereby, the guanine usually paired to this residue is left orphan and a large cleft is created in the DNA that is filled in part with solvent.[44] Finally, in the M.*Taq*I – DNA structure (TCG<u>A</u>), the orphan thymidine residue is shifted towards the center of the double

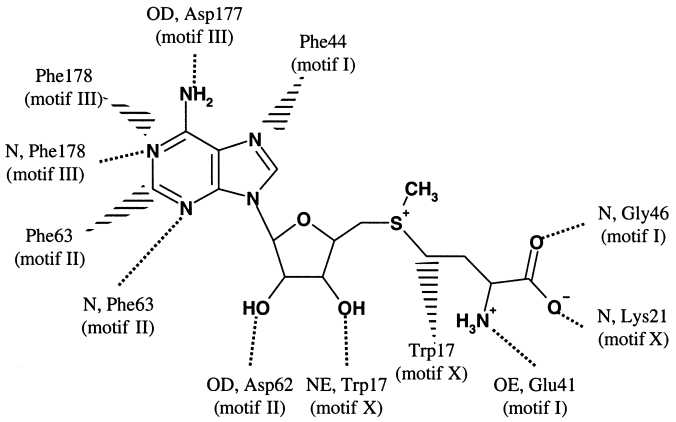
helix in a manner leads to a compression of the DNA backbone. In its new position, the thymidine in part occupies the space of the rotated adenine.[50] Thus, the structural adaptations of the enzyme – DNA complex after base flipping are dependent on the protein and the sequence of the DNA and are different in each case. For example, in the M.*Hae*III – DNA complex the orphan guanine can only "steal" a cytosine for base pairing, because the recognition sequence of M.*Hae*III is GGCC.

2.1.4. The AdoMet binding site

The AdoMet binding site is remarkable conserved in all DNA (and also non-DNA) MTases. It is created by residues from the motifs I – III and X, which form conserved contacts to almost every hydrogen-bond donor and acceptor of the AdoMet and, in addition, several hydrophobic interactions to the cofactor (Scheme 2). (A detailed compilation of the atomic contacts between MTases and AdoMet as seen in different MTase/AdoMet co-crystal structures can be found in ref. [47].) The roles of many of these residues have been confirmed by mutagenesis experiments (see refs. [7, 55], and references cited therein). There is just one exception to this general similarity: the Phe residue in motif I (DXFXGXG) is not present in γ-type N-MTases, like M.*Taq*I. In these enzymes, a highly conserved Phe in motif V closely approaches the position of the missing Phe residue of motif I and functionally replaces it. Many DNA MTases bind so strongly to AdoMet that it is copurified over several chromatography steps. In M.*Taq*I it has been shown that the product of the methylation reaction *S*-adenosyl-L-homocysteine (AdoHcy), which does not carry a positive charge at the sulfur center, binds to the MTase in a different conformation than AdoMet, such that its homocysteine moiety interacts with the active site residues located in motif IV.[56] A similar result was obtained with sinefungin (adenosyl-L-ornithine), a natural inhibitor of DNA MTases, which has an inverted charge configuration at the $\rangle CH{-}NH_3^+$ center as compared with the $\rangle S^+{-}CH_3$ center of AdoMet, but otherwise is isoelectronic to AdoMet. These results explain why most MTases show a pronounced product inhibition by AdoHcy and a strong inhibition by sinefungin. Also AdoMet can bind to the binding pockets of MTases in two different modes (see ref. [57], and references cited therein). The functional relevance of this observation is still not known.

More than one AdoMet binding site has been observed in some DNA MTases.[46, 58–61] It has been shown in several instances that binding of a second AdoMet can allosterically influence the active site of the MTase. However it is not known if the second AdoMet has mechanistic relevance, for example, if it can it be channeled into the active site. Since the cofactor binding site of MTases is embedded in the enzyme – DNA complex, there has to be a channel for the cofactor to diffuse through if cofactor exchange is possible when the DNA is bound. The efficiency of such exchange processes would be greatly enhanced if these channels had binding sites for the AdoMet that have a lower affinity than the active-site binding pocket. One could speculate that such sites are observed in the above-mentioned cases, which is in agreement with the finding that the affinities of the allosteric sites for AdoMet are usually significantly lower than that of the catalytic binding site.



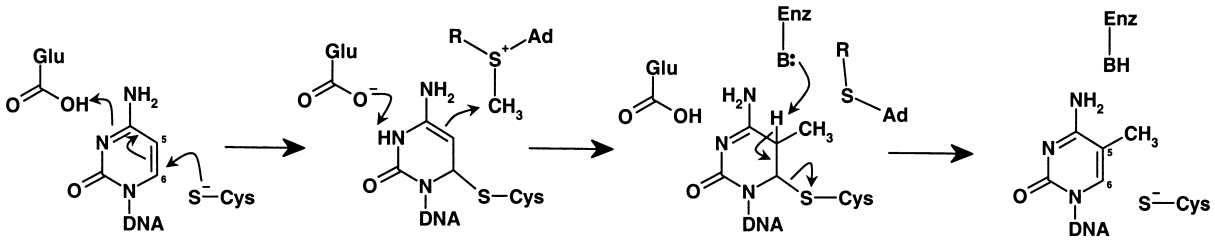


Scheme 2. *Structure of the cofactor of the DNA MTases, AdoMet, and its potent competitor sinefungin. In the lower panel the AdoMet binding site of the DpnM MTase is shown schematically.[47] The positions of atoms (N = main-chain nitrogen atom; OD, OE, NE = delta (D) and epsilon (E) side-chain oxygen and nitrogen atoms of Asp, Glu, and Trp, respectively) and numbers of amino acid residues in contact with the AdoMet are indicated, the numbers of the motifs are given in parenthesis. Hydrogen bonds are indicated by dotted lines and hydrophobic contacts by dashed wedges.*

2.2. The mechanism of DNA methyltransferases

2.2.1. Catalytic mechanism of cytosine-C5 MTases

Although AdoMet is a very effective donor for methyl groups, methylation of cytosine residues at position 5 is not a trivial reaction, because cytosine is an electron-poor heterocyclic aromatic ring system and position 5 of cytosine is not capable of making a nucleophilic attack on the methyl group of AdoMet. Therefore, the reaction catalyzed by the enzyme follows the reaction pathway of a Michael addition (Scheme 3).[62, 63] Initially a cysteine SH group from the active site of the enzyme makes a nucleophilic attack at position 6 of the cytosine



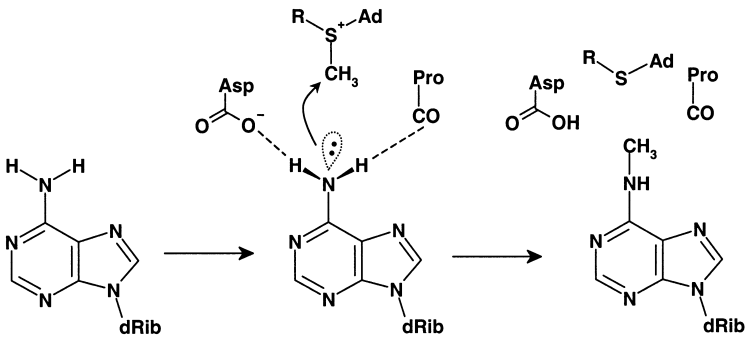
Scheme 3. *Catalytic mechanism of C-MTases. The figure is based on the structure of the M.*Hha*I – DNA complex.*[49, 65] *The cysteine residue is from motif IV (PCQ), the glutamic acid residue is from motif VI (ENV) of the C-MTase.*

ring to yield a covalent complex between the DNA and the enzyme as an intermediate of the methylation reaction. The attacking cysteine residue is located in the highly conserved PCQ motif (motif IV; Cys 81 in M.*Hha*I). The enzyme facilitates the nucleolytic attack on the C6 atom by a transient protonation of the cytosine ring at the endocyclic nitrogen atom N3;[64] this protonation is mediated by Glu 119 from the highly conserved motif VI (ENV) with the help of Arg 165 from motif VIII in the case of M.*Hha*I.[65] Thereby, position 5 of the cytosine is strongly activated and attacks the methyl group. The covalent enzyme – DNA complex is resolved by deprotonation at position 5 which leads to the elimination of the cysteine SH group and reestablishes aromaticity. The nature of the proton abstracting base is not known with certainty, involvement of a phosphodiester group has been discussed.[65] This reaction mechanism is analogous to that of 2′-deoxyuridine monophosphate methylation by thymidylate synthetase.[66] An interesting feature of this mechanism is that the enzyme requires access to both sides of the aromatic ring system of the cytosine, because the initial attack of the SH group on the cytosine and the nucleophilic attack of the cytosine ring system on the AdoMet methyl group occur at different sides of the ring system. Therefore, the catalytic mechanism of C-MTases can explain why DNA MTases developed a base-flipping mechanism.

2.2.2. Catalytic mechanism of N-MTases

The methylation reaction at the exocyclic amino groups of adenine and cytosine rings proceeds with inversion of the configuration of the methyl group in an S_N2 reaction without a covalent intermediate.[67] Like methylation of cytosine residues at position 5, methylation of the exocyclic amino groups of cytosine and adenine is not an easy task, because neither adenine nor cytosine displays nucleophilicity at the exocyclic amino group, since their free electron pairs are conjugated with the aromatic systems. In fact, detectable nucleophilicity is only observed at the N1 and N3 positions of adenine and the N3 position of cytosine, which in DNA are not the targets for enzyme-mediated methylation reactions. Given the observation that the specificity of adenine-N^6 and cytosine-N^4 MTases overlaps (adenine-N^6 MTases also accept cytosine as a target for methylation, cytosine-N^4 MTases also accept adenine),[68, 69] it is quite likely that the reaction mechanisms of both types of enzymes are very similar. This conclusion is supported by the finding that cytosine-N^4 MTases are found in the α, β, and γ groups of N-MTases.[40, 70] Moreover, both these findings suggest that the target-base specificity of N-MTases has been changed several times during molecular evolution.[68, 70]

Adenine-N^6 and cytosine-N^4 MTases are characterized by a conserved (D/N/S)PP(Y/F) motif (motif IV), that is located in the active site of the enzyme at a position structurally analogous to the PCQ motif of the C-MTases. Mutations within this region strongly reduce the catalytic activity of different DNA MTases.[55, 71–74] According to the structure of the adenine-N^6 MTase M.*Taq*I,[50] the most important function of this tetrapeptide is that the side chain of the D/N/S residue and the main-chain carbonyl group of the first proline serve as hydrogen-bond acceptors for the protons of the exocyclic amino group. Since the acceptor groups are presented in a tetrahedral geometry, a change in hybridization of the nitrogen atom from sp^2 to sp^3 is induced which localizes the free electron pair at the N^6 position (Scheme 4). It is likely that the D/N/S residue also functions as proton acceptor during the reaction,[46, 48] because in M.*Pvu*II the serine residue is connected to a glutamic acid residue by a charge relay system, that might allow the serine to act as a base.[46] Most likely, the D/N/S residue is only transiently protonated and immediately transfers the proton to other residues and finally to water. Given the pK_a values of the N^6 atom of the adenine or the N^4 atom of cytosine and an aspartic acid, asparagine, or serine residue, it is likely that the proton transfer occurs late in the reaction when the N–CH_3 bond has been almost completely formed. Therefore, a cationic transition state has to be envisaged which is further stabilized by cation – π



Scheme 4. *Catalytic mechanism of N-MTases. The figure is based on the structure of the M.*Taq*I – DNA complex.*[50] *The aspartate and proline residues in the figure are from motif IV of the N-MTase (D/N/S)PP(Y/F).*

interactions with surrounding aromatic amino acid residues.[75] This model is supported by the observation that the active sites of all N-MTases contain at least one or even more aromatic amino acid residues that contact the flipped base,[76–79] like the Y/F residue in the catalytic tetrapeptide, whereas C-MTases usually do not show aromatic amino acid residues near the active site. It has been shown in several cases that an exchange of these aromatic residues to a nonaromatic residue reduces the catalytic efficiency much more than an exchange by another aromatic residue,[55, 72, 74, 76, 80, 81] as one would predict from the cation–π interaction model.

2.2.3. Kinetic mechanism of DNA MTases

DNA MTases form a ternary complex with both substrates, the DNA and the cofactor AdoMet. In principle, these substrates can bind to the enzyme either randomly or sequentially ordered, in the latter case either the DNA or the cofactor can bind first. So far, all of these mechanisms have been observed with different DNA MTases: M.*Eco*P15I[82] and M.*Eca*I[83] follow a random mechanism, M.*Eco*RI,[84] M.*Eco*RV,[85] and M.*Rsr*I[86] follow a sequential mechanism in which the AdoMet is bound first. Recent results for T4 dam MTase suggest that the enzyme exits in two conformations: one can bind AdoMet and DNA in a random order, the other first binds to AdoMet and then to DNA.[87] In contrast, M.*Hha*I,[63, 88] M.*Msp*I,[89] and *E. coli* dam[90] follow a sequential mechanism where the DNA binds first. The mechanism for M.*Hha*I has recently been challenged and evidence for a random mechanism has been provided.[91] Perhaps here, as in T4 dam, alternative conformations of the enzyme require different orders of substrate binding.

2.2.4. Processivity of DNA methylation

Since DNA MTases have the ability to slide on the DNA by linear diffusion,[85, 92, 93] the question arises of whether they are able to catalyze the methylation of several recognition sites located on one substrate molecule in a processive manner without leaving the DNA. This issue has been investigated for some MTases that are components of bacterial RM systems (M.*Hha*I, M.*Hpa*II, M.*Eco*RI, and M.*Eco*RV). In all cases it turned out that the enzymes were not able to catalyze several rounds of DNA methylation on one target molecule in a processive manner.[85, 93, 94] In contrast, the *Ccr*M and M.*Sss*I MTases were shown to methylate DNA processively[94, 95] and similar results were recently obtained with the dam MTase;[90] these are solitary MTases that are not part of an RM system. This difference could be an important adaptation of the enzymes to the biological function of RM systems which is to cleave invading phage DNA by the restriction endonuclease. Since methylation protects the DNA from cleavage, it is important that the endonuclease reaches a restriction site on the phage DNA, before it is modified by the corresponding MTase. However, since the phage DNA usually contains more than one recognition site and the endonuclease is able to scan the DNA by linear diffusion, methylation of some recognition sites does not prevent restriction of the phage DNA. The only way to protect the invading DNA from cleavage would be to modify all recognition sites before an endonuclease has bound to the substrate. However, this event (undesirable, from the point of view of the RM system) is efficiently prevented by the distributive mechanism of the DNA MTases, because after each turnover they first have to release the DNA and exchange the cofactor, and only then can they rebind to the DNA. Given this line of arguments it is likely that other DNA MTases that are parts of type II RM systems also will show a distributive mechanism of DNA methylation.

2.2.5. DNA binding and DNA recognition

Like other types of enzymes that specifically recognize DNA (for example, restriction endonucleases[19, 20]) DNA MTases interact with the DNA in a multistep reaction: after nonspecific binding the DNA is scanned by linear diffusion for the target sites.[85, 93, 96] Specific binding is accompanied by large conformational changes of the enzyme–DNA complex (see below) that finally activate the enzyme and lead to methyl group transfer and product dissociation. There are many reports of conformational changes of DNA MTases during DNA binding: substantial conformational changes are visible in the structures of M.*Hha*I and M.*Taq*I, so far the only enzymes whose structures are solved in both the DNA-bound and DNA-free forms,[43, 45, 49, 50] and DNA bending has been demonstrated by different techniques for many DNA MTases.[97–100] Therefore, DNA binding is accompanied by large conformational changes of the DNA.[85, 101, 102] Finally, it has been shown that the M.*Eco*RV and T4 dam enzymes exist in open and closed conformations that are in slow equilibrium.[85, 87, 103] The dynamics of target-base flipping are discussed in Section 2.2.6.

Like other DNA interacting enzymes, DNA MTases achieve sequence specificity by a multitude of direct and water-mediated contacts of the enzyme to the DNA substrate. M.*Hha*I contacts the DNA in the major groove with two recognition loops located in the small domain of the enzyme. With the exception of one contact in the minor groove (formed by Glu 76 located in motif IV of the catalytic domain), M.*Hae*III also contacts the DNA from the major-groove side with residues located in the small domain of the MTase. Biochemical results suggest a similar contact pattern for other C-MTases.[104] In contrast, the M.*Taq*I enzyme directly contacts the DNA both from the major-groove side with residues of the small domain and also from the minor-groove side with residues located in the large domain in motif IV and immediately before motif V. Therefore, in N-MTases the large domain appears to be intimately involved in sequence recognition, a fact which is also shown by biochemical results with M.*Eco*RV.[105]

Restriction endonucleases, which are a paradigm of enzymes interacting with DNA in a highly specific manner, usually saturate all possible hydrogen-bond contacts in the major groove of the recognition sequence.[19, 20] In contrast to this, the number of specific interactions observed between the MTases and the DNA is smaller. This may explain the general observation that the specificity of DNA MTases is not as high as that of restriction enzymes. DNA methylation studies have shown that many DNA MTases (*Dpn*A,[106] M.*Eco*RI,[107] M.*Fok*I,[78] and M.*Eco*RV[105]) methylate sites that differ by one base pair from their canonical sites

with rates only reduced by factors of 5 – 10. Such sites usually are cleaved at least 100 – 1000 times more slowly by restriction enzymes. Moreover, even sites that differ in more than one base pair from the specific site are modified with rates only reduced by 1 – 2 orders of magnitude (*Eco*RV,[105] M.*Fok*I[78])—similar results were never obtained with a restriction enzyme.

2.2.6. Base flipping

Apart from the three crystal structure analyses already mentioned there is now biochemical evidence confirming that DNA MTases in general make use of a base-flipping mechanism. This includes the demonstration of very high cross-linking yields of reactive base analogues located at the target position,[77, 78] high accessibility of the target base towards chemical probes like permanganate,[108] the observation of an increase in fluorescence of 2-aminopurine when it is located at the position of the target base,[61, 86, 102, 109, 110] and the observation that many DNA MTases bind more strongly to substrates in which the target base has been replaced by a base analogue or another base such that a base mismatch is created.[78, 86, 101, 111–114] Although these studies were carried out with different enzymes and none of them finally proves that a base rotating mechanism is operative, when taken together the available evidence strongly suggests that all DNA MTases make use of base flipping as an integral part of their reaction mechanism.

The thermodynamics of base flipping is illustrated by the preference of several MTases for binding to substrates which contain a base mismatch at the target position. The most straightforward explanation for this observation is that base flipping requires disruption of the Watson/Crick hydrogen bonds of the target base, which makes flipping more favorable if the target base is not engaged in a regular base pair. On the other hand the rotated target base, as well as the partner base in the DNA helix, could be contacted by the MTase, a process that should provide interaction energy and specificity, because such contacts are diminished in cases where the nature of one of these bases is altered. The general observation that the balance of these energetic contributions is almost always favorable with mismatched substrates demonstrates that no strong, specific interactions are formed between the enzyme and the target base pair. The absence of a specific interaction with the target base is also demonstrated by the following two observations: 1) Structures of M.*Hha*I with oligonucleotides containing mismatches at the target position show that, in addition to cytosine, adenine and uracil are also flipped by the enzyme and bound in the same hydrophobic pocket.[115] 2) Several adenine-N^6 and cytosine-N^4 MTases can modify adenine and cytosine target bases[68, 69] and cytosine-C5 MTases can also methylate uracil, albeit at low efficiency.[112, 113]

The dynamics of base flipping have been most intensively studied with the M.*Hha*I cytosine-C5 MTase and the M.*Eco*RI adenine-N^6 MTase. For M.*Hha*I it has been shown that base flipping comprises at least two steps: first, the base is rotated out of the DNA helix, but not tightly bound by the enzyme and then it is inserted into the hydrophobic binding pocket of the enzyme.[116] This result is in good agreement to the general finding that base flipping is not very specific (see above). Also, in M.*Eco*RI, base flipping is a two-step process[111] with very fast rate constants, of the order of 100 s^{-1}, for the first step.[117]

The mechanism of base flipping is particularly enlightened by the structure of a complex of the M.*Hha*I MTase with a substrate containing an abasic nucleotide at the target position.[118] Even in this case the sugar moiety is flipped out of its normal position and located outside of the DNA helix, in a very similar manner to that observed if a base is present. In the light of this result, "nucleotide flipping" would be a more appropriate designation for the phenomenon. It shows that no interaction between the flipped base and the enzyme is required for base flipping. Since the conformational adaptations that lead to a stabilization of the orphan base in the DNA are different in all the MTase structures, the "relaxation" of the complex structure after nucleotide rotation is also not likely to play a vital role in the process. This presumption has been verified in part for M.*Hha*I, where it has been shown that Gln 237, the amino acid residue that is inserted into the DNA to fill the place of the flipped cytosine and that forms all contacts to the partner guanine, can be replaced by other residues with moderate loss of activity.[119] So, it is most likely that contacts of the enzyme to the phosphodiester groups on either side of the flipped nucleotide are involved in base flipping.[52] During base flipping, the phosphodiester groups are rotated and the distance between them must be transiently increased. Both of these processes could be induced or supported by contacts of the enzyme to the nonbridging oxygen atoms which do not necessarily have to persist in the final conformation where "back-rotation" is prevented by the structural adaptations of the complex. Interaction of the enzyme with the phosphodiester groups flanking the flipped base could also serve to synchronize nucleotide rotation and conformational changes of the protein.

3. The Role of DNA Methylation in Eukaryotes

3.1. DNA methylation in higher eukaryotes

In higher eukaryotes DNA methylation is the only known covalent modification of the DNA. It has a plethora of important different functions and many biological processes including human diseases are influenced by DNA methylation. So far, in metazoa only cytosine-C5 methylation has been found in DNA; this methylation mainly occurs at CG sequences, about 60 – 90% of which are modified in mammals (corresponding to 3 – 8% of all cytosine residues). Thereby, a pattern of modified and unmodified CG sites is created. Thus, unlike the case in prokaryotes where the DNA sequence alone determines the site of methylation, in eukaryotes a pattern of modified and nonmodified recognition sites exists (Figure 3). Since methylation takes place in both DNA strands at palindromic sites, DNA replication transforms the pattern of unmodified and fully methylated sites into a pattern comprising unmodified and hemimethylated CG sites. Therefore, after DNA replication the information encoded in the pattern of DNA methylation is still available and the initial pattern of methylation can be reestablished by a maintenance MTase that specifically modifies hemi-

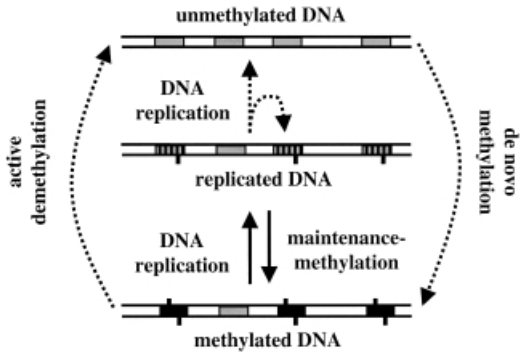


Figure 3. *Dynamics of DNA methylation in mammals. Unmethylated CG sites are indicated by gray boxes. Fully methylated CG sites are colored black, with the methyl groups are represented by two small, black rectangles. Hemimethylated CG sites are colored gray and black and carry only one black rectangle.*

methylated but not unmethylated target sites. Therefore, the pattern of DNA methylation is stable over cell divisions and somatically inherited (occasionally even through the germ line). Nevertheless, it can be edited, either by de novo methylation or by demethylation, which makes DNA methylation a unique way to encode information in a stable but reversible manner. Given these properties, DNA methylation is ideally suited to control processes like cellular differentiation or development.

Certainly, DNA methylation is a central mechanism in epigenetic inheritance (that is, transmission of inheritable information not encoded in the DNA sequence). Epigenetics is a rapidly developing field with a great impact on gene regulation, development, and important therapeutic perspectives[120]. Other epigenetic mechanisms include covalent modifications of the histone proteins (like acetylation, methylation, and phosphorylation) which influence the accessibility of the chromatin and gene expression[121] and the expression of noncoding RNA molecules that lead to specific gene silencing.[122] Although the detailed relation of DNA methylation, silencing by noncoding RNAs, and histone modification is not entirely clear, evidence is accumulating that these processes are interdependent and act in a synergistic fashion.[123–128] However, DNA methylation, so far, is the only epigenetic process whose mechanism of inheritance (maintenance methylation) is understood at a molecular level.

Currently, it is unknown how the pattern of DNA methylation is generated and edited by a combination of specific de novo methylation and demethylation. In principle, there are four alternative mechanisms which do not exclude each other: 1) The MTases and demethylases could have an intrinsic specificity for certain target regions, as is the rule for prokaryotic DNA MTases. However, so far there is no evidence in favor of such a mechanism being available. 2) The MTases and demethylases could be directed by other proteins to the sites of methylation and demethylation. This model is supported by the observation that all active DNA MTases known in mammals have large N-terminal regions that interact with other proteins and are involved in targeting the enzymes to certain cellular locations (see below). 3) Binding of other proteins could protect regions of the DNA from de novo methylation or demethylation, a model that is also supported by convincing experimental evidence.[129–132] 4) The accessibility of the chromatin might control which regions of the DNA are subject to methylation or demethylation. This model is supported by the general relation between DNA methylation and chromatin remodeling. Moreover, chromatin boundary elements which restrict methylation and demethylation to certain domains of the DNA have recently been identified.[133]

Unlike in mammals, DNA methylation in plants is also observed at CNG sites (where N is any base), a fact that suggests the presence of two independent codes of DNA methylation in this kingdom of life.[134] It has been shown recently, that CNG methylation, like CG methylation (see below), serves to repress gene expression.[135] Methylation of non-CG (and non-CNG) sequences has also been observed in mammals, in particular in early phases of development (see ref. [136], and references cited therein). The biological relevance of this type of methylation, which does not have a straightforward mechanism of maintenance as it does not occur at palindromic sites, is not known yet. Finally, very recently evidence for methylation of CCWGG sites (where W is A or T) in human DNA has been reported (see ref. [137], and references cited therein).

3.2. DNA demethylation

In mammals the pattern of DNA methylation is set not only by DNA MTases but also by DNA demethylation.[138–140] In this process 5-methylcytosine is converted back into cytosine. Demethylation occurs if a second round of DNA replication takes place before maintenance methylation has been completed (passive demethylation). In this case one of the daughter strands is hemimethylated and one is unmethylated. Therefore, passive demethylation is a rather slow process (5 replication cycles are required to reduce the methylation level to $< 5\,\%$). In contrast to this, in active demethylation the removal of 5-methylcytosine is initiated by a demethylase enzyme. So far, only one active mechanism of DNA demethylation has been established biochemically. In this mechanism the methylated base is excised by a glycosidase and the resulting lesion is repaired by the cellular repair machinery. However, the observation of a genome-wide demethylation within hours that is not accompanied by large amounts of DNA replication and repair[141] suggests that it is also possible to convert methylcytosine directly into cytosine, although breaking a C–C bond is energetically difficult. This process could follow a mechanism that resembles the methylation reaction itself and would be initiated by a nucleophilic attack of the enzyme on the C6 position of the 5-methylcytosine and formation of a covalent enzyme – DNA intermediate.[142] Alternatively, an oxidative mechanism appears possible, that would finally release the methyl group as carbon dioxide.

3.3. Dynamics of the DNA methylation pattern in mammals

During the development of mammals DNA methylation and demethylation are orchestrated to achieve major changes in the methylation pattern.[143] The DNA is methylated in oocytes and sperms. After fertilization, before the onset of the first cleavage divisions, the paternal DNA is rapidly demethylated by an active

mechanism within hours. The maternal genome is also demethylated but by a passive mechanism during the first cleavage divisions, because Dnmt1 is excluded from the nucleus in this period. Despite an almost complete demethylation of large parts of the genome the imprint is maintained even during this phase, which requires the presence of Dnmt1o, an oocyte specific isoform of Dnmt1.[144] Later in embryogenesis, remethylation by de novo methylation of DNA occurs. In this phase, high expression levels of the Dnmt3a and 3b de novo MTases are observed. In germ cell development a similar reprogramming (almost complete demethylation followed by remethylation) takes place which erases the original and sets a new, sex-specific imprint on the DNA.

3.4. Deamination of 5-methylcytosine and CG islands

The presence of 5-methylcytosine in the DNA of higher eukaryotes has the important disadvantage that it is potentially mutagenic, because it promotes deamination of cytosine to uracil.[145] Hydrolytic deamination of cytosine residues in double-stranded DNA occurs at a relatively slow rate with a half life of about 30,000 years. It is initiated by a hydroxy ion attack on the C4 position in a cytosine base that is protonated at the N3 position. 5-methylcytosines are deaminated 2 – 4 times more rapidly than cytosines.[146, 147] The rate of deamination of cytosine and 5-methylcytosine residues can be accelerated by DNA C-MTases, in particular at low AdoMet concentrations.[148–151] This is understandable in the light of the mechanism of DNA MTases, which rotate the target base out of the DNA helix such that it becomes more accessible and protonate the extrahelical cytosine at the N3 position. However, similar enzyme-catalyzed deamination reactions are not observed with the human enzyme Dnmt1.[152] If this difference is due to the different assay systems used in the various reports or if it reflects a special property of the Dnmt1 enzyme remains to be seen.

The mutational damage of deamination of 5-methylcytosine is augmented by the lower repair efficiency of T/G mismatches (arising from deamination of 5-methylcytosine) as compared to U/G mismatches (arising from deamination of unmethylated cytosine). Uracil is a nonnatural base in DNA and is efficiently recognized and excised by the repair enzyme uracil-DNA glycosylase.[153] In contrast, thymidine is a natural component of DNA that cannot be repaired by a global mechanism. However, at least two repair enzymes exist for T/G mismatches in a CG context: one mismatch specific thymine glycosylase[154] and the 5-methylcytosine binding protein MBD4.[155] Despite these repair systems, methylation-mediated mutagenesis events apparently had a strong influence on the genome of vertebrates, because most CG sequences have been removed from the genome during evolution. In mammals, CG dinucleotides are 5 – 10 fold underrepresented with respect to their expected normal frequency. No CG depletion is observed in CG islands, which are found in the promotor regions of about 60% of all genes including most housekeeping genes and 40% of the tissue-specific genes.[156] The CG sequences in CG islands are not methylated in normal cells and in the germ line.[157–159]

Deamination of 5-methylcytosine is also a predominant cause of mutations observed in somatic tissues. For example, methylated CG dinucleotides are the single most important mutational target in the p53 tumor-suppressor gene,[160] which is affected in approximately 50% of all human cancers. Given this mutational burden, the fact that DNA methylation is still observed in most animals, plants, and fungi indicates that it must play important biological roles.

3.5. The functions of DNA methylation in higher eukaryotes

Methylation of CG sites is involved in gene regulation, with methylation of CG sites in the promotor regions of genes usually leading to a reduction of gene expression.[161, 162] Repression of gene expression occurs at three levels of control: 1) Several transcription factors, like AP-2, c-Myc/Myn, E2F, and NFκB, are not able to bind to methylated target sites. 2) DNA methylation recruits 5-methylcytosine binding proteins that act as repressors of gene transcription. 3) DNA methylation triggers histone deacetylation and thereby induces chromatin condensation which leads to a strong and stable repression of gene expression. Therefore, DNA methylation is a general tool for gene regulation that can be considered as an evolutionary device involved in many different biological functions.[11] In mammals, genomic imprinting and X-chromosome inactivation are mediated by DNA methylation, as described in detail in the following sections. However, these are specialized processes that, in this form, are characteristic for mammals. Therefore, these functions cannot explain the almost ubiquitous occurrence of DNA methylation in biology. Three other models on the general function of DNA methylation have been put forward: It has been suggested that DNA methylation contributes to the protection of the genome against selfish genetic elements, that it is involved in developmental processes, or that it is a tool to reduce transcriptional noise in organisms with a complex genome. All three models agree that DNA methylation serves as a general mechanism for gene repression. However, they attribute different main functions to this process: in the noise-reduction model general gene repression itself is the main function of DNA methylation, in the genome-protection model silencing of selfish genetic elements by methylation is considered the main function of methylation, and in the development model gene and cell type specific methylation and demethylation of promotor regions is the most important function of DNA methylation.

3.5.1. DNA methylation and genomic imprinting

In mammals, a small number of genes (≈50) carry an imprint, that allows the paternal and maternal copies of these genes to be distinguished. Imprinted genes are only expressed from one chromosome, for example, only the maternal copy of H19 and only the paternal copy of Igf2 are active.[163, 164] The involvement of DNA methylation in imprinting is shown by several observations: differences in the pattern of methylation of imprinted genes are observed in many cases, Dnmt1 knock-out mice which show a strongly reduced level of DNA methylation have lost their imprint,[165] and imprinting in plants also requires DNA methyl-

ation.[166] The mechanism of imprinting of the H19/Igf2 pair has recently been determined: it involves one enhancer whose effect is directed to the nearby H19 in the maternal genome. Its influence on the Igf2 gene is prevented by binding of the CTCF protein between the enhancer and the gene; CTCF functions as a chromatin boundary element. In the paternal genome, the H19 promotor sequence and the CTCF binding site are methylated and thereby inactivated. Under these conditions, CTCF does not bind and the enhancer acts on the Ifg2 gene.[167–169] The imprint is transmitted as a certain pattern of methylation of imprinted genes which is set in the gonads during spermatogenesis and oogenesis. After fertilization the imprint persists in the somatic cells for the whole life of the individual. In contrast, in the germ cells the imprint must be erased and reset, because the new imprint depends on the sex of the individual. So, the paternal and maternal copies of imprinted genes will obtain a female imprint in oocytes and a male imprint in sperms. Therefore, this process depends on both characteristic features of DNA methylation: stable and inheritable silencing of genes and the possibility to alter the encoded information if required.

3.5.2. DNA methylation and X-chromosome inactivation

In mammals, females carry two X chromosomes while males have only one. Therefore, dose compensation is required for the genes encoded on the X chromosome. To this end, one X chromosome is inactivated in a process that involves specific expression of the Xist RNA from the inactivated chromosome, as well as dense methylation and histone deacetylation of the inactivated chromosome.[170, 171] All these processes act synergistically to ensure the extreme stability of the inactive state,[124] but it is known that DNA methylation is required for X-chromosome inactivation[172, 173] and it has been shown that mouse embryos deficient in Dnmt1 do not show stable maintenance of X-chromosome inactivation.[174] The initial choice of which X chromosome is inactivated is made in early embryogenesis. In embryonic tissues one chromosome is selected for inactivation in a random fashion whereas in extraembryogenic tissues the paternal chromosome is always chosen for inactivation. Therefore, X-chromosome inactivation is also an example of parental imprinting. Interestingly, X-chromosome inactivation, like imprinting, depends on the CTCF repressor protein that detects the methylation state of the DNA.[268, 269]

3.5.3. DNA methylation and protection from selfish genetic elements

The human genome is challenged by different types of selfish genetic elements, like transposons, retrotransposons, and viruses. Surprisingly $>40\%$ of our genome is made of such selfish DNA.[175] Since a random integration of transposable elements into the genome is an important source for mutations, prevention of transposon mobility by transcriptional silencing of these sequences is crucial for life. It has been suggested that DNA methylation is involved in protection of the genome against genetic parasites.[176] Today, a role of DNA methylation in genome protection is beyond reasonable doubt: transposons and other repetitive DNA sequences are usually relatively rich in CG sequences and heavily methylated[175] and an increase in transcription of transposons is observed in Dnmt1 knock-out embryonic stem cells[177] and cell lines.[178] These cells also show an elevated rate of mutations involving gene rearrangements.[179] Similar observations were made in plants where reduction of DNA methylation leads to the expression and mobilization of transposons.[180] Moreover, the function of DNA methylation in the fungus *Ascobulus immersus* is also to protect the genome against foreign DNA in a process called MIP (methylation induced premitotically).[181]

However, other mechanisms also serve to protect the genome against parasitic DNA. One such mechanism is RNA interference.[122, 182] At least in plants, RNA interference appears to be connected to DNA methylation,[183–185] a fact suggesting the presence of a programmable DNA methyltransferase which so far has not been identified. However, RNA interference is also observed in *Saccharomyces cerevisiae* and *Caenorhabditis elegans*, which do not methylate DNA. Thus, at least in these species, methods for genome protection exist that are independent of DNA methylation. In other species, genome protection is unlikely to be a function of DNA methylation because invertebrates, insects, and many fungi do not show the dense, genome-wide methylation that is required for genome protection, and it has been shown that repetitive DNA is not methylated in *Chiona intestinalis*, an invertebrate chordate.[186] Thus, whereas it is beyond doubt that DNA methylation contributes to genome protection in many species including mammals and plants, it is unlikely that this is the only function of DNA methylation.

3.5.4. DNA methylation and the reduction of transcriptional noise

It has been proposed that DNA methylation serves to reduce transcriptional noise in organisms with complex genomes.[187] This model is supported by the findings that in mammals most CG sequences are methylated and therefore the default expression status of most genes is "off". In apparent agreement to this model, in Dnmt1 (and p53) knock-out cell lines the expression of 4 – 10% of the detectable genes is induced, while only 1 – 2% are repressed;[178] this, of course, does not rule out a specific role of DNA methylation and demethylation during development.

However the noise-reduction model is based on older estimates of the number of genes in different species, which recently have been revised considerably. According to the results of the genome-sequencing projects, the number of genes in *C. elegans* (19000) and *Drosophila melanogaster* (13000) is not dramatically lower than in *Arabidopsis thaliana* (20000) and humans (35000 – 40000). Nevertheless, in humans and *A. thaliana* most of the CG sequences are methylated (in *A. thaliana*, CNG sequences are also modified) but *C. elegans* does not have DNA methylation and *D. melanogaster* does not carry genome-wide methylation, although some methylcytosine is present.[188, 189] Although the best way to measure the "complexity" of a genome may be disputable, the question arises of whether a

two – threefold increase in the total number of genes when comparing *C. elegans* and humans makes an additional level of gene control—with all its disadvantages (see above)—necessary.

3.5.5. DNA methylation and development

It was suggested more than 25 years ago that DNA methylation as an inheritable but flexible epigenic mark is involved in development.[190, 191] Although this is still under debate,[12] there now is firm evidence from in vivo and knock-out studies in different species that DNA methylation is involved in gene silencing during development. Dnmt1 and Dnmt3b knock-out mice die during early embryogenesis, Dnmt3a knock-out mice are runted and die shortly after birth.[192, 193] It is interesting that Dnmt1 knock-out embryonic stem cells are viable despite a reduced level of methylation but die after induction of differentiation,[178, 192] a fact which also supports a role of DNA methylation in development. *Xenopus laevis* eggs depleted in Dnmt1 show disregulation of gene expression and premature activation of genes.[194] In zebra fish DNA methylation is required during gastrulation and somite patterning.[195]

Recently, evidence has been provided that protein binding protects the DNA from de novo methylation.[131, 132] This result suggests that DNA methylation could be a tool to freeze the transcriptional state of a cell for the future and transmit it to following cell generations. In agreement with this model, it has been shown that demethylation of DNA is involved in the activation of the M-lysozyme gene during differentiation of murine macrophage cells[196, 197] and the liver specific tyrosine aminotransferase (Tat) gene.[198] The Tat gene is selectively activated in response to the release of glucocorticoid hormones before birth. After the glucocorticoid stimulus, a rapid chromatin remodeling is observed within minutes and this is followed by demethylation after 2 – 3 days. In contrast to the chromatin remodeling, the demethylation is stable after hormone withdrawal. Furthermore, demethylation recruits additional transcription factors to the DNA. This result shows that demethylation serves to memorize a regulatory event during development and that it contributes to the fine tuning of gene expression. An example of the involvement of DNA methylation in tissue-specific gene expression is the endothelial nitric oxide synthase. The exclusive expression of this enzyme in the endothel is accompanied by a selective demethylation of its promoter region only in endothelial cells (P. A. Marsden, personal communication). It is likely that many more examples of tissue-specific alterations of the DNA-methylation pattern will be found during the genome-wide analyses of the methylation status of gene promotor regions in different tissues that are currently in progress.

3.6. DNA methylation and disease

Alterations in the pattern of DNA methylation are frequently observed in cancerous tissues.[199, 200] Most often a general hypomethylation of the DNA is accompanied by a hypermethylation at specific loci. Both of these processes can have cancer-promoting effects. Hypomethylation leads to genomic instability often observed in cancer cells and may lead to an activation of retrotransposons. In addition, the expression of oncogenes may become stimulated. Hypermethylation is often found in the promotor regions of tumor-suppressor genes and there it has the same effect as a mutation in the gene itself. Depending on the tumor type, epigenetic inactivation of tumor-suppressor genes can be the predominant method of functional gene loss in tumor cells. It should be noticed that methylation defects in cancer cells, at least in principle, are reversible, which makes DNA methylation a promising target for a new generation of anticancer drugs. A genome-wide demethylation is also observed during aging and may contribute to the loss of gene regulation in aging cells.[201]

Two genetic diseases are due to mutations in proteins related to DNA methylation.[14] ICF is a rare autosomal, recessive disease that is associated with immunodeficiency, centromere instability, and facial abnormalities. It has been shown that mutations in the *dnmt3b* gene are responsible for this syndrome. In ICF patients, the DNA is hypomethylated mainly at the satellite regions 2 and 3 of chromosomes 1, 9, and 16.[193, 202, 203] Moreover, mutations in the MeCP2 protein cause Rett syndrome, a severe neurological regression starting 6 – 18 months after birth.[204] MeCP2 specifically recognizes 5-methylcytosine and thereby is one of the proteins that read the methylation code and translate it into biological effects.[205] Rett syndrome is the second most common genetic disorder among Caucasian females, occurring once every 10 000 – 15 000 live births. It is X-chromosome linked and dominant, such that patients have one mutated and one wild-type allele. Males having an affected X-chromosome are most likely not viable.

4. DNA Methyltransferases in Higher Eukaryotes

The distribution of DNA MTases in different species may also shed light on the function of DNA methylation. While the unicellular fungus *S. cerevisiae* does not have DNA methylation, other multicellular fungi do have it (*Neurospora crassa*, *Ascobulus immersus*). DNA methylation is found in some insects, but the amount of methylation and the number of enzymes involved in methylation is quite low. (*D. melanogaster* contains only one putatitve cytosine-C5 MTase and the amount of 5-methylcytosine is about tenfold lower than in mammals.) Evidence for DNA methylation also exists for several other insects (see ref. [188], and references cited therein) and at least one annelid.[207] Higher levels of CG methylation and homologues of the Dnmt1 maintenance MTase are found in the deuterostomia branch of Coelomatae, including *Paracentrotus lividus* (sea urchin), *Danio rerio* (zebra fish), *Xenopus laevis* (claw frog), and mammals. In vertebrates, a genome-wide methylation is observed, whereas in invertebrates the genomes are predominantly unmodified with a minor fraction of methylated DNA.[208] In contrast to all these examples, the nematode *C. elegans* does not have any DNA methylation. So far, this is the only case of a multicellular organism clearly devoid of DNA methylation. Interestingly, nematodes have a strictly deterministic way of development where the fate of each cell is genetically determined. In contrast

most animals follow a flexible way of ontogenesis that is characterized by a position-dependent development of cells and capabilities of regeneration. It is noticable that nematodes show this unusual way of development and that they have renounced DNA methylation, which is an important tool for development in other species. Perhaps, DNA methylation is required for flexible development as an epigenetic device for cell programming but it is dispensable for a deterministic development.

4.1. Mammalian DNA MTases

So far, three active DNA MTases (Dnmt1, Dnmt3a, and Dnmt3b) and one candidate protein (Dnmt2) have been identified in mammals (Table 1, Figure 4). The properties of these enzymes are reviewed in the next sections.

4.1.1. Dnmt1

Dnmt1 is a very large protein which comprises 1620 amino acid residues (Table 1, Figure 4). In vivo, alternative start codons as well as different splicing isoforms have been described.[209–212] The Dnmt1b isoform carries 48 additional amino acid residues in the very N-terminal part of the protein. Its methylation activity is similar to that of the previously described form of Dnmt1 in vitro.[211, 212] An oocyte-specific form of Dnmt1 (Dnmt1o) lacks about 150 amino acid residues at the N terminus. It is also active in vivo and in vitro.[209, 210] This isoform has been implicated in the maintenance of the parental imprint during the first cleavage divisions, where a global demethylation occurs.[144]

The C-terminal part of Dnmt1 forms the catalytic domain and contains all the amino acid sequence motifs characteristic for prokaryotic DNA cytosine-C5 MTases (Figure 4).[213, 214] The N-terminal part of the enzyme contains a nuclear localization signal[1] and a region that directs the enzyme to replication foci.[215] Within this region a major phosphorylation site of Dnmt1 has been identified.[216] Furthermore, the N-terminal part of Dnmt1 contains a zinc-binding domain[217, 218] of the CXXC zinc-finger type. This region contains eight conserved cysteine residues and binds to zinc.[218] Similar sequences are observed in other proteins that bind to 5-methylcytosine and HRX-related transcription factors. The N-terminal part of Dnmt1 also contains a region homologous to the polybromo-1 protein from chicken[1] which is built up of two BAH domains. It might be involved in protein – protein interactions. In Dnmt1, it is also implicated in directing the enzyme to replication foci.[219] The N-terminal part of Dnmt1 has been shown to interact with several other proteins, like PCNA,[220] the transcriptional co-repressor DMAP1,[221] the histone deacetylases HDAC1[222, 223] and HDAC2,[221] and the transcription factor E2F1,[223] as well as the Rb tumor-suppressor protein.[223] Thus, it appears that the N-terminal part of Dnmt1 serves as a platform for assembly of various proteins involved in chromatin condensation and gene regulation. However, it should be noticed that so far it has not been shown in any of these cases that these interactions have a role in vivo whatsoever.

Dnmt1 shows a significant preference for hemimethylated DNA,[224–228] a fact suggesting a role for the enzyme in maintenance methylation in vivo (Figure 5). Dnmt1 knock-out embryonic stem cells and mice show a strongly reduced level of DNA methylation but cells still display de novo MTase activity.[192] The Dnmt1 knock-out animals show a recessive, lethal phenotype; embryos were stunted, delayed in development, and did not survive past mid-gestation. In contrast, Dnmt1 knock-out

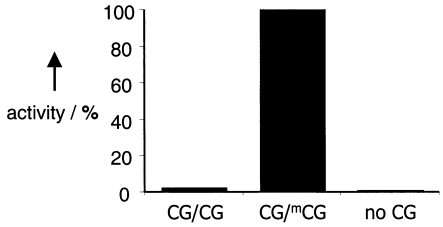


Figure 5. *Specificity of Dnmt1. Catalytic activities are determined with duplex oligonucleotide substrates containing one hemimethylated CG site, one unmethylated CG site, or no CG sites at all (data were taken from ref. [228]).*

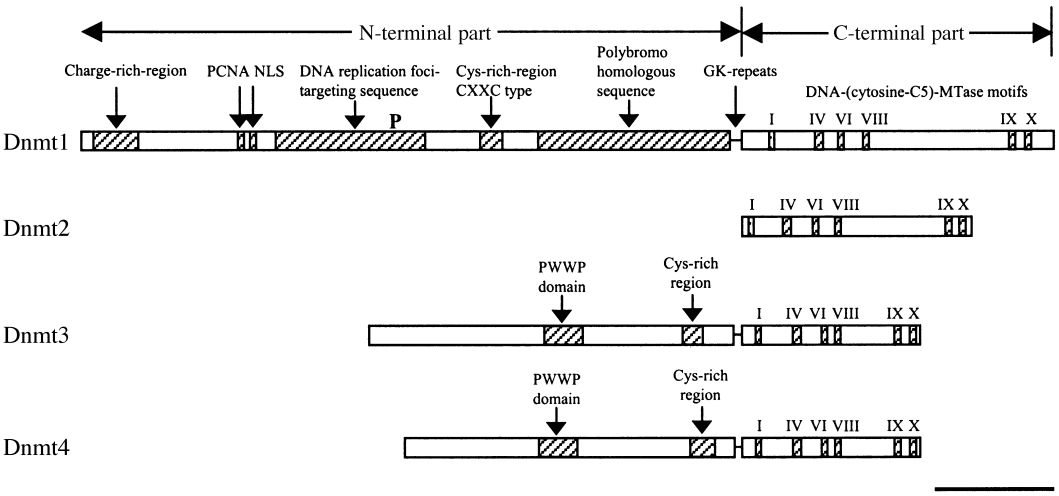


Figure 4. *Schematic drawing of the primary structures and domain arrangement of Dnmt1, Dnmt2, Dnmt3a, and Dnmt3b. The scale bar (bottom right) indicates the length corresponding to 200 amino acid residues.*

embryonic stem cells lines are viable, despite the reduced level of DNA methylation. Several observations indicate that Dnmt1 also participates in cell cycle control in mammals. It has been observed that Dnmt1 is involved in the Jun/Fos signal transduction pathway as a downstream effector.[229] Moreover, depletion of cells from Dnmt1 leads to p53-driven apoptosis[178, 230] and to inhibition of DNA replication.[231] If these effects are due to demethylation of the DNA or if Dnmt1 has additional functions in the cell that do not depend on its catalytic activity is not yet known.

The catalytic domain of Dnmt1 is under tight allosteric control by the N-terminal part of the enzyme, as shown by the observations that the isolated catalytic domain does not methylate DNA in vivo or in vitro.[228, 232, 233] Interestingly, Dnmt1 is allosterically activated to methylate unmodified target sites by binding to methylated DNA.[228, 234, 235] Binding of methylated DNA occurs within the N-terminal part of the enzyme, most likely to the zinc domain, which forms a direct protein–protein contact to the catalytic domain of the enzyme.[228] After allosteric stimulation, Dnmt1 has a similar activity on unmethylated and hemimethylated DNA, which suggests that this enzyme could also have a role in de novo methylation of DNA (Figure 6). The allosteric activation mechanism of Dnmt1 makes DNA methylation behave in an all-or-none fashion, because some methylation will always attract more methylation. This explains the observation that methylation tends to spread from heavily methylated regions of the DNA into neighboring unmethylated regions (spreading of methylation[236]). After extensive spreading, only completely unmethylated and fully methylated regions of the DNA that are separated by chromatin boundary elements coexist.[133] This all-or-none behavior might increase the efficiency of switching on and off gene expression by DNA methylation.

4.1.2. Dnmt2

Dnmt2 belongs to a large family of proteins conserved from *Schizosaccharomyces pombe* to man.[237, 238] The Dnmt2 enzymes are relatively small (the mouse enzyme comprises 391 amino acid residues) and resemble prokaryotic MTases in that they do not have a large non-MTase N-terminal domain (Table 1, Figure 4). These proteins contain all the sequence motifs characteristic for cytosine-C5 MTases. However, catalytic activity has not yet been shown for any protein belonging to this family[238, 239] and a knock-out of Dnmt2 in embryonic stem cells does not show a phenotype.[240] In accordance with the sequence homology of Dnmt2 to prokaryotic MTases, the structure of this protein in complex with AdoHcy strongly resembles that of the M.*Hha*I enzyme (both structures can be superimposed with a root mean square deviation of <1 Å for the C^{α} positions).[239] Interestingly, Dnmt2 was found to interact with DNA in a denaturant-resistant, most likely covalent manner; Dnmt2–DNA complexes could not be disrupted in sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE).[239] However, it is not known if irreversible binding to DNA also occurs in vivo or if it only takes place in vitro, because the enzyme is not able to complete a catalytic cycle of DNA methylation such that the covalent intermediate is trapped. A role of Dnmt2 in one of the methylation dependent processes in the cell can only be excluded by generation of a Dnmt2 knock-out mouse strain and careful investigation of the animals over several generations. In any case, Dnmt2 is the first MTase-like protein from higher eukaryotes whose structure has been solved.

S. pombe contains a gene (*pmt1*) for a Dnmt2 homologue that, like the other Dnmt2 proteins, is inactive in vitro.[241] However, the Pmt1 protein could be transformed into an active DNA MTase that modified DNA at CCWGG sequences in vitro by just one mutation.[242] This result, which certainly needs further validation, suggests that other Dnmt2 proteins also might recognize CCWGG sequences.

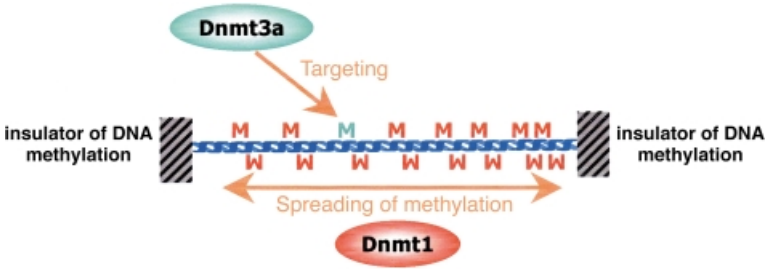


Figure 6. *Model of functional cooperation of Dnmt3a and Dnmt1 in de novo methylation of DNA. This model is based on two central observations: 1) Dnmt3a does not modify DNA in a processive reaction, which makes a fast methylation of one domain of the DNA by this enzyme difficult, and 2) Dnmt1 is stimulated by binding to methylated DNA. M = methyl group.*

4.1.3. Dnmt3a and Dnmt3b

The Dnmt3 MTases were discovered in 1998.[243] Dnmt3a and 3b are heavily expressed in embryonic tissues, whereas only low expression is observed in differentiated cells.[243, 244] The gene products from mice comprise 908 and 859 amino acid residues, respectively, and share 36% amino acid sequence identity with each other (>80% in the C-terminal part of the proteins; Table 1, Figure 4). There are several alternative splice variants of Dnmt3b,[243] two of which are active, one is not.[245] The Dnmt3a and 3b enzymes contain a cysteine-rich region that is similar to the ATRX zinc-finger. Moreover, they contain a PWWP domain which occurs in proteins that play a role in cell growth and differentiation. The structure of the PWWP domain from Dnmt3b has been solved and it has been shown that this domain interacts with DNA.[270] The ATRX domain of Dnmt3a associates with the histone deacetylase HDAC1.[246] Moreover, Dnmt3a interacts with RP58, a DNA-binding transcriptional repressor protein found at transcriptionally silent heterochromatin.[246] Mutations in human Dnmt3b cause ICF, a severe, hereditary disease.[193, 202, 203] In ICF patients, hypomethylation of the DNA is observed at the classical satellite regions 2 and 3 of chromosomes 1, 9 and, 16, a fact suggesting that Dnmt3b is involved in the methylation of these regions.[193, 202, 203] Transgenic mice lacking Dnmt3a and Dnmt3b, singly and in combination, are hypomethylated and die in the embryonic stages (Dnmt3a$^{-/-}$/Dnmt3b$^{-/-}$, Dnmt3b$^{-/-}$) or shortly after birth (Dnmt3a$^{-/-}$); this

indicates the critical role of these enzymes during development.[193] The observation that the phenotype of a double knock-out is more severe than any of the single knock-outs suggests that Dnmt3a and 3b have partially overlapping functions. However, both proteins are essential, neither of them can completely replace the other. Therefore, there must be also distinctive roles for both enzymes. Since in ICF patients and Dnmt3b knock-out mice demethylation is mainly observed at satellite 2 and 3 DNA, one role of Dnmt3b might be methylation of these repetitive parts of the human genome.

Baculovirus-expressed Dnmt3a and Dnmt3b proteins, as well as *E. coli*-expressed Dnmt3a, methylate CG dinucleotides without significant preference for hemimethylated DNA,[206, 243] which suggests a role of these enzymes in de novo methylation. De novo methylation of DNA by Dnmt3a and 3b was also demonstrated in vivo after expression in human cell lines[247] and by expression of Dnmt3a in transgenic *D. melanogaster*.[248] Dnmt3a and 3b have a lower preference for CG sites than Dnmt1 and also methylate DNA at asymmetric non-CG sites (CA > CC, CT).[136, 206, 245] The biological relevance of this non-CG methylation is not known so far.

Dnmt3a expressed in *E. coli* catalyzes the methylation of DNA in a distributive manner.[206] This result, together with the observation that Dnmt1 is allosterically activated by binding to methylated DNA, suggests the interesting model that Dnmt3a and Dnmt1 might functionally cooperate in de novo methylation of DNA. In this model, Dnmt3a would be directed toward a target region of the DNA and introduce one or a few methyl groups. Then, Dnmt1 might be recruited and spread the methylation over the whole domain of the DNA (Figure 6).[206, 228]

4.2. Plant DNA MTases

DNA methylation in plants also occurs at CNG sites, leading to the modification of up to 40% of all cytosine residues (6% in *Arabidopsis thaliana*).[134, 249, 250] A much higher number of DNA MTases are known in plants than in animals (compare, for example, 11 in *A. thaliana* with 4 in mice;[134] Table 1). In *A. thaliana* four enzymes of the Dnmt1 family, one enzyme of the Dnmt3 family, one member of the Dnmt2 family, three members of the family of chromomethylases, and one member of the masc1 family, an MTase family first identified in the fungus *Ascobulus immersus*, are known. Interestingly, the Dnmt3 homologue in *A. thaliana* contains the conserved motifs of cytosine-C5 MTases in a different order (a domain-rearranged methyltransferase or DRM),[251] which suggests a circular permutation of the protein, similar to that observed in prokaryotic N-MTases.[40, 41] Chromomethylases are only present in plants. They are characterized by the insertion of a chromodomain between the conserved motifs II and IV of the DNA MTase.[252] Chromodomains were first identified in Drosophila proteins where they direct proteins to the heterochromatin. Chromomethylases have been shown to be responsible for methylation at CNG sites, which only occurs in plants and contributes to epigenetic gene silencing.[135, 253, 254]

The abundance of DNA MTases in plants, as well as the presence of at least two independent methylation codes (CG and CNG), might be related to the fact that plants as sessile organisms cannot escape from unfavorable environmental conditions. Therefore, they require a fine adaptation of their genetic program to the microenvironment at each habitat. It is clearly favorable if this adaptation can be inherited since most offspring will seed nearby. However, different adaptations are required at different locations, such that any change should be reversible—conditions that are ideally met by epigenetic changes mediated by DNA methylation. In fact, it is well documented that DNA methylation is involved in genetically stable epimutations, that is, inheritable but also reversible phenotypic changes in plants. For example, methylated alleles of the *A. thaliana* SUPERMAN locus in genetic crosses behave like unstable mutant alleles, with a frequency of spontaneous reversion of > 1%.[255] Different epimutations affect other morphological traits like leaf structure, time of flowering, or flower structure.[249, 256, 257]

4.3. DNA MTases in fungi

Fungi typically have low levels of DNA methylation, approximately 1.5% of all cytosine residues are modified in *Neurospora crassa*[258] and 0.25% in *Aspergillus flavus*.[259] In fungi methylation is restricted mostly to transposons and other repeats.[250] In some species like *S. cerevisiae* and *S. pombe*, DNA methylation is not detectable at all. Nevertheless, *S. pombe* contains a gene for a Dnmt2 homologue (Pmt1) that could be activated by a single mutation (see above).[242] This result suggests that the endogenous protein also might have some catalytic activity that escaped detection in vitro. Then, *S. pombe* might contain some 5-methylcytosine, perhaps only at certain developmental stages. In *Neurospora crassa*, DNA methylation is involved in a process called repeat induced point mutation (RIP) that serves to silence foreign DNA. In this organism a high de novo methylation activity is detectable.[260] Methylation in *N. crassa* is due to only one MTase, the dim-2 enzyme.[261] The enzymology of fungal DNA MTases is also well investigated in *A. immersus*, the first organism in which the presence of a bona fide de novo DNA MTase has been shown, the Masc1 enzyme.[262] Masc1 is a short protein, that lacks the large N-terminal extension typical for mammalian DNA MTase. It is involved in a methylation-dependent inactivation of repetitive DNA, MIP (methylation induced premitotically). The enzyme turned out to be catalytically inactive in vitro and disruption of Masc1 had no effect on viability of *A. immersus* and its maintenance methylation but it did prevent de novo methylation of DNA repeats. Knock-out cells are arrested at early stages of sexual reproduction, which indicates that the Masc1 protein is required for this developmental process, a result that again demonstrates that DNA methylation is involved in development. A second DNA MTase of *A. immersus*, Masc2, is a large protein comprising 1356 amino acid residues. This enzyme is catalytically active in vitro, but a knock-out did not reveal any phenotype in vivo.[263, 264] Therefore, an additional DNA MTase responsible for maintenance methylation must be present in *A. immersus*.

5. Future Perspectives

5.1. Prokaryotic MTases

There are numerous interesting open questions regarding the enzymology and biology of prokaryotic DNA MTases:

- Definitely more structures are needed. Structures of related MTases with different recognition sites will help to understand DNA recognition by these enzymes. Structures of MTases in complex with nonspecific DNA and with specific DNA in an unflipped state would shed light on the conformational changes of the enzyme – DNA complex during enzymatic turnover.
- The physicochemical mechanism of base flipping is not well understood. Transient interactions, like interactions of the MTases with the phosphate groups flanking the target, have to be defined and investigated by experiments.
- Prokaryotic DNA MTases provide a unique tool to study the molecular evolution of DNA recognition. However, the available sequence information needs to be complemented by enzyme – DNA structures of related enzymes with different recognition sites and by biochemical studies to elucidate the functional relevance of the contacts between the enzyme and the DNA.
- Medical applications: Adenine methylation has been shown to be important for bacteria and in certain cases involved in pathogenicity. Since, as far as we know, adenine MTases are not present in mammals, these enzymes could be an interesting drug target in the future.

5.2. Eukaryotic MTases

Many aspects of the biology of DNA methylation in higher eukaryotes are still mysterious. In addition, we are only just beginning to collect information on the enzymology of the eukaryotic MTases (apart from Dnmt1 which was cloned over 10 years ago and has been well investigated); this is certainly a prerequisite for a detailed understanding of the whole process of DNA methylation. Future issues include:

- The mechanism of DNA demethylation in eukaryotes and the enzymes involved need to be better understood.
- The generation of the pattern of DNA methylation during embryogenesis is not understood at a molecular level. Also, we are only just beginning to understand the mechanisms of directed changes of the pattern of methylation during development by de novo methylation and demethylation.
- It is still not clear if Dnmt1 has a role in de novo methylation of DNA in vivo although some evidence suggests this. Moreover, it has been discussed that it might have roles independent of its methylation activity. Many interaction partners of Dnmt1 have been found, but so far these interactions have not been shown to be important for the in vivo function of the enzyme. Also, the molecular enzymology of the Dnmt3a and 3b enzymes has to be investigated in much more detail.
- DNA methylation at nonpalindromic sites (non-CG and non-CNG) is observed in mammals and plants. In mammals, Dnmt3a is involved in this process. However, the biological function of this process is still completely unknown. Does it contribute to gene silencing? Is there a mechanism to maintain asymmetric methylation?
- Do CG and CNG (and perhaps also CCWGG) represent independent methylation codes with distinctive biological functions or are they functionally connected?
- Aberrant DNA methylation is often involved in carcinogenesis and is also implicated in aging. Since DNA methylation, in principle, is reversible, the possibility exists to correct errors in the methylation pattern and thereby revert a pathological phenotype. In fact, it is possible to inhibit carcinogenesis in mice and to revert a transformed phenotype in a cell culture by inhibition of Dnmt1 and reduction of DNA methylation.[265, 266]
- Understanding the biology of DNA methylation is very important for cloning experiments, because if mammals are cloned from a differentiated cell, the methylation pattern must be reset. It is likely, that methylation errors are causative to the low yields of cloning of mammals and developmental abnormalities of cloned embryos.[143, 267]

Work in the authors' laboratory has been supported by the Deutsche Forschungsgemeinschaft (JE 252/2-3, JE 252/1-2), the BMBF (BioFuture program), the European Union (BIO4 CT98 0328), and the Justus-Liebig Universität. It is a pleasure to thank A. Pingoud for his advice and generous support. Comments and suggestions on the manuscript by C. Beck, H. Gowher, and A. Hermann are gratefully acknowledged.

[1] T. H. Bestor, G. L. Verdine, *Curr. Opin. Cell Biol.* **1994**, *6*, 380.
[2] X. Cheng, *Curr. Opin. Struct. Biol.* **1995**, *5*, 4.
[3] X. Cheng, *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 293.
[4] E. B. Faumann, R. M. Blumenthal, X. Cheng in *S-adenosylmethionine-dependent methyltransferases: Structures and functions* (Eds.: X. Cheng, R. M. Blumenthal), World Scientific Publishing, Singapore, **1999**, p. 1.
[5] X. Cheng, R. J. Roberts, *Nucleic Acids Res.* **2001**, *29*, 3784.
[6] M. Noyer-Weidner, T. A. Trautner in *DNA Methylation: Molecular Biology and Biological Significance* (Eds.: J. P. Jost, H. P. Saluz), Birkhauser, Basel, **1993**, p. 40.
[7] D. T. F. Dryden, in *S-Adenosylmethionine-dependent Methyltransferases: Structures and Functions* (Eds.: X. Cheng, R. M. Blumenthal), World Scientific Publishing, Singapore, **1999**, p. 283.
[8] T. A. Bickle, D. H. Kruger, *Microbiol. Rev.* **1993**, *57*, 434.
[9] R. L. P. Adams, *Bioessays* **1995**, *17*, 139.
[10] P. M. Vertino in *S-Adenosylmethionine-dependent Methyltransferases: Structures and Functions* (Eds.: X. Cheng, R. M. Blumenthal), World Scientific Publishing, Singapore, **1999**, p. 341.
[11] V. Colot, J. L. Rossignol, *Bioessays* **1999**, *21*, 402.
[12] T. H. Bestor, *Hum. Mol. Genet.* **2000**, *9*, 2395.
[13] K. D. Robertson, A. P. Wolffe, *Nat. Rev. Genet.* **2000**, *1*, 11.
[14] B. Hendrich, *Curr. Biol.* **2000**, *10*, R60.
[15] P. A. Jones, D. Takai, *Science* **2001**, *293*, 1068.
[16] G. G. Wilson, N. E. Murray, *Annu. Rev. Genet.* **1991**, *25*, 585.
[17] J. Heitman in *Genetic Engineering* (Ed.: J. K. Setlow), Plenum Press, New York, **1993**, p. 57.
[18] R. J. Roberts, S. E. Halford in *Nucleases* (Eds.: S. M. Linn, R. S. Lloyd, R. J. Roberts), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, **1993**, p. 35.
[19] A. Pingoud, A. Jeltsch, *Eur. J. Biochem.* **1997**, *246*, 1.
[20] A. Pingoud, A. Jeltsch, *Nucleic Acids Res.* **2001**, *29*, 3705.
[21] R. J. Roberts, D. Macelis, *Nucleic Acids Res.* **2001**, *29*, 268.
[22] W. Messer, M. Noyer-Weidner, *Cell* **1988**, *54*, 735.
[23] F. Barras, M. G. Marinus, *Trends Genet.* **1989**, *5*, 139.
[24] I. Stancheva, T. Koller, J. M. Sogo, *EMBO J.* **1999**, *18*, 6542.

[25] M. G. Marinus, *Annu. Rev. Genet.* **1987**, *21*, 113.
[26] J. L. Campbell, N. Kleckner, *Cell* **1990**, *62*, 967.
[27] M. Lu, J. L. Campbell, E. Boye, N. Kleckner, *Cell* **1994**, *77*, 413.
[28] S. Goldman, Y. Navon, F. Fish, *Microb Pathog.* **1987**, *2*, 327.
[29] W. B. Hale, M. W. van der Woude, B. A. Braaten, D. A. Low, *Mol. Genet. Metab.* **1998**, *65*, 191.
[30] D. M. Heithoff, R. L. Sinsheimer, D. A. Low, M. J. Mahan, *Science* **1999**, *284*, 967.
[31] M. Enserink, *Science* **1999**, *284*, 883.
[32] F. Garcia-Del Portillo, M. G. Pucciarelli, J. Casadesus, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 11578.
[33] D. M. Heithoff, R. L. Sinsheimer, D. A. Low, M. J. Mahan, *Philos. Trans. R. Soc. Lond. Ser. B* **2000**, *355*, 633.
[34] C. Bucci, A. Lavitola, P. Salvatore, L. Del Giudice, D. R. Massardo, C. B. Bruni, P. Alifano, *Mol. Cell* **1999**, *3*, 435.
[35] A. Reisenauer, L. S. Kahng, S. McCollum, L. Shapiro, *J. Bacteriol.* **1999**, *181*, 5135.
[36] C. Stephens, A. Reisenauer, R. Wright, L. Shapiro, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 1210.
[37] D. C. Wahnon, V. K. Shier, S. J. Benkovic, *J. Am. Chem. Soc.* **2001**, *123*, 976.
[38] G. G. Wilson, *Methods Enzymol.* **1992**, *216*, 259.
[39] S. Kumar, W. S. Johnson, M. Tomasz, *Biochemistry* **1993**, *32*, 1364.
[40] T. Malone, R. M. Blumenthal, X. Cheng, *J. Mol. Biol.* **1995**, *253*, 618.
[41] A. Jeltsch, *J. Mol. Evol.* **1999**, *49*, 161.
[42] S. Kumar, X. Cheng, S. Klimasauskas, M. Sha, J. Posfai, R. J. Roberts, G. G. Wilson, *Nucleic Acids Res.* **1994**, *22*, 1.
[43] X. Cheng, S. Kumar, J. Posfai, J. W. Pflugrath, R. J. Roberts, *Cell* **1993**, *74*, 299.
[44] K. M. Reinisch, L. Chen, G. L. Verdine, W. N. Lipscomb, *Cell* **1995**, *82*, 143.
[45] J. Labahn, J. Granzin, G. Schluckebier, D. P. Robinson, W. E. Jack, I. Schildkraut, W. Saenger, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 10957.
[46] W. Gong, M. O'Gara, R. M. Blumenthal, X. Cheng, *Nucleic Acids Res.* **1997**, *25*, 2702.
[47] P. H. Tran, Z. R. Korszun, S. Cerritelli, S. S. Springhorn, S. A. Lacks, *Structure* **1998**, *6*, 1563.
[48] R. D. Scavetta, C. B. Thomas, M. A. Walsh, S. Szegedi, A. Joachimiak, R. I. Gumport, M. E. A. Churchill, *Nucleic Acids Res.* **2000**, *28*, 3950.
[49] S. Klimasauskas, S. Kumar, R. J. Roberts, X. Cheng, *Cell* **1994**, *76*, 357.
[50] K. Goedecke, M. Pignot, R. S. Goody, A. J. Scheidig, E. Weinhold, *Nat. Struct. Biol.* **2001**, *8*, 121.
[51] R. J. Roberts, *Cell* **1995**, *82*, 9.
[52] X. Cheng, R. M. Blumenthal, *Structure* **1996**, *4*, 639.
[53] R. S. Lloyd, X. Cheng, *Biopolymers* **1997**, *44*, 139.
[54] R. J. Roberts, X. Cheng, *Annu. Rev. Biochem.* **1998**, *67*, 181.
[55] M. Roth, S. Helm-Kruse, T. Friedrich, A. Jeltsch, *J. Biol. Chem.* **1998**, *273*, 17333.
[56] G. Schluckebier, M. Kozak, N. Bleimling, E. Weinhold, W. Saenger, *J. Mol. Biol.* **1997**, *265*, 56.
[57] M. O'Gara, X. Zhang, R. J. Roberts, X. Cheng, *J. Mol. Biol.* **1999**, *287*, 201.
[58] G. M. Adams, R. M. Blumenthal, *Biochemistry* **1997**, *36*, 8284.
[59] A. Bergerat, W. Guschlbauer, *Nucleic Acids Res.* **1990**, *18*, 4369.
[60] A. Bergerat, W. Guschlbauer, G. V. Fazakerley, *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 6394.
[61] E. G. Malygin, A. A. Evdokimov, V. V. Zinoviev, L. G. Ovechkina, W. M. Lindstrom, N. O. Reich, S. L. Schlagman, S. Hattman, *Nucleic Acids Res.* **2001**, *29*, 2361.
[62] D. V. Santi, A. Norment, C. E. Garrett, *Proc. Natl. Acad. Sci. USA* **1984**, *81*, 6993.
[63] J. C. Wu, D. V. Santi, *J. Biol. Chem.* **1987**, *262*, 4778.
[64] L. Chen, A. M. MacMillan, G. L. Verdine, *J. Am. Chem. Soc.* **1993**, *115*, 5318.
[65] M. O'Gara, S. Klimasauskas, R. J. Roberts, X. Cheng, *J. Mol. Biol.* **1996**, *261*, 634.
[66] J. C. Wu, D. V. Santi in *Biochemistry and Biology of DNA Methylation* (Eds.: G. L. Cantoni, A. Razin, A. R. Liss), Wiley-Liss, New York, **1986**, p. 119.
[67] A. L. Pogolotti, A. Ono, R. Subramaniam, D. V. Santi, *J. Biol. Chem.* **1988**, *263*, 7461.
[68] A. Jeltsch, F. Christ, M. Fatemi, M. Roth, *J. Biol. Chem.* **1999**, *274*, 19538.
[69] A. Jeltsch, *Biol. Chem.* **2001**, *382*, 707.
[70] J. M. Bujnicki, M. Radlinska, *Nucleic Acids Res.* **1999**, *27*, 4501.
[71] H. Sugisaki, K. Yamamoto, M. Takanami, *J. Biol. Chem.* **1991**, *266*, 13952.
[72] D. F. Willcock, D. T. F. Dryden, N. E. Murray, *EMBO J.* **1994**, *13*, 3902.
[73] J. B. Guyot, J. Grassi, U. Hahn, W. Guschlbauer, *Nucleic Acids Res.* **1993**, *21*, 3183.
[74] H. Kong, C. L. Smith, *Nucleic Acids Res.* **1997**, *25*, 3687.
[75] G. Schluckebier, J. Labahn, J. Granzin, W. Saenger, *Biol. Chem.* **1998**, *379*, 389.
[76] T. Friedrich, M. Roth, S. Helm-Kruse, A. Jeltsch, *Biol. Chem.* **1998**, *379*, 475.
[77] B. Holz, N. Dank, J. E. Eickhoff, G. Lipps, G. Krauss, E. Weinhold, *J. Biol. Chem.* **1999**, *274*, 15066.
[78] A. Jeltsch, M. Roth, T. Friedrich, *J. Mol. Biol.* **1999**, *285*, 1121.
[79] D. L. Wong, N. O. Reich, *Biochemistry* **2000**, *39*, 15410.
[80] H. Pues, N. Bleimling, B. Holz, J. Wolcke, E. Weinhold, *Biochemistry* **1999**, *38*, 1426.
[81] M. Roth, A. Jeltsch, *Nucleic Acids Res.* **2001**, *29*, 3137.
[82] D. N. Rao, M. G. P. Page, T. A. Bickle, *J. Mol. Biol.* **1989**, *209*, 599.
[83] L. Szilak, A. Der, F. Deak, P. Venetianer, *Eur. J. Biochem.* **1993**, *218*, 727.
[84] N. O. Reich, N. Mashhoon, *Biochemistry* **1991**, *30*, 2933.
[85] H. Gowher, A. Jeltsch, *J. Mol. Biol.* **2000**, *303*, 93.
[86] S. S. Szegedi, N. O. Reich, R. I. Gumport, *Nucleic Acids Res.* **2000**, *28*, 3962.
[87] A. A. Evdokimov, V. V. Zinoviev, E. G. Malygin, S. L. Schlagman, S. Hattman, *J. Biol. Chem.* **2001**, *276*, 30.
[88] W. M. Lindstrom, J. Flynn, N. O. Reich, *J. Biol. Chem.* **2000**, *275*, 4912.
[89] S. K. Bhattacharya, A. K. Dubey, *J. Biol. Chem.* **1999**, *274*, 14743.
[90] S. Urig, H. Gowher, A. Hermann, C. Beck, A. Humeny, A. Jeltsch, **2002**, submitted.
[91] G. Vilkaitis, E. Merkiene, S. Serva, E. Weinhold, S. Klimasauskas, *J. Biol. Chem.* **2001**, *276*, 20924.
[92] G. Nardone, J. George, J. G. Chirikjian, *J. Biol. Chem.* **1986**, *261*, 12128.
[93] M. A. Surby, N. O. Reich, *Biochemistry* **1996**, *35*, 2201.
[94] P. Renbaum, A. Razin, *FEBS Lett.* **1992**, *313*, 243.
[95] A. J. Berdis, I. Lee, J. K. Coward, C. Stephens, R. Wright, L. Shapiro, S. J. Benkovic, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 2874.
[96] M. A. Surby, N. O. Reich, *Biochemistry* **1996**, *35*, 2209.
[97] S. Cal, B. A. Connolly, *J. Biol. Chem.* **1996**, *271*, 1008.
[98] R. A. Garcia, C. J. Bustamante, N. O. Reich, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 7618.
[99] A. K. Dubey, S. K. Bhattacharya, *Nucleic Acids Res.* **1997**, *25*, 2025.
[100] T. Rasko, C. Finta, A. Kiss, *Nucleic Acids Res.* **2000**, *28*, 3083.
[101] S. Cal, B. A. Connolly, *J. Biol. Chem.* **1997**, *272*, 490.
[102] Y. V. Reddy, D. N. Rao, *J. Mol. Biol.* **2000**, *298*, 597.
[103] A. Jeltsch, T. Friedrich, M. Roth, *J. Mol. Biol.* **1998**, *275*, 747.
[104] A. Kiss, G. Posfai, G. Zsurka, T. Rasko, P. Venetianer, *Nucleic Acids Res.* **2001**, *29*, 3188.
[105] C. Beck, S. Cranz, M. Solmaz, M. Roth, A. Jeltsch, *Biochemistry* **2001**, *40*, 10956.
[106] S. Cerritelli, S. S. Springhorn, S. A. Lacks, *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 9223.
[107] N. O. Reich, C. Olsen, F. Osti, J. Murphy, *J. Biol. Chem.* **1992**, *267*, 15802.
[108] S. Serva, E. Weinhold, R. J. Roberts, S. Klimasauskas, *Nucleic Acids Res.* **1998**, *26*, 3473.
[109] B. W. Allan, N. O. Reich, *Biochemistry* **1996**, *35*, 14757.
[110] B. Holz, S. Klimasauskas, S. Serva, E. Weinhold, *Nucleic Acids Res.* **1998**, *26*, 1076.
[111] B. W. Allan, J. M. Beechem, W. M. Lindstrom, N. O. Reich, *J. Biol. Chem.* **1998**, *273*, 2368.
[112] S. Klimasauskas, R. J. Roberts, *Nucleic Acids Res.* **1995**, *23*, 1388.
[113] A. S. Yang, J.-C. Shen, J.-M. Zingg, S. Mi, P. A. Jones, *Nucleic Acids Res.* **1995**, *23*, 1380.
[114] D. R. Mernagh, I. A. Taylor, G. G. Kneale, *Biochem. J.* **1998**, *336*, 719.
[115] M. O'Gara, J. R. Horton, R. J. Roberts, X. Cheng, *Nat. Struct. Biol.* **1998**, *5*, 872.
[116] S. Klimasauskas, T. Szyperski, S. Serva, K. Wüthrich, *EMBO J.* **1998**, *17*, 317.
[117] B. W. Allan, N. O. Reich, J. M. Beechem, *Biochemistry* **1999**, *38*, 5308.
[118] M. O'Gara, J. R. Horton, R. J. Roberts, X. Cheng, *Nature Struct. Biol.* **1998**, *5*, 872.
[119] S. Mi, D. Alonso, R. J. Roberts, *Nucleic Acids Res.* **1995**, *23*, 620.
[120] A. P. Wolffe, M. A. Matzke, *Science* **1999**, *286*, 481; see also: *Science* **2001**, *293* (5532), pp. 1063 – 1102, which is completely devoted to epigenetics.
[121] T. Jenuwein, C. D. Allis, *Science* **2001**, *293*, 1074.
[122] M. Matzke, A. J. Matzke, J. M. Kooter, *Science* **2001**, *293*, 1080.

[123] P. Chiurazzi, M. G. Pomponi, R. Pietrobono, C. E. Bakker, G. Neri, B. A. Oostra, *Hum. Mol. Genet.* **1999**, *8*, 2317.
[124] G. Csankovszki, A. Nagy, R. Jaenisch, *J. Cell. Biol.* **2001**, *153*, 773.
[125] F. Magdinier, A. P. Wolffe, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 4990.
[126] J. R. Dobosy, E. U. Selker, *Cell. Mol. Life Sci.* **2001**, *58*, 721.
[127] X. Yang, D. L. Phillips, A. T. Ferguson, W. G. Nelson, J. G. Herman, N. E. Davidson, *Cancer Res.* **2001**, *61*, 7025.
[128] H. Tamaru, E. U. Selker, *Nature* **2001**, *414*, 277.
[129] M. Brandeis, D. Frank, I. Keshet, Z. Siegfried, M. Mendelsohn, A. Nemes, V. Temper, A. Razin, H. Cedar, *Nature* **1994**, *371*, 435.
[130] D. Macleod, J. Charlton, J. Mullins, A. P. Bird, *Genes Dev.* **1994**, *8*, 2282.
[131] I. G. Lin, C. L. Hsieh, *EMBO Rep.* **2001**, *2*, 108.
[132] L. Han, I. G. Lin, C. L. Hsieh, *Mol. Cell. Biol.* **2001**, *21*, 3416.
[133] D. S. Millar, C. L. Paul, P. L. Molloy, S. J. Clark, *J. Biol. Chem.* **2000**, *275*, 24893.
[134] E. J. Finnegan, K. A. Kovac, *Plant Mol. Biol.* **2000**, *43*, 189.
[135] A. M. Lindroth, X. Cao, J. P. Jackson, D. Zilberman, C. M. McCallum, S. Henikoff, S. E. Jacobsen, *Science* **2001**, *292*, 2077.
[136] B. H. Ramsahoye, D. Biniszkiewicz, F. Lyko, V. Clark, A. P. Bird, R. Jaenisch, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 5237.
[137] M. C. Lorincz, M. Groudine, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 10034.
[138] A. P. Wolffe, P. L. Jones, P. A. Wade, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 5894.
[139] J. P. Jost, E. J. Oakeley, S. Schwarz in *S-adenosylmethionine-dependent methyltransferases: Structures and functions* (Eds.: X. Cheng, R. M. Blumenthal), World Scientific Publishing, Singapore, **1999**, p. 373.
[140] C. Kress, H. Thomassin, T. Grange, *FEBS Lett.* **2001**, *494*, 135.
[141] J. Oswald, S. Engemann, N. Lane, W. Mayer, A. Olek, R. Fundele, W. Dean, W. Reik, J. Walter, *Curr. Biol.* **2000**, *10*, 475.
[142] H. Cedar, G. L. Verdine, *Nature* **1999**, *397*, 568.
[143] W. Reik, W. Dean, J. Walter, *Science* **2001**, *293*, 1089.
[144] C. Y. Howell, T. H. Bestor, F. Ding, K. E. Latham, C. Mertineit, J. M. Trasler, J. R. Chaillet, *Cell* **2001**, *104*, 829.
[145] G. P. Pfeifer, M.-S. Tang, M. F. Denissenko in *DNA methylation and cancer* (Eds.: P. A. Jones, P. K. Vogt), Springer, New York, **2000**, p. 1.
[146] T. Lindahl, *Nature* **1993**, *362*, 709.
[147] J. Shen, W. M. I. Rideout, P. A. Jones, *Nucleic Acids Res.* **1994**, *22*, 972.
[148] J. Shen, W. M. I. Rideout, P. A. Jones, *Cell* **1992**, *71*, 1073.
[149] W. Wyszynski, S. Garbara, A. S. Bhagwat, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 1574.
[150] C. C. Yang, B. K. Baxter, M. D. Topal, *Biochemistry* **1994**, *33*, 14918.
[151] M. J. Yebra, A. S. Bhagwat, *Biochemistry* **1995**, *34*, 14752.
[152] M. F. Chan, R. van Amerongen, T. Nijjar, E. Cuppen, P. A. Jones, P. W. Laird, *Mol. Cell. Biol.* **2001**, *21*, 7587.
[153] R. D. Wood, M. Mitchell, J. Sgouros, T. Lindahl, *Science* **2001**, *291*, 1284.
[154] P. Neddermann, P. Gallinari, T. Lettieri, D. Schmid, O. Truong, J. J. Hsuan, K. Wiebauer, J. Jiricny, *J. Biol. Chem.* **1996**, *271*, 12767.
[155] S. Tweedie, H.-H. Ng, A. L. Barlow, B. M. Turner, B. Hendrich, A. Bird, *Nat. Genet.* **1999**, *23*, 389.
[156] S. H. Cross, J. A. Charlton, X. Nan, A. P. Bird, *Nat. Genet.* **1994**, *6*, 236.
[157] A. Bird, M. Taggart, M. Frommer, O. J. Miller, D. Macleod, *Cell* **1985**, *40*, 91.
[158] A. Bird, *Cell* **1992**, *70*, 5.
[159] D. F. Schorderet, S. M. Gartler, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 957.
[160] P. W. Laird, R. Jaenisch, *Annu. Rev. Genet.* **1996**, *30*, 441.
[161] S. U. Kass, D. Pruss, A. P. Wolffe, *Trends Genet.* **1997**, *13*, 444.
[162] Z. Siegfried, H. Cedar, *Curr. Biol.* **1997**, *7*, R305.
[163] S. M. Tilghman, *Cell* **1999**, *96*, 185.
[164] W. Reik, J. Walter, *Nat. Rev. Genet.* **2001**, *2*, 21.
[165] E. Li, C. Beard, R. Jaenisch, *Nature* **1993**, *366*, 362.
[166] M. D. Adams et al., *Science* **2000**, *287*, 2185.
[167] W. Reik, A. Murrell, *Nature* **2000**, *405*, 408.
[168] A. C. Bell, G. Felsenfeld, *Nature* **2000**, *405*, 482.
[169] A. T. Hark, C. J. Schoenherr, D. J. Katz, R. S. Ingram, J. M. Levorse, S. M. Tilghman, *Nature* **2000**, *405*, 486.
[170] P. Avner, E. Heard, *Nat. Rev. Genet.* **2001**, *2*, 59.
[171] Y. K. Kang, D. B. Koo, J. S. Park, Y. H. Choi, A. S. Chung, K. K. Lee, Y. M. Han, *Nat. Genet.* **2001**, *28*, 173.
[172] C. Beard, E. Li, R. Jaenisch, *Genes Dev.* **1995**, *9*, 2325.
[173] B. Panning, R. Jaenisch, *Genes Dev.* **1996**, *10*, 1991.
[174] T. Sado, M. H. Fenner, S. S. Tan, P. Tam, T. Shioda, E. Li, *Dev. Biol. (Orlando, FL, USA)* **2000**, *225*, 294.
[175] A. F. Smit, *Curr. Opin. Genet. Dev.* **1999**, *9*, 657.
[176] J. A. Yoder, C. P. Walsh, T. H. Bestor, *Trends Genet.* **1997**, *13*, 335.
[177] C. P. Walsh, J. R. Chaillet, T. H. Bestor, *Nat. Genet.* **1998**, *20*, 116.
[178] L. Jackson-Grusby, C. Beard, R. Possemato, M. Tudor, D. Fambrough, G. Csankovszki, J. Dausman, P. Lee, C. Wilson, E. Lander, R. Jaenisch, *Nat. Genet.* **2001**, *27*, 31.
[179] R. Z. Chen, U. Pettersson, C. Beard, L. Jackson-Grusby, R. Jaenisch, *Nature* **1998**, *395*, 89.
[180] A. Miura, S. Yonebayashi, K. Watanabe, T. Toyama, H. Shimada, T. Kakutani, *Nature* **2001**, *411*, 212.
[181] G. Faugeron, *Curr. Opin. Microbiol.* **2000**, *3*, 144.
[182] T. Tuschl, *ChemBioChem* **2001**, *2*, 239.
[183] M. Wassenegger, *Plant Mol. Biol.* **2000**, *43*, 203.
[184] M. A. Matzke, M. F. Mette, A. J. Matzke, *Plant. Mol. Biol.* **2000**, *43*, 401.
[185] J. Bender, *Cell* **2001**, *106*, 129.
[186] M. W. Simmen, S. Leitgeb, J. Charlton, S. J. Jones, B. R. Harris, V. H. Clark, A. Bird, *Science* **1999**, *283*, 1164.
[187] A. P. Bird, *Trends Genet.* **1995**, *11*, 94.
[188] H. Gowher, O. Leismann, A. Jeltsch, *EMBO J.* **2000**, *19*, 6918.
[189] F. Lyko, B. H. Ramsahoye, R. Jaenisch, *Nature* **2000**, *408*, 538.
[190] A. D. Riggs, *Cytogenet. Cell. Genet.* **1975**, *14*, 9.
[191] R. Holliday, J. E. Pugh, *Science* **1975**, *187*, 226.
[192] E. Li, T. H. Bestor, R. Jaenisch, *Cell* **1992**, *69*, 915.
[193] M. Okano, D. W. Bell, D. A. Haber, E. Li, *Cell* **1999**, *99*, 247.
[194] I. Stancheva, R. R. Meehan, *Genes. Dev.* **2000**, *14*, 313.
[195] C. C. Martin, L. Laforest, M. A. Akimenko, M. Ekker, *Dev. Biol. (Orlando, FL, USA)* **1999**, *206*, 189.
[196] A. Schmitz, M. Short, O. Ammerpohl, C. Asbrand, J. Nickel, R. Renkawitz, *J. Biol. Chem.* **1997**, *272*, 20850.
[197] O. Ammerpohl, A. Schmitz, L. Steinmüller, R. Renkawitz, *Nucleic Acids Res.* **1998**, *26*, 5256.
[198] H. Thomassin, M. Flavin, M. L. Espinas, T. Grange, *EMBO J.* **2001**, *20*, 1974.
[199] P. M. Warnecke, T. H. Bestor, *Curr. Opin. Oncol.* **2000**, *12*, 68.
[200] S. B. Baylin, J. G. Herman, *Trends Genet.* **2000**, *16*, 168.
[201] J. P. Issa, *Curr. Top. Microbiol. Immunol.* **2000**, *249*, 101.
[202] G.-L. Xu, T. H. Bestor, D. Bourc'his, C.-L. Hsieh, N. Tommerup, M. Bugge, M. Hulten, X. Qu, J. J. Russo, E. Viegas-Pequignot, *Nature* **1999**, *402*, 187.
[203] R. S. Hansen, C. Wijmenga, P. Luo, A. M. Stanek, T. K. Canfield, C. M. Weemaes, S. M. Gartler, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 14412.
[204] R. E. Amir, I. B. Van den Veyver, M. Wan, C. Q. Tran, U. Francke, H. Y. Zoghbi, *Nat. Genet.* **1999**, *23*, 185.
[205] E. Ballestar, A. P. Wolffe, *Eur. J. Biochem.* **2001**, *268*, 1.
[206] H. Gowher, A. Jeltsch, *J. Mol. Biol.* **2001**, *309*, 1201.
[207] R. del Gaudio, R. Di Giaimo, N. Potenza, M. Branno, F. Aniello, G. Geraci, *FEBS Lett.* **1999**, *460*, 380.
[208] S. Tweedie, J. Charlton, V. Clark, A. Bird, *Mol. Cell. Biol.* **1997**, *17*, 1469.
[209] F. Gaudet, D. Talbot, H. Leonhardt, R. Jaenisch, *J. Biol. Chem.* **1998**, *273*, 32725.
[210] C. Mertineit, J. A. Yoder, T. Taketo, D. W. Laird, J. M. Trasler, T. H. Bestor, *Development (Cambridge, UK)* **1998**, *125*, 889.
[211] D. W. Hsu, M. J. Lin, T. L. Lee, S. C. Wen, X. Chen, C. J. Shen, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 9751.
[212] C. Bonfils, N. Beaulieu, E. Chan, J. Cotton-Montpetit, A. R. MacLeod, *J. Biol. Chem.* **2000**, *275*, 10754.
[213] T. Bestor, A. Laudano, R. Mattaliano, V. Ingram, *J. Mol. Biol.* **1988**, *203*, 971.
[214] S. Kumar, J. R. Horton, G. D. Jones, R. T. Walker, R. J. Roberts, X. Cheng, *Nucleic Acids Res.* **1997**, *25*, 2773.
[215] H. Leonhardt, A. W. Page, H. U. Weier, T. H. Bestor, *Cell* **1992**, *71*, 865.
[216] J. F. Glickman, J. G. Pavlovich, N. O. Reich, *J. Biol. Chem.* **1997**, *272*, 17851.
[217] T. H. Bestor, *EMBO J.* **1992**, *11*, 2611.
[218] L. S.-H. Chuang, H.-H. Ng, J.-N. Chia, B. F. L. Li, *J. Mol. Biol.* **1996**, *257*, 935.
[219] Y. Liu, E. J. Oakeley, L. Sun, J.-P. Jost, *Nucleic Acids Res.* **1998**, *26*, 1038.
[220] L. S. H. Chuang, H.-I. Ian, T. W. Koh, H.-H. Ng, G. Xu, B. F. L. Li, *Science* **1997**, *277*, 1996.
[221] M. R. Rountree, K. E. Bachman, S. B. Baylin, *Nat. Genet.* **2000**, *25*, 269.
[222] F. Fuks, W. A. Burgers, A. Brehm, L. Hughes-Davies, T. Kouzarides, *Nat. Genet.* **2000**, *24*, 88.
[223] K. D. Robertson, S. Ait-Si-Ali, T. Yokochi, P. A. Wade, P. L. Jones, A. P. Wolffe, *Nat. Genet.* **2000**, *25*, 338.
[224] T. O. Tollefsbol, C. A. I. Hutchison, *J. Biol. Chem.* **1995**, *270*, 18543.
[225] T. O. Tollefsbol, C. A. I. Hutchison, *J. Mol. Biol.* **1997**, *269*, 494.

[226] J. Flynn, J. F. Glickman, N. O. Reich, *Biochemistry* **1996**, *35*, 7308.
[227] S. Pradhan, A. Bacolla, R. D. Wells, R. J. Roberts, *J. Biol. Chem.* **1999**, *274*, 33 002.
[228] M. Fatemi, A. Hermann, S. Pradhan, A. Jeltsch, *J. Mol. Biol.* **2001**, *309*, 1189.
[229] A. V. Bakin, T. Curran, *Science* **1999**, *283*, 387.
[230] I. Stancheva, C. Hensey, R. R. Meehan, *EMBO J.* **2001**, *20*, 1963.
[231] J. D. Knox, F. D. Araujo, P. Bigey, A. D. Slack, G. B. Price, M. Zannis-Hadjopoulos, M. Szyf, *J. Biol. Chem.* **2000**, *275*, 17 986.
[232] C. Zimmermann, E. Guhl, A. Graessmann, *Biol. Chem.* **1997**, *378*, 393.
[233] J. B. Margot, A. M. Aguirre-Arteta, B. V. Di Giacco, S. Pradhan, R. J. Roberts, M. C. Cardoso, H. Leonhardt, *J. Mol. Biol.* **2000**, *297*, 293.
[234] A. Bacolla, S. Pradhan, R. J. Roberts, R. D. Wells, *J. Biol. Chem.* **1999**, *274*, 33 011.
[235] J. Flynn, N. Reich, *Biochemistry* **1998**, *37*, 15 162.
[236] M. S. Turker, *Semin. Cancer Biol.* **1999**, *9*, 329.
[237] J. A. Yoder, T. H. Bestor, *Hum. Mol. Genet.* **1998**, *7*, 279.
[238] I. Van den Wyngaert, J. Sprengel, S. U. Kass, W. H. M. L. Luyten, *FEBS Lett.* **1998**, *426*, 283.
[239] A. Dong, J. A. Yoder, X. Zhang, L. Zhou, T. H. Bestor, X. Cheng, *Nucleic Acids Res.* **2001**, *29*, 439.
[240] M. Okano, S. Xie, E. Li, *Nucleic Acids Res.* **1998**, *26*, 2536.
[241] C. R. M. Wilkinson, R. Bartlett, P. Nurse, A. P. Bird, *Nucleic Acids Res.* **1995**, *23*, 203.
[242] E. Pinarbasi, J. Elliott, D. P. Hornby, *J. Mol. Biol.* **1996**, *257*, 804.
[243] M. Okano, S. Xie, E. Li, *Nat. Genet.* **1998**, *19*, 219.
[244] K. D. Robertson, E. Uzvolgyi, G. Liang, C. Talmadge, J. Sumegi, F. A. Gonzales, P. A. Jones, *Nucleic Acids Res.* **1999**, *27*, 2291.
[245] A. Aoki, I. Suetake, J. Miyagawa, T. Fujio, T. Chijiwa, H. Sasaki, S. Tajima, *Nucleic Acids Res.* **2001**, *29*, 3506.
[246] F. Fuks, W. A. Burgers, N. Godin, M. Kasai, T. Kouzarides, *EMBO J.* **2001**, *20*, 2536.
[247] C.-L. Hsieh, *Mol. Cell. Biol.* **1999**, *19*, 8211.
[248] F. Lyko, B. H. Ramsahoye, H. Kashevsky, M. Tudor, M. A. Mastrangelo, T. L. Orr-Weaver, R. Jaenisch, *Nat. Genet.* **1999**, *23*, 363.
[249] E. J. Finnegan, R. K. Genger, W. J. Peacock, E. S. Dennis, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1998**, *49*, 223.
[250] R. A. Martienssen, V. Colot, *Science* **2001**, *293*, 1070.
[251] X. Cao, N. M. Springer, M. G. Muszynski, R. L. Phillips, S. Kaeppler, S. E. Jacobsen, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 4979.
[252] S. Henikoff, L. Comai, *Genetics* **1998**, *149*, 307.
[253] L. Bartee, F. Malagnac, J. Bender, *Genes Dev.* **2001**, *15*, 1753.
[254] C. M. Papa, N. M. Springer, M. G. Muszynski, R. Meeley, S. M. Kaeppler, *Plant Cell* **2001**, *13*, 1919.
[255] S. E. Jacobsen, E. M. Meyerowitz, *Science* **1997**, *277*, 1100.
[256] T. Kakutani, J. A. Jeddeloh, S. K. Flowers, K. Munakata, E. J. Richards, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 12 406.
[257] P. Cubas, C. Vincent, E. Coen, *Nature* **1999**, *401*, 157.
[258] H. M. Foss, C. J. Roberts, K. M. Claeys, E. U. Selker, *Science* **1993**, *262*, 1737.
[259] H. Gowher, K. C. Ehrlich, A. Jeltsch, *FEMS Microbiol. Lett.* **2001**, *205*, 151.
[260] V. P. Miao, M. Freitag, E. U. Selker, *J. Mol. Biol.* **2000**, *300*, 249.
[261] E. Kouzminova, E. U. Selker, *EMBO J.* **2001**, *20*, 4309.
[262] F. Malagnac, B. Wendel, C. Goyon, G. Faugeron, D. Zickler, J.-L. Rossignol, M. Noyer-Weidner, P. Vollmayr, T. A. Trautner, J. Walter, *Cell* **1997**, *91*, 281.
[263] A. V. Chernov, P. Vollmayr, J. Walter, T. A. Trautner, *Biol. Chem.* **1997**, *378*, 1467.
[264] F. Malagnac, A. Gregoire, C. Goyon, J.-L. Rossignol, G. Faugeron, *Mol. Microbiol.* **1999**, *31*, 331.
[265] A. R. MacLeod, M. Szyf, *J. Biol. Chem.* **1995**, *270*, 8037.
[266] S. Ramchandani, A. R. MacLeod, M. Pinard, E. von Hofe, M. Szyf, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 684.
[267] W. M. I. Rideout, K. Eggan, R. Jaenisch, *Science* **2001**, *293*, 1093.
[268] W. Chao, K. D. Huynh, R. J. Spencer, L. S. Davidow, J. T. Lee, *Science* **2002**, *295*, 345.
[269] I. Percec, M. S. Bartolomei, *Science* **2002**, *295*, 287.
[270] C. Qiu, K. Sawada, X. Zhang, X. Cheng, *Nat. Struct. Biol.* **2002**, in press.

Received: August 22, 2001 [A 287]