DNA methylation and polyamines in embryonic development and cancer

OLLE HEBY*

Department of Cellular and Developmental Biology, University of Umeå, Umeå, Sweden

ABSTRACT **Mammalian DNA contains relatively large amounts of a modified base, 5-methylcytosine (m^5C). Methylation of cytosine is catalyzed by DNA(cytosine-5)methyltransferase (DNA MTase). DNA methylation seems to play an important role in the regulation of gene expression during development. Thus, m^5C may inhibit transcription by preventing the binding of transcription factors and/or by altering chromatin structure. The DNA methylation patterns of the male and female pronuclei are erased in the morula and early blastula, and when the blastocyst forms, most of the DNA has become demethylated. Following implantation, however, there is a surge of *de novo* methylation affecting the entire genome, and already by gastrulation DNA is methylated to an extent characteristic of that of the adult animal. During subsequent development, tissue-specific genes undergo programmed demethylation, which may cause their activation. Site-directed mutagenesis of the DNA MTase gene, has recently shown that DNA methylation is absolutely required for normal development of the early mouse embryo. DNA methylation and polyamine synthesis depend on a common substrate, S-adenosylmethionine (AdoMet). As a consequence, changes in cellular polyamine levels may affect the degree of DNA methylation. When the first step in the polyamine biosynthetic pathway is blocked, F9 teratocarcinoma stem cells accumulate large amounts of decarboxylated AdoMet, the aminopropyl group donor in polyamine synthesis, and go through terminal differentiation into parietal endoderm cells. The accumulation of decarboxylated AdoMet is a direct consequence of the polyamine-depleted state of the cell. Although the decarboxylated AdoMet molecule contains a methyl group, it does not act as a methyl group donor in DNA methylation. Instead it acts as a competitive inhibitor of DNA MTase. A consequence of polyamine depletion is therefore genome-wide loss of DNA methylation due to insufficient maintenance methylation during successive rounds of DNA replication. Our recent finding that prevention of the accumulation of decarboxylated AdoMet counteracts the differentiative effect lends further support to the hypothesis proposed.**

KEY WORDS: *DNA methylation, DNA methyltransferase, polyamines, embryogenesis, cancer*

Introduction

The enzymatic methylation of bases in DNA is an essential element of genomic function in organisms ranging from bacteria to man. There are 3 major classes of enzymes involved. Two carry out methylation on exocyclic nitrogens, converting adenine to N6-methyladenine or cytosine to N4-methylcytosine. The third, which is discussed in this review, methylates the 5 carbon of cytosine to yield 5-methylcytosine (m^5C).

In prokaryotes, DNA methylation directs the mismatch repair and restriction-modification systems, which correct errors of replication and prevent transformation by non-self DNA. In eukaryotes, methylation of the nuclear DNA has been suggested to play a regulatory role in many important events, such as tissue-specific gene expression during development, X chromosome inactivation, genomic imprinting, virus latency, mutagenesis, tumorigenesis and senescence.

Cytosine methylation is catalyzed by DNA(cytosine-5)methyltransferase (DNA MTase, EC 2.1.1.37), and most m^5C residues occur in the palindrome 5'-CpG-3' (Fig. 1), with methylation on both strands (Razin and Riggs, 1980). Many early studies in this field demonstrated an inverse correlation between DNA methylation and gene expression. The conclusion derived from these studies, that DNA methylation is directly involved in the suppression of transcription, is supported by the fact that genes, transfected into cells, are not expressed if their promoters have

Abbreviations used in this paper: AbeAdo, 5´-{[(Z)-4-amino-2-butenyl]methylamino}-5´-deoxyadenosine; AdoHcy, *S*-adenosylhomocysteine; AdoMet, *S*-adenosylmethionine; DFMO, α-difluoromethylornithine; DNA MTase, DNA(cytosine-5)methyltransferase; ES cells, embryonic stem cells; HIV, human immunodeficiency virus; HSV TK, herpes simplex virus thymidine kinase; LTR, long terminal repeat; m^5C, 5-methylcytosine; MoMuLV, Moloney murine leukemia virus.

***Address for reprints**: Department of Cellular and Developmental Biology, University of Umeå, S-901 87 Umeå, Sweden. FAX: 46.90166691.

0214-6282/95/$03.00

been methylated *in vitro* (Razin and Cedar, 1991). Conversely, endogenous genes can be activated from a repressed state by treatment with 5-azacytidine, a demethylation agent (Jones, 1984). Recent gene targeting experiments, which introduced partial loss-of-function mutations in the DNA MTase gene of embryonic stem (ES) cells of the mouse, attest to the importance of DNA methylation for the normal development of the early mouse embryo by demonstrating that even modest perturbations of methylation patterns cause prenatal lethality, preventing homozygous mutant embryos from progressing past the 20 somite stage in mid-gestation (Li *et al.,* 1992).

DNA methylation and polyamine biosynthesis depend on a common substrate, *S*-adenosylmethionine (AdoMet). As a consequence, changes in cellular polyamine levels may affect the degree of DNA methylation. Thus, it is proposed that teratocarcinoma stem cell differentiation, induced by polyamine depletion, may be due to interference with DNA MTase activity and genome-wide loss of methylation.

Presence of m^5C in cell genomes

m^5C seems to be almost universally present in the DNA of eukaryotes with large (>10^8 base pair) genomes (Bestor, 1990). This DNA modification may have evolved to reduce the effective size of the expanded genomes of complex higher eukaryotes. In the mammalian DNA approximately 1% of the bases are m^5C. The m^5Cs are found almost exclusively at CpG dinucleotides. The haploid mammalian genome contains approximately 5x10^7 CpG dinucleotides, and about 60% of these are methylated. The m^5Cs are distributed throughout the genome in cell-type-specific patterns, which are faithfully transmitted by clonal inheritance during DNA replication. They compartmentalize mammalian genomes into a large methylated part that is propagated in the condensed state, and a small, constitutively unmethylated part (forming the so-called CpG islands) that is accessible to diffusible regulatory factors (Bird, 1987; Bestor, 1990). This arrangement may guide regulatory proteins to their target sites in DNA, and may provide an immune function for the genome by restricting the activity of invasive sequences such as proviral DNA and transposable elements. There is ample evidence to suggest that this type of DNA methylation may serve as a global suppressor of gene expression.

CpG islands in mammalian genomes

All mammalian species appear to have CpG islands. These are short (1-2 kb) regions of GC-rich (60-70%) DNA containing densely clustered, unmethylated CpG dinucleotides (Bird, 1987; Cross and Bird, 1995). Taken together the CpG islands account for about 2% of the genome. The remaining 98% of the genome has a GC content of about 40%, is methylated at its CpGs and exhibits a CpG dinucleotide frequency of less than a quarter of that expected from base composition. The depletion of CpGs probably results from deamination of m^5C to thymine. Thus m^5CpGs are mutated to TpGs and CpAs. Accordingly, the excess of the latter two dinucleotides in the genome is equivalent to the deficiency of CpGs.

Ribosomal DNA resembles CpG island DNA in the sense that it has a high GC content and is mainly nonmethylated (Bird *et al.,* 1981a).

CpG islands always seem be associated with genes, but all genes do not have CpG islands. The proportion of genes that are associated with CpG islands has been estimated to 56% in human and 47% in mouse (Antequera and Bird, 1993). It was also estimated that the number of CpG islands in the haploid genome of human and mouse is 45,000 and 37,000, respectively, even though both species have the same number of genes, i.e. some 80,000 (22,000 housekeeping genes and 58,000 tissue-specific genes) per haploid genome. It appears that both human and mouse are losing CpG islands over evolutionary time due to *de novo* CpG methylation in the germ line and subsequent mutation of the m^5CpGs to TpGs and CpAs, but this process is more rapid in mouse. Antequera and Bird (1993) have advanced the hypothesis that CpG islands arose with the vertebrates; the genes of an invertebrate ancestor may have been embedded in entirely nonmethylated DNA (as are all current invertebrate genes), but with the appearance of the vertebrates, DNA methylation may have spread through the genome, sparing promoters which otherwise may have turned off the expression of the associated genes. With time, an increasing number of genes may have become methylated, and therefore lost their CpG islands, without losing their capacity for expression. Only tissue-specific genes seem to have accomplished this transition, because all known housekeeping genes are still associated with CpG islands.

CpG islands consistently appear at the 5´ ends of genes. This is true for all housekeeping genes and for 40% of the tissue-specific genes (Larsen *et al.,* 1992). It has been estimated that 50-60% of mammalian genes have CpG islands associated with their 5´ ends. A CpG island is often seen to include the first exon of the gene (Cross and Bird, 1995). From the above it may be concluded that one half of all CpG islands in the human genome (approximately 22,000) is associated with housekeeping genes, whereas the other half is associated with tissue-specific genes. Notably, CpG islands at the 5´ ends of tissue-specific genes are constitutively unmethylated (like those of housekeeping genes) but they are transcriptionally inactive in most cell types (unlike those of housekeeping genes).

Several mechanisms by which CpG islands remain free of methylation in the otherwise heavily methylated mammalian genome, have been proposed. One involves local *cis*-acting elements, e.g. Sp1 sites (GC boxes) located in the 5´ end of the CpG island (Brandeis *et al.,* 1994; Macleod *et al.,* 1994). Thus, when such Sp1 sites of a CpG island were deleted or mutated in a transgenic mouse assay, the CpG island became *de novo* methylated, whereas the intact CpG island remained unmethylated. However, this effect does not seem to be simply attributable to steric hindrance by transcription factors, excluding DNA MTase from the CpG island (which co-localizes with the promoter) and protecting it from methylation. Cross and Bird (1995) propose that protein factors antagonize DNA MTase at a crucial developmental stage when the global DNA methylation pattern is established, and that the resulting pattern is subsequently retained by maintenance methylation. Some *in vitro* experiments suggest that CpG islands are a poor substrate for DNA MTase, either because of their high density of CpGs or their GC richness, even though the CpG dinucleotide is the target for the enzyme (Carotti *et al.,* 1989). However, this model does not explain why CpG islands are methylated on the inactive X chro-

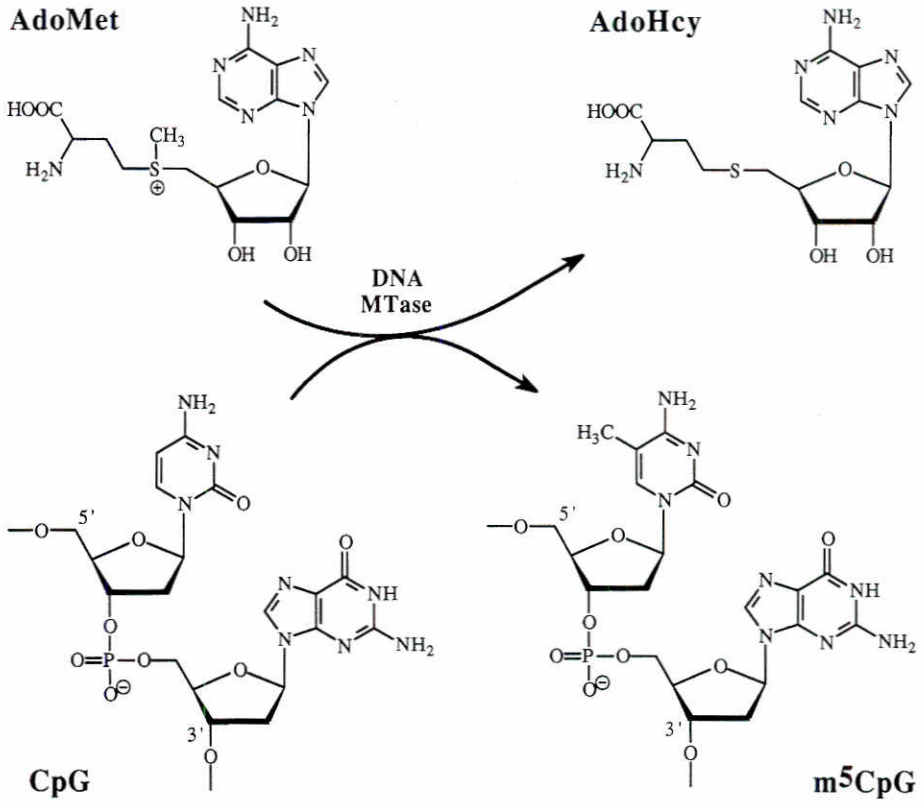

Fig. 1. DNA MTase-catalyzed methylation of the cytosine residue in a CpG-dinucleotide of nuclear DNA. *DNA MTase recognizes cytosines located 5′ to guanines and catalyzes their methylation. The enzyme appears to act independently of the sequences surrounding the substrate CpG sites. In somatic tissues of mammals approximately 4% of the cytosines are methylated and the vast majority of these occur in CpG-containing sequences. Methylcytosines appear symmetrically on both DNA strands except in newly replicated DNA, where the daughter strand is unmethylated. DNA MTase has strong preference for this hemimethylated DNA. S-Adenosylmethionine (AdoMet) acts as the methyl group donor, yielding m^{5}CpG and S-adenosylhomocysteine (AdoHcy) as the reaction products. The chemistry of the methylation reaction is shown in Fig. 5.*

mosome, at imprinted genes and at non-essential genes in cell cultures (Antequera *et al.*, 1990). Yet another possibility, for which there is experimental support, is that CpG islands are methylated *de novo*, but that the methylation is removed by a 5′ CpG-island-specific demethylating system operating in the early embryo (Frank *et al.*, 1991).

Using fluorescent *in situ* hybridization and *in situ* suppression hybridization to metaphase chromosomes it has been shown that the distribution of CpG islands in the human genome corresponds to the reverse (R) banding pattern (Craig and Bickmore, 1993, 1994). The R bands represent chromosomal regions that are early-replicating, comparatively GC-rich and less condensed than those revealed by Giemsa (G) banding. The highest concentration of CpG islands is in a subset of R bands, known as T bands (of which about half are at telomeric locations in the human karyotype). The CpG islands show a strikingly uneven distribution both along and between chromosomes, paralleling the differences in the numbers of genes assigned to various chromosomes and chromosome regions. It appears that >80% of CpG islands are in the 45% of the human genome that corresponds to R bands (Craig and Bickmore, 1994). This figure is similar to the percentage of human genes, attained in an analysis of 1,000 genes in the databases, that map to R bands (Craig and Bickmore, 1993, 1994).

There are exceptions to the general rule that CpG islands are nonmethylated in somatic cells, for example the CpG islands of the inactive X chromosome, where CpG island hypermethylation seems to stabilize the inactive state of the associated genes (Riggs and Pfeifer, 1992). Other examples include the CpG islands of the testis-specific *H2B* histone gene (Choi and Chae, 1991) and several imprinted genes such as *H19* (Ferguson-Smith *et al.*, 1993), and *Igf2r* (Stöger *et al.*, 1993). However, all these CpG islands appear to be nonmethylated in the germ line (Choi and Chae, 1991; Brandeis *et al.*, 1993).

Catalysis of m^{5}C formation by a maintenance DNA MTase following DNA replication

The genomic pattern of cytosine methylation is faithfully inherited from generation to generation by means of a post-replicative mechanism (Fig. 2). This process, known as maintenance DNA methylation, is catalyzed by DNA MTase, which is specific for hemimethylated DNA generated during DNA replication. There is a short delay of approximately 1 min between the incorporation of deoxycytidine into the DNA and its conversion to m^{5}C (Gruenbaum *et al.*, 1983). This lag may be a result of limiting accessibility to the DNA at the replication fork, because the methylase activity present in the nucleus permits immediate methylation.

Maintenance DNA methylation, which requires the recognition and methylation of all hemimethylated CpG sites prior to the subsequent S phase, may be involved in re-establishing not only the original pattern of DNA methylation but also the original chromatin structure including the packaging of the newly synthesized DNA into a nucleosome structure. It may provide the first signal in the sequence of events that generates an inactive chromatin.

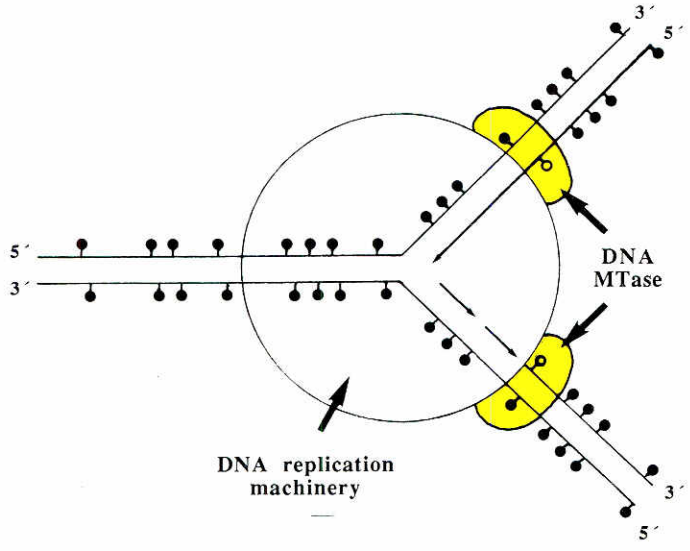

Fig. 2. Maintenance DNA methylation at a replication fork. *Tissue-specific patterns of cytosine methylation in the mammalian genome are faithfully inherited from generation to generation by means of a post-replicative mechanism called maintenance DNA methylation. The process is catalyzed by DNA MTase, which is specific for the hemimethylated DNA generated during DNA replication. For the pattern of DNA methylation of the parental DNA to be conserved in the daughter DNA all hemimethylated CpG sites have to be recognized and methylated prior to the subsequent S phase. Methylation normally takes place approximately 1 min after the deoxycytidine has been incorporated into DNA.* ●, m^5C. ○, *cytosine to be methylated in the active site of DNA MTase.*

At the time of replication, chromatin proteins are no longer tightly associated with the DNA thus allowing DNA methylation to occur before their reassociation to the nascent DNA.

To replicate their large genomes, mammalian cells initiate pairs of replication forks (replication origins) at between 10,000 to 100,000 different sites scattered throughout the chromosomal DNA. Replication origins are activated in clusters called DNA replication foci, each focus containing at least 300 origins (Nakamura *et al.*, 1986; Mills *et al.*, 1989; Nakayasu and Berezney, 1989). Origins remain tightly clustered in such groups throughout the S phase. Replication foci are assembled in the cell nucleus at the beginning of the S phase, and associated with these structures are the many proteins that are involved in replicating the DNA. DNA MTase is also associated with the DNA replication foci and colocalize with these throughout the S phase of the cell cycle (Leonhardt *et al.*, 1992). It appears that nucleoplasmic DNA MTase is recruited to replication foci when cells enter the S phase. As cells progress through mid- and late-S, DNA MTase becomes associated with larger structures that are enriched in centromeric heterochromatin. However, it has not been conclusively demonstrated that these large replication foci only contain sites of replication for γ satellite DNA. During the other phases of the cell cycle, the enzyme is evenly distributed in the nucleoplasm and seems to be largely excluded from the nucleoli. The only instances in which DNA MTase has been found in the cytoplasmic compartment are in mitotic cells, where the nuclear membrane has been disassembled, and in oocytes and cells of the preimplantation embryo (Carlson *et al.*, 1992).

Structures and functions of DNA MTases

Mammalian DNA MTases

Although different cell types contain different methylation patterns they appear to have a single form of DNA MTase (Fig. 3). The primary translation product of the single DNA MTase gene transcript present in adult mouse cells has an M_r of 190,000 (Bestor *et al.*, 1988). The M_r 175,000 and 150,000 species that also have been observed are probably generated by proteolytic processing of the extremely protease-sensitive sequences near the N-terminus (the first 350 amino acids). There are many observations which suggest that yet another form of DNA MTase, capable of methylating DNA *de novo*, may be present in germ cells and early embryos. However, the fact that the maintenance enzyme does catalyze *de novo* methylation of unmethylated DNA, although inefficiently (Gruenbaum *et al.*, 1982; Bestor and Ingram, 1983), should be taken into account.

A key feature of mammalian DNA MTases is their strong preference for hemimethylated sites (Bestor, 1992). Thus, mammalian DNA MTase is stimulated 30- to 40-fold by hemimethylated substrates (Bestor and Ingram, 1983). A DNA MTase, purified from mouse plasmacytoma cells, was found to be active not only on hemimethylated but also on unmethylated DNA substrates (Hitt *et al.*, 1988). For the *de novo* methylation of duplex DNA, however, single-stranded regions or large amounts of enzyme were required.

Mammalian DNA MTase is likely to have arisen by fusion of two genes coding for a DNA binding protein and a bacterial-like type II restriction DNA MTase (Bestor *et al.*, 1988; Bestor, 1990). The corresponding N-terminal and C-terminal domains of the protein contain regulatory sequences and the catalytic site, respectively (Fig. 3). The link between the two domains is constituted by 13 alternating lysine and glycine residues that are encoded by 39 consecutive purine nucleotides. Since cleavage between the domains has been found to strongly stimulate the rate of methylation of unmethylated DNA, without changing the rate of methylation of hemimethylated DNA, it is believed that the N-terminal domain is responsible for the clonal propagation of methylation patterns through suppression of the inherent *de novo* DNA methylation activity of the C-terminal domain (Bestor, 1992). An interesting possibility is that proteolytic cleavage between the regulatory and catalytic domain could stimulate *de novo* activity and thus contribute to the process of ectopic methylation observed in the DNA of aging animals, tumors and cultured cell lines.

The N-terminal domain (Fig. 3), which is without counterpart in the bacterial enzymes, appears to have many functions. It contains a nuclear localization signal, because, at variance with prokaryotic DNA MTases, it has to enter the nucleus to reach its target. The N-terminal domain also targets the enzyme to the replication machinery, regulates the enzymatic activity of the catalytic domain and may be involved in sensing the pre-existing methylation state of the DNA (Bestor, 1992; Leonhardt *et al.*, 1992; Bestor and Verdine, 1994).

The nuclear localization signal as well as the replication focus signal are located near the N-terminus of the enzyme (Leonhardt

et al., 1992). Together with the strong preference of DNA MTase for hemimethylated DNA, these targeting signals ensure the clonal propagation of tissue-specific methylation patterns. They can be considered to increase the rate of methylation of hemimethylated sites generated by the replication forks. Within the N-terminal domain, but closer to the C-terminal domain, there is also a region that suppresses *de novo* DNA methylation, i.e. discriminates between unmethylated and hemimethylated CpG sites (Fig. 3). The *de novo* activity of DNA MTase has little sequence specificity beyond the CpG dinucleotide and it remains a major challenge to establish how tissue-specific methylation patterns are acquired. Finally, the N-terminal domain contains a cysteine-rich region that binds zinc ions (Fig. 3) (Bestor, 1992). This zinc-binding region may be an important DNA-binding element. Interestingly, a similar motif is present in ALL-1/HRX, a mammalian homolog of the *Drosophila* trithorax (Ma *et al.,* 1993).

The human DNA MTase that has been isolated and characterized (Yen *et al.,* 1992) exhibits essentially the same properties as the mouse enzyme. The single 5.2 kb transcript that is found in all human tissues tested has an open reading frame of 4485 nucleotides encoding a protein of 1495 amino acid residues with an estimated M_r of 169,000. The size of the gene is at least 93 kb, as indicated by the distance spanned by a series of DNA MTase-positive cosmids. These hybridized to chromosome 19p13.2-p13.3. In the mouse, DNA MTase has been mapped to chromosome 9.

Before the catalytic mechanism of cytosine methylation was solved, the major problem was to understand how to assemble an active site around a reaction center that is deeply buried within the structure of double-helical DNA.

Bacterial modification DNA MTases

Analyses of bacterial modification DNA MTases (Klimasauskas *et al.,* 1994; Reinisch *et al.,* 1995) have recently contributed to the elucidation of the function of the closely related carboxy-terminal catalytic domain of eukaryotic DNA MTases. High-resolution structural data on M. *Hha* I, in both free and DNA-bound forms, have revealed that the enzyme is organized into a larger domain, containing most of the conserved amino acid residues implicated in catalysis, and a smaller domain, containing residues implicated in sequence-specific DNA-recognition (Chen *et al.,* 1993; Cheng *et al.,* 1993; Klimasauskas *et al.,* 1994). The enzyme binds DNA as a monomer, with the DNA located in a large cleft formed at the junction of the two domains (Fig. 4). The smaller target recognition domain has a highly unusual structure and varies widely both in size and sequence among DNA MTases, probably allowing different enzymes to recognize different DNA sequences (Bestor and Verdine, 1994). This smaller domain interacts with DNA from the major groove side, and may select the right cytosine to be methylated. The larger catalytic domain interacts primarily with the minor groove to open the GC base pair, swing out the cytosine from the double helix, and insert it into the active-site pocket of the enzyme, where AdoMet, the methyl group donor, is also bound (Figs. 4 and 5). The space in the DNA helix, vacated by the completely expelled cytosine, is occupied from the minor groove side by Ser-87 and from the major groove side by Gln-237 (Fig. 4). By penetrating deeply into the helix, forming hydrogen bonds to both Ser 87 and the "orphan" guanine, the side chain of Gln-237

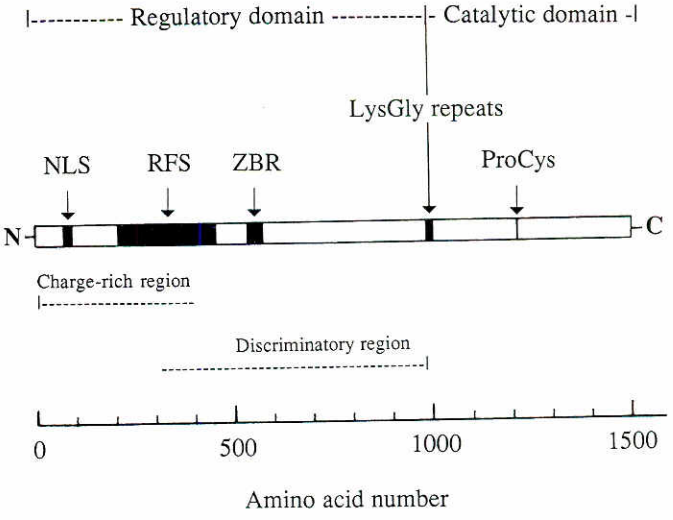

Fig. 3. Functional domains and sequence elements in mouse DNA MTase. *The mouse DNA MTase has a large N-terminal regulatory domain (amino acids 1-989) and a smaller C-terminal catalytic domain (amino acids 1003-1502). It appears to be a fusion protein, in which 13 alternating Lys and Gly residues (LysGly repeats; amino acids 990-1002), encoded by 39 consecutive purine nucleotides, join the N-terminal domain (probably derived from a sequence-specific DNA-binding protein) and the C-terminal domain (probably derived from a bacterial type II restriction DNA MTase). Proteolytic processing of the primary translation product (190 kDa) from the N-terminal end is believed to generate the smaller products (175 and 150 kDa) that are sometimes observed. This should cause no loss of activity, because the charge-rich N-terminal region has been shown to be dispensable for enzyme activity. However, this domain contains a nuclear localization signal (NLS; amino acids 72-92) and a replication focus signal (RFS; amino acids 207-455), which are essential for the targeting of the enzyme to the replication forks during the S phase of the cell cycle. The Cys-rich (8 Cys within 39 residues) Zn-binding region (ZBR; amino acids 537-575) is part of a discriminatory region, believed to be involved in the discrimination between unmethylated and hemimethylated CpG sites. The strong preference for hemimethylated sites ensures the clonal inheritance of tissue-specific methylation patterns through inhibition of the de novo activity of the catalytic domain, rendering the enzyme a maintenance DNA MTase. This regulatory domain has no counterpart in the bacterial DNA MTases. The catalytic domain, however, exhibits strong amino acid sequence similarity to the 50 known bacterial type II restriction DNA MTases (Kumar et al., 1994). The invariant ProCys dipeptide (amino acids 1213-1214) conforms to the Pro-Cys-X_6-Gly consensus sequence for DNA MTases (Sznyter et al., 1987). It is part of the catalytic center, where the Cys residue acts as the catalytic nucleophile (Fig. 5). Another motif that has been assigned a functional role is Phe-X-Gly-X-Gly, which is presumed to be at the AdoMet-binding site (Kumar et al., 1994). Features are drawn approximately to scale. Data from Bestor (1992) and Bestor et al. (1988).*

ensures the stack in the DNA, thus preventing the helix from collapsing (Klimasauskas *et al.,* 1994). However, the structure of M. *Hha* I reveals little about the mode of methyl group recognition on the complementary strand.

In a recent analysis of the crystal structure of another bacterial DNA MTase (M. *Hae*III from *Haemophilus aegyptius*) covalently bound to a DNA duplex, the target cytosine was also found to be flipped out of the helix (Reinisch *et al.,* 1995). However, the adjacent base pairs in the DNA helix were unstacked and thus

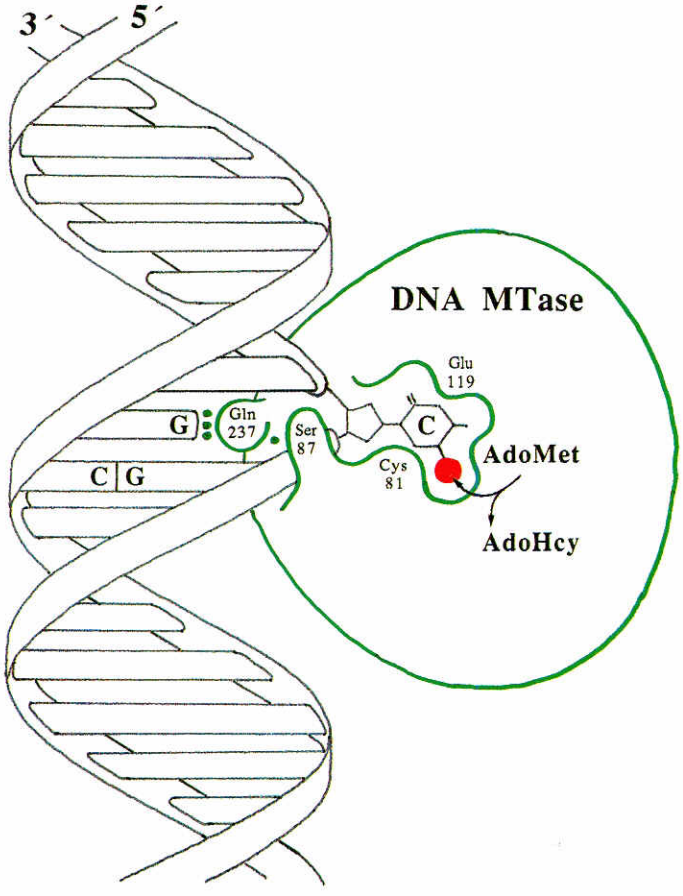

Fig. 4. A simplified view of the target cytosine flipped-out of the DNA double helix and buried in the active site cavity of the DNA MTase. *A spectacular feature of DNA MTase, is that it forces its substrate DNA to flip out the cytosine base it is going to methylate to an extrahelical position (Klimasauskas et al., 1994). There are only small distortions to the adjacent base pairs. The amino acids indicated are those of M. Hha I, a DNA MTase that forms part of a type II restriction-modification system from Haemophilus haemolyticus (Klimasauskas et al., 1994). M. Hha I methylates the first cytosine in its recognition sequence 5´-GCGC-3´ in double-stranded DNA. In place of the cytosine, the DNA is infiltrated by a glutamine (Gln-237) and a serine (Ser-87) side chain of the DNA MTase (from the major and minor groove side, respectively), and the DNA helix is thus prevented from collapsing (Klimasauskas et al., 1994). The preferred substrate is hemimethylated DNA, generated during DNA replication (Fig. 2). The cytosine is buried in the active-site pocket (where the Cys-81 and Glu-119 side chains of the enzyme have important roles), adjacent to AdoMet, the methyl group donor. Transfer of the methyl group (red circle) to cytosine is followed by expulsion of AdoHcy. Then m^5C has to swing back from its extrahelical position, re-establish Watson-Crick base-pairing with the "orphan" guanine, and again get buried in duplex DNA, but these events have not yet been experimentally documented.*

considerably more distorted than the corresponding base pairs in the M. *Hha* I-DNA complex (Klimasauskas *et al.*, 1994). There is also recent crystallographic evidence suggesting that enzymes other than DNA MTases, notably DNA repair enzymes, may flip bases out of DNA to gain access to their target residues (Roberts, 1995).

Although these structural analyses have revealed a novel and remarkable mechanism for protein-DNA interaction at a high resolution, many questions remain unanswered. For instance, how does the DNA MTase initially recognize its substrate, and how does it return the methylated cytosine into the helix? Is it conceivable that the enzyme merely captures a spontaneously extrahelical cytosine? In fact, looped-out bases have been seen in DNA duplexes containing unpaired bases (Joshua-Tor *et al.*, 1992). Moreover, the rate of spontaneous base-pair opening is 4-5 magnitudes greater than the rate of substrate turnover by DNA MTases (Leroy *et al.*, 1988), which are extremely slow enzymes with turnover numbers as low as 1 methyl transfer per min at 37°C (Friedman, 1985; Wu and Santi, 1987). Base-pairs open one at a time, with lifetimes in the range of milliseconds at 35°C. The lifetimes are shorter for A·T (1-7 ms) than for G·C pairs (7-40 ms) (Leroy *et al.*, 1988). Finally it should be emphasized that the bacterial DNA MTases, which are part of a simple immune system, do not discriminate between unmethylated and hemimethylated DNA.

A mechanism by which DNA MTase may recognize its substrate *in vivo*, reconcilable with its lack of sequence specificity *in vitro*, is suggested by the finding that secondary structures induced in DNA by local supercoiling dramatically altered the *in vitro* sequence specificity of a mammalian DNA MTase (Bestor, 1987).

Effects of m^5C on the properties and conformation of DNA

The methyl group at the carbon-5 position of cytosine reaches into the major groove of the DNA double helix much like that of thymidine, and it increases the thermal stability (T_m) of C·G base pairs in DNA molecules by several degrees. Cytosine methylation can induce local distortions in the structure of duplex DNA (Hodges-Garcia and Hagerman, 1992). Conversely, the interaction of DNA MTases with their substrates is stimulated by distortion (Erlanson *et al.*, 1993) and structural abnormalities (Smith *et al.*, 1991) of the DNA.

DNA methylation may regulate gene function by altering the conformation of DNA, e.g. by facilitating the transition from the right-handed B-DNA form to the left-handed Z-DNA form (Behe and Felsenfeld, 1981; Rich *et al.*, 1984), or by stabilizing triple-stranded DNA helix structures (Zacharias, 1993). It remains largely unknown, however, to what extent these unusual structures play a role in the regulation of gene expression in the animal.

Interference of m^5C with the binding of transcription factors to DNA

Methyl moieties at CpG residues probably suppress transcription by affecting DNA-protein interactions, thus altering the accessibility of genes to *trans*-acting factors. In fact, when analyzing the consequence of methylation of CpG sites, it has been consistently shown that those sites present in the 5' flanking region of the gene are the ones that upon methylation can interfere with the initiation of its transcription.

By forming a protrusion from the major groove of the DNA double helix, the methyl group at the carbon-5 position of cytosine may interfere with protein binding much the same way that

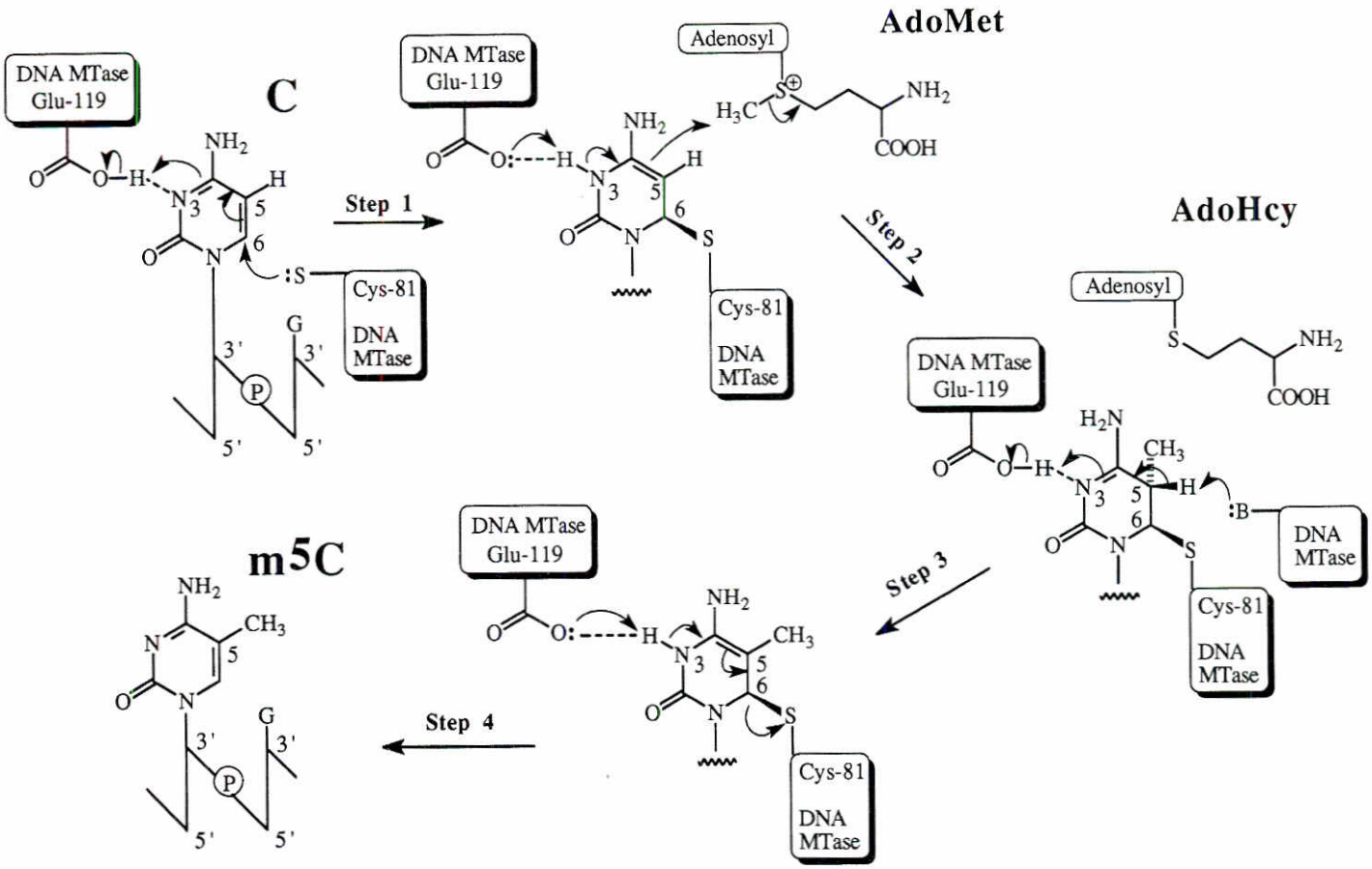

Fig. 5. The chemical process used by bacterial DNA MTases for the covalent addition of a methyl group to cytosine residues in DNA *(based on the work of Wu and Santi, 1987; Chen et al., 1991, 1993; Cheng et al., 1993; Erlanson et al., 1993; Klimasauskas et al., 1994). The amino acids indicated are those of M. Hha I (Klimasauskas et al., 1994). The reaction is initiated by a nucleophilic attack of a cysteinyl thiolate [probably Cys-81 in the Pro-Cys motif that is conserved in all known DNA MTases (Lauster et al., 1989)] on carbon-6 of cytosine, generating a covalent enzyme-DNA intermediate (Step 1). As a means of avoiding the formation of a high-energy carbanion, it is believed that this step is accompanied by protonation of nitrogen-3 by an acidic group (Glu-119) of the enzyme. The resulting enamine attacks the methyl group of AdoMet, transferring it to carbon-5 (Step 2). Thus, a methylated covalent intermediate and AdoHcy are formed. Abstraction of the proton at carbon-5 again yields an enamine (Step 3), which undergoes conjugate elimination to liberate m^5C-containing DNA and regenerate free enzyme (Step 4). The mechanism described implies that the enzyme must cause transient disruption of Watson-Crick base-pairing. In fact, the enzyme swings out the cytosine residue through the minor groove to an extrahelical position during the methylation reaction. The strand separation event is highly localized, possibly involving only the actual C-G pair. The catalytic base (B) that abstracts a proton from carbon-5 remains unknown. It might be an H_2O molecule that diffuses into the active site or the neighboring sulphur atom of Cys-81. This enzymatic DNA methylation is an unusually slow process. Figure modified from Bestor and Verdine (1994) and Verdine (1994).*

it blocks the action of restriction endonucleases when present at their sequence-specific site of cleavage. Some transcription factors, which recognize sequences that contain CpG residues, are prevented from binding when the CpG site is methylated, and thus cannot stimulate transcription. These include E2F, an adenovirus *E2* promoter binding factor (Kovesdi *et al.*, 1987), CREB, cAMP-responsive element binding proteins (Iguchi-Ariga and Schaffner, 1989), AP-2 (Comb and Goodman, 1990), NF-κB (Bednarik *et al.*, 1991), and c-*Myc/Myn* (Prendergast *et al.*, 1991). There is also evidence to suggest that DNA methylation may directly inhibit the action of chloroplast RNA polymerase (Kobayashi *et al.*, 1990).

At variance with the transcription factors described above, Sp1 elements bind DNA and activate transcription even when its consensus binding site is CpG methylated (Harrington *et al.*, 1988; Höller *et al.*, 1988). This suggested the possibility that the sequence-specific binding of Sp1, which is often associated with promoters of housekeeping genes, prevents methylation of CpG islands (Höller *et al.*, 1988; Brandeis *et al.*, 1994; Macleod *et al.*, 1994). Sp1 also counteracts H1-mediated repression, but activates transcription even in the absence of H1 (Croston *et al.*, 1991).

Results obtained by genomic sequencing and *in vivo* dimethylsulfate footprinting methodology, indicate that methylation of CpG sites may be involved in establishing cell type-specific gene expression by preventing the binding of ubiquitous transcription factors to their target sequences in cell types that are not supposed to express the gene in question (Becker *et al.*,

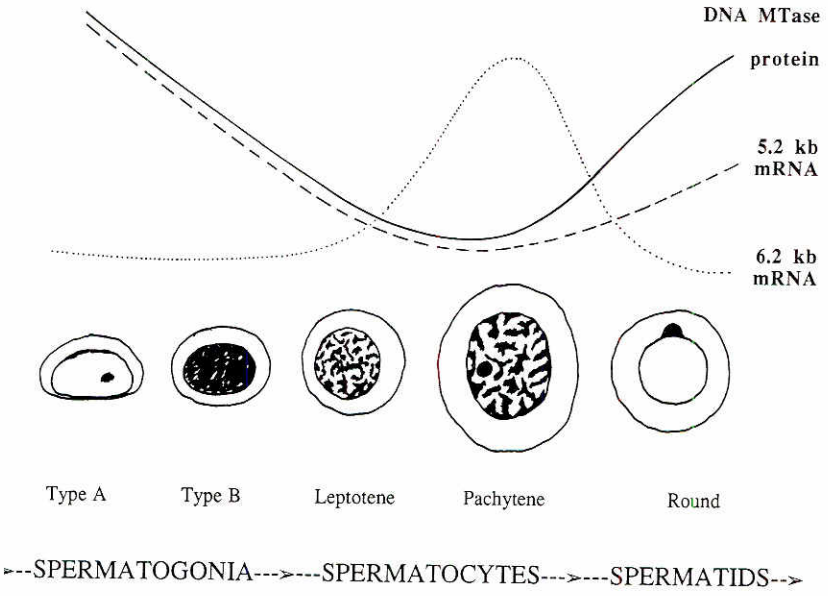

Fig. 6. DNA MTase expression during spermatogenesis in the mouse. *The DNA methylation patterns characteristic of the somatic tissues of adult mammalian genomes are established by de novo methylation during gametogenesis and embryogenesis. During spermatogenesis in the mouse, the MTase content (solid line) is high in type A spermatogonia and in haploid round spermatids. Whether there is de novo DNA MTase activity at these stages [resulting from translation of a 5.2 kb DNA MTase mRNA (dashed line)], or in pachytene spermatocytes and post-meiotic round spermatids [resulting from translation of a 6.2 kb DNA MTase mRNA (dotted line)], remains an open question. Figure based on data from Trasler et al. (1992) and Benoit and Trasler (1994).*

1987; Choi and Chae, 1991). It has been suggested that one of the functions of DNA methylation is to suppress the low levels of basal expression that could result from the exposure of promoter sequences to the multitude of general transcription factors present in each cell (Eden and Cedar, 1994).

Inhibitory effects of sequence-specific methylation in promoter regions have been documented for a large number of DNA-dependent RNA polymerase II-transcribed eukaryotic genes, but also for some RNA polymerase I-transcribed (Bird *et al.*, 1981a,b) and RNA polymerase III-transcribed genes (Jüttermann *et al.*, 1991). In the *Xenopus* embryo, there is a progressive loss of methyl groups in rDNA spacers, beginning when embryonic rRNA synthesis is initiated (at the end of the blastula stage) and accompanying rRNA synthesis as it increases in rate during gastrulation (Bird *et al.*, 1981b).

Effects of DNA methylation on chromatin structure

Another means by which DNA methylation may interfere with gene activity is by altering the overall chromatin structure, thus making genes less accessible to the transcription machinery. Experiments using DNA-mediated transfer of genes, either methylated or unmethylated, into mouse L cells have shown that DNA methylation exerts a general effect on local chromosome structure (Keshet *et al.*, 1986). Thus, unmethylated DNA sequences integrate into the mouse genome in a DNase I-sensitive conformation, whereas CpG-methylated sequences become resistant to DNase I, and form an inactive chromatin structure. The finding that *in vitro* methylated herpes simplex virus thymidine kinase (HSV TK) DNA is transcribed immediately after microinjection into the nuclei of TK^- rat 2 cells, but that the gene becomes inactive upon chromatin formation, suggests that the m^5C groups have to interact with chromatin proteins in order to inhibit gene activity (Buschhausen *et al.*, 1987). Apparently, m^5C groups cannot prevent the binding of RNA polymerase II and transcription factors to promoter sequences until chromatin formation occurs.

Histone H1, being primarily associated with transcriptionally inactive DNA, is considered to be a general repressor of basal RNA polymerase II transcription (Croston *et al.*, 1991). CpG island chromatin, which usually includes the promoter regions of active genes, is structurally different from bulk inactive chromatin (Tazi and Bird, 1990). It exhibits an "open" chromatin structure, characterized by histone H1-depletion (or altered histone H1-binding), by histone H3- and H4-hyperacetylation, and by a nucleosome-free region. This chromatin conformation may allow, or be a consequence of, the binding of transcription factors to promoter regions. Using polyclonal antibodies against m^5C, it has been estimated that at least 80% of the m^5C is localized in nucleosomes that contain histone H1 (Ball *et al.*, 1983). Histone H1 causes stronger inhibition of transcription with a methylated template than with an unmethylated template, probably by preferential binding to the methylated DNA (Levine *et al.*, 1993). This inhibitory effect is exerted at the stage of initiation and not elongation, and the extent of inhibition depends on the density of methyl groups in the promoter region.

Methyl-CpG binding proteins

A methyl-CpG binding protein (MeCP1), abundant in somatic mammalian cells but barely detectable in early embryonic cells, has recently been discovered (Meehan *et al.*, 1989, 1992). Interestingly, it binds strongly to DNA sequences only if there are enough (at least 15) symmetrically methylated CpG moieties for each molecule of MeCP1. Having no apparent nucleotide sequence specificity, MeCP1 should be able to exert its effect in a variety of sequence contexts. Indeed, 4 different methylated promoters were shown to be repressed through interaction with MeCP1 *in vitro* (Boyes and Bird, 1991). The fact that MeCP1-deficient cells show much reduced repression of methylated genes, indicates that MeCP1-mediated repression also occurs in the living cell. Subsequent transfection studies with HeLa cells demonstrated that even a low density of methyl-CpG can repress a weak promoter, apparently by weak binding

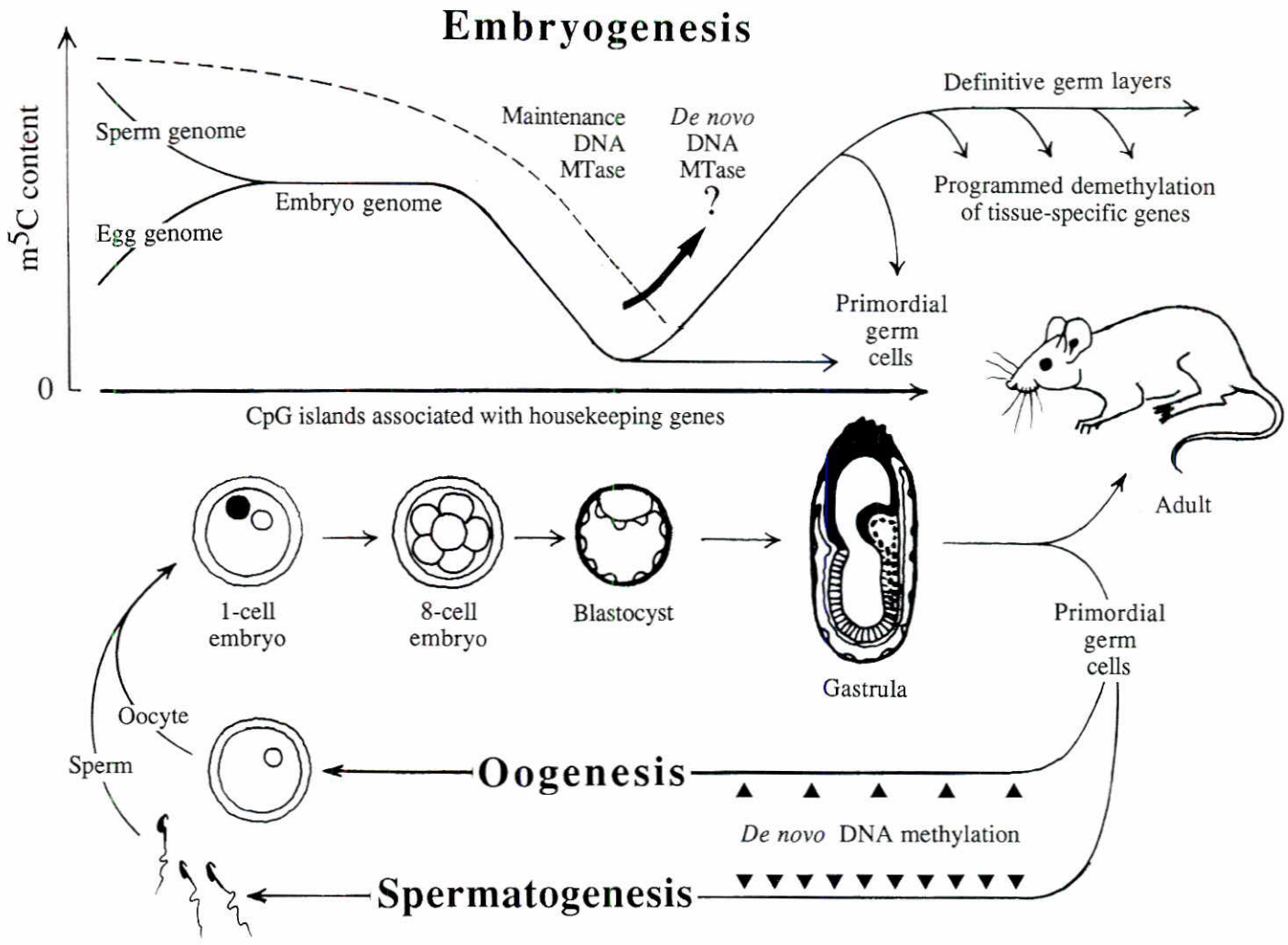

Fig. 7. DNA methylation in mouse development. *The sperm contributes essentially no DNA MTase to the embryo, but the level of maternally inherited enzyme is extremely high in the egg. Nevertheless, the sperm genome is highly methylated whereas the egg genome is undermethylated. The extent of methylation remains constant during early cleavage, but after the 8-cell stage there is a global decrease, and in blastocysts most of the DNA is unmethylated. Following implantation there is a surge of de novo DNA methylation, possibly accounted for by a de novo DNA MTase. The primordial germ cell DNA is markedly undermethylated compared to all other DNAs. It is either protected from the surge of de novo methylation or becomes specifically demethylated. After having reached the developing gonads by day 11, the primordial germ cells begin to re-establish their genomic methylation patterns. In the somatic cells, the bulk of the genome becomes methylated, but CpG islands associated with housekeeping genes are left unmethylated. At later stages of development, tissue-specific genes are subject to programmed demethylation in those cell types where they are expressed. Exceptions to the general rule that CpG islands are unmethylated include the allele-specific CpG island methylation that is associated with imprinted genes, and CpG island methylation in the inactive X chromosome. These DNA methylation events may represent separate regulatory processes, taking place independently of the surge of de novo DNA methylation that begins at implantation. Figure based on data from Monk et al. (1987, 1991) and Carlson et al. (1992).*

of MeCP1 (Boyes and Bird, 1992). It was also shown that repression of the transfected genes is alleviated if the strength of the promoter is increased by including an SV40 enhancer in the gene construct. Nevertheless, it was possible to regain the repressed state by increasing the density of methylation in the promoter region, thereby increasing its affinity for MeCP1. In view of these studies it may be suggested that there is competition between transcription factors and MeCP1 for binding to promoters. When the methylated gene constructs were instead transfected into F9 teratocarcinoma stem cells, their transcription was not inhibited to the same extent as in HeLa cells, probably because the F9 cells are deficient in MeCP1 (Boyes and Bird, 1991).

MeCP2, another methyl-CpG binding protein, is more abundant than MeCP1 (Meehan *et al.*, 1992) and is concentrated in pericentromeric heterochromatin, which contains about 40% of all genomic m^5C, and in euchromatic arms of rodent metaphase chromosomes (Lewis *et al.*, 1992). Unlike MeCP1, MeCP2 can bind to DNA that contains a single methyl-CpG pair. The expression of MeCP2, like that of MeCP1, is strong in somatic mam-

malian cells, but weak in early embryonic cells (Meehan *et al.*, 1992).

Yet another methylated-DNA-binding protein (MDBP), ubiquitous in mammalian cells, has been discovered (Ehrlich and Ehrlich, 1990). It is sequence-specific and appears to interact with a highly degenerate 14-bp consensus sequence.

De novo DNA methylation and expression of viral genomes – a function of the developmental stage

The introduction of foreign cellular and viral genomes into early mouse embryos has been used to study the regulation of gene expression in mammalian development (Jähner *et al.*, 1982; Palmiter *et al.*, 1982). The existence of a *de novo* DNA MTase in early mouse embryos was suggested by the finding that sequences in a fusion plasmid (containing a mouse metallothionein promoter joined to the HSV TK gene) became extensively methylated after microinjection into fertilized mouse eggs (Palmiter *et al.*, 1982).

Early mouse embryos and teratocarcinoma stem cells are non-permissive for DNA and RNA virus gene expression (Lehman *et al.*, 1975; Stewart *et al.*, 1982). When retroviral genomes are introduced into mouse zygotes (by microinjection of proviral DNA) or into morula (4-16-cell) stage preimplantation mouse embryos (by infection with Moloney murine leukemia virus, MoMuLV), the proviral DNAs become *de novo* methylated upon integration into the host genome, and are blocked in expression (Jähner *et al.*, 1982). In contrast, there is no *de novo* methylation and no restriction of virus expression upon infection of post-implantation embryos (day 8 of gestation). The methylation patterns of the proviral genomes are then faithfully preserved throughout the life of the mouse by maintenance DNA methylation. Consequently, there seem to be times in mammalian development when DNA MTase provides an immunological function where unmethylated sites, representing nonself (e.g. viral) DNA, become methylated and the associated gene(s) inactivated. Hemimethylated sites, representing self (genomic) DNA immediately after replication, likewise become methylated. In this scenario, fully methylated sites represent mature self (genomic) DNA.

F9 teratocarcinoma stem cells apparently have an efficient *de novo* methylation activity, because proviral genomes of MoMuLV become highly methylated upon integration (Stewart *et al.*, 1982). In contrast, the same proviral genomes do not become methylated when integrated into the DNA of differentiated derivatives of PSA-4 teratocarcinoma stem cells. Virus expression is also blocked in F9 teratocarcinoma stem cells carrying a single, integrated copy of the SV40 genome, but not in their differentiated counterparts induced by retinoic acid (Linnenbach *et al.*, 1980). A correlation between DNA methylation and loss of infectivity of the viral genome is also seen in studies showing that the germ line-transmitted copy of the viral genome is methylated and non-infectious (Harbers *et al.*, 1981), whereas the somatically acquired copies (derived from the germ line copy) are hypomethylated and infectious (Stuhlmann *et al.*, 1981).

The fact that viral genes introduced into early embryos or teratocarcinoma stem cells become heavily methylated and suppressed, may explain why teratocarcinoma stem cells (when compared to differentiated cells) exhibit such a low transformation frequency in DNA transfection experiments (Pellicer *et al.*, 1980; Stewart *et al.*, 1982).

Methylation patterns of Epstein-Barr virus (EBV) genomes seem to depend on the host cell infected (Minarovits *et al.*, 1991). Thus, Burkitt´s lymphoma biopsy cells have been found to carry highly methylated EBV genomes whereas a lymphoblastoid cell line, transformed by the virus *in vitro*, carries hypomethylated EBV genomes. This difference does not seem to be due to integration of the viral genome into different regions of the host cell DNA, because integration is a rare event in EBV-infected cells. As a rule, the viral DNA is present in the form of circular episomes, which co-replicate with the cellular genome. Despite the major difference in their extent of methylation, both EBV genomes were maintained in a latent state (Minarovits *et al.*, 1991). This clearly suggests that mechanisms other than high overall methylation of the viral episomes contribute to EBV latency.

At variance with the EBV episome, MoMuLV provirus DNA remains unmethylated as long as it is not integrated into the host genome (Jähner *et al.*, 1982). Soon after integration into the genome of F9 cells, however, the viral DNA becomes *de novo* methylated. This may be relevant for the observation that DNA microinjected into mouse zygotes or into *Xenopus* eggs is expressed as long as it remains in an episomal state. Unmethylated plasmid DNA molecules microinjected into *Xenopus* eggs, replicate extrachromosomally and remain unmethylated in progeny molecules (Harland, 1982).

There is ample evidence to suggest that cells of the early embryo possess an efficient *de novo* DNA methylation activity that inactivates any DNA that is integrated into the genome, including proviral DNAs derived from RNA viruses. It has been suggested that this mechanism evolved to protect the developing embryo against the deleterious consequences of virus infections (Jähner *et al.*, 1982).

The DNA of frog virus 3 (FV3) is the most highly methylated of all known animal viruses. Nevertheless, cells infected with this iridovirus are capable of overriding the inhibitory effect that methylation exerts on transcription of viral sequences (Thompson *et al.*, 1986). When plasmid constructs, containing an adenovirus promoter (E1a or E2aE) linked to the gene for chloramphenicol acetyltransferase (CAT), were methylated and transfected into vertebrate cells, essentially no CAT activity was detected. However, upon infection with FV3, expression of the CAT gene is strongly promoted. In contrast, adenovirus type 5 (Ad5) infection of similarly transfected cells does not induce transcription from the methylated CAT construct. It is important to notice that the transfected plasmid DNA from FV3-infected cells was not demethylated. The conclusion that has been drawn, is that the FV3 genome, although methylated at every cytosine residue, apparently encodes a factor that activates transcription and allows the virus to propagate in animal cells (Thompson *et al.*, 1986).

The literature also contains some other examples of viral genomes that can be transcribed despite the presence of methyl moieties that would normally inhibit RNA synthesis. The methylation-inactivated late E2A promoter of Ad2, for example, is reactivated by E1A functions without going through demethylation (Weisshaar *et al.*, 1988). Moreover, the transcriptional block, caused by methylation of two CpG sites in the long terminal

repeat (LTR) of the human immunodeficiency virus (HIV), is overcome by the presence of the HIV *trans*-activator *tat*, without significant demethylation of the HIV LTR (Bednarik *et al.*, 1990).

Establishment of DNA methylation patterns during gametogenesis

Genomic methylation patterns are erased in primordial germ cells, but are re-established by *de novo* methylation during early differentiation of the germ cells in the fetal ovary and testis, and by further modification during postnatal gametogenesis (Groudine and Conkin, 1985; Trasler *et al.*, 1990, 1992; Ariel *et al.*, 1991; Chaillet *et al.*, 1991; Kafri *et al.*, 1992). The DNA of mature gametes contains sex- and gene-specific DNA methylation patterns, with an overall level of methylation that is lower in oocytes than in spermatozoa (Monk *et al.*, 1987). These patterns seem to be important in developmental gene regulation and in genomic imprinting.

In the mouse testis, DNA MTase is expressed at a very high level (Trasler *et al.*, 1992; Benoit and Trasler, 1994). The 5.2 kb DNA MTase mRNA, characteristic of all somatic cells, is also found in the male germ cells. It is present in all meiotic cell types, but decreases in amount during differentiation – from type A spermatogonia to post-replicative pachytene spermatocytes (Fig. 6). However, in pachytene spermatocytes, which are known to be active in *de novo* methylation, there is a major increase in DNA MTase mRNA content, attributable to a novel type of message. This DNA MTase mRNA is about 1 kb longer than that of all other cell types, with the exception of mouse F9 teratocarcinoma stem cells which also express a larger form (L. Frostesjö and O. Heby, unpublished results). The 6.2 kb message disappears and the 5.2 kb form reappears in haploid round spermatids (Trasler *et al.*, 1992; Benoit and Trasler, 1994). Despite the presence of DNA MTase mRNA in pachytene spermatocytes, very little immunoreactive protein is seen (Benoit and Trasler, 1994). In postmeiotic round spermatids, however, the protein is abundant.

The fact that only 15% of the 6.2 kb transcript was associated with polysomes, as compared to 100% of the 5.2 kb transcript, suggests that the 6.2 kb transcript is less efficiently translated and may be subject to regulation at the translational level (Trasler *et al.*, 1992). Notably, the 6.2 kb transcript does not encode a larger enzyme than the 5.2 kb transcript. The two transcripts are believed to arise from the same gene, and their size difference may be due to an alternative transcription start site, alternate splicing and/or an alternate polyadenylation signal.

The large transcript present in non-replicating pachytene spermatocytes has been suggested to encode an enzyme capable of *de novo* methylation of DNA, making spermatozoal DNA more highly methylated than oocyte DNA (Trasler *et al.*, 1992). Such a mechanism could provide a way to differentially regulate the expression of DNA MTase in replicating versus non-replicating spermatogenic cells. These data are consistent with a role for DNA methylation events, not only during DNA replication in premeiotic germ cells, but also during meiotic prophase and postmeiotic development (Trasler *et al.*, 1992; Benoit and Trasler, 1994). It may be relevant to emphasize that the 6.2 kb DNA MTase transcript appears at a time when the male germ cells undergo chromosome pairing and recombination, events that may generate DNA structures specifically recognized by the DNA MTase.

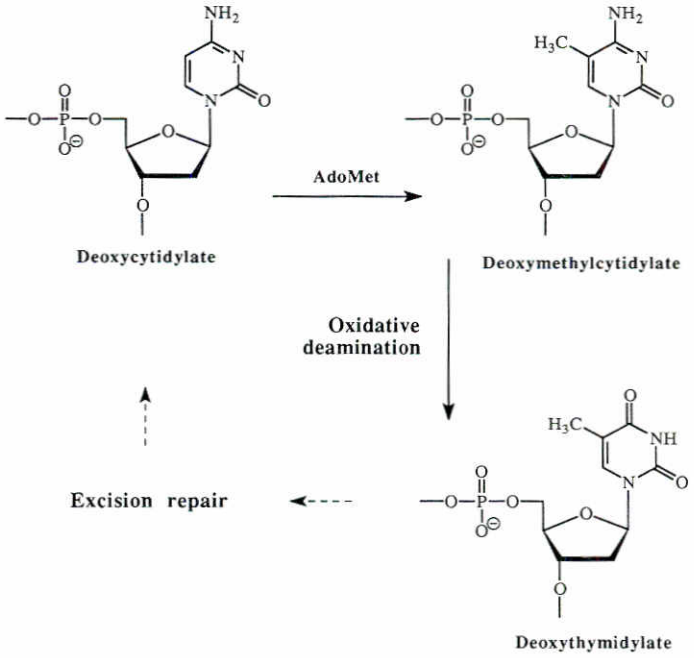

Fig. 8. Conversion of m^{5}C to T by oxidative deamination at the N4 position of m^{5}C. *This reaction may explain why the C→T transition (at CpG dinucleotides) is the most frequent mutation in human DNA. Alternatively, deamination of cytosine may change the base to uracil, which is fixed as T if not repaired prior to DNA replication. Mammalian cells possess a specific mismatch repair pathway, which counteracts the mutagenic effects of these deamination reactions by correcting about 90% of the G·T mismatches back to G·C base pairs. This repair is mediated by excision of the aberrant thymidine monophosphate residue, followed by gap-filling to re-establish the original G·C pair.*

DNA demethylation and methylation during embryogenesis

The cell-type-specific patterns of DNA methylation that are seen in the adult organism are established in a programmed manner during development (Fig. 7). The importance of DNA methylation in early development was clearly demonstrated by the finding that even a modest reduction in DNA methylation, resulting from disruption of the DNA MTase gene by homologous recombination in ES cells and establishment of the mutation in the germline of mice, caused severe developmental abnormalities and death of the mouse embryos during organogenesis (mid-gestation) (Li *et al.*, 1992). In the absence of the large maternal stores of DNA MTase present in the unfertilized egg, it is conceivable that adverse developmental effects would have been established at an earlier time in the homozygous mutant embryos. The homozygous DNA MTase deletion, generated by consecutive targeting of both wild-type alleles, has not been established as a null mutation but may be a partial loss of function mutation, because the m^{5}C content was reduced to only about one third (on day 10) of the normal level. This means that

the homozygous mutant embryos had about $1x10^7$ methylated CpG sites left (per haploid genome). Nevertheless, these embryos exhibited a recessive lethal phenotype. Homozygous mutant ES cell lines were viable and showed normal morphology and proliferation rates in culture, despite a reduction of the genomic m^5C content to about the same extent, i.e. one third of the wild-type level (Li *et al.,* 1992). Whether the continued presence of m^5Cs in the genome of the homozygous mutant embryos and ES cell lines is due to the existence of additional species of DNA MTases (for which there is some evidence), or the production of a truncated DNA MTase with some enzymatic activity, remains to be established.

Methylation patterns established in the oocyte and in the sperm during gametogenesis are subject to dramatic changes during the preimplantation period (Monk *et al.,* 1987) (Fig. 7). The egg genome is undermethylated and the sperm genome relatively methylated. In the 8-cell embryo the extent of DNA methylation corresponds to a mixture of undermethylated maternal and methylated paternal DNA, indicating that there is no loss of genomic methylation during early cleavage. However, between the 8-cell stage and the blastocyst stage there is a global decrease in the extent of methylation in both maternal and paternal DNA (with the paternal DNA being more affected). The cells of the blastocyst exhibit a lower level of m^5C in their DNA than the cells of any other developmental stage, with the possible exception of primordial germ cells. In fact, most of their DNA is unmethylated. This is probably a result of demethylation, but could be due to prevention of maintenance methylation after each DNA replication cycle in the cleaving embryo, for example by lack of AdoMet, the methyl group donor.

Following implantation, there is a surge of *de novo* DNA methylation affecting the entire embryo (Fig. 7) (Monk *et al.,* 1987). However, embryonic and extraembryonic lineages are methylated independently and to different degrees. The DNAs of the extraembryonic lineages in the postimplantation conceptuses (trophectoderm, primary endoderm, extraembryonic ectoderm, chorion) are more methylated than blastocyst DNA, yet undermethylated compared with the 7.5-day (late primitive streak stage) embryonic DNA. The mechanisms that establish these differences in DNA methylation patterns are largely unknown. By gastrulation the extent of DNA methylation reaches the level that characterizes the adult animal. Nevertheless, DNA methylation continues postgastrulation. This *de novo* methylation of the DNA, beginning in the late blastocyst, may be involved in directing the majority of tissue-specific genes to an inactive conformation. The bulk of the genome is methylated, while CpG islands associated with housekeeping genes are left unmethylated. At later stages of embryonic development, tissue-specific genes are subject to programmed demethylation in those cell types where they are expressed. This programmed demethylation may be part of the differentiation process, as in B cells where the κ enhancer seems to regulate differentiation by inducing demethylation and by promoting tissue-specific transcription (Lichtenstein *et al.,* 1994). The adult DNA methylation pattern is characterized by full methylation of the inactive genes and undermethylation of the transcriptionally active genes.

Both in male and female mouse embryos, the germ line is established from cells originating in the epiblast. These primordial germ cells are first seen in the yolk sac (by day 7, i.e. early primitive streak stage), before migrating to the genital ridge. Despite the surge of *de novo* DNA methylation occurring in the embryo prior to gastrulation (by day 6) the primordial germ cells, both male and female, seem to escape from getting their DNA methylated. However, the possibility remains that germ cell DNA is first *de novo* methylated and then demethylated in a cell-specific manner. Nevertheless, when the primordial germ cells reach the developing gonads by day 11, their DNA is markedly undermethylated. The female germ cells enter the prophase of the first meiotic division by day 13. At about the same time the previously inactive X chromosome is reactivated (Monk and McLaren, 1981). The male germ cells, on the other hand, enter mitotic arrest and do not start their mitotic and meiotic cycles until after birth.

Despite these marked changes in the extent of DNA methylation, the maternal and paternal genomes seem to maintain the differences in methylation patterns that may be involved in parental imprinting (see below and Ohlsson and Franklin, this volume).

The activity of maternally inherited DNA MTase is extremely high in the unfertilized egg (Fig. 7) (Monk *et al.,* 1991). The enzyme activity remains constant (on a per-embryo-basis) throughout the first 3 cleavage divisions, but decreases 10-fold between the 8-cell stage and the blastocyst stage. On a per-cell-basis, the decrease in DNA MTase activity between the one-cell and the blastocyst stage is several thousand-fold (Carlson *et al.,* 1992). Despite this overall loss of DNA MTase activity, which may partly result from its degradation and partly from its dilution during cleavage divisions, the cells of the preimplantation embryos clearly contain very high DNA MTase activities relative to somatic cells. In fact, unfertilized eggs and blastocysts were calculated to have 15,000-fold and 40-fold more DNA MTase per nucleus than mouse 3T3 fibroblasts. There is also evidence that oocytes and early embryos have a form of DNA MTase that differs from that of somatic tissues. It may be translated from an alternative start site or escape post-translational modifications, but without losing its catalytic properties. It has not yet been established whether embryos have additional species of DNA MTase with *de novo* methylation activity. It also remains to be determined at what stage of development that the embryonic gene for DNA MTase is activated.

Not only the activity, but also the distribution of DNA MTase within cells, changes during development (Carlson *et al.,* 1992). In oocytes and in early embryos (up to the 4-cell stage) most of the DNA MTase protein is found in the peripheral cytoplasm. In 4- and 8-cell embryos it is seen in cytoplasmic granules, and in 8-cell embryos it is also seen in large amounts in the nuclei. In blastocysts, the DNA MTase present in cytoplasmic granules is unchanged whereas the nuclear amount of the enzyme protein is reduced to very low levels, both in cells of the inner cell mass and in trophoblast cells. Notably, any significant amounts of DNA MTase are found in the nucleus only at the 8-cell stage. These changes in compartmentalization are likely to affect the functional activity of the enzyme, but the mechanism by which the undermethylated state of the blastocyst is attained, remains unclear.

Ectopic *de novo* DNA methylation

Proteolytic processing of mammalian DNA MTase has been shown to stimulate the inherent *de novo* methylation capacity of the enzyme (Bestor, 1992). This phenomenon may contribute to

the process of ectopic *de novo* methylation observed in the DNA of cultured cells (Antequera *et al.*, 1990), tumors (de Bustros *et al.*, 1988; Silverman *et al.*, 1989) and aging animals (Uehara *et al.*, 1989). Gene inactivation by ectopic *de novo* DNA methylation may lead to tumor development if a tumor suppressor locus is affected. If it occurs in the germ line it could contribute to genetic disease as may occur in the case of a gene at the fragile-X locus in humans. The novel DNA methylation pattern, resulting from *de novo* methylation events, is transmitted to daughter cells by maintenance methylation.

Role of DNA methylation in X chromosome inactivation

In somatic cells of adult female eutherian mammals only one of the two X chromosomes is active. DNA methylation is believed to play a role in the X chromosome inactivation process, which causes silencing of the housekeeping genes on one X chromosome, whereas their homologs on the other X chromosome are kept active in the same cell. DNA methylation may not play a role in the primary events of X chromosome inactivation, which take place in the blastula, but may be part of a secondary, tissue-specific mechanism for maintaining the transcriptionally silent state of many genes on the inactive X chromosome (Singer-Sam *et al.*, 1990). The promoter regions of most genes on the inactive X chromosome are highly methylated and lack bound transcription factors, and the genes are functionally inactive (Pfeifer *et al.*, 1990a,b).

During mouse embryo development, inactivation of an X chromosome occurs first in extraembryonic lineages (trophectoderm and primitive endoderm) and later in embryonic tissue (primitive ectoderm) (Monk and Harper, 1979). Preferential inactivation of the paternal X chromosome in the early extraembryonic lineages suggests that paternal and maternal genomes are distinguishable. At variance, inactivation in the early embryonic tissue (primitive ectoderm), is random.

On the inactive X chromosome even the CpG islands are methylated. Once a specific DNA methylation pattern has been achieved on an X chromosome in the post-implantation embryo, the normal maintenance DNA MTase activity will automatically propagate the pattern, and possibly the inactive state, to all sister cells.

The *Xist* (X-inactivating specific transcript) gene, which is expressed exclusively from the inactive X chromosome, may be responsible for the initiation of X inactivation. Since the transcript is not translated into a protein, the RNA itself is believed to be involved in the X inactivation process. It has been recently shown that the *Xist* gene has a CpG island that is methylated at the silent allele and unmethylated at the expressed allele (Kay *et al.*, 1994; Norris *et al.*, 1994). In fact, it was proposed that DNA methylation has a primary causal role in the X-inactivation process. Moreover, the preferential inactivation of the paternal X chromosome, and expression of the paternal *Xist* allele, in some mouse tissues, suggests an imprinting effect.

DNA methylation and the fragile-X syndrome

Fragile-X syndrome is the most common genetic cause of mental retardation in humans after trisomy 21 and the most common cause of hereditary mental retardation (Sutherland, 1985; Reiss *et al.*, 1995). The syndrome is associated with a chromosomal fragile site (*FRAXA*) that can be induced *in vitro* by deoxynucleotide pool perturbation (Sutherland, 1985). The fragile site is observed as an isochromatic gap of poorly staining chromatin at Xq27.3.

The target of mutations that cause fragile-X mental retardation syndrome is a CGG trinucleotide repeat in a gene termed *FMR-1* (fragile-X mental retardation-1) (Verkerk *et al.*, 1991), encoding an RNA-binding protein (Siomi *et al.*, 1993) that is predominantly present in the cytoplasm (Devys *et al.*, 1993; Verheij *et al.*, 1993). The *FMR-1* gene is 38 kb long and has 17 exons (Eichler *et al.*, 1993). Within its first exon there is a tract of CGG tandem repeats which is polymorphic, varying from 2 to 60 triplets among normal individuals. This repeat element is confined to the long (≥318 nt) 5'-untranslated region of the primary transcript (Ashley *et al.*, 1993). As a result of alternative splicing, the latter may generate 12 distinct mRNA products encoding 12 *FMR-1* isoproteins (Ashley *et al.*, 1993; Devys *et al.*, 1993; Verheij *et al.*, 1993; Verkerk *et al.*, 1993). The $(CGG)_n$-repeat being noncoding is consistent with the finding that it is not conserved in evolution. In fact, the CGG repeats in the *FMR-1* mRNA may be an exclusively mammalian phenomenon (Siomi *et al.*, 1995). The alternative splicing allows functional diversity of the *FMR-1* gene (Verkerk *et al.*, 1993), but it remains

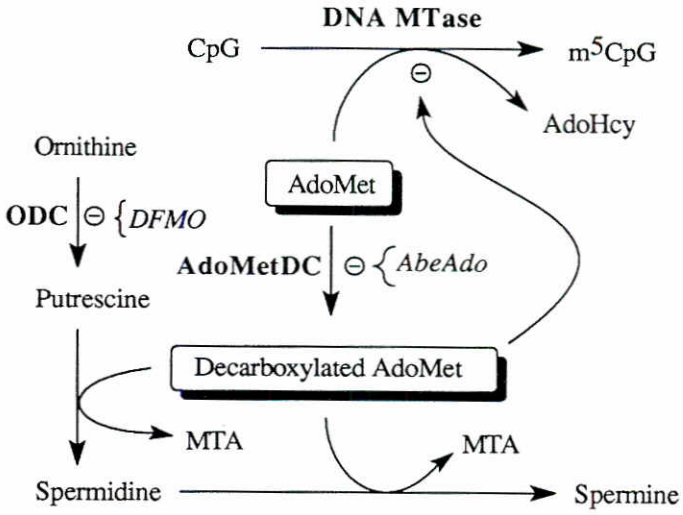

Fig. 9. Mechanism by which polyamine depletion may cause inhibition of DNA methylation in F9 teratocarcinoma stem cells. *When ornithine decarboxylase (ODC) is irreversibly inactivated by treatment with DFMO there is a precipitous drop in the cellular putrescine and spermidine content and a concomitant accumulation of decarboxylated AdoMet. The latter is a consequence of the lack of aminopropyl group acceptors (putrescine and spermidine) for decarboxylated AdoMet, which is normally the aminopropyl group donor in spermidine and spermine synthesis. In these reactions, 5'-methylthioadenosine (MTA) is also formed. At the high concentrations reached, decarboxylated AdoMet interferes with AdoMet in the DNA methylation reaction, where AdoMet acts as the methyl group donor. By inhibiting the AdoMet decarboxylase (AdoMetDC) activity with AbeAdo (Fig. 10) it is possible to prevent the accumulation of decarboxylated AdoMet and the inhibition of DNA methylation.*

to be determined whether all 12 possible proteins are synthesized and whether they are functionally active. Thus far, the transcripts have been shown to translate into at least 4 discrete proteins (Verkerk *et al.,* 1993), which exhibit no tissue specificity. Interestingly, the CGG repeat is on the same exon as the ATG start codon and cannot be spliced out without removing the translational start signal.

Clinically affected individuals exhibit a major expansion of the repeat number with over 230 tandem repeats and often exceeding 1000. Since the translational start signal is downstream of the CGG-repeat (Ashley *et al.,* 1993; Devys *et al.,* 1993; Verheij *et al.,* 1993), the *FMR-1* proteins should be unaffected by the amplification. However, the massive expansion of the CGG trinucleotide repeat is associated with transcriptional silencing of the *FMR-1* gene, apparently as a result of methylation of all CpGs both in and surrounding the tract of repeats, including an unmethylated CpG island within the promoter region of the *FMR-1* gene, approximately 250 nucleotides upstream of the repeat (Verkerk *et al.,* 1991; Hornstra *et al.,* 1993). Methylation may be elicited by the tremendous density of CpG sites present in the fully expanded CGG-repeat (Sutcliffe *et al.,* 1992). It should be emphasized that it is not the expansion of the CGG repeat as such, but instead the methylation of the CpG island and the CGG repeat, that causes transcriptional silencing of the *FMR-1* gene (Sutcliffe *et al.,* 1992), thereby generating at least the major features of the fragile-X phenotype.

It has recently become evident that not only transcriptional silencing, but also translational suppression may be the cause of *FMR-1* protein deficiency. Thus, in a study of fibroblast subclones from a patient, it was found that when the trinucleotide repeat expansion exceeded 200 repeats, the 40S ribosomal subunits stalled in the 5´-untranslated region of the message, thus preventing further *FMR-1* protein synthesis (Feng *et al.,* 1995).

The phenotype of *FMR-1* knockout mice, lacking normal *FMR-1* protein, and exhibiting macroorchidism, learning deficits and abnormal behavior, lends support to the contention that *FMR-1* has a central role in fragile-X syndrome (Bakker *et al.,* 1994).

The 4.4 kb *FMR-1* transcript is abundant in tissues of phenotypic importance such as brain and testis (Hinds *et al.,* 1993) yet it exhibits no absolute tissue specificity. It is predicted to encode a 70 kDa protein, whose function remains elusive (Verkerk *et al.,* 1991; Devys *et al.,* 1993; Siomi *et al.,* 1993; Verheij *et al.,* 1993). Presumably the absence of *FMR-1* protein in males with a full mutation contributes to the clinical phenotype associated with fragile-X syndrome. The syndrome may also result from the expression of a mutant protein that is impaired in RNA binding (Siomi *et al.,* 1994). Although there is a strong connection between fragile-X syndrome and the RNA-binding activity of *FMR-1*, the cognate RNA target of *FMR-1* and its precise functions remain elusive.

The CGG repeat appears to be a protein-binding motif to which a nuclear $p(CCG)_n$-binding protein (CCB-BP1) may bind (Richards *et al.,* 1993). Methylation of the repeat sequence prevents the *in vitro* binding of this protein, instead allowing a different protein (possibly MeCP1) to bind (Richards *et al.,* 1993). If this protein is MeCP1 then its binding would further suppress *FMR-1* transcription since MeCP1 is a known repressor of transcription (Boyes and Bird, 1991). This finding may explain the mechanism by which methylation shuts down transcription of the *FMR-1* gene (Richards *et al.,* 1993).

Remarkably, the expansion of a premutation to a full mutation in *FMR-1* occurs only on the maternal X, apparently in a postzygotic stage (between day 5 and day 20 of gestation) (Reyniers *et al.,* 1993). While the expanded CGG repeat mutation is observed in both the chorionic villi and fetus, the methylation of the CpG island is limited to the fetal DNA (Sutcliffe *et al.,* 1992). *FMR-1* gene expression is repressed in the fetal tissues where the CpG island is methylated, while the undermethylated chorionic villi expressed *FMR-1* (Sutcliffe *et al.,* 1992). Since the genetic background of these tissues is identical, including the fragile-X chromosome, these data indicate that the abnormal methylation of the *FMR-1* CpG-island is responsible for the loss of *FMR-1* transcription and suggests that the methylation occurs around the 10th week of gestation (Sutcliffe *et al.,* 1992). The finding that *FMR-1* mRNA is present in the unmethylated chorionic villi sample, despite an expanded full mutation, suggests that the lack of transcription observed in the fetal tissue is a consequence of the DNA methylation (Sutcliffe *et al.,* 1992). These data suggest that induced demethylation, e.g. by 5-azacytidine treatment, may restore mRNA production and *FMR-1* protein synthesis, and may provide a potential means of therapeutic intervention in fragile-X syndrome.

Another fragile site *(FRAXE)*, 600 kb distal to *FRAXA*, is also associated with mental retardation, but the affected gene has yet to be reported. It has recently been shown to possess amplifications of a GCC repeat unit adjacent to a CpG island in Xq28 of the human X chromosome (Knight *et al.,* 1993). Normal alleles have between 6 and 25 copies of the GCC repeat whereas affected individuals have between 115 and 850. In the families affected the *FRAXA* CGG amplification is absent and the amplified fragile site *FRAXE* is distal to *FRAXA*. Since the *FRAXE* repeat sequence (GC**CpG**CC) also contains the dinucleotide CpG it can be methylated. The expanded GCC repeat and the adjacent CpG island are hypermethylated in affected males but unmethylated in normal males (Knight *et al.,* 1993). It is likely that a mechanism similar to that in fragile-X syndrome is operating and that large expansions at the *FRAXE* locus lead to methylation of the CpG sites of the repeat sequence and the CpG island, inactivating an associated gene and thereby causing mental retardation.

An autosomal homolog (*FXR-1*) of the *FMR*-1 gene has recently been discovered (Siomi *et al.,* 1995). The FXR-1 protein, like the FMR-1 protein, contains sequence motifs characteristic of RNA binding proteins, including two KH domains and an RGG box. In fact, the two proteins have similar RNA binding properties *in vitro*. Although the two proteins are closely related (probably members of a gene family) the *FXR-1* protein does not seem to be able to complement the lack of *FMR-1* protein function.

DNA methylation in genomic imprinting

Experiments in which pronuclei were transferred between mouse zygotes clearly demonstrated that development cannot proceed properly if an embryo is derived from two maternal or two paternal pronuclei (Solter, 1987; Surani *et al.,* 1988). Gynogenetic embryos are small and morphologically normal, but

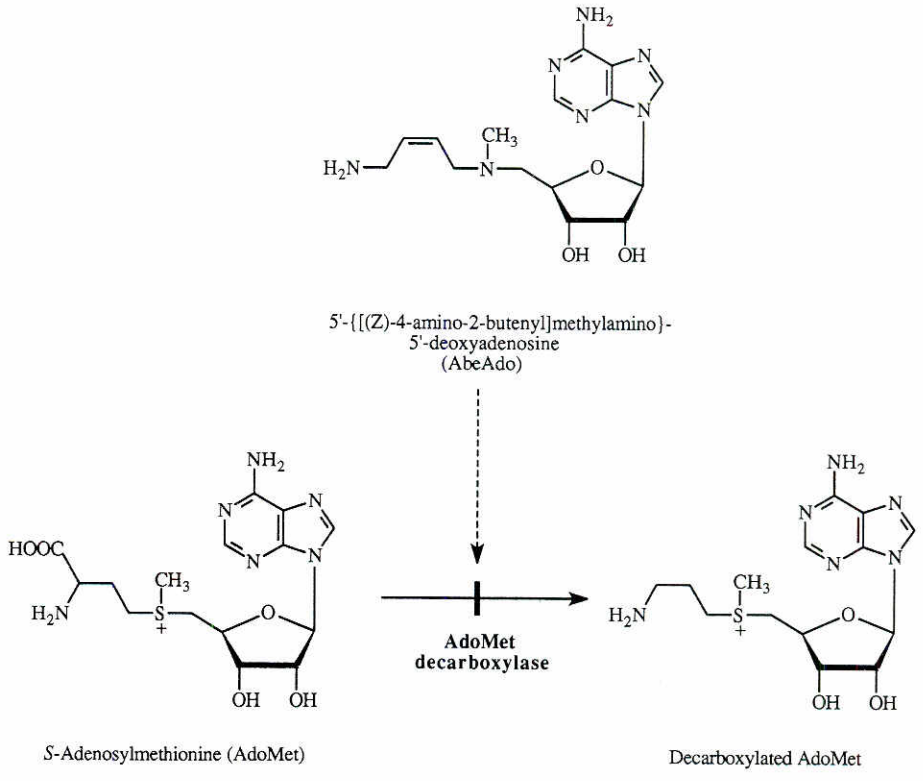

Fig. 10. Structural formula of AbeAdo as compared to those of AdoMet and decarboxylated AdoMet. *AbeAdo is an enzyme-activated irreversible inhibitor of AdoMet decarboxylase (Danzin et al., 1990), which catalyzes a key step in polyamine biosynthesis (Fig. 9).*

do not fully develop extraembryonic membranes, and therefore abort. Androgenetic embryos, on the other hand, develop morphologically normal extraembryonic membranes, but the embryonic part is greatly reduced. From these experiments it may be concluded that the expression patterns of the parental genomes are different and complementary. Genomic imprinting is the term most frequently used to denote a form of developmental gene regulation, in which only one of the parental alleles is expressed (Barlow, 1994).

A number of imprinted genes have been identified in both mice and humans. Insulin-like growth factor 2 (*Igf2*), for example, is expressed exclusively from the paternal allele whereas insulin-like growth factor 2 receptor (*Igf2r*) and *H19* , which encodes an untranslated RNA, are transcribed only from their maternal allele in somatic cells. The imprint seems to be established in the gametes, maintained during embryogenesis and in adult somatic tissues, and erased in the early germ line (Razin and Cedar, 1994). DNA methylation, as a heritable and reversible modification process, is clearly a good candidate for marking and distinguishing individual parental alleles at the molecular level. In fact, several transgenes have been shown to be inherited in the unmethylated state from parents of one sex and in the methylated state from parents of the other sex (Swain *et al.*, 1987; Surani *et al.*, 1988).

Recently, a direct genetic proof has been obtained for the involvement of DNA methylation in the maintenance of allele-specific imprinted gene expression (Li *et al.*, 1993). As a result of low levels of maintenance methylation in mouse embryos homozygous for a DNA MTase mutation, the *H19* gene (which normally is expressed exclusively from the maternal allele) becomes active at both alleles, but the genes for *Igf2* and *Igf2r* become completely silenced. The latter finding may be explained by assuming that m^5Cs in the *Igf2* and *Igf2r* promoters normally prevent the binding of a repressor, and that demethylation reverses this effect.

For further information and additional references, see Ohlsson and Franklin (this volume).

DNA methylation in mutagenesis and cancer

On the basis of a large number of studies, documenting various changes in DNA methylation patterns or DNA MTase expression in cancer cells, it has long been speculated that DNA methylation plays a role in malignant transformation. Both DNA hypomethylation and hypermethylation have been associated with cancer (for references, see Baylin *et al.*, 1991; Jones *et al.*, 1992; Laird and Jaenisch, 1994). Only recently, however, a more direct and possibly causal link between DNA methylation and cancer has been established.

A coupling between hypomethylation and overexpression of specific growth-promoting genes (such as proto-oncogenes) has been observed in a wide variety of tumors (for references, see Laird and Jaenisch, 1994), and has frequently been extrapolated to indicate causality. Likewise, there have been many reports showing a correlation between hypermethylation and inactivation of growth-inhibiting genes (such as tumor-suppressor genes), and a chromosomal "hot spot" for hypermethylation has been identified in the short arm of chromosome 11, in an area known to harbor tumor-suppressor genes (de Bustros *et al.*, 1988).

In established cell lines, CpG islands associated with nonessential genes may become accidentally methylated, and

the genes inactivated (Antequera *et al.,* 1990). This may result in a selective advantage for a cell under culture conditions. It is conceivable that an analogous mechanism may lead to inactivation of tumor-suppressor genes during tumorigenesis. There is data to suggest epigenetic silencing, by DNA hypermethylation of normally unmethylated CpG islands in 5´ flanking regions, of the *RB* gene in retinoblastoma (Greger *et al.,* 1989), the von Hippel-Lindau (*VHL*) tumor-suppressor gene in renal carcinoma (Herman *et al.,* 1994), the *bcr-abl* locus in chronic myelogenous leukemia (Zion *et al.,* 1994), and the estrogen receptor gene in colorectal tumors (Issa *et al.,* 1994). Interestingly, a correlation between estrogen receptor CpG island methylation and aging was also found, and it was suggested that methylation of this CpG island in aging colorectal cells could be the first step in the development of colorectal tumors. These studies are all consistent with the idea that aberrant promoter methylation of tumor-suppressor genes represents an epigenetic inactivation pathway contributing to tumorigenesis. The possibility remains, however, that the hypermethylation of tumor-suppressor genes seen in cancer cells is merely a secondary event, helping to maintain gene inactivity acquired by other means.

The basis for hypermethylation of genes may be an increase in *de novo* DNA methylation capacity. DNA MTase gene expression has been shown to increase in early stages of human colonic neoplasia and to continue increasing during tumor progression (El-Deiry *et al.,* 1991). Moreover, constitutive overexpression of an exogenous mouse DNA MTase has been found to cause a marked increase in overall DNA methylation and to induce tumorigenic transformation of NIH 3T3 mouse fibroblasts in culture (Wu *et al.,* 1993). Conversely, the expression of antisense DNA MTase RNA has been shown to cause DNA demethylation and inhibition of tumorigenesis (MacLeod and Szyf, 1995). These findings are consistent with the idea of DNA MTase being a mutator enzyme. In fact, it has been reported that the enzyme itself (at least a prokaryotic DNA MTase) can be highly mutagenic when AdoMet levels are limiting (Shen *et al.,* 1992). AdoMet limiting conditions may occur in early stages of tumor development, leading to high rates of DNA MTase-mediated CpG mutagenesis. The frequently observed methionine auxotrophy of cancer cells and the tumorigenic effects of methyl-deficient diets are consistent with such a mechanism. However, it is also conceivable that methyl-deficient diets promote carcinogenesis by stimulating proto-oncogene expression. Thus, AdoMet depletion has been shown to induce DNA hypomethylation within sequences specifying the c-*myc*, c-*fos* and c-Ha-*ras* genes and, possibly as a consequence, to stimulate the expression of these genes (Wainfan and Poirier, 1992). The hypomethylation observed in proto-oncogenes, however, could also be due to active demethylation, e.g. by nucleotide excision repair (Jost, 1993; Jost *et al.,* 1995).

Cytosine methylation *per se* creates a pre-mutagenic site, because hydrolytic deamination of m^5C yields thymine (Fig. 8). It has been estimated that such C→T transition mutations account for at least 30% of all germ line point mutations in humans (Cooper and Youssoufian, 1988; Sved and Bird, 1990; Jones *et al.,* 1992), and m^5C residues are therefore considered mutation hot spots. Thus, at the same time as DNA methylation is essential for normal embryonic development (Li *et al.,* 1992) it places a heavy mutational load on the mammalian genome.

Not only m^5C, but also cytosine residues are subject to spontaneous deamination, but the rate is much lower than for m^5C, and the end product is uracil instead of thymine. Thymine, being a normal component of DNA, is not easily recognized as the mutated residue in a G·T mismatch. Uracil, on the other hand, not being a normal component of DNA, is more easily recognized and repaired. Spontaneous hydrolytic deamination of m^5C to thymine has been estimated to generate up to 12 G·T mismatched base pairs in the human genome per day (Wiebauer and Jiricny, 1989). However, mammalian cells possess a specific mismatch repair pathway, which counteracts the mutagenic effects of this deamination by correcting G·T mismatches back to the G·C pairs that were lost through deamination of m^5Cs (Brown and Jiricny, 1987, 1988). This repair is mediated by excision of the aberrant thymidine monophosphate residue, followed by gap-filling to generate a G·C pair (Fig. 8) (Wiebauer and Jiricny, 1989). Thus, about 90% of these mutations have been estimated to be repaired. Nevertheless, spontaneous deamination of m^5C residues seems to be responsible for the high incidence of C→T transitions found for example in the p53 tumor-suppressor gene.

In view of their finding, that the fidelity by which human DNA MTase copies a methylation pattern is comparable to the fidelity by which mammalian DNA polymerases copy a DNA sequence, Smith *et al.* (1992) concluded that the high frequency of C→T transitions at CpG sites in human DNA is unlikely to be due to the normal enzymatic maintenance of methylation patterns. The mutation rate of the CpG dinucleotide is 10-40-fold that of other dinucleotides (Cooper and Youssoufian, 1988; Sved and Bird, 1990; Jones *et al.,* 1992). The possibility that this high rate of mutation at CpG dinucleotides is due to DNA MTase-facilitated deamination of m^5C residues is discussed by Laird and Jaenisch (1994). On the basis of present knowledge it may be concluded that DNA methylation plays an important role in the generation of mutations in human tumors.

Absence of m^5C in small genome metazoans

Although DNA methylation may be one of the mechanisms underlying differential programming of cell lineages in mammalian development there must be other mechanisms involved in the corresponding developmental processes in animals that have no detectable m^5Cs in their DNA, e.g. *Drosophila melanogaster* (Urieli-Shoval *et al.,* 1982) and *Caenorhabditis elegans* (Simpson *et al.,* 1986). There is evidence for heritable gene repression in these species. Thus, regulatory factors may catalyze the assembly of ubiquitous chromosomal proteins on specific sequences and thereby establish a repressed chromatin state. These protein complexes can be transmitted by mitotic inheritance and persist for many cell generations after the disappearance of the regulatory factors, as in the regulation of homeotic gene expression in *Drosophila* (Orlando and Paro, 1995; Simon, 1995). Thus in early embryogenesis, products of maternal and segmentation genes help to assemble inhibitory Polycomb group (Pc-G) proteins, or activating trithorax group (trx-G) proteins, on homeotic genes. The pattern established during the transient presence of the maternal and segmentation gene products is maintained by mitotic inheritance and is used as a means of specifying cellular identity along the ante-

rior-posterior axis by directing the expression of homeotic genes. By analogy, methylation of a mammalian promoter may direct the binding of ubiquitous proteins (recognizing methylated sequences) to a particular gene locus. Such proteins (MeCP1 and MeCP2 (Meehan *et al.,* 1989, 1992), being two obvious candidates) may then catalyze the assembly of repressing factors in the promoter region, which in their turn will generate a condensed, and transcriptionally repressed, chromatin state.

Recently, some very interesting sequence similarities were discovered in DNA MTase and *ALL-1/HRX*, a mammalian homolog of the *Drosophila* trithorax (Ma *et al.,* 1993). These proteins share a very unusual cysteine-rich region, which is not discernible in *Drosophila*, however. Possible implications of this finding, which reveals a link between two systems known to be involved in heritable gene expression during development, are discussed in a recent review by Bestor and Verdine (1994).

DNA hypomethylation caused by polyamine deprivation

When the first step in the polyamine biosynthetic pathway is blocked, F9 teratocarcinoma stem cells accumulate large amounts of decarboxylated AdoMet, the aminopropyl group donor in polyamine synthesis (Fig. 9) (Frostesjö *et al.,* manuscript in preparation). This accumulation is a direct consequence of the polyamine-depleted state of the cell; firstly the polyamine levels become limiting and the polyamines can no longer serve as acceptor molecules for the aminopropyl groups, i.e. decarboxylated AdoMet is not consumed (Oredsson *et al.,* 1986); secondly, the lack of polyamines causes feedback stimulation of AdoMet decarboxylase (Stjernborg *et al.,* 1993), the enzyme that catalyzes the formation of decarboxylated AdoMet. Although the decarboxylated AdoMet molecule contains a methyl group, it does not serve as a methyl group donor in DNA methylation (Heby *et al.,* 1988). Instead it acts as a competitive inhibitor of DNA MTase. The consequence of long-lasting polyamine depletion is therefore genome-wide loss of DNA methylation due to insufficient maintenance methylation during successive rounds of DNA replication (Frostesjö *et al.,* in preparation). Of particular interest is that the ultimate effect of polyamine depletion is terminal differentiation of the F9 cells into parietal endoderm cells (Oredsson *et al.,* 1985). Our recent finding that prevention of the accumulation of decarboxylated AdoMet, by irreversibly inhibiting the AdoMet decarboxylase activity with 5´-{[(Z)-4-amino-2-butenyl]methylamino}-5´-deoxyadenosine (AbeAdo) (Figs. 9 and 10) (Danzin *et al.,* 1990), counteracts the differentiative effect (Frostesjö *et al.,* in preparation) lends further support to the hypothesis proposed.

In the presence of elevated dcAdoMet levels some CpG sites may not be methylated after DNA replication. If these sites remain unmethylated during the subsequent S phase, this will lead to a heritable loss of methylation at these sites. This might well be the mechanism for the process of DNA demethylation that occurs during DFMO-mediated teratocarcinoma stem cell differentiation (Frostesjö *et al.,* in preparation). Teratocarcinoma stem cells have been shown to exhibit high levels of DNA methylation (Singer *et al.,* 1979) and they may actually possess a *de novo* DNA MTase activity (Stewart *et al.,* 1982).

Acknowledgments

The author's laboratory is supported by grants from the Swedish Natural Science Research Council, the Swedish Cancer Society, Lion's Cancer Research Foundation at Umeå University, and the J.C. Kempe Foundation. I am grateful to Dr. Timothy H. Bestor for critical reading of the manuscript.

References

ANTEQUERA, F. and BIRD, A. (1993). Number of CpG islands and genes in human and mouse. *Proc. Natl. Acad. Sci. USA 90:* 11995-11999.

ANTEQUERA, F., BOYES, J. and BIRD, A. (1990). High levels of de novo methylation and altered chromatin structure at CpG islands in cell lines. *Cell 62:* 503-514.

ARIEL, M., McCARREY, J. and CEDAR, H. (1991). Methylation patterns of testis-specific genes. *Proc. Natl. Acad. Sci. USA 88:* 2317-2321.

ASHLEY, C.T., SUTCLIFFE, J.S., KUNST, C.B., LEINER, H.A., EICHLER, E.E., NELSON, D.L. and WARREN, S.T. (1993). Human and murine *FMR-1:* alternative splicing and translational initiation downstream of the CGG-repeat. *Nature Genet. 4:* 244-251.

BAKKER, C.E. and 18 coauthors (1994). *Fmr1* knockout mice: a model to study fragile X mental retardation. *Cell 78:* 23-33.

BALL, D.J., GROSS, D.S. and GARRARD, W.T. (1983). 5-methylcytosine is localized in nucleosomes that contain histone H1. *Proc. Natl. Acad. Sci. USA 80:* 5490-5494.

BARLOW, D.P. (1994). Imprinting: a gamete's point of view. *Trends Genet. 10:* 194-199.

BAYLIN, S.B., MAKOS, M., WU, J., YEN, R-W.C., DE BUSTROS, A., VERTINO, P. and NELKIN, B.D. (1991). Abnormal patterns of DNA methylation in human neoplasia: Potential consequences for tumor progression. *Cancer Cells 3:* 383-390.

BECKER, P.B., RUPPERT, S. and SCHÜTZ, G. (1987). Genomic footprinting reveals cell type-specific DNA binding of ubiquitous factors. *Cell 51:* 435-443.

BEDNARIK, D.P., COOK, J.A. and PITHA, P.M. (1990). Inactivation of the HIV LTR by DNA CpG methylation: evidence for a role in latency. *EMBO J. 9:* 1157-1164.

BEDNARIK, D.P., DUCKETT, C., KIM, S.U., PEREZ, V.L., GRIFFIS, K., GUENTHER, P.C. and FOLKS, T.M. (1991). DNA CpG methylation inhibits binding of NF-kappa B proteins to the HIV-1 long terminal repeat cognate DNA motifs. *New Biol. 3:* 969-976.

BEHE, M. and FELSENFELD, G. (1981). Effects of methylation on a synthetic polynucleotide: the B-Z transition in poly(dG-m^{5}dC)·poly(dG-m^{5}dC). *Proc. Natl. Acad. Sci. USA 78:* 1619-1623.

BENOIT, G. and TRASLER, J.M. (1994). Developmental Expression of DNA methyltransferase messenger ribonucleic acid, protein, and enzyme activity in the mouse testis. *Biol. Reprod. 50:* 1312-1319.

BESTOR, T. (1987). Supercoiling-dependent sequence specificity of mammalian DNA methyltransferase. *Nucleic Acids Res. 15:* 3835-3843.

BESTOR, T.H. (1990). DNA methylation: evolution of a bacterial immune function into a regulator of gene expression and genome structure in higher eukaryotes. *Philos. Trans. R. Soc. Lond. Biol. 326:* 179-187.

BESTOR, T.H. (1992). Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain. *EMBO J. 11:* 2611-2617.

BESTOR, T.H. and INGRAM, V.M. (1983). Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. *Proc. Natl. Acad. Sci. USA 80:* 5559-5563.

BESTOR, T.H. and VERDINE, G.L. (1994). DNA methyltransferases. *Curr. Opin. Cell Biol. 6:* 380-389.

BESTOR, T., LAUDANO, A., MATTALIANO, R. and INGRAM, V. (1988). Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *J. Mol. Biol. 203:* 971-983.

BIRD, A.P. (1987). CpG islands as gene markers in the vertebrate nucleus. *Trends Genet. 3:* 342-347.

BIRD, A.P., TAGGART, M.H. and GEHRING, C.A. (1981a). Methylated and unmethylated ribosomal RNA genes in the mouse. *J. Mol. Biol. 152:* 1-17.

BIRD, A., TAGGART, M. and MACLEOD, D. (1981b). Loss of rDNA methylation accompanies the onset of ribosomal gene activity in early development of *X. laevis*. *Cell 26:* 381-390.

BOYES, J. and BIRD, A. (1991). DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. *Cell 64:* 1123-1134.

BOYES, J. and BIRD, A. (1992). Repression of genes by DNA methylation depends on CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein. *EMBO J. 11:* 327-333.

BRANDEIS, M., FRANK, D., KESHET, I., SIEGFRIED, Z., MENDELSOHN, M., NEMES, A., TEMPER, V., RAZIN, A. and CEDAR, H. (1994). Sp1 elements protect a CpG island from *de novo* methylation. *Nature 371:* 435-438.

BRANDEIS, M., KAFRI, T., ARIEL, M., CHAILLET, J.R., McCARREY, J., RAZIN, A. and CEDAR, H. (1993). The ontogeny of allele-specific methylation associated with imprinted genes in the mouse. *EMBO J. 12:* 3669-3677.

BROWN, T.C. and JIRICNY, J. (1987). A specific mismatch repair event protects mammalian cells from loss of 5-methylcytosine. *Cell 50:* 945-950.

BROWN, T.C. and JIRICNY, J. (1988). Different base/base mispairs are corrected with different efficiencies and specificities in monkey kidney cells. *Cell 54:* 705-711.

BUSCHHAUSEN, B., WITTIG, B., GRAESSMANN, M. and GRAESSMANN, A. (1987). Chromatin structure is required to block transcription of the methylated herpes simplex virus thymidine kinase gene. *Proc. Natl. Acad. Sci. USA 84:* 1177-1181.

CARLSON, L.L., PAGE, A.W. and BESTOR, T.H. (1992). Properties and localization of DNA methyltransferase in preimplantation mouse embryos: implications for genomic imprinting. *Genes Dev. 6:* 2536-2541.

CAROTTI, D., PALITTI, F., LAVIA, P. and STROM, R. (1989). *In vitro* methylation of CpG-rich islands. *Nucleic Acids Res. 17:* 9219-9229.

CHAILLET, J.R., VOGT, T.F., BEIER, D.R. and LEDER, P. (1991). Parental-specific methylation of an imprinted transgene is established during gametogenesis and progressively changes during embryogenesis. *Cell 66:* 77-83.

CHEN, L., MacMILLAN, A.M. and VERDINE, G.L. (1993). Mutational separation of DNA binding from catalysis in a DNA cytosine methyltransferase. *J. Am. Chem. Soc. 115:* 5318-5319.

CHEN, L., MacMILLAN, A.M., CHANG, W., EZAZ-NIKPAY, K., LANE, W.S. and VERDINE, G.L. (1991). Direct identification of the active-site nucleophile in a DNA (cytosine-5)-methyltransferase. *Biochemistry 30:* 11018-11025.

CHENG, X., KUMAR, S., POSFAI, J., PFLUGRATH, J.W. and ROBERTS, R.J. (1993). Crystal structure of the HhaI DNA methyltransferase complexed with S-adenosyl-L-methionine. *Cell 74:* 299-307.

CHOI, Y-C. and CHAE, C-B. (1991). DNA hypomethylation and germ cell-specific expression of testis-specific H2B histone gene. *J. Biol. Chem. 266:* 20504-20511.

COMB, M. and GOODMAN, M. (1990). CpG methylation inhibits proenkephalin gene expression and binding of the transcription factor AP-2. *Nucleic Acids Res. 18:* 3975-3982.

COOPER, D.N. and YOUSSOUFIAN, H. (1988). The CpG dinucleotide and human genetic disease. *Hum. Genet. 78:* 151-155.

CRAIG, J.M. and BICKMORE, W.A. (1993). Chromosome bands - flavours to savour. *BioEssays 15:* 349-354.

CRAIG, J.M. and BICKMORE, W.A. (1994). The distribution of CpG islands in mammalian chromosomes. *Nature Genet. 7:* 376-382.

CROSS, S.H. and BIRD, A.P. (1995). CpG islands and genes. *Curr. Opin. Genet. Dev. 5:* 309-314.

CROSTON, G.E., KERRIGAN, L.A., LIRA, L.M., MARSHAK, D.R. and KADONAGA, J.T. (1991). Sequence-specific antirepression of histone H1-mediated inhibition of basal RNA polymerase II transcription. *Science 251:* 643-649.

DANZIN, C., MARCHAL, P. and CASARA, P. (1990). Irreversible inhibition of rat *S*-adenosylmethionine decarboxylase by 5´-{[(Z)-4-amino-2-butenyl]methylamino}-5´-deoxy-adenosine. *Biochem. Pharmacol. 40:* 1499-1503.

DE BUSTROS, A., NELKIN, B.D., SILVERMAN, A., EHRLICH, G., POIESZ, B. and BAYLIN, S.B. (1988). The short arm of chromosome 11 is a "hot spot" for hypermethylation in human neoplasia. *Proc. Natl. Acad. Sci. USA 85:* 5693-5697.

DEVYS, D., LUTZ, Y., ROUYER, N., BELLOCQ, J-P. and MANDEL, J-L. (1993). The FMR-1 protein is cytoplasmic, most abundant in neurons and appears normal in carriers of a fragile X premutation. *Nature Genet. 4:* 335-340.

EDEN, S. and CEDAR, H. (1994). Role of DNA methylation in the regulation of transcription. *Curr. Opin. Genet. Dev. 4:* 255-259.

EHRLICH, K.C. and EHRLICH, M. (1990). Highly repeated sites in the apolipoprotein(a) gene recognized by methylated DNA-binding protein, a sequence-specific DNA-binding protein. *Mol. Cell. Biol. 10:* 4957-4960.

EICHLER, E.E., RICHARDS, S., GIBBS, R.A. and NELSON, D.L. (1993). Fine structure of the human FMR1 gene. *Hum. Mol. Genet. 2:* 1147-1153.

EL-DEIRY, W.S., NELKIN, B.D., CELANO, P., YEN, R-W.C., FALCO, J.P., HAMILTON, S.R. and BAYLIN, S.B. (1991). High expression of the DNA methyltransferase gene characterizes human neoplastic cells and progression stages of colon cancer. *Proc. Natl. Acad. Sci. USA 88:* 3470-3474.

ERLANSON, D.A., CHEN, L. and VERDINE, G.L. (1993). DNA methylation through a locally unpaired intermediate. *J. Am. Chem. Soc. 115:* 12583-12584.

FENG, Y., ZHANG, F., LOKEY, L.K., CHASTAIN, J.L., LAKKIS, L., EBERHART, D. and WARREN, S.T. (1995). Translational suppression by trinucleotide repeat expansion at *FMR1* . *Science 268:* 731-734.

FERGUSON-SMITH, A.C., SASAKI, H., CATTANACH, B.M. and SURANI, M.A. (1993). Parental-origin-specific epigenetic modification of the mouse *H19* gene. *Nature 362:* 751-755.

FRANK, D., KESHET, I., SHANI, M., LEVINE, A., RAZIN, A. and CEDAR, H. (1991). Demethylation of CpG islands in embryonic cells. *Nature 351:* 239-241.

FRIEDMAN, S. (1985). The irreversible binding of azacytosine-containing DNA fragments to bacterial DNA(cytosine-5)methyltransferases. *J. Biol. Chem. 260:* 5698-5705.

GREGER, V., PASSARGE, E., HÖPPING, W., MESSMER, E. and HORSTHEMKE, B. (1989). Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma. *Hum. Genet. 83:* 155-158.

GROUDINE, M. and CONKIN, K.F. (1985). Chromatin structure and de novo methylation of sperm DNA: implications for activation of the paternal genome. *Science 228:* 1061-1068.

GRUENBAUM, Y., CEDAR, H. and RAZIN, A. (1982). Substrate and sequence specificity of a eukaryotic DNA methylase. *Nature 295:* 620-622.

GRUENBAUM, Y., SZYF, M., CEDAR, H. and RAZIN, A. (1983). Methylation of replicating and post-replicated mouse L-cell DNA. *Proc. Natl. Acad. Sci. USA 80:* 4919-4921.

HARBERS, K., SCHNIEKE, A., STUHLMANN, H., JÄHNER, D. and JAENISCH, R. (1981). DNA methylation and gene expression: endogenous retroviral genome becomes infectious after molecular cloning. *Proc. Natl. Acad. Sci. USA 78:* 7609-7613.

HARLAND, R.M. (1982). Inheritance of DNA methylation in microinjected eggs of *Xenopus laevis*. *Proc. Natl. Acad. Sci. USA 79:* 2323-2327.

HARRINGTON, M.A., JONES, P.A., IMAGAWA, M. and KARIN, M. (1988). Cytosine methylation does not affect binding of transcription factor Sp1. *Proc. Natl. Acad. Sci. USA 85:* 2066-2070.

HEBY, O., PERSSON, L. and SMITH, S.S. (1988). Polyamines, DNA methylation and cell differentiation. In *Progress in Polyamine Research. Novel Biochemical, Pharmacological, and Clinical Aspects* (Eds. V. Zappia and A.E. Pegg). Plenum Press, New York, pp. 291-299.

HERMAN, J.G., LATIF, F., WENG, Y., LERMAN, M.I., ZBAR, B., LIU, S., SAMID, D., DUAN, D-S.R., GNARRA, J.R., LINEHAN, W.M. and BAYLIN, S.B. (1994). Silencing of the *VHL* tumor-suppressor gene by DNA methylation in renal carcinoma. *Proc. Natl. Acad. Sci. USA 91:* 9700-9704.

HINDS, H.L., ASHLEY, C.T., SUTCLIFFE, J.S., NELSON, D.L., WARREN, S.T., HOUSMAN, D.E. and SCHALLING, M. (1993). Tissue specific expression of FMR-1 provides evidence for a functional role in fragile X syndrome. *Nature Genet. 3:* 36-43.

HITT, M.M., WU, T-L., COHEN, G. and LINN, S. (1988). *De novo* and maintenance DNA methylation by a mouse plasmacytoma cell DNA methyltransferase. *J. Biol. Chem. 263:* 4392-4399.

HODGES-GARCIA, Y. and HAGERMAN, P.J. (1992). Cytosine methylation can induce local distortions in the structure of duplex DNA. *Biochemistry 31:* 7595-7599.

HÖLLER, M., WESTIN, G., JIRICNY, J. and SCHAFFNER, W. (1988). Sp1 transcription factor binds DNA and activates transcription even when the binding site is CpG methylated. *Genes Dev. 2:* 1127-1135.

HORNSTRA, I.K., NELSON, D.L., WARREN, S.T. and YANG, T.P. (1993). High resolution methylation analysis of the FMR1 gene trinucleotide repeat region in fragile X syndrome. *Hum. Mol. Genet. 2:* 1659-1665.

IGUCHI-ARIGA, S.M.M. and SCHAFFNER, W. (1989). CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. *Genes Dev. 3:* 612-619.

ISSA, J-P.J., OTTAVIANO, Y.L., CELANO, P., HAMILTON, S.R., DAVIDSON, N.E. and BAYLIN, S.B. (1994). Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon. *Nature Genet. 7:* 536-540.

JÄHNER, D., STUHLMANN, H., STEWART, C.L., HARBERS, K., LÖHLER, J., SIMON, I. and JAENISCH, R. (1982). *De novo* methylation and expression of retroviral genomes during mouse embryogenesis. *Nature 298:* 623-628.

JONES, P.A. (1984). Gene activation by 5-azacytidine. In *DNA methylation. Biochemistry and Biological Significance* (Eds. A. Razin, H. Cedar and A.D. Riggs). Springer Verlag, New York, pp. 165-187.

JONES, P.A., RIDEOUT III, W.R., SHEN, J-C., SPRUCK, C.H. and TSAI, Y.C. (1992). Methylation, mutation and cancer. *BioEssays 14:* 33-36.

JOSHUA-TOR, L., FROLOW, F., APPELLA, E., HOPE, H., RABINOVICH, D. and SUSSMAN, J.L. (1992). Three-dimensional structures of bulge-containing DNA fragments. *J. Mol. Biol. 225:* 397-431.

JOST, J-P. (1993). Nuclear extracts of chicken embryos promote an active demethylation of DNA by excision repair of 5-methyldeoxycytidine. *Proc. Natl. Acad. Sci. USA 90:* 4684-4688.

JOST, J-P., SIEGMANN, M., SUN, L. and LEUNG, R. (1995). Mechanisms of DNA demethylation in chicken embryos. Purification and properties of a 5-methylcytosine-DNA glycosylase. *J. Biol. Chem. 270:* 9734-9739.

JÜTTERMANN, R., HOSOKAWA, K., KOCHANEK, S. and DOERFLER, W. (1991). Adenovirus type 2 VAI RNA transcription by polymerase III is blocked by sequence-specific methylation. *J. Virol. 65:* 1735-1742.

KAFRI, T., ARIEL, M., BRANDEIS, M., SHEMER, R., URVEN, L., McCARREY, J., CEDAR, H. and RAZIN, A. (1992). Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. *Genes Dev. 6:* 705-714.

KAY, G.F., BARTON, S.C., SURANI, M.A. and RASTAN, S. (1994). Imprinting and X chromosome counting mechanisms determine *Xist* expression in early mouse development. *Cell 77:* 639-650.

KESHET, I., LIEMAN-HURWITZ, J. and CEDAR, H. (1986). DNA methylation affects the formation of active chromatin. *Cell 44:* 535-543.

KLIMASAUSKAS, S., KUMAR, S., ROBERT, R.J. and CHENG, X. (1994). Hhal methyltransferase flips its target base out of the DNA helix. *Cell 76:* 357-369.

KNIGHT, S.J.L., FLANNERY, A.V., HIRST, M.C., CAMPBELL, L., CHRISTODOULOU, Z., PHELPS, S.R., POINTON, J., MIDDLETON-PRICE, H.R., BARNICOAT, A., PEMBREY, M.E., HOLLAND, J., OOSTRA, B.A., BOBROW, M. and DAVIES, K.E. (1993). Trinucleotide repeat amplification and hypermethylation of a CpG island in *FRAXE* mental retardation. *Cell 74:* 127-134.

KOBAYASHI, H., NGERNPRASIRTSIRI, J. and AKAZAWA, T. (1990). Transcriptional regulation and DNA methylation in plastids during transitional conversion of chloroplasts to chromoplasts. *EMBO J. 9:* 307-313.

KOVESDI, I., REICHEL, R. and NEVINS, J.R. (1987). Role of an adenovirus *E2* promoter binding factor in E1A-mediated coordinate gene control. *Proc. Natl. Acad. Sci. USA 84:* 2180-2184.

KUMAR, S., CHENG, X., KLIMASAUSKAS, S., MI, S., POSFAI, J., ROBERTS, R.J. and WILSON, G.G. (1994). The DNA (cytosine-5) methyltransferases. *Nucleic Acids Res. 22:* 1-10.

LAIRD, P.W. and JAENISCH, R. (1994). DNA methylation and cancer. *Hum. Mol. Genet. 3:* 1487-1495.

LARSEN, F., GUNDERSEN, G., LOPEZ, R. and PRYDZ, H. (1992). CpG islands as gene markers in the human genome. *Genomics 13:* 1095-1107.

LAUSTER, R., TRAUTNER, T.A. and NOYER-WEIDNER, M. (1989). Cytosine-specific type II DNA methyltransferases. A conserved enzyme core with variable target-recognizing domains. *J. Mol. Biol. 206:* 305-312.

LEHMAN, J.M., KLEIN, I.B. and HACKENBERG, R.M. (1975). The response of murine teratocarcinoma cells to infection with DNA and RNA viruses. In *Teratomas and Differentiation* (Eds. M.I. Sherman and D. Solter). Academic Press, New York, pp. 289-301.

LEONHARDT, H., PAGE, A.W., WEIER, H-U. and BESTOR, T.H. (1992). A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell 71:* 865-873.

LEROY, J.L., KOCHOYAN, M., HUYNH-DINH, T. and GUÉRON, M. (1988). Characterization of base-pair opening in deoxynucleotide duplexes using catalyzed exchange of the imino proton. *J. Mol. Biol. 200:* 223-238.

LEVINE, A., YEIVIN, A., BEN-ASHER, E., ALONI, Y. and RAZIN, A. (1993). Histone H1-mediated inhibition of transcription initiation of methylated templates *in vitro*. *J. Biol. Chem. 268:* 21754-21759.

LEWIS, J.D., MEEHAN, R.R., HENZEL, W.J., MAURER-FOGY, I., JEPPESEN, P., KLEIN, F. and BIRD, A. (1992). Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell 69:* 905-914.

LI, E., BEARD, C. and JAENISCH, R. (1993). Role for DNA methylation in genomic imprinting. *Nature 366:* 362-365.

LI, E., BESTOR, T.H. and JAENISCH, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell 69:* 915-926.

LICHTENSTEIN, M., KEINI, G., CEDAR, H. and BERGMAN, Y. (1994). B cell-specific demethylation: a novel role for the intronic κ chain enhancer sequence. *Cell 76:* 913-923.

LINNENBACH, A., HUEBNER, K. and CROCE, C.M. (1980). DNA-transformed murine teratocarcinoma cells: regulation of expression of simian virus 40 tumor antigen in stem versus differentiated cells. *Proc. Natl. Acad. Sci. USA 77:* 4875-4879.

MA, Q., ALDER, H., NELSON, K.K., CHATTERJEE, D., GU, Y., NAKAMURA, T., CANAANI, E., CROCE, C.M., SIRACUSA, L.D. and BUCHBERG, A.M. (1993). Analysis of the murine *All-1* gene reveals conserved domains with human *ALL-1* and identifies a motif shared with DNA methyltransferases. *Proc. Natl. Acad. Sci. USA 90:* 6350-6354.

MACLEOD, A.R. and SZYF, M. (1995). Expression of antisense to DNA methyltransferase mRNA induces DNA demethylation and inhibits tumorigenesis. *J. Biol. Chem. 270:* 8037-8043.

MACLEOD, D., CHARLTON, J., MULLINS, J. and BIRD, A.P. (1994). Sp1 sites in the mouse *aprt* gene promoter are required to prevent methylation of the CpG island. *Genes Dev. 8:* 2282-2292.

MEEHAN, R.R., LEWIS, J.D. and BIRD, A.P. (1992). Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. *Nucleic Acids Res. 20:* 5085-5092.

MEEHAN, R.R., LEWIS, J.D., McKAY, S., KLEINER, E.L. and BIRD, A.P. (1989). Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell 58:* 499-507.

MILLS, A.D., BLOW, J.J., WHITE, J.G., AMOS, W.B., WILCOCK, D. and LASKEY, R.A. (1989). Replication occurs at discrete foci spaced throughout nuclei replicating *in vitro*. *J. Cell Sci. 94:* 471-477.

MINAROVITS, J., MINAROVITS-KORMUTA, S., EHLIN-HENRIKSSON, B., FALK, K., KLEIN, G. and ERNBERG, I. (1991). Host cell phenotype-dependent methylation patterns of Epstein-Barr virus DNA. *J. Gen. Virol. 72:* 1591-1599.

MONK, M. and HARPER, M.I. (1979). Sequential X chromosome inactivation coupled with cellular differentiation in early mouse embryos. *Nature 281:* 311-313.

MONK, M. and McLAREN, A. (1981). X-chromosome activity in foetal germ cells of the mouse. *J. Embryol. Exp. Morphol. 63:* 75-84.

MONK, M., ADAMS, R.L.P. and RINALDI, A. (1991). Decrease in DNA methylase activity during preimplantation development in the mouse. *Development 112:* 189-192.

MONK, M., BOUBELIK, M. and LEHNERT, S. (1987). Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development 99:* 371-382.

NAKAMURA, H., MORITA, T. and SATO, C. (1986). Structural organizations of replicon domains during DNA synthetic phase in the mammalian nucleus. *Exp. Cell Res. 165:* 291-297.

NAKAYASU, H. and BEREZNEY, R. (1989). Mapping replicational sites in the eucaryotic cell nucleus. *J. Cell Biol. 108:* 1-11.

NORRIS, D.P., PATEL, D., KAY, G.F., PENNY, G.D., BROCKDORFF, N., SHEARDOWN, S.A. and RASTAN, S. (1994). Evidence that random and imprinted *Xist* expression is controlled by preemptive methylation. *Cell 77:* 41-51.

OREDSSON, S.M., BILLGREN, M. and HEBY, O. (1985). Induction of F9 embryonal carcinoma cell differentiation by inhibition of polyamine synthesis. *Eur. J. Cell Biol. 38:* 335-343.

OREDSSON, S.M., KANJE, M., MAMONT, P.S., WAGNER, J. and HEBY, O. (1986). Polyamine depletion increases cellular ribonucleotide levels. *Mol. Cell. Biochem. 70:* 89-96.

ORLANDO, V. and PARO, R. (1995). Chromatin multiprotein complexes involved in the maintenance of transcription patterns. *Curr. Opin. Genet. Dev. 5:* 174-179.

PALMITER, R.D., CHEN, H.Y. and BRINSTER, R.L. (1982). Differential regulation of metallothionein-thymidine kinase fusion genes in transgenic mice and their offspring. *Cell 29:* 701-710.

PELLICER, A., WAGNER, E.F., EL KAREH, A., DEWEY, M.J., REUSER, A.J., SILVERSTEIN, S., AXEL, R. and MINTZ, B. (1980). Introduction of a viral thymidine kinase gene and the human β-globin gene into developmentally multipotential mouse teratocarcinoma cells. *Proc. Natl. Acad. Sci. USA 77:* 2098-2102.

PFEIFER, G.P., STEIGERWALD, S.D., HANSEN, R.S., GARTLER, S.M. and RIGGS, A.D. (1990a). Polymerase chain reaction-aided genomic sequencing of an X chromosome-linked CpG island: Methylation patterns suggest clonal inheritance, CpG site autonomy, and an explanation of activity state stability. *Proc. Natl. Acad. Sci. USA 87:* 8252-8256.

PFEIFER, G.P., TANGUAY, R.L., STEIGERWALD, S.D. and RIGGS, A.D. (1990b). *In vivo* footprint and methylation analysis by PCR-aided genomic sequencing: comparison of active and inactive X chromosomal DNA at the CpG island and promoter of human PGK-1. *Genes Dev. 4:* 1277-1287.

PRENDERGAST, G.C., LAWE, D. and ZIFF, E.B. (1991). Association of myn, the murine homolog of max, with c-Myc stimulates methylation-sensitive DNA binding and Ras cotransformation. *Cell 65:* 395-407.

RAZIN, A. and CEDAR, H. (1991). DNA methylation and gene expression. *Microbiol. Rev. 55:* 451-458.

RAZIN, A. and CEDAR, H. (1994). DNA methylation and genomic imprinting. *Cell 77:* 473-476.

RAZIN, A. and RIGGS, A.D. (1980). DNA methylation and gene function. *Science 210:* 604-610.

REINISCH, K.M., CHEN, L., VERDINE, G.L. and LIPSCOMB, W.N. (1995). The crystal structure of HaeIII methyltransferase covalently complexed to DNA: an extrahelical cytosine and rearranged base pairing. *Cell 82:* 143-153.

REISS, A.L., ABRAMS, M.T., GREENLAW, R., FREUND, L. and DENCKLA, M.B. (1995). Neurodevelopmental effects of the FMR-1 full mutation in humans. *Nature Med. 1:* 159-167.

REYNIERS, E., VITS, L., DE BOULLE, K., VAN ROY, B., VAN VELZEN, D., DE GRAAFF, E., VERKERK, A.J.M.H., JORENS, H.Z.J., DARBY, J.K., OOSTRA, B. and WILLEMS, P.J. (1993). The full mutation in the FMR-1 gene of male fragile X patients is absent in their sperm. *Nature Genet. 4:* 143-146.

RICH, A., NORDHEIM, A. and WANG, A.H-J. (1984). The chemistry and biology of left-handed Z-DNA. *Annu. Rev. Biochem. 53:* 791-846.

RICHARDS, R.I., HOLMAN, K., YU, S. and SUTHERLAND, G.R. (1993). Fragile X syndrome unstable element, p(CCG)n, and other simple tandem repeat sequences are binding sites for specific nuclear proteins. *Hum. Mol. Genet. 2:* 1429-1435.

RIGGS, A.D. and PFEIFER, G.P. (1992). X-chromosome inactivation and cell memory. *Trends Genet. 8:* 169-174.

ROBERTS, R.J. (1995). On base flipping. *Cell 82:* 9-12.

SHEN, J-C., RIDEOUT III, W.M. and JONES, P.A. (1992). High frequency mutagenesis by a DNA methyltransferase. *Cell 71:* 1073-1080.

SILVERMAN, A.L., PARK, J-G., HAMILTON, S.R., GAZDAR, A.F., LUK, G.D. and BAYLIN, S.B. (1989). Abnormal methylation of the calcitonin gene in human colonic neoplasms. *Cancer Res. 49:* 3468-3473.

SIMON, J. (1995). Locking in stable states of gene expression: transcriptional control during *Drosophila* development. *Curr. Opin. Cell Biol. 7:* 376-385.

SIMPSON, V.J., JOHNSON, T.E. and HAMMEN, R.F. (1986). *Caenorhabditis elegans* DNA does not contain 5-methylcytosine at any time during development or aging. *Nucleic Acids Res. 14:* 6711-6719.

SINGER, J., ROBERTS-EMS, J., LUTHARDT, F.W. and RIGGS, A.D. (1979). Methylation of DNA in mouse early embryos, teratocarcinoma cells and adult tissues of mouse and rabbit. *Nucleic Acids Res. 7:* 2369-2385.

SINGER-SAM, J., GRANT, M., LEBON, J.M., OKUYAMA, K., CHAPMAN, V., MONK, M. and RIGGS, A.D. (1990). Use of a *Hpa*II-polymerase chain reaction assay to study DNA methylation in the *Pgk*-1 CpG island of mouse embryos at the time of X-chromosome inactivation. *Mol. Cell. Biol. 10:* 4987-4989.

SIOMI, H., CHOI, M., SIOMI, M.C., NUSSBAUM, R.L. and DREYFUSS, G. (1994). Essential role for KH domains in RNA binding: impaired RNA binding by a mutation in the KH domain of FMR1 that causes fragile X syndrome. *Cell 77:* 33-39.

SIOMI, H., SIOMI, M.C., NUSSBAUM, R.L. and DREYFUSS, G. (1993). The protein product of the fragile X gene, *FMR1*, has characteristics of an RNA-binding protein. *Cell 74:* 291-298.

SIOMI, M.C., SIOMI, H., SAUER, W.H., SRINIVASAN, S., NUSSBAUM, R.L. and DREYFUSS, G. (1995). FXR1, an autosomal homolog of the fragile X mental retardation gene. *EMBO J. 14:* 2401-2408.

SMITH, S.S., KAN, J.L.C., BAKER, D.J., KAPLAN, B.E. and DEMBEK, P. (1991). Recognition of unusual DNA structures by human DNA(cytosine-5)methyltransferase. *J. Mol. Biol. 217:* 39-51.

SMITH, S.S., KAPLAN, B.E., SOWERS, L.C. and NEWMAN, E.M. (1992). Mechanism of human methyl-directed DNA methyltransferase and the fidelity of cytosine methylation. *Proc. Natl. Acad. Sci. USA 89:* 4744-4748.

SOLTER, D. (1987). Inertia of the embryonic genome in mammals. *Trends Genet. 3:* 23-27.

STEWART, C.L., STUHLMANN, H., JÄHNER, D. and JAENISCH, R. (1982). *De novo* methylation, expression, and infectivity of retroviral genomes introduced into embryonal carcinoma cells. *Proc. Natl. Acad. Sci. USA 79:* 4098-4102.

STJERNBORG, L., HEBY, O., MAMONT, P. and PERSSON, L. (1993). Polyamine-mediated regulation of *S*-adenosylmethionine decarboxylase expression in mammalian cells – Studies using 5'-{[(Z)-4-amino-2-butenyl]methylamino}-5'-deoxyadenosine, a suicide inhibitor of the enzyme. *Eur. J. Biochem. 214:* 671-676.

STÖGER, R., KUBICKA, P., LIU, C.-G., KAFRI, T., RAZIN, A., CEDAR, H. and BARLOW, D.P. (1993). Maternal-specific methylation of the imprinted mouse *Igf2r* locus identifies the expressed locus as carrying the imprinting signal. *Cell 73:* 61-71.

STUHLMANN, H., JÄHNER, D. and JAENISCH, R. (1981). Infectivity and methylation of retroviral genomes is correlated with expression in the animal. *Cell 26:* 221-232.

SURANI, M.A., REIK, W. and ALLEN, N.D. (1988). Transgenes as molecular probes for genomic imprinting. *Trends Genet. 4:* 59-62.

SUTCLIFFE, J.S., NELSON, D.L., ZHANG, F., PIERETTI, M., CASKEY, C.T., SAXE, D. and WARREN, S.T. (1992). DNA methylation represses *FMR-1* transcription in fragile X syndrome. *Hum. Mol. Genet. 1:* 397-400.

SUTHERLAND, G.R. (1985). The enigma of the fragile X chromosome. *Trends Genet. 1:* 108-112.

SVED, J. and BIRD, A. (1990). The expected equilibrium of the CpG dinucleotide in vertebrate genomes under a mutation model. *Proc. Natl. Acad. Sci. USA 87:* 4692-4696.

SWAIN, J.L., STEWART, T.A. and LEDER, P. (1987). Parental legacy determines methylation and expression of an autosomal transgene: A molecular mechanism for parental imprinting. *Cell 50:* 719-727.

SZNYTER, L.A., SLATKO, B., MORAN, L., O'DONNELL, K.H. and BROOKS, J.E. (1987). Nucleotide sequence of the *Dde*I restriction-modification system and characterization of the methylase protein. *Nucleic Acids Res. 15:* 8249-8266.

TAZI, J. and BIRD, A. (1990). Alternative chromatin structure at CpG islands. *Cell 60:* 909-920.

THOMPSON, J.P., GRANOFF, A. and WILLIS, D.B. (1986). Trans-activation of a methylated adenovirus promoter by a frog virus 3 protein. *Proc. Natl. Acad. Sci. USA 83:* 7688-7692.

TRASLER, J.M., ALCIVAR, A.A., HAKE, L.E., BESTOR, T. and HECHT, N.B. (1992). DNA methyltransferase is developmentally expressed in replicating and non-replicating male germ cells. *Nucleic Acids Res. 20:* 2541-2545.

TRASLER, J.M., HAKE, L.E., JOHNSON, P.A., ALCIVAR, A.A., MILLETTE, C.F. and HECHT, N.B. (1990). DNA methylation and demethylation events during meiotic prophase in the mouse testis. *Mol. Cell. Biol. 10:* 1828-1834.

UEHARA, Y., ONO, T., KURISHITA, A., KOKURYU, H. and OKADA, S. (1989). Age-dependent and tissue-specific changes of DNA methylation within and around the c-*fos* gene in mice. *Oncogene 4:* 1023-1028.

URIELI-SHOVAL, S., GRUENBAUM, Y., SEDAT, J. and RAZIN, A. (1982). The absence of detectable methylated bases in *Drosophila melanogaster* DNA. *FEBS Lett. 146:* 148-152.

VERDINE, G.L. (1994). The flip side of DNA methylation. *Cell 76:* 197-200.

VERHEIJ, C., BAKKER, C.E., DE GRAAFF, E., KEULEMANS, J., WILLEMSEN, R., VERKERK, A.J.M.H., GALJAARD, H., REUSER, A.J.J., HOOGEVEEN, A.T. and OOSTRA, B.A. (1993). Characterization and localization of the *FMR-1* gene product associated with fragile X syndrome. *Nature 363:* 722-724.

VERKERK, A.J.M.H., DE GRAAFF, E., DE BOULLE, K., EICHLER, E.E., KONECKI, D.S., REYNIERS, E., MANCA, A., POUSTKA, A., WILLEMS, P.J., NEL-

SON, D.L. and OOSTRA, B.A. (1993). Alternative splicing in the fragile X gene *FMR1. Hum. Mol. Genet. 2:* 399-404.

VERKERK, A.J.M.H., PIERETTI, M., SUTCLIFFE, J.S., FU, Y-H., KUHL, D.P.A., PIZZUTI, A., REINER, O., RICHARDS, S., VICTORIA, M.F., ZHANG, F., EUSSEN, B.E., VAN OMMEN, G-J.B., BLONDEN, L.A.J., RIGGINS, G.J., CHASTAIN, J.L., KUNST, C.B., GALJAARD, H., CASKEY, C.T., NELSON, D.L., OOSTRA, B.A. and WARREN, S.T. (1991). Identification of a gene (*FMR-1*) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell 65:* 905-914.

WAINFAN, E. and POIRIER, L.A. (1992). Methyl groups in carcinogenesis: effects on DNA methylation and gene expression. *Cancer Res. 52:* 2071s-2077s.

WEISSHAAR, B., LANGNER, K.-D., JÜTTERMANN, R., MÜLLER, U., ZOCK, C., KLIMKAIT, T. and DOERFLER, W. (1988). Reactivation of the methylation-inactivated late E2A promoter of adenovirus type 2 by E1A (13S) functions. *J. Mol. Biol. 202:* 255-270.

WIEBAUER, K. and JIRICNY, J. (1989). *In vitro* correction of G·T mispairs to G·C pairs in nuclear extracts from human cells. *Nature 339:* 234-236.

WU, J.C. and SANTI, D.V. (1987). Kinetic and catalytic mechanism of *Hha*I methyltransferase. *J. Biol. Chem. 262:* 4778-4786.

WU, J., ISSA, J-P., HERMAN, J., BASSETT, D.E.,JR., NELKIN, B.D. and BAYLIN, S.B. (1993). Expression of an exogenous eukaryotic DNA methyltransferase gene induces transformation of NIH 3T3 cells. *Proc. Natl. Acad. Sci. USA 90:* 8891-8895.

YEN, R-W.C., VERTINO, P.M., NELKIN, B.D., YU, J.J., EL-DEIRY, W., CUMARASWAMY, A., LENNON, G.G., TRASK, B.J., CELANO, P. and BAYLIN, S.B. (1992). Isolation and characterization of the cDNA encoding human DNA methyltransferase. *Nucleic Acids Res. 20:* 2287-2291.

ZACHARIAS, W. (1993). Methylation of cytosine influences the DNA structure. In *DNA Methylation: Molecular Biology and Biological Significance* (Eds. J.P. Jost and H.P. Saluz). Birkhäuser Verlag, Basel, pp. 27-38.

ZION, M., BEN-YEHUDA, D., AVRAHAM, A., COHEN, O., WETZLER, M., MELLOUL, D. and BEN-NERIAH, Y. (1994). Progressive *de novo* DNA methylation at the *bcr-abl* locus in the course of chronic myelogenous leukemia. *Proc. Natl. Acad. Sci. USA 91:* 10722-10726.