Annu. Rev. Biochem. 2005. 74:481–514
doi: 10.1146/annurev.biochem.74.010904.153721

First published online as a Review in Advance on March 11, 2005

EUKARYOTIC CYTOSINE METHYLTRANSFERASES

Mary Grace Goll and Timothy H. Bestor
Department of Genetics and Development, College of Physicians and Surgeons of Columbia University, New York, New York 10032; email: MGG22@columbia.edu, THB12@columbia.edu

Key Words 5-methylcytosine, DNA methyltransferase, transposons, genomic imprinting

■ **Abstract** Large-genome eukaryotes use heritable cytosine methylation to silence promoters, especially those associated with transposons and imprinted genes. Cytosine methylation does not reinforce or replace ancestral gene regulation pathways but instead endows methylated genomes with the ability to repress specific promoters in a manner that is buffered against changes in the internal and external environment. Recent studies have shown that the targeting of de novo methylation depends on multiple inputs; these include the interaction of repeated sequences, local states of histone lysine methylation, small RNAs and components of the RNAi pathway, and divergent and catalytically inert cytosine methyltransferase homologues that have acquired regulatory roles. There are multiple families of DNA (cytosine-5) methyltransferases in eukaryotes, and each family appears to be controlled by different regulatory inputs. Sequence-specific DNA-binding proteins, which regulate most aspects of gene expression, do not appear to be involved in the establishment or maintenance of genomic methylation patterns.

CONTENTS

INTRODUCTION

The modified base 5-methylcytosine (m^5C) is present in the DNA of all vertebrates and flowering plants; some fungal, invertebrate, and protist taxa; and many bacterial species. Cytosine methylation is common to all large-genome eukaryotes but is present in only some small-genome eukaryotes. Cytosine methylation is mediated by a conserved group of proteins called DNA (cytosine-5) methyltransferases. The biological roles of cytosine methylation have long been controversial, but sequence-based studies and forward and reverse genetic approaches have greatly improved our understanding of the form and function of genomic methylation patterns. This review addresses the relationship of the cytosine methyltransferases of eukaryotes and the diverse inputs that control them.

The biological functions of cytosine methylation are fundamentally different in prokaryotes and eukaryotes. DNA methylation has a central role in host restriction of phage DNA in bacteria, but there is no compelling evidence that eukaryotes restrict viruses in this way. Nor is there any evidence that cytosine methylation has a role in strand discrimination during mismatch repair in eukaryotes, as does adenine methylation in some bacterial taxa. It has been almost 30 years since Riggs (1) and Holliday & Pugh (2) predicted the existence of two general classes of DNA methyltransferases: de novo enzymes that would establish methylation patterns at specific sequences early in development and maintenance enzymes that would preserve methylation patterns during cell division by specific methylation of hemimethylated CpG dinucleotides produced by semiconservative DNA replication. Other factors were predicted to regulate demethylation or de novo methylation of regulatory sequences to activate and repress genes during cellular differentiation. However, the discovery that differentiation is regulated by mechanisms that are conserved between organisms with methylated genomes (vertebrates, flowering plants, and some fungi) and those whose genomes have little or no cytosine methylation (*Caenorhabditis elegans*, *Drosophila melanogaster*) made the regulatory hypothesis much less tenable; the evidence against this hypothesis is summarized elsewhere (3). However, the major innovation of the hypotheses of Riggs and of Holliday & Pugh (the somatic inheritance of genomic methylation patterns in mammals) has been abundantly confirmed.

Wigler et al. (4) found that arbitrary patterns of cytosine methylation imposed on plasmid substrates were stably maintained for many cell cycles when the plasmids were integrated into the genomes of transfected cells. More recently, it was shown that methylation patterns were maintained essentially unchanged for 80 cell divisions in a system that controlled for the effects of copy number and integration site (5). The findings from studies of cultured cells are consistent with findings from the study of human pathological conditions arising from the early gain or loss

of methylation at the promoters of certain imprinted human genes, wherein the early methylation defect is propagated through the course of development and into adult genomes. To date, cytosine methylation represents the only chromatin modification for which a means of stable propagation through cell division has been identified. Although today it is believed that patterns of histone modifications and aspects of chromatin structure other than DNA sequence and methylation patterns are subject to replication during S phase, there is little evidence that this is so, and no plausible mechanism exists by which higher-order chromatin structures could be subject to passive clonal inheritance in the absence of an ongoing signal. The histone modifications characteristic of active genes may be a consequence of active transcription rather than a cause; it is not at all clear that histone modifications convey regulatory information, and they instead may be slaved to the activity of sequence-specific DNA-binding proteins.

Cytosine Methylation in Host Defense and Genome Stability

Until very recently, little was known of the large-scale organization of genomic methylation patterns in eukaryotes, and methylation patterns were largely overlooked in the major genome sequencing projects. It is now known that the majority of cytosine methylation in plants and mammals and almost all cytosine methylation in the ascomycete fungus *Neurospora crassa* resides in repetitive elements. Much of this methylation is in transposons, which are interspersed repeated sequences that constitute more than 45% of the human genome (6). Transposons threaten the stability of the genome through insertional mutagenesis, chimeric transcripts arising from transposon promoters, antisense transcripts that activate RNAi pathways, and dysregulated gene expression mediated by the activation of transposon promoters within and around genes (7–9). Transposons can only proliferate in the genomes of sexual populations, where the fitness of the transposon is greater than that of the host. Sexual populations are therefore under selective pressures to develop systems that oppose transposon action (10), and it is now clear that methylation represents the primary mechanism of transposon suppression in host genomes. Most genomic m^5C resides in transposons, and transposons are reanimated in the demethylated genomes of the mouse (3, 11, 12). Genome demethylation in plants (which can survive larger reductions in genomic m^5C than can mammals) also causes greatly increased rates of transposon insertion (12–15). Transposons are methylated in the genomes of mammalian germ cells, and over time cytosine methylation in transposable elements leads to their irreversible inactivation through accumulation of $C \rightarrow T$ transition mutations arising by deamination of m^5C to thymine (16). An accelerated version of this mutagenesis is mediated by a methyltransferase homologue in *N. crassa*, where repeated sequences trigger not only cytosine methylation but also an active mutation of cytosine to thymine (17).

Cytosine Methylation and Gene Regulation

CpG islands, which are associated with the promoter regions of 76% of human genes (18, 19), are 0.4–3 kb in length, are relatively rich in G + C (>55%), and

are enriched in the CpG dinucleotide relative to the remainder of the genome (the observed/expected ratio is >0.6). Methylated CpG islands are strongly and heritably repressed. Islands are associated with both tissue-specific and housekeeping genes and are unmethylated, or lightly and variably methylated, in all tissues. Monoallelic methylation is seen at some CpG islands associated with imprinted mammalian genes and genes on the inactive X chromosome, and CpG islands associated with tissue-specific genes in established lines of cultured cells tend to be methylated at positions not methylated in any tissue (3). The fact that islands are unmethylated in both expressing and nonexpressing tissues indicates that their activity is not controlled by methylation. It is nonetheless widely believed that cytosine methylation regulates development (the only function attributed to cytosine methylation in Reference 20 concerned developmental gene control). Most of the expression-methylation studies involve non-CpG island genes, which tend to have light and variable cytosine methylation that may be less in cells that express the gene. The demethylation reported in these studies may not be regulatory but rather a consequence of gene activation; the binding of certain transcription factors or even the *Escherichia coli* lac repressor in cells transfected with a methylated *lac* operator can cause the loss of methylation from sequences in the vicinity of the protein-binding site in dividing cells (21, 22). A convincing example of demethylation of a heavily methylated CpG island during development has not appeared, and in no case has a specific methylation pattern at a tissue-specific gene in a nonexpressing tissue been shown to repress transcription of that gene in a cell type normally capable of expressing the gene.

A direct assessment of the role of cytosine methylation in tissue specific-gene regulation came from the analysis of mouse embryos that had severely reduced levels of cytosine methylation. Such mouse embryos developed normally until 8.5 days post coitus (dpc), far later than would be expected if massive gene dysregulation had occurred (23). Genes reported to be regulated in a methylation-dependent manner in cell culture systems were unaffected by the reduction of cytosine methylation in mouse embryos (3), and microarray analysis of a fibroblast cell line lacking significant levels of cytosine methylation identified only five tissue-specific genes with greater than fivefold increases in gene expression (24). Although a specialized role cannot be excluded, a general role for cytosine methylation in regulation of tissue or developmentally specific genes is increasingly implausible. This conclusion is supported by genetic data from plants, which have a DNA-methylating system that is closely related to that of mammals. *Arabidopsis thaliana* strains that lack one or more DNA methyltransferases are viable, with marked phenotypes appearing only after several generations in the demethylated state. Those phenotypes have been attributed primarily to reactivated transposons (25), with some phenotypes arising from disruption of imprinted gene expression in endosperm.

Cytosine methylation is required in both plants and mammals for the monoallelic expression of imprinted genes, which are normally expressed from only one of two identical alleles according to the sex of the parent that contributed the allele.

Removal of cytosine methylation in somatic cells or a failure to establish methylation patterns at imprinted loci in germ cells causes biallelic expression of imprinted genes in somatic tissues. Defects in imprinting at specific loci are responsible for a number of human disorders [reviewed in (26)]. Imprinting may occur in either the male or female germ line, although most imprinted genes undergo de novo methylation in oogenesis (27). X-chromosome inactivation, in which transcription of genes on one of the two X chromosomes in females is silenced as a means of dosage compensation, also depends on cytosine methylation (28).

Flowering plants display imprinted gene expression in triploid extraembryonic tissues of endosperm and not in the embryo proper (29). Imprinting in *A. thaliana* requires cytosine methyltransferases and the helicase homologue decrease in DNA methylation 1 (DDM1), and it also requires the activity of the DNA glycosylase DEMETER (30, 31), which has been proposed to demethylate the active allele by excision of 5-methylcytosine followed by replacement with cytosine. A second DNA glycosylase, ROS1, may also play a role in removing methylation marks in *A. thaliana* (32). It appears that in plants methylation of imprinted genes on both alleles may be the default state, whereas in mammals, the opposite is true. The stability of methylation patterns is likely to be required for heritable silencing phenomena such as genomic imprinting. The extant data indicate that only methylated genomes can maintain long-term restriction of gene expression in the absence of sequence rearrangements or an ongoing stimulus.

Conservation Between Bacterial and Eukaryotic Cytosine Methyltransferases

Methylation of the vinyl carbon at the 5 position of cytosine residues in neutral aqueous solution has been termed a chemically improbable reaction (33). Cytosine methyltransferases overcome the low reactivity of the cytosine C5 by means of a covalent catalysis mechanism that is similar to that of thymidylate synthetase (34) (Figure 1*a*). The cysteine thiolate of a conserved prolylcysteinyl (PC) dipeptide in motif IV forms a covalent bond with the C6 of cytosine, as proposed by Santi and colleagues and modified by Verdine and colleagues (35). This cysteine is invariant in eukaryotic cytosine methyltransferases, and its substitution has been shown to result in a loss of activity by bacterial restriction methyltransferases and by the eukaryotic cytosine methyltransferases Dnmt3A and Dnmt3B (36). Approach trajectories of catalytic amino acid side chains to the 5 and 6 positions of pyrimidines in duplex DNA are occluded by neighboring bases. This steric embarrassment is resolved by eversion of the target cytosine from the DNA helix and insertion into the active site pocket of the large domain (Figure 1*b*) (37). Base eversion was first observed in the bacterial restriction methyltransferase M.HhaI and is now known to be common among reactions that involve modification or removal of bases from duplex DNA [reviewed in (38)].

Eukaryotic cytosine methyltransferases share the 10 sequence motifs that are conserved within the bacterial (cytosine-5) methyltransferases (42–44),

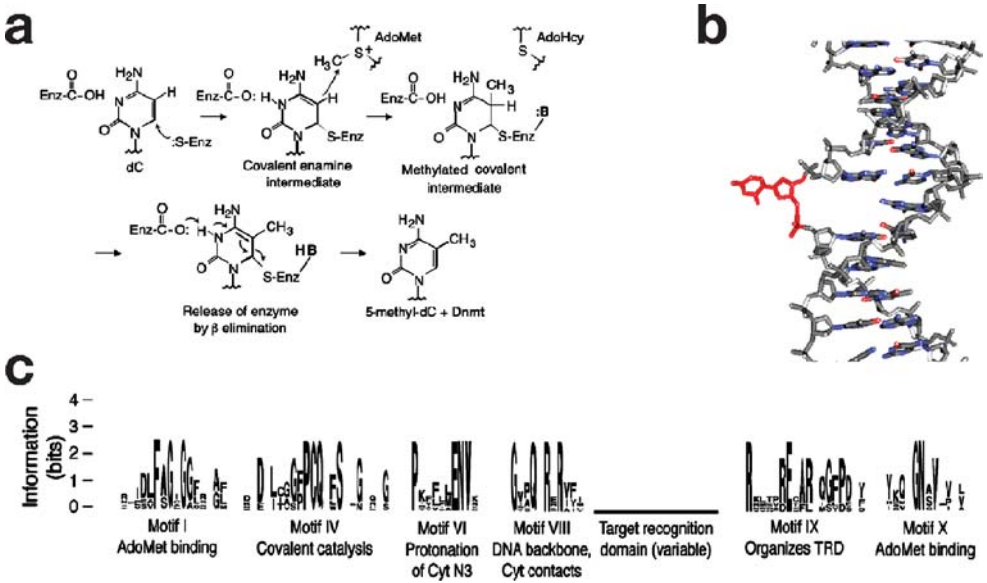


Figure 1 Catalytic mechanism and conserved motifs in DNA (cytosine-5) methyltransferases. (*a*) Catalytic mechanism as initially proposed by Santi et al. and modified by Verdine and colleagues. Covalent addition of an enzyme nucleophile (the cysteine thiolate of the conserved PC motif) to the cytosine-6 position and protonation of the N3 position produces the 4,5 enamine intermediate that attacks the methyl group of S-adenosyl L-methionine (AdoMet). The demethylated cofactor is S-adenosyl L-homocysteine (AdoHcy). Following methyltransfer, an unidentified enzyme base or water molecule (35, 39) abstracts a proton from the cytosine-5 position, which allows release of free enzyme by beta elimination. (*b*) Eversion of the target cytosine from the DNA during catalysis. The DNA shown is from the M.HhaI-DNA-AdoHcy cocrystal structure. (*c*) Conserved motifs in DNA (cytosine-5) methyltransferases. The six most highly conserved motifs are shown in LOGOS (40); functions of each motif are given at bottom. The input to LOGOS was a ClustalW (41) alignment of the C-terminal catalytic domain of mouse Dnmt1 and the bacterial restriction methyltransferases M.AluI, M.DdeI, M.HhaI, M.HaeIII, M.HpaII, and M.SssI. The pattern of conserved residues shown here has proven, in most cases, a reliable predictor of cytosine methyltransferases.

(Figure 1*c*). The conserved motifs have high predictive value in the identification of new DNA cytosine methyltransferases, and almost all of the eukaryotic cytosine methyltransferase homologues were initially identified by the content of these motifs (34, 45–47). The functions of all 10 motifs are known from crystallographic studies of transition-state intermediates and from mutagenesis studies. A region between motifs VIII and IX makes sequence-specific contacts with base edges in the major groove and confers sequence specificity to bacterial cytosine methyltransferases (48, 49). This region has been termed the target recognition domain. Crystal structures of the bacterial cytosine methyltransferases M.HhaI (50), and M.HaeIII (51), as well as the enigmatic cytosine methyltransferase homologue DNMT2 (52), revealed strong sequence and structural conservation among cytosine methyltransferases. The general architecture consists of a strongly conserved large domain, which includes the binding site for the cofactor AdoMet and the active site motifs, and a small domain, which is poorly conserved and is largely represented by the target recognition domain. Bacterial DNA (cytosine-5) methyltransferases have defined recognition sequences of 2 to 8 nucleotides, and all cognate sequences in the host genome are normally methylated. As discussed below, target selection by eukaryotic cytosine methyltransferases is not a function of innate sequence specificity.

THE CYTOSINE METHYLTRANSFERASE FAMILIES OF EUKARYOTES

A large number of eukaryotic cytosine methyltransferase homologues have been identified by sequence similarity since the first was reported in 1988 (53). Few have been shown to methylate DNA in vitro, although many have shown evidence of involvement in cytosine methylation by genetic tests. Most cytosine methyltransferases can be grouped into four distinct families based on sequence homology within their C-terminal catalytic domains, although the fungal enzymes show greater divergence (Figure 2). All organisms that possess proteins from the DNA methyltransferase-1 (Dnmt1) family appear to also have at least one Dnmt3 homologue (Figure 3). Dnmt2 homologues are present in all organisms known to contain Dnmt1 and Dnmt3 homologues as well as in a number of additional organisms in which Dnmt2 is the only cytosine methyltransferase homologue. The chromomethylase family is unique to the plant kingdom. Some eukaryotes (notably *C. elegans* and *Saccharomyces cerevisiae*) lack detectable cytosine methylation in their genomes and have no sign of any cytosine methyltransferase coding sequence. Other organisms, such as *A. thaliana*, have 10 or more cytosine methyltransferase homologues. The major families of eukaryotic cytosine methyltransferases are diagrammed in Figure 2*a*, and their phylogenetic distributions are shown in Figure 3. The methyltransferase families are discussed below.

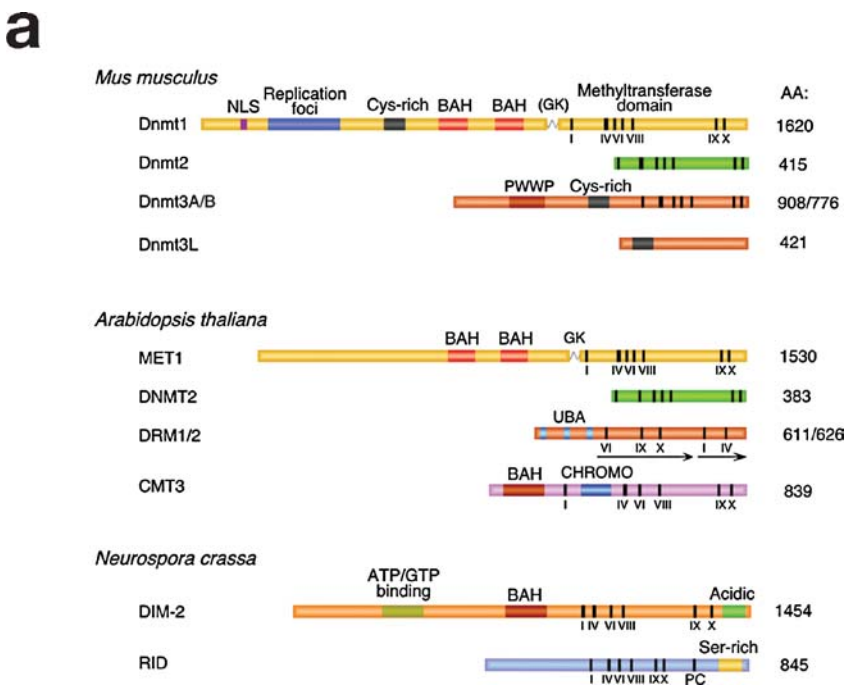
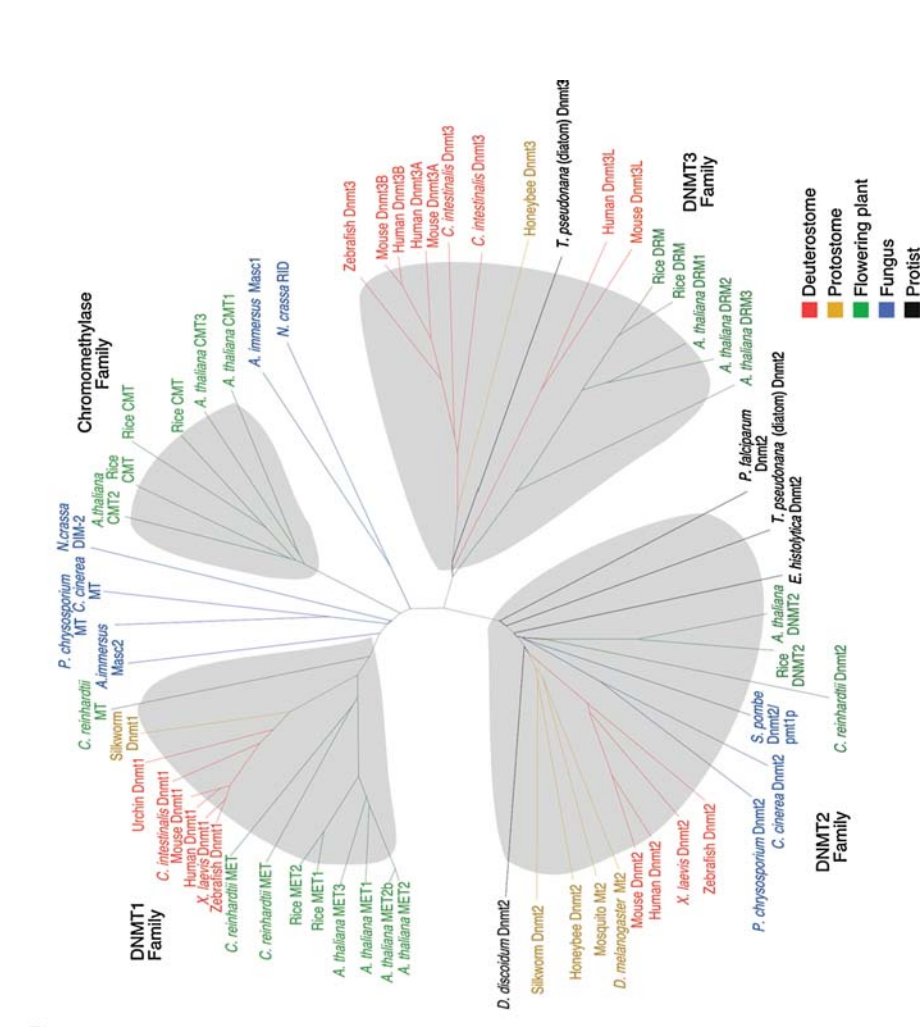


Figure 2 Organization and sequence relationships of DNA cytosine methyltransferases. (*a*) Classes of cytosine methyltransferases and cytosine methyltransferase homologues in mouse, *A. thaliana*, and the ascomycete fungus *N. crassa*. The positions of sequence motifs are indicated. (*b*) Distribution of cytosine methyltransferases and cytosine methyltransferase homologues in eukaryotes. Sequence alignments and dendrograms were prepared in ClustalW and included only C-terminal catalytic domains of homologs for which the full C-terminal sequence was available. Major taxa are indicated by the color scheme shown at bottom. Sequence divergence between the Dnmt3 group and the domains rearranged methyltransferase (DRM) group is less than indicated because of an alignment penalty incurred by the circular permutation of the catalytic motifs in the DRM proteins.



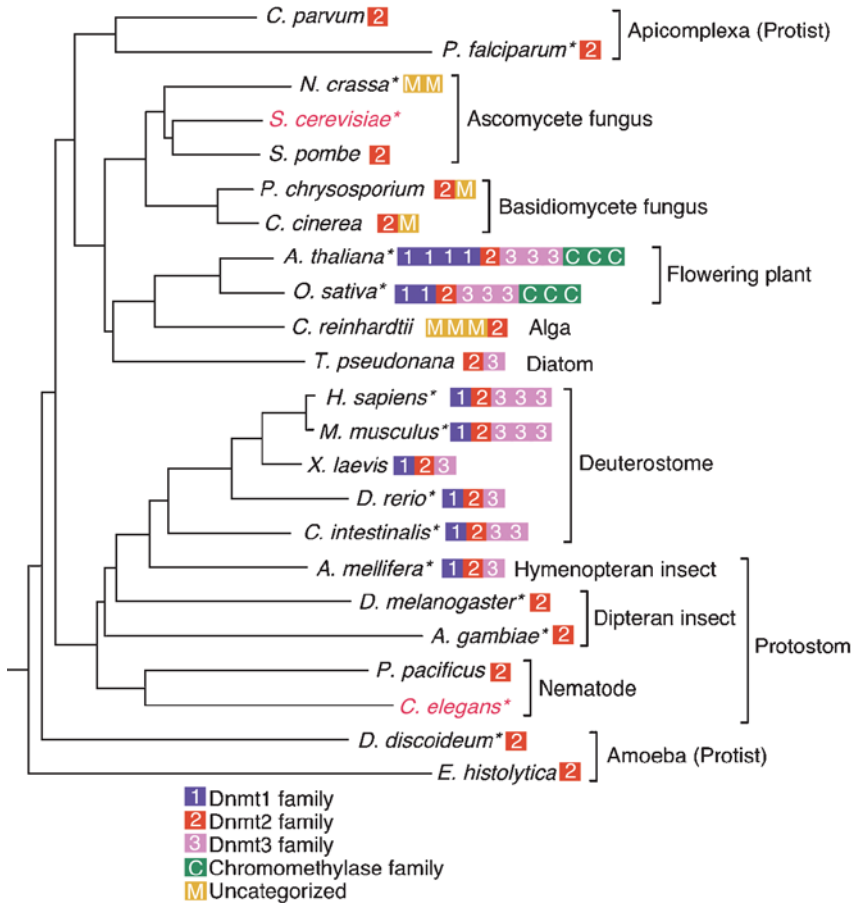


Figure 3 Distribution of the major cytosine methyltransferase families in eukaryotes. The sequences of 18s rDNA were used to construct the dendrogram in ClustalW. BLAST was used to identify cytosine methyltransferase homologues as in Figure 2*b*. The color scale at bottom indicates membership in the cytosine methyltransferase families. Asterisks indicate organisms for which there is a nominally complete draft genome sequence; others are only partially sequenced, or the methyltransferase sequences were represented by expressed sequence tags. The compilation is not exhaustive due to the lack of complete sequence coverage, and additional DNA methyltransferase homologues are likely to exist within the organisms shown. The genomes of *C. elegans* and *S. cerevisiae* lack any sequences that bear the methyltransferase motifs of Figure 1 and are shown in red. Note that Dnmt2 is always present when the Dnmt1 and Dnmt3 families are both represented, but many organisms contain only Dnmt2 homologues. Two members of the bacterial genus Geobacter also have Dnmt2 homologues, but this is the only prokaryotic taxon known to contain members of this family.

Dnmt1

The first eukaryotic DNA methyltransferase to be purified and cloned was later named Dnmt1 (53). Cedar, Razin, and colleagues (54) showed that hemimethylated DNA was methylated more rapidly than unmethylated DNA in nuclear extracts of cultured mammalian cells. It was later shown that purified Dnmt1 from mouse cells methylated hemimethylated substrates at an initial rate 5- to 30-fold greater than the unmethylated substrate (55); the exact difference depended on the sequence of the substrates. Maintenance methylation provides heritability to genomic methylation patterns in a way that has no counterpart outside of DNA replication itself. The preference of Dnmt1 for hemimethylated DNA caused it to be assigned a function in maintenance methylation, although Dnmt1 remains the only eukaryotic DNA methyltransferase to have been purified and cloned on the basis of its activity as a de novo cytosine methyltransferase (53). Also, the specific activity of Dnmt1 on unmethylated DNA substrates is greater than that of Dnmt3A and Dnmt3B, which are held to be the sole de novo DNA methyltransferases (46). Whether faithful maintenance methylation is enforced by other factors that inhibit the de novo activity of Dnmt1 in vivo (no such factors have been described) or the protein has de novo activity in vivo remains an unresolved issue.

Dnmt1 from Friend murine erythroleukemia (MEL) cells, which express high levels of Dnmt1 activity, was cloned by chromatographic purification of the protein, fragmentation of the homogenous protein with cyanogen bromide, and vapor-phase Edman degradation of purified peptides. Oligonucleotide probes, with a codon order corresponding to the amino acid sequences, were used to identify homologous cDNAs. The sequence of mouse Dnmt1 cDNA revealed a protein of 1620 amino acids that had a C-terminal domain of ~500 amino acids with clear similarities to the bacterial restriction methyltransferase M.DdeI (the only bacterial cytosine methyltransferase in the sequence databases at the time); a region of alternating glycine and lysine residues joined the C-terminal domain to a long (1100 amino acid) N-terminal domain, as shown in Figure 2. The *Dnmt1* gene was later found to have undergone amplification in MEL cells, probably as a result of coamplification with the nearby erythropoietin receptor, which is the target of the Friend virus transforming protein. This coamplification was the cause of the high-level expression of Dnmt1 in MEL cells.

The N-terminal domain of Dnmt1 contains a number of functional domains that have accreted over the course of evolution (Figure 2*a*). Experiments have identified a sequence required for import of Dnmt1 into nuclei and a second sequence required for association with replication foci, which are micrometer-scaled structures in which DNA synthesis occurs within mammalian nuclei (56). Dnmt1 has a diffuse nucleoplasmic distribution in G1 phase but associates with replication foci during S phase. Dnmt1 is present at only very low levels in noncycling cells. Sequences very near the N terminus have been shown to interact with DMAP1 (DNA methyltransferase associated protein-1) (57). The importance of this interaction is not clear because deletion of the region of Dnmt1 that interacts with DMAP1 produces no

overt phenotype in mutant mice (58). The N-terminal domain also has a role in coupling stabilization of DNA to the growth state of cells. Full-length Dnmt1 is degraded in G_0 cells, but when 118 N-terminal amino acids of the protein are removed by forcing translation to initiate at the second ATG codon, the protein is stabilized in G_0 cells (58). Mice that express only this shortened form of Dnmt1 were made by targeted deletion of the first ATG and found to be viable and fertile (58).

The form of Dnmt1 present in mouse oocytes (where it accumulates to very high levels) is the degradation-resistant shortened form; an oocyte-specific promoter and 5′ exon remove the first ATG and cause translation to initiate at the second ATG. This oocyte-specific form of Dnmt1 is called Dnmt1o; the truncation is thought to allow accumulation of high levels of Dnmt1o protein in noncycling oocytes. The removal of the N-terminal 118 amino acids of full-length Dnmt1 causes Dnmt1o to bind to annexin V, a calcium-sensitive phospholipid-binding protein that retains Dnmt1o in the cytoplasm of mouse oocytes and early embryos (59). ATRX (alpha thalassemia and mental retardation on the X chromosome) also binds to annexin V, and ATRX syndrome also involves methylation abnormalities at repeated sequences (60). The mechanism is currently unknown but may involve interaction between Dnmt1 and ATRX. The sex-specific promoters and splicing patterns of Dnmt1 are described in Reference 61.

Dnmt1 has a number of sequence motifs shared with other proteins but of unknown or unconfirmed function. Dnmt1 contains two bromo-adjacent homology (BAH) domains, which are also found in origin recognition complex proteins and other proteins involved in chromatin regulation (62). The BAH motif has been proposed to act as a protein-protein interaction module. Near the center of the N-terminal domain is a cysteine-rich region that binds zinc ions. The function of this cysteine-rich region is unknown, but it is present in all confirmed mammalian cytosine methyltransferases, known mammalian proteins affecting cytosine methylation, as well as the methyl-binding domain (MBD) proteins MBD1 and CpG binding protein. It has not been seen in any methylation-related protein from other organisms. A lysine- and glycine-rich sequence connects the N-terminal domain to the C-terminal catalytic domain. This sequence has some similarity to the N-terminal tail of histone H4 and may be the site of posttranslational modification, although very little is known of the role of posttranslational modification in the regulation of eukaryotic DNA methyltransferases.

Although Dnmt1 can be considered a replication factor, the protein is present at high levels in the cytoplasm of postmitotic neurons in mouse brain (63). What function Dnmt1 might play in neurons is not known. Deletion of the *Dnmt1* gene after the completion of neurogenesis has little effect on brain function (64).

The biological role of Dnmt1 has been addressed by extensive genetic studies. The *Dnmt1* gene was disrupted by homologous recombination in embryonic stem cells in 1992 (23), and a series of conditional, gain of function, and partial loss-of-function alleles has since been derived. Simple loss of function alleles of *Dnmt1* produce several novel phenotypes. First, the genome is severely demethylated in mice or embryonic stem (ES) cells that are homozygous for null alleles of *Dnmt1*,

although methylation persists at a level ~5% of wild type (65). Second, there is biallelic expression of most imprinted genes in homozygous embryos (66). Some imprinted genes, such as *H19* and *Kcnq1ot1*, are expressed from both alleles, whereas others, such as *Igf2* and *Kcnq1*, are expressed from neither. Third, there is inactivation of all X chromosomes in homozygous mutant embryos due to the demethylation and activation of *Xist* (28). Fourth, there is a modest (<10-fold) increase in mutation rates at exogenous marker genes in ES cells that lack Dnmt1 and an increase in rates of loss of heterozygosity because of mitotic recombination (67). A small defect in mismatch repair has recently been reported in Dnmt1-deficient ES cells (68). Fifth, there is the novel lethal differentiation phenotype in which ES cells that lack Dnmt1 grow normally in the undifferentiated state but die by apoptosis when induced to differentiate in vivo or in vitro (23). Cell-autonomous apoptosis is also the cause of death in Dnmt1-deficient embryos. The mechanism that underlies the lethal differentiation phenotype is unknown. Last, loss of Dnmt1 causes the demethylation and expression of very high levels of transposons of the intracisternal A particle (IAP) class, which are LTR retroposons (3). *Dnmt1* is the only gene known to be required for the repression of transposons in mammalian somatic cells.

Genetic studies of Dnmt1 provided a link between cytosine methylation and tumorigenesis. Heterozygosity for mutations in *Dnmt1*, or treatment of mice with the demethylating drug 2′ deoxy 5-azacytidine, reduced the number of colonic polyps in mice heterozygous for Apc^{min}, a mutation that predisposes to colon neoplasia (69). It was unclear as to whether heterozygosity for *Dnmt1* mutations actually cause measurable demethylation because no effect was seen in earlier studies (23). A knock-in mutation that caused mice to express Dnmt1 at a level of ~10% of wild type resulted in animals that were viable but stunted and had very high rates of lethal T-cell lymphomas (70). Leukemogenesis is likely to be the result of the demethylation and mobilization of endogenous retroviruses, which cause the majority of leukemias and lymphomas in most mouse strains. Overexpression of Dnmt1 in transgenic animals caused de novo methylation of the normally unmethylated allele at imprinted loci, which further argues against the designation of Dnmt1 as an obligate maintenance cytosine methyltransferase. The de novo methylation caused by overexpression of Dnmt1 was lethal when the increase in level of expression was more than a few-fold (71).

DNMT1 was reported to be overexpressed in human colon cancer by as much as 200-fold in comparison to normal mucosa (72), and on the basis of this report, the inhibition of DNMT1 has been proposed as a promising approach to the control of colorectal and other cancers. However, complete inhibition of DNMT1 would be expected to kill all dividing cells, and partial inhibition may cause genome instability and the development of other neoplastic diseases. Furthermore, other laboratories were unable to detect significant overexpression of DNMT1 in colorectal cancer (73, 74). There is a net loss of m^5C in many tumor cell genomes, which is inconsistent with increased DNMT1 levels. Mutation or amplification of the *DNMT1* gene in cancer has not been reported, and there is no independent

evidence that *DNMT1* is an oncogene. There are no genetic data that implicate any DNA methyltransferase or related factor in carcinogenesis (75).

MET1 of *A. thaliana*

The first plant DNA methyltransferase to be identified was cloned from *A. thaliana* by virtue of sequence similarity to Dnmt1 (76) and named MET1. The organization of the MET1 protein is similar to that of Dnmt1, with a C-terminal domain related to bacterial restriction methyltransferases and a long N-terminal domain joined via a glycine- and lysine-rich sequence, although this sequence is not a tract of simple alternating glycine and lysine as in Dnmt1. The C-terminal domains of Dnmt1 and MET1 are 50% identical, and the N-terminal domains are 24% identical. MET1 also has BAH domains, although it lacks the cysteine-rich region common to vertebrate Dnmt1 family members. Even though Dnmt1 represents the only member of its family in mammals, in *A. thaliana* MET1 is encoded by one of four similar genes that have conserved intron positions (76, 77). MET1 represents the major member of this family and is the only member that has appeared in phenotype-based forward genetic screens. It is not known whether all members of the plant kingdom possess multiple MET1 homologues, but two MET1 homologues have been reported in carrot and maize (78, 79).

Although MET1 has not been shown to be capable of methylating DNA in vitro and it is not known whether it is stimulated by hemimethylated substrates in vivo, genetic evidence does indicate that MET1 is a cytosine methyltransferase. *A. thaliana* plants that are homozygous for mutations at *MET1* or that express an antisense construct against *MET1* show genome demethylation and a set of phenotypic abnormalities that include homeotic transformation of floral organs, abnormalities in flowering time, and defects in vernalization responses. Demethylation in *met1* mutants is largely limited to CpG dinucleotides, and methylation of CpNpG trinucleotides and asymmetric sites at some sequences is actually increased over wild type. The phenotype of *met1* mutants is likely to arise from the combined effects of de novo methylation at non-CpG sites and the loss of CpG methylation. MET1 is clearly homologous to Dnmt1 and is considered to be a maintenance cytosine methyltransferase, but there are data which indicate that it is also involved in some de novo methylation events (80).

An unusual feature of genomic demethylation in plants is its inheritance across generations. Demethylation induced by homozygous mutations in *MET1* (or antisense/RNAi inhibition of MET1) or a second gene (*DDM1*) persists in descendent heterozygous mutant and wild-type segregants. Methylation levels only return to wild-type levels after several generations (81–83). This identifies another striking difference between genomic methylation patterns in plants and mammals in that methylation patterns are largely reset in each generation in mammals but are transmitted through meiosis with only modest alterations in plants. Phenotypes created in plants by demethylation can be transmitted to wild-type offspring with no sign of abnormality in any DNA sequence. This represents the first case in which an

epigenetic defect was transmitted through meiosis and provides a striking example of an epigenetic effect in which genotype cannot predict phenotype unless the history of the genome is also known.

The Dnmt3 Family

The mammalian genome encodes two functional cytosine methyltransferases of the Dnmt3 family, Dnmt3A and Dnmt3B, which primarily methylate CpG dinucleotides, and a third homologue, Dnmt3L, which lacks cytosine methyltransferase activity and functions as a regulatory factor in germ cells. *A. thaliana* members of the Dnmt3 family, or the domains rearranged methyltransferase (DRM) family, exhibit a circular permutation of the methyltransferase motifs and methylate DNA at asymmetric sites. A member of the Dnmt3 family has recently been found in the genome of the hymenopteran insect *Apis mellifera*, the first protostome shown to possess a member of the Dnmt3 family.

DNMT3A AND DNMT3B IN MAMMALS As shown in Figure 2, Dnmt3A and Dnmt3B are closely related proteins which bear N-terminal tails that contain a PWWP domain and a cysteine-rich zinc-binding region related to that of Dnmt1, ATRX, and some of the MBDs (methyl-binding domain proteins) (84). Both recombinant proteins transfer methyl groups to hemimethylated and unmethylated substrates at equal rates and without evidence of intrinsic sequence specificity beyond the CpG dinucleotide (46); Dnmt3A has also been reported to methylate CpA sites (85). Dnmt3A and Dnmt3B are expressed in a range of adult tissues but at lower levels than Dnmt1.

Reverse genetic studies in mice have identified some of the key functions of Dnmt3A and Dnmt3B. Dnmt3A-deficient mice survive to term but are runted and die in early adulthood with signs of aganglionic megacolon and a loss of germ cells in males. Global methylation patterns appear to be intact in Dnmt3A-deficient mice (86). Conditional alleles of *Dnmt3A* in which there was a preferential deletion of the *Dnmt3A* gene in germ cells showed that Dnmt3A was required in male germ cells for the establishment of methylation imprints at the differentially methylated regions (DMRs) of *H19* and *Gtl2-Dlk1* but dispensable for de novo methylation at the DMR of *Rasgrf* (87). The normal methylation of *Rasgrf* indicates that Dnmt3A cannot be the sole de novo DNA methyltransferase, although the agent responsible for de novo methylation of *Rasgrf* is unknown. At this time it is not known whether the demethylation defect is restricted to DMRs of imprinted genes or affects other regions of the genome, and it is not known whether transposons are demethylated and reanimated in Dnmt3A-deficient male germ cells. The large majority of de novo methylation in the germ line affects transposons and pericentric repeats. The fact that adult male mice that had a germ cell-specific loss of Dnmt3A progressed to complete azoospermia suggests that there is more widespread demethylation because a simple failure to methylate imprinted genes might not be lethal to germ cells. It should be noted that prospermatogonia, the male germ cell type in which

most de novo methylation occurs, contains a distinct isoform of Dnmt3A called Dnmt3A2. The Dnmt3A2 mRNA initiates at a promoter between exons 6 and 7 of the mouse *Dnmt3A* gene. Mouse Dnmt3A2 lacks 219 amino acids that are part of full-length Dnmt3A. The functional difference between Dnmt3A and Dnmt3A2 is not yet known, although different subcellular localization patterns have been reported (88).

Point mutations in human *DNMT3B* are responsible for the rare autosomal recessive human disorder known as ICF (immunodeficiency, centromere instability, and facial anomalies) syndrome (89). ICF syndrome is caused by a specific loss of methylation of classical satellite DNA (also known as satellites 2 and 3) at the pericentromeric regions of chromosomes 1, 9, and 16. The demethylated chromosomal regions cause a very high rate of gains and losses of long arms, and single chromosomes can have multiple long arms and single short arms. The cytogenetic abnormalities seen in ICF syndrome lymphocytes are unique to this disorder and among the most extreme of any human syndrome. Although classical satellite DNA is demethylated in all tissues of ICF syndrome patients, the chromosome instability is most prominent in stimulated T lymphocytes, and the major clinical feature is variable combined immunodeficiency. Lymphocytes must express factors that destabilize demethylated sequences or fail to express factors that stabilize demethylated DNA. Additional demethylation is seen at D4Z4 repeats in which changes in copy number have been implicated in the etiology of facioscapulohumeral muscular dystrophy (90). However, individuals with this disorder do not share symptoms with ICF syndrome. ICF syndrome patients have not shown evidence of tumor predisposition, although the syndrome is rare and most patients succumb to infectious disease at early ages. *Dnmt3B* mutant mice die around 9.5 dpc with demethylation of minor satellite repeats (86). It is not known if Dnmt3B is required for de novo methylation in germ cells of either sex.

Dnmt3A$^{-/-}$ *Dnmt3B*$^{-/-}$ mouse embryos die at ~8.5 dpc with global demethylation of their genomes, although the extent of demethylation is less than in *Dnmt1*$^{-/-}$ mutant embryos (86). Double mutant ES cells are unable to methylate newly integrated retroviral DNA (86). Although these results led to the classification of these two enzymes as the only de novo methyltransferases, *Dnmt3A*$^{-/-}$ *Dnmt3B*$^{-/-}$ ES cells can perform de novo methylation on partially methylated DNA integrated as single-copy sequence at defined genomic sites (91). In addition, ES cells from which Dnmt3A and Dnmt3B have been removed show mild demethylation, which requires that these enzymes have a role in the perpetuation of methylation patterns or that maintenance methylation is inefficient in ES cells (86, 92).

DNMT3L Dnmt3L (DNA methyltransferase 3-like) is the sole DNA methyltransferase homologue that is expressed specifically in germ cells (93). Dnmt3L is related to Dnmt3A and Dnmt3B in both N- and C-terminal domains and retains the cysteine-rich domain but lacks the PWWP domain (Figure 3). Similarity with Dnmt3A and Dnmt3B is seen in framework regions, but key residues within

catalytic motifs have been subject to nonconservative substitutions. In addition, the protein has not been shown to possess methyltransferase activity. However, Dnmt3L is essential for establishment of a subset of methylation patterns in both male and female germ cells (94).

Dnmt3L is essential for establishment of maternal genomic imprints in the growing oocyte and at dispersed repeated sequences in the prospermatogonia (a precursor to spermatogonial stem cells that exists for a short period around the time of birth). Although *Dnmt3L* homozygous null mice are viable, both males and females are sterile (94). Heterozygous embryos derived from homozygous *Dnmt3L* null oocytes die around 9 dpc and show a lack of maternal methylation imprints, with biallelic expression of imprinted genes normally expressed only from the allele of paternal origin. The paternally methylated and imprinted *H19* gene is imprinted normally, which confirms that imprint establishment rather than postfertilization maintenance is affected. The lack of Dnmt3L does not notably affect maternal meiosis, and there is only a small effect on global genome methylation. The demethylation effect is restricted to the DMRs of maternally imprinted genes. Comparison of the expression patterns of Dnmt3L and the phenotypes of Dnmt3L-deficient male and female mice reveals deep sexual dimorphism. Dnmt3L is expressed specifically in growing oocytes, which are tetraploid cells arrested in the dictyate stage of meiosis I. Meiotic recombination occurs prior to the appearance of Dnmt3L, and no mitotic divisions will take place until fertilization. Dnmt3L is expressed in male germ cells only in prospermatogonia, nondividing diploid cells present only at around the time of birth. Prospermatogonia differentiate into spermatogonial stem cells, which will undergo many mitotic divisions (as many as 100) before entering meiosis as spermatocytes. The loss of Dnmt3L does not interfere with oogenesis or early development in heterozygous embryos derived from homozygous *Dnmt3L* mutant oocytes, but Dnmt3L-deficient male germ cells display meiotic catastrophe with nonhomologous synapsis, asynapsis, and the accumulation of highly abnormal synaptonemal complexes (11). Abnormal synapsis triggers an apoptotic checkpoint, and no spermatocytes are observed to progress to the pachytene stages. Dnmt3L is dispensable for female meiosis (in which germ cells have completed the pachytene stage long before Dnmt3L is expressed) but required for male meiosis, even though Dnmt3L is not expressed during male meiosis and meiosis can occur many years after the brief perinatal period of Dnmt3L expression. The sexual dimorphism observed in the visible phenotypes of *Dnmt3L* mutant mice extends to de novo methylation as well. Dnmt3L-deficient oocytes are methylated essentially normally at transposons and tandem repeats; the defect is largely restricted to singly copy sequences associated with maternally imprinted genes. Dnmt3L-deficient male germ cells lack methylation at transposons and express very high levels of LINE-1 and IAP transposon RNAs. Abnormal synapsis is likely to be a secondary effect of transposon demethylation. The methylation of imprinted genes is affected much less in male germ cells than in female germ cells; there is normal methylation of the DMR of *Dlk1-Gtl2* and only partial demethylation of the DMR of *H19* in Dnmt3L-deficient male germ cells,

whereas there is nearly complete demethylation of single-copy DMR sequences in oocytes that lack Dnmt3L (11, 94).

The similar phenotypes seen in the germ line-specific *Dnmt3A* mutant mouse and the Dnmt3L-deficient female mouse, together with a report that Dnmt3L can stimulate the activity of Dnmt3A in an episome-based assay, argue that Dnmt3L and Dnmt3A proteins interact in the establishment of maternal imprints in germ cells (87, 95). Dnmt3L and Dnmt3A are, at present, the only factors known to be required for de novo methylation in germ cells. Additional factors could be identified by isolation of the genes mutated in the recessive maternal-effect disorder known as human familial biparental hydatidiform mole (FBHM), which is characterized by a complete lack of maternal imprints. FBHM conceptuses are similar in clinical presentation to spontaneous androgenetic complete hydatidiform moles despite the biparental inheritance of the chromosome complement. At least two loci that independently cause FBHM have been mapped to moderate resolution (96–98).

Dnmt3L homologues in mice and humans appear to be diverging rapidly when compared with other DNA methyltransferases (Figure 2*b*). Rapid evolution often reflects an evolutionary chase in which a parasite evolves at a high rate to evade host defense systems, which are then brought under selective pressures to counter the innovation of the parasite. Transposons represent the most rapidly diverging sequences within host genomes as a result of incessant selective pressures to evade host defense mechanisms. The rate of evolution of DNA methyltransferases is constrained by the requirement to preserve enzymatic activity. This constraint will limit the diversification of these enzymes and favor the evolution of adapter proteins free of this constraint and therefore capable of evolution at a much greater rate. It is suggested that Dnmt3L arose from an enzymatically active Dnmt3 family member in this way.

DNMT3 FAMILY MEMBERS IN INVERTEBRATES The recent accumulation of genomic and cDNA sequences from a broad range of species has enabled searches for sequences that contain the conserved cytosine methyltransferase motifs. All deuterostomes examined to date contain members of the Dnmt1, Dnmt2, and Dnmt3 families. A Dnmt3 homologue can be detected in the zebrafish genome. The C-terminal catalytic domain of the inferred sequence of this protein is very similar to that of mammalian Dnmt3A and Dnmt3B, but the N-terminal sequences bear no obvious similarity to any other protein (84). The genome of the sea squirt *Ciona intestinalis* (a protochordate) contains two Dnmt3 homologues, both of which show stronger conservation in the C-terminal catalytic region to Dnmt3A than to Dnmt3B.

Until very recently there was no evidence of Dnmt3 family members in protostome invertebrates. Homologues of Dnmt1 and Dnmt3 are present in the genome of honeybee (*A. mellifera*, a hymenopteran insect). Although sequencing of the lepidopteran insect *Bombyx mori* (silkworm) is not complete, a Dnmt1 homologue has also been detected in its genome. It is predicted that further sequencing of the *B. mori* genome will produce a Dnmt3 homologue. The genomes of the

dipteran insects *D. melanogaster* and *Anopheles gambiae* have been sequenced more thoroughly than those of *A. mellifera* or *B. mori*, but the dipteran genomes show evidence only of Dnmt2. Cytosine methylation levels in *B. mori* were estimated to be between 0.158% and 0.198% by high-pressure liquid chromatography analysis, whereas levels reported in the same study for *D. melanogaster* methylation were less than 0.003% (99). This report is contradicted by two more recent studies (100, 101), which report low but detectable levels of m^5C in *D. melanogaster*; although one reports almost constant levels of m^5C during development, while the other reports that adult *D. melanogaster* lacks this modification. If there are small amounts of DNA methylation in dipterans, it is likely to have a different form and function than in organisms that have Dnmt1 and Dnmt3 family members.

THE DRM FAMILY: RNA-GUIDED METHYLTRANSFERASES IN PLANTS The DRMs represent a divergent group within the Dnmt3 family that is not known to exist outside the flowering plants. DRM proteins show an average 28% amino acid identity to mammalian Dnmt3A and Dnmt3B proteins in their C-terminal domains with strong conservation at catalytic motifs. The DRM N-terminal domains lack recognizable motifs other than ubiquitin-associated domains (Figure 3), which are of unknown function and are not present in other cytosine methyltransferases (47). The cysteine-rich domain characteristic of the mammalian Dnmt1 and Dnmt3 families is not discernible in the DRM proteins or any other plant DNA methyltransferase homologue. DRM proteins show circular permutation of the cytosine methyltransferase motifs with motifs VI through X preceding motifs I–V (47). Examination of the crystal structures of M. HhaI, M. HaeIII, and DNMT2 suggests this circular permutation might still allow the general methyltransferase domain structure to be maintained because motifs I and X make up the S-adenosyl L-methionine-binding site and are located in close proximity in the folded structure (Figure 1). At least one bacterial cytosine methyltrasferase homolog has been identified, which demonstrates both permutated motifs and m^5C activity (102).

The *A. thaliana* genome encodes two DRM proteins, DRM1 and DRM2 (47). These proteins are responsible for the initial establishment of CpN, CpG, and CpNpG methylation by a process called RNA directed DNA methylation (RdDM) (103, 104). RdDM appears to rely on the generation of short RNAs as an initial signal for methylation of homologous DNA sequences. DRM1 and DRM2 have also been shown to work with CMT3 (see below) to perpetuate methylation at asymmetric sites (103) but have no effect on maintenance of CpG methylation. Additional DRM genes (*DMT10* and *DMT106*), which lack the critical PC and ENV motifs, have been identified in *A. thaliana* and *Zea mays*, respectively. DMT10 shows 34% identity with DRM2 in the C-terminal domain. The role of this protein remains to be identified, but its function may be analogous to that of mammalian Dnmt3L in that it represents a regulatory factor derived from a cytosine methyltransferase that is released from constraint and can evolve rapidly in pursuit of transposons, which rapidly evolve under pressures of host defense.

DIM-2 and RID in *N. crassa*

The deepest investigation of the form, function, and regulation of genomic methylation patterns in any organism has been performed by Selker (105) in the ascomycete *N. crassa.* All apparent cytosine methylation in vegetative cells of *N. crassa* can be attributed to a single methyltransferase homologue, defective in methylation-2 (DIM-2). DIM-2 was identified in a forward genetic screen for methylation mutants; mutations were found in a sequence that contained the diagnostic cytosine methyltransferase motifs. DIM-2 deficient strains have no detectable cytosine methylation but are viable and fertile (105). The *dim-2* gene encodes a protein of 1454 amino acids with an N-terminal tail that bears a degenerate BAH domain and an ATP/GTP-binding motif (Figure 3). DIM-2 has an acidic region C-terminal of the methyltransferase domain. Within the cytosine methyltransferase domain, DIM-2 shows distant similarity to MET1 and Dnmt1 proteins and is likely to be a highly diverged member of the Dnmt1 family. Cytosine methylation in *N. crassa* occurs in all sequence contexts, and preferential methylation at symmetrical sequences has not been reported. It is not known whether the DIM-2 protein has a preference for hemimethylated DNA or whether there is somatic inheritance of methylation patterns in *N. crassa*. As discussed below, DNA methylation mediated by DIM-2 depends on trimethylation of lysine 9 of histone H3 (H3K9).

Repeated sequences are silenced and methylated at cytosine residues in all sequence contexts in a premeiotic genome scanning process that occurs after fusion of hyphae of different mating types but prior to fusion of haploid nuclei. This gene silencing pathway is known as RIP (repeat-induced point mutation); the affected sequences become irreversibly inactivated by large numbers of C $\rightarrow$ T transition mutations. Once subject to RIP, the mutated sequences remain methylated in vegetative cells even after segregation into the single copy state. Sequences that have been subjected to RIP are again methylated after they are demethylated by cloning and reintroduced as single copy sequences.The increased density of the TpA dinucleotide after RIP is thought to be the cue for methylation by DIM-2 (106, 107).

The isolation and characterization of the methylated compartment of the *N. crassa* genome showed that virtually all m^5C was in the remnants of transposons and that all transposon remnants were heavily mutated by RIP (108). *N. crassa* is the only sexual organism known to lack active transposons (109). *N. crassa* is very aggressive in its response to repeated sequences and will destroy essential genes if present in more than one copy. The extreme intolerance of repeated sequences by *N. crassa* has eliminated all active transposons but has also prevented the evolution of gene families, and the evolutionary potential of this organism may have been constrained by an inability to tolerate the duplication and divergence of existing genes. This may be the price of complete protection against transposons (110).

Dim-2 mutants retain the ability to perform RIP (105). A second methyltransferase homologue was identified in *N. crassa* genomic sequence by its content of cytosine methyltransferase catalytic motifs (17). This homologue, RIP defective

(RID), has both an N-terminal tail that lacks recognizable conserved motifs and a C-terminal extension to the methyltransferase domain that contains a PC dipeptide and a serine-rich domain (Figure 2*a*). There are two possible functions of RID in RIP. RID may methylate cytosines prior to deamination by unidentified factors to drive C → T mutagenesis, or RID might be a cytosine deaminase that catalyses C → U mutations that are converted to U → T mutations during replication. Under some conditions, cytosine methyltransferases themselves can act as cytosine deaminases (111). RID is the only factor known to be involved in RIP, which is itself one of the most unusual genetic phenomena ever described.

Although this process of repeat-induced gene silencing by methylation was discovered and is best understood in *N. crassa*, a similar process exists in the ascomycete *Ascobolus immersus* and in the distantly related basidiomycete *Coprinus cinereus*, which contains several cytosine methyltransferase homologues (112). Several methyltransferase homologues are also present in the basidiomycete *Phanerochaete chrysosporium*. It is striking that those model organisms that have come to be most favored in genetic studies (notably *C. elegans*, *D. melanogaster*, *Schizosaccharomyces pombe*, and *S. cerevisiae*) have no evidence of a methylation-dependent gene silencing system, which complicates studies of transmission genetics.

Chromomethylases

Methylation of the symmetrical sequence CpNpG is seen in plants but is rare in vertebrates; this methylation depends on the plant-specific chromomethylase family. Methylated CpNpG sites present symmetrical hemimethylated sites after replication, and their methylation could be perpetuated by maintenance methylation. Forward genetic screens in *A. thaliana* identified mutations that reactivated cellular genes, which had been methylated at CpNpG sites as a result of loss of MET1. These mutations caused a genome-wide loss of CpNpG methylation (113, 114). The mutations were found to inactivate a chromomethylase gene.

Henikoff & Comai (114) identified the chromomethylases as a new family of cytosine methyltransferases unique to flowering plants. They are distinguished by the presence of a chromodomain between methyltransferase motifs II and IV (Figure 2*a*) (114). Chromodomains were first described in Polycomb group proteins and are implicated in targeting proteins to heterochromatin (115). Chromodomains are not found in other cytosine methyltransferases. The *A. thaliana* genome includes three chromomethylase homologues: CMT1, CMT2, and CMT3. Additional family members have been identified in rice and broccoli (116, 117), but none have been identified outside of the plant kingdom. Although the functions of CMT1 and CMT2 are unknown and retroelement insertions into *CMT1* in some ecotypes of *A. thaliana* suggest this *CMT* gene is dispensable in the wild (114), mutations at the CMT3 locus result in a severe decrease in CpNpG methylation (113, 118). Variable sequence-dependent reductions in asymmetric methylation and minor effects on CpG methylation were also reported, although the cause of

these effects is unclear. Inactivation of CMT3 leads to transposon reanimation; however, the severe defect in CpNpG methylation does not translate into any obvious morphological phenotypes even after multiple generations of self-fertilization (113, 118). *A. thaliana* is largely self-fertilizing and has few active transposons (8), and reactivation of transposons under CMT3 deficiency is not expected to have immediate phenotypic consequences. In plants, it is probable that CpNpG methylation replaces or reinforces CpG methylation in sequences without large numbers of CpG targets. Indeed, CACTA transposons are reactivated to a significantly higher degree in *met1 cmt3* double mutants than in either single mutant (12). CpNpG methylation may have evolved to counter transposons that had developed CpG-free promoters and were therefore immune to CpG methylation, which was probably the methylated dinucleotide in the common ancestor of vertebrates and flowering plants. CpNpG methylation reinforces the CpG methylation system, and transposons, reanimated by demethylation of CpG and CpNpG sites, produce RNA transcripts that target methylation and silencing to the source elements through the RNA- and DRM-mediated pathway.

Dnmt2

The most strongly conserved, most widely distributed, and most enigmatic of the cytosine methyltransferase homologues are those of the Dnmt2 family. Dnmt2 was identified as an expressed sequence tag that bore the characteristic DNA methyltransferase motifs (45). The inferred protein sequence contained all 10 diagnostic motifs in the canonical order. A related protein had earlier been identified in the fission yeast *S. pombe* and named pmt1p. However, mammalian Dnmt2 has the canonical PC motif at motif IV, whereas pmt1p had a PSC sequence at that motif. (The basidiomycete fungus, *C. cinereus*, has a PAC tripeptide at motif IV; only fungal Dnmt2 homologues have been found to contain an interrupted motif IV.) Deletion of the central serine in *S. pombe* Dnmt2/pmt1p and expression of the mutant protein in *E. coli* was reported to confer enzyme activity to recombinant pmt1p, with a sequence specificity of CCWGG (119). This recognition sequence is identical to that of Dcm, a cytosine methyltransferase common to many laboratory strains of *E. coli*. Sensitive mechanism-based assays failed to detect any DNA methyltransferase activity in Dnmt2 (45, 52, 120), although other laboratories have reported levels of methyltransferase activity in Dnmt2 that are close to the background of the assays used (121, 122). The balance of the available data indicates that Dnmt2 does not methylate single- or double-stranded DNA to any significant extent; this conclusion is supported by genetic data described below.

As shown in Figures 2 and 3, Dnmt2 is the most widely distributed DNA methyltransferase homologue. Organisms that contain members of the Dnmt1 and Dnmt3 families invariably contain Dnmt2, and in a number of species, Dnmt2 is the sole DNA methyltransferase homologue. Dnmt2 (and all other DNA methyltransferase homologues) is absent from *C. elegans* but is present in the related nematode *Pristionchus pacificus* and in *D. melanogaster*. Dnmt2 homologues are also present in

two species within the bacterial genus *Geobacter* but not in any other bacterium whose genome has been sequenced. Although Dnmt1 and Dnmt3 frequently form multigene families in single species, Dnmt2 has been found only as a singleton gene. Northern blot analysis showed that it is expressed in many adult mammalian tissues (45, 123).

Almost every detail of sequence and structure indicates that Dnmt2 should be a DNA cytosine methyltransferase. Sequence comparisons show that the catalytic motifs are well conserved, as is a region between motifs VIII and IX that normally comprises the target recognition domain. This motif is centered on a CFTXXYXXY (CFT) motif that is nearly invariant within the Dnmt2 family but absent from any other DNA methyltransferase homologue. The crystal structure of human DNMT2, completed by X. Cheng and colleagues (52), shows that the structures of DNMT2 and the bacterial restriction methyltransferase M.HhaI are essentially superimposable over the large domain and that all the conserved methyltransferase motifs are present in a well-organized active site with every side chain in the correct orientation (Figure 4*a*,*b*). Both the M.HhaI-DNA and DNMT2 structures contained the demethylated cofactor S-adenosyl L-homocysteine, and the positions of the cofactors in the two structures are also superimposable (Figure 4*b*). The CFT motif spans the interface between the large and small domains in a region that makes sequence-independent DNA contacts in M.HhaI and M.HaeIII. The DNA from the M.HhaI-DNA cocrystal can be modeled into the DNMT2 structure with accommodation of the everted cytosine and with only one significant steric conflict. This involves the second tyrosine in the CFT motif, which clashes with the nontarget strand of double stranded DNA (Figure 4*d*). This tyrosine is nearly invariant in Dnmt2 homologues of flowering plants and metazoa. In M.HhaI, this position in the structure is occupied by glycine 257, and extensive random mutagenesis of M.HhaI showed that no other residue could be tolerated at this site (124). Structural analysis shows that this glycine interacts with the G4′ of the DNA through a main chain nitrogen atom and has ϕ-ψ angles that are incompatible with the presence of a side chain (124). The second tyrosine in the CFT motif is actually positioned to make hydrogen bonds with a base in the strand complementary to that involved in the clash. This finding makes it unlikely that the substrate for Dnmt2 is normal duplex DNA, although the real substrate has remained elusive.

Dnmt2 homologues are unique among eukaryotic methyltransferases in that they lack N-terminal extensions. In this respect they more closely resemble the bacterial cytosine methyltransferases. However, some Dnmt2 homologues possess additional sequence between motif VIII and the TRD region (52). The sequence of this region is poorly conserved between deuterostomes and plants and completely absent from Dnmt2 homologues of insects, fission yeast, and protozoa. This insertion shows no homology to other known proteins by BLAST (52), and its function is unknown.

The function of Dnmt2 has been addressed by genetic studies. ES cells homozygous for disruption alleles of Dnmt2 do not show methylation abnormalities

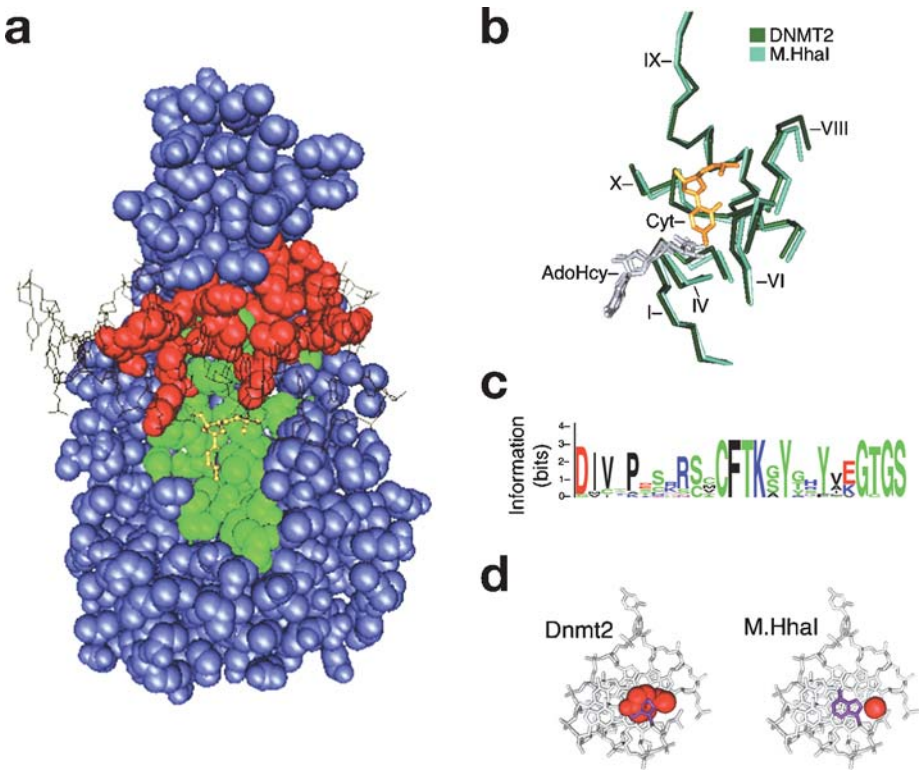


Figure 4 Distinctive sequence and structural features of Dnmt2 homologues. (*a*) Crystal structure of human DNMT2 with the DNA from the M.HhaI-DNA-AdoHcy cocrystal modeled onto the DNMT2 structure by the NCBI Vector Alignment Search Tool (VAST). Green indicates the conserved motifs shown in Figure 1; red indicates the CFT motif that is diagnostic of the Dnmt2 family. Most of the face of DNMT2 that by analogy with M.HhaI and M.HaeIII would interact with DNA is represented by the CFT motif. (*b*) Isostery of catalytic motifs in DNMT2 and M.HhaI. Optimal overall superimpositions of DNMT2 and M.HhaI were obtained from NCBI VAST, and the catalytic motifs shown in Figure 1 were displayed in NCBI Cn3D. Both DNMT2 and M.HhaI contained AdoHcy; only the M.HhaI structure contained DNA. Note the very similar trajectories of the peptides that contain the catalytic motifs in DNMT2 (*dark green*) and M.HhaI (*light green*) and note the close superimposition of AdoHcy in the two structures. (*c*) Conservation of the CFT motif in Dnmt2 homologues of vertebrates and flowering plants. ClustalW alignments are shown in LOGOS. (*d*) Steric clash of a conserved tyrosine in DNMT2 with double stranded DNA. The structural model contained DNA from the M.HhaI-DNA-AdoHcy shown in (*a*). The tyrosine side chain is shown in red and can be seen to be in conflict with the nontarget DNA strand. The glycine that occupies this position in M.HhaI does not clash, and mutagenesis experiments have shown that substitutions at this position are incompatible with the enzymatic activity of M.HhaI (124).

and are of normal phenotype (120). Mice derived from a second deletion allele of Dnmt2 are of normal phenotype and have normal genomic methylation patterns after seven generations as homozygotes (M.G. Goll, J.A. Yoder, and T.H. Bestor, unpublished results). A point mutation that produces a highly nonconservative E → K substitution in the nearly invariant ENV tripeptide of motif VI of the *A. thaliana* Dnmt2 homologue did not produce a detectable phenotype. No defect in gene silencing, centromere function, imprinted mating type switching, or other functions was observed in *pmt1* deletion strains of *S. pombe* (M.G. Goll and T.H. Bestor, unpublished results). The *D. melanogaster Dnmt2/Mt2* gene was disrupted by insertion via homologous recombination of stop codons into an internal exon. Mutated flies are fertile and show no obvious developmental abnormalities (K. Maggert, K. Golic, M.G. Goll, and T.H. Bestor, unpublished information). The strong conservation of sequence and structure, wide phylogenetic distribution, and ubiquitous expression of Dnmt2 seem to demand that it have a function, but the genetic and biochemical data indicate that it is not important for survival under laboratory conditions and that it is not a conventional DNA cytosine methyltransferase. Dnmt2 presents an unusual case of a protein that appears to be an authentic DNA cytosine methyltransferase by every detail of sequence and structure, but has failed to show any evidence of such a function in genetic and biochemical tests.

METHYLATED DNA BINDING PROTEINS

Methylated promoters are usually inactive, but the factors that mediate this repression are unknown. Mammalian proteins that have been reported to show some selective binding to methylated CpG sites have been proposed to inhibit the transcription of methylated promoters. The most prominent of these is methylcytosine-binding protein 2 (MeCP2), which has been promoted as the global repressor of methylated promoters (125). Methyl-CpG-binding domain protein 2 (Mbd2) has also been reported to bind with higher affinity to methylated CpG sites than to unmethylated sites (126). MeCP2 has been proposed to mediate its inhibitor effects by direct recruitment of histone deacetylase 1 (HDAC1) and Sin3A to deacetylate histones in the vicinity of methylated CpG sites (127). However, more recent data suggest that Sin3A and MeCP2 do not form a stable complex (128).

Early biochemical and overexpression experiments were consistent with the view that MeCP2 represses methylated promoters. However, genetic data clearly show that MeCP2 and Mbd2 are not required for the repression of methylated promoters, which remain silent in mice that lack either or both proteins. Furthermore, there are no detectable abnormalities of histone acetylation in the absence of MeCP2, as would be expected if MeCP2 targets HDAC1 to inactive chromatin (129). Development to term is normal in the absence of MeCP2 or Mbd2, and the Mbd2 mutation causes only small effects on maternal behavior (130, 131). Mutations in MeCP2 cause the human neurodevelopmental disorder called Rett syndrome and a closely related disorder in MeCP2 mutant mice, but there is very

little evidence that the reactivation of methylated promoters occurs or has any role in the etiology of the disorder (131, 132). There is little or no similarity in the phenotypes of mice that lack cytosine methyltransferases and those that lack methylcytosine binding proteins. The key phenotypes of demethylating mutations (lethality, transposon reactivation, and loss of allele-specific transcription) have not been reported to occur in MeCP2- or MBD2-deficient mice.

REGULATORY INPUTS THAT CONTROL SEQUENCE-SPECIFIC DNA METHYLATION

As mentioned above, eukaryotic DNA methyltransferases have little or no innate sequence specificity, and the mechanism by which sequence specificity is conferred to the methylation reaction has been one of the outstanding problems in the field of epigenetics. Recent results indicate that target specificity is the result of multiple inputs and that different cytosine methyltransferase families respond to different stimuli. The emerging view of the regulatory pathways is shown in Figure 5.

Interaction of Repeated Sequences

The large majority of cytosine methylation in eukaryotic genomes is in repeated sequences. All cytosine methylation in *N. crassa* is within RIPed and mutated transposons, and the large majority (>90%) of cytosine methylation in the genomes of vertebrates and flowering plants resides in dispersed repeats in the form of transposons and tandem repeats in the form of pericentromeric satellite DNA.

Transposons must increase in copy number if they are to spread within the host population, and genome scanning mechanisms have evolved to allow the host to identify and repress sequences that are increasing in copy number. RIP was the first such host defense mechanism to be identified; even one supernumery sequence is sufficient to trigger silencing and active mutagenesis of both copies regardless of sequence. Transcription is not known to be involved in the detection of repeats, and there is no evidence of RNA involvement in RIP. Repeat-dependent methylation and silencing in mammals occurs in germ cells; the de novo methylation of dispersed repeats is dependent on Dnmt3L in male germ cells of the mouse. Repeated sequences, however, are methylated normally in Dnmt3L-deficient oocytes (11, 94). Factors required for de novo methylation of dispersed repeats in oocytes have not been identified.

Possible mechanisms of repeat-dependent de novo methylation have been presented (133). Of particular interest are the homology-heterology junctions that form when dispersed repeats undergo strand exchange reactions as part of a homology search process. The rate of formation of such junctions will follow first-order kinetics and will scale as the square of copy number, thereby providing the possibility of a threshold effect in which an increase in copy number above a certain point causes de novo methylation and silencing of all copies of the repeated

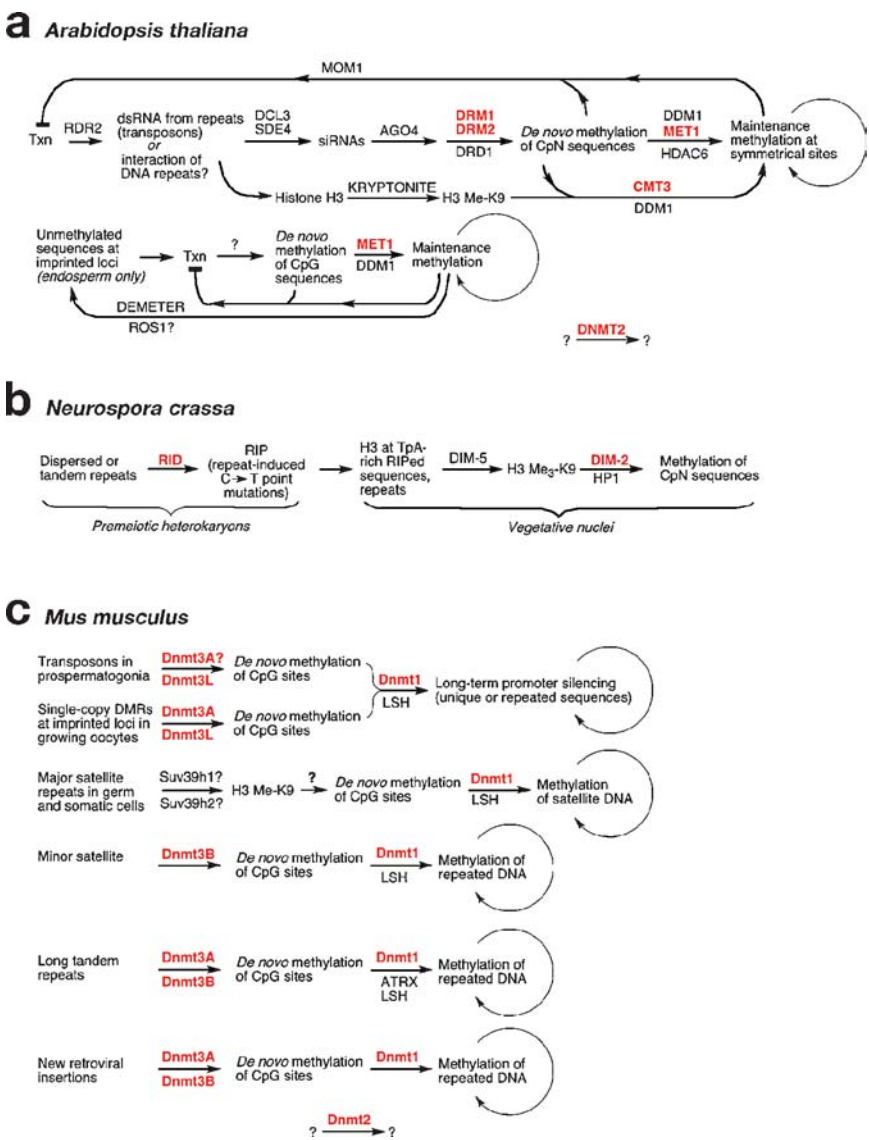


Figure 5 Regulation of cytosine methylation in eukaryotes. Genetic interactions were identified in forward genetic screens in *A. thaliana* and *N. crassa*, although RID was characterized in a targeted mutagenesis experiment. The interactions shown in (*c*) were largely the result of reverse genetic experiments that involved targeted gene disruption in mice. DNA methyltransferase homologues are shown in red even if no enzymatic activity has been observed. Not all the factors involved in the indicated pathways have been identified, and the diagram will require revision in the future. To date no sequence-specific DNA binding protein has been shown to be involved in the establishment or maintenance of genomic methylation patterns.

sequences. Such a process has been proposed to operate in prospermatogonia (11). A similar mechanism in which the stimulus for de novo methylation is the presence of hemimethylated sites in strand exchange intermediates will also cause a common methylation pattern to spread through a network of nonallelic repeats with first-order kinetics and the possibility of a threshold effect. As mentioned above in connection with repeat silencing in *N. crassa*, some tolerance of repeated sequences is necessary to allow the evolution of new functions by the duplication and divergence of existing genes. It is suggested that large-genome organisms tolerate repeated sequences until their number exceeds a threshold value, at which point they are identified as proliferating transposons and all copies are methylated and silenced.

Regulation of Cytosine Methylation by Histone H3 Methylation

Histone methylation was first shown to be required for targeting cytosine methylation in *N. crassa*. A forward screen for demethylating mutants uncovered a gene related to the Su(var)3-9 H3K9 methyltransferases. This gene was named *dim-5* and was shown to be essential for all detectable cytosine methylation in vegetative cells (134). The demethylation phenotype was confirmed to result from the lack of histone tail modifications by construction of a K9 → R mutation in the singleton histone H3 gene of *N. crassa*, which was found to phenocopy the demethylation phenotype of *dim-2* mutants. Mutations in the gene for heterochromatin protein 1 (*HP1*), which has been shown to bind H3K9, also abolish methylation in *N. crassa* (135). These results provided very strong evidence of an essential role for DIM-5-mediated histone H3K9 in DNA methylation in vegetative cells of *N. crassa*.

Methylation in *N. crassa* is distinct from methylation in other organisms because cytosines in all sequence contexts are methylated and because both de novo and maintenance methylation appear to depend on a single methyltransferase, DIM-2. In *A. thaliana*, loss of KRYPTONITE, a relative of Su(var)3-9, phenocopies mutations in the chromomethylase *CMT3*. *KRYPTONITE* mutants show a severe loss of methylation at CpNpG sites, whereas methylation at CpG sites, which represents the bulk of genomic methylation in *A. thaliana*, is not markedly affected (136). However, there is at least one exception to the rule that CMT3 requires H3K9 methylation for CpNpG methylation. At the phosphoribosyl anthranilate isomerase (PAI) locus, single *PAI* genes are unmethylated in *KRYPTONITE* mutants; however, an inverted repeat at the locus retains CpNpG methylation (137). In this system, a requirement of HP1 was not observed for methylation at either the single gene or the inverted repeat. In mice, mutation of the two Su(var)3-9 homologs has a relatively minor effect on cytosine methylation and caused only partial demethylation of major satellite regions of pericentromeric heterochromatin (138). The specificity of the demethylation defect and the fact that major satellite repeats were not nearly as demethylated as in Dnmt1-deficient embryos suggest a specialized role for H3K9 methylation in targeting of cytosine methylation in mammals. H3K9 is clearly essential for targeting cytosine methylation in vegetative

cells of *N. crassa* and is involved in the chromomethylase pathway of CpNpG methylation in *A. thaliana*, but it has less apparent importance in vertebrates, in which there is little non-CpG methylation.

Swi2/Snf2 Helicase Homologues and Cytosine Methylation

Although their exact role has not been elucidated, some members of the Swi2/Snf2 family of ATP-dependent chromatin-remodeling factors are clearly important for cytosine methylation (Figure 5). DDM1 was first identified in a forward genetic screen for genes that affect global cytosine methylation patterns in *A. thaliana*. Plants mutant for *DDM1* show a 70% decrease in methylation (primarily at repetitive DNA), but the mutant plants are viable and fertile (139). Defects increase in severity in later generations, and morphological phenotypes become apparent. De novo methylation of the *SUPERMAN* locus is possible in *ddm1* mutant plants (140). Mutations in lymphoid-specific helicase (*Lsh*), the mammalian homologue of DDM1, cause a similar loss of methylation in mice. *Lsh* mutant mice die in the perinatal period with defects in peripheral T cells and kidney abnormalities (141). Additionally, these mice show a 50% reduction in cytosine methylation at repetitive sequences and some demethylation of single copy sequences (142).

Additional members of the Swi2/Snf2 family have also been implicated in cytosine methylation pathways. Mutations in *ATRX* lead to demethylation of several repeated sequences, including the rDNA arrays, a Y-specific satellite, and subtelomeric repeats (143, 144). A plant-specific Swi2/Snf2 homologue named DRD1 (145) is specifically required for non-CpG methylation during the process of RdDM, which suggests that Swi2/ Snf2 family members are involved in multiple methylation pathways in both plants and mammals.

RNA Directed DNA Methylation

RdDM was first identified in transgenic tobacco plants. Transgenes, consisting of the cDNA of potato spindle tuber viroid, became methylated after viroid infection, even though there is no DNA phase in the viroid replication cycle (146). These experiments provided the first link between RNA and the control of DNA methylation. Establishment of methylation by RdDM is mediated by the DRM methyltransferases (104, 147) and is dependent on a number of components of the RNAi pathway (148). The fact that methylation is confined to regions of RNA-DNA sequence similarity (149) argues for a mechanism in which the RNA molecule itself targets methylation to a specific sequence through a base pairing mechanism. It is likely that the hypermethylation of non-CpG sites, observed at specific loci in globally demethylated *met1* mutant plants, can be attributed to RdDM. Global loss of CpG methylation is predicted to relieve transcriptional repression of transposable elements and other repeated sequences. This generates aberrant double stranded RNAs, which activate the RdDM pathway to impose de novo methylation on homologous DNA sequences. Local de novo methylation in a context of global genome demethylation has also been proposed to occur in some cancer cell genomes.

It is not clear that RdDM is a major mechanism for de novo methylation outside of the plant kingdom. RdDM does not appear to play a major role in *N. crassa.* DNA methylation and HP1 localization occur normally in the absence of elements of the RNAi machinery (135). Proteins shown to be essential for the RdDM pathway in *A. thaliana* appear to be absent from mammals. The mammalian genome does not contain a homologue of an RNA-dependent RNA polymerase or the putative SNF2 chromatin-remodeling protein DRD1, which are both required for RdDM in plants (145, 148), and outside of the C-terminal catalytic domains, the DRM proteins and mammalian DNMT3 homologues are highly divergent. Moreover, although there are multiple Dicer genes in plants and some evidence suggests that siRNAs can be generated both in the nucleus and the cytoplasm (150, 151), thus far it appears that the single functional Dicer protein in mammals is cytoplasmic. However, there have been recent reports of RNA-dependent cytosine methylation in mammalian cells (152, 153). The importance of this pathway in mammalian cytosine methylation awaits further study.

CONCLUSION

It has recently become clear that the establishment and maintenance of genomic methylation patterns do not depend on the recognition of specific sequences but are instead controlled by the sensing of repeated sequences, the interaction of RNA and DNA, and, in some organisms, by histone methylation. The apparent lack of involvement of sequence-specific DNA binding proteins in methylation-dependent gene silencing violates the accepted view of gene regulation. De novo methylation is imposed on target sequences at particular points in the development of the organism, and phenotype can be predicted by genotype only if both the sequence and history of that genotype are known.

ACKNOWLEDGMENTS

We thank K. Anderson, D. Bourc'his, M. Damelin, M. Matzke, and E. Selker for comments and discussion. This work was supported by grants from the National Institutes of Health.

The *Annual Review of Biochemistry* is online at http://biochem.annualreviews.org

LITERATURE CITED

1. Riggs AD. 1975. *Cytogenet. Cell Genet.* 14:9–25
2. Holliday R, Pugh JE. 1975. *Science* 187:226–32
3. Walsh CP, Bestor TH. 1999. *Genes Dev.* 13:26–34
4. Wigler M, Levy D, Perucho M. 1981. *Cell* 24:33–40

5. Schubeler D, Lorincz MC, Cimbora DM, Telling A, Feng YQ, et al. 2000. *Mol. Cell Biol.* 20:9103–12
6. Smit AF, Riggs AD. 1996. *Proc. Natl. Acad. Sci. USA* 93:1443–48
7. Yoder JA, Walsh CP, Bestor TH. 1997. *Trends Genet.* 13:335–40
8. Takahara T, Ohsumi T, Kuromitsu J, Shibata K, Sasaki N, et al. 1996. *Hum. Mol. Genet.* 5:989–93
9. Chang-Yeh A, Mold DE, Huang RC. 1991. *Nucleic Acids Res.* 19:3667–72
10. Bestor TH. 2003. *Trends Genet.* 19:185–90
11. Bourc'his D, Bestor TH. 2004. *Nature* 431:96–99
12. Kato M, Miura A, Bender J, Jacobsen SE, Kakutani T. 2003. *Curr. Biol.* 13:421–26
13. Miura A, Yonebayashi S, Watanabe K, Toyama T, Shimada H, Kakutani T. 2001. *Nature* 411:212–14
14. Singer T, Yordan C, Martienssen RA. 2001. *Genes Dev.* 15:591–602
15. Hirochika H, Okamoto H, Kakutani T. 2000. *Plant Cell* 12:357–69
16. Schorderet DF, Gartler SM. 1992. *Proc. Natl. Acad. Sci. USA* 89:957–61
17. Freitag M, Williams RL, Kothe GO, Selker EU. 2002. *Proc. Natl. Acad. Sci. USA* 99:8802–7
18. Marino-Ramirez L, Spouge JL, Kanga GC, Landsman D. 2004. *Nucleic Acids Res.* 32:949–58
19. Davuluri RV, Grosse I, Zhang MQ. 2001. *Nat. Genet.* 29:412–17
20. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, et al. 2001. *Science* 291:1304–51
21. Lin IG, Tomzynski TJ, Ou QL, Hsieh CL. 2000. *Mol. Cell Biol.* 20:2343–49
22. Matsuo K, Silke J, Georgiev O, Marti P, Giovannini N, Rungger D. 1998. *EMBO J.* 17:1446–53
23. Li E, Bestor TH, Jaenisch R. 1992. *Cell* 69:915–26
24. Jackson-Grusby L, Beard C, Possemato R, Tudor M, Fambrough D, et al. 2001. *Nat. Genet.* 27:31–39
25. Kankel MW, Ramsey DE, Stokes TL, Flowers SK, Haag JR, et al. 2003. *Genetics* 163:1109–22
26. Paulsen M, Ferguson-Smith AC. 2001. *J. Pathol.* 195:97–110
27. Reik W, Walter J. 2001. *Nat. Genet.* 27: 255–56
28. Panning B, Jaenisch R. 1996. *Genes Dev.* 10:1991–2002
29. Martienssen R. 1998. *Curr. Opin. Genet. Dev.* 8:240–44
30. Xiao W, Gehring M, Choi Y, Margossian L, Pu H, et al. 2003. *Dev. Cell* 5:891–901
31. Kinoshita T, Miura A, Choi Y, Kinoshita Y, Cao X, et al. 2004. *Science* 303:521–23
32. Gong ZH, Morales-Ruiz T, Ariza RR, Roldan-Arjona T, David L, Zhu JK. 2002. *Cell* 111:803–14
33. Chen L, MacMillan AM, Chang W, Ezaz-Nikpay K, Lane WS, Verdine GL. 1991. *Biochemistry* 30:11018–25
34. Santi DV, Garrett CE, Barr PJ. 1983. *Cell* 33:9–10
35. Bestor TH, Verdine GL. 1994. *Curr. Opin. Cell Biol.* 6:380–89
36. Hsieh CL. 1999. *Mol. Cell Biol.* 19:8211–18
37. Horton JR, Ratner G, Banavali NK, Huang N, Choi Y, et al. 2004. *Nucleic Acids Res.* 32:3877–86
38. Cheng X, Roberts RJ. 2001. *Nucleic Acids Res.* 29:3784–95
39. O'Gara M, Klimasauskas S, Roberts RJ, Cheng X. 1996. *J. Mol. Biol.* 261:634–45
40. Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. *Genome Res.* 14:1188–90
41. Thompson JD, Higgins DG, Gibson TJ. 1994. *Nucleic Acids Res.* 22:4673–80
42. Posfai J, Bhagwat AS, Posfai G, Roberts RJ. 1989. *Nucleic Acids Res.* 17:2421–35
43. Lauster R, Trautner TA, Noyer-Weidner M. 1989. *J. Mol. Biol.* 206:305–12
44. Kumar S, Cheng X, Klimasauskas S, Mi S, Posfai J, et al. 1994. *Nucleic Acids Res.* 22:1–10
45. Yoder JA, Bestor TH. 1998. *Hum. Mol. Genet.* 7:279–84

46. Okano M, Xie S, Li E. 1998. *Nat. Genet.* 19:219–20
47. Cao X, Springer NM, Muszynski MG, Phillips RL, Kaeppler S, Jacobsen SE. 2000. *Proc. Natl. Acad. Sci. USA* 97: 4979–84
48. Wilke K, Rauhut E, Noyer-Weidner M, Lauster R, Pawlek B, et al. 1988. *EMBO J.* 7:2601–9
49. Trautner TA, Balganesh TS, Pawlek B. 1988. *Nucleic Acids Res.* 16:6649–58
50. Cheng X, Kumar S, Posfai J, Pflugrath JW, Roberts RJ. 1993. *Cell* 74:299–307
51. Reinisch KM, Chen L, Verdine GL, Lipscomb WN. 1995. *Cell* 82:143–53
52. Dong AP, Yoder JA, Zhang X, Zhou L, Bestor TH, Cheng XD. 2001. *Nucleic Acids Res.* 29:439–48
53. Bestor T, Laudano A, Mattaliano R, Ingram V. 1988. *J. Mol. Biol.* 203:971–83
54. Stein R, Gruenbaum Y, Pollack Y, Razin A, Cedar H. 1982. *Proc. Natl. Acad. Sci. USA* 79:61–65
55. Yoder JA, Soman NS, Verdine GL, Bestor TH. 1997. *J. Mol. Biol.* 270:385–95
56. Leonhardt H, Page AW, Weier HU, Bestor TH. 1992. *Cell* 71:865–73
57. Rountree MR, Bachman KE, Baylin SB. 2000. *Nat. Genet.* 25:269–77
58. Ding F, Chaillet JR. 2002. *Proc. Natl. Acad. Sci. USA* 99:14861–66
59. Doherty AS, Bartolomei MS, Schultz RM. 2002. *Dev. Biol.* 242:255–66
60. Ohsawa K, Imai Y, Ito D, Kohsaka S. 1996. *J. Neurochem.* 67:89–97
61. Mertineit C, Yoder JA, Taketo T, Laird DW, Trasler JM, Bestor TH. 1998. *Development* 125:889–97
62. Callebaut I, Courvalin JC, Mornon JP. 1999. *FEBS Lett.* 446:189–93
63. Goto K, Numata M, Komura JI, Ono T, Bestor TH, Kondo H. 1994. *Differentiation* 56:39–44
64. Fan G, Beard C, Chen RZ, Csankovszki G, Sun Y, et al. 2001. *J. Neurosci.* 21:788–97
65. Lei H, Oh SP, Okano M, Juttermann R, Goss KA, et al. 1996. *Development* 122:3195–205
66. Li E, Beard C, Jaenisch R. 1993. *Nature* 366:362–65
67. Chen RZ, Pettersson U, Beard C, Jackson-Grusby L, Jaenisch R. 1998. *Nature* 395:89–93
68. Guo G, Wang W, Bradley A. 2004. *Nature* 429:891–95
69. Laird PW, Jackson-Grusby L, Fazeli A, Dickinson SL, Jung WE, et al. 1995. *Cell* 81:197–205
70. Gaudet F, Hodgson JG, Eden A, Jackson-Grusby L, Dausman J, et al. 2003. *Science* 300:489–92
71. Biniszkiewicz D, Gribnau J, Ramsahoye B, Gaudet F, Eggan K, et al. 2002. *Mol. Cell. Biol.* 22:2124–35
72. El-Deiry WS, Nelkin BD, Celano P, Yen RW, Falco JP, et al. 1991. *Proc. Natl. Acad. Sci. USA* 88:3470–74
73. Eads CA, Danenberg KD, Kawakami K, Saltz LB, Danenberg PV, Laird PW. 1999. *Cancer Res.* 59:2302–6
74. Lee PJ, Washer LL, Law DJ, Boland CR, Horon IL, Feinberg AP. 1996. *Proc. Natl. Acad. Sci. USA* 93:10366–70
75. Baylin S, Bestor TH. 2002. *Cancer Cell* 1:299–305
76. Finnegan EJ, Dennis ES. 1993. *Nucleic Acids Res.* 21:2383–88
77. Genger RK, Kovac KA, Dennis ES, Peacock WJ, Finnegan EJ. 1999. *Plant Mol. Biol.* 41:269–78
78. Bernacchia G, Primo A, Giorgetti L, Pitto L, Cella R. 1998. *Plant J.* 13:317–29
79. Teerawanichpan P, Chandrasekharan MB, Jiang Y, Narangajavana J, Hall TC. 2004. *Planta* 218:337–49
80. Aufsatz W, Mette MF, Matzke AJ, Matzke M. 2004. *Plant Mol. Biol.* 54:793–804
81. Kakutani T, Munakata K, Richards EJ, Hirochika H. 1999. *Genetics* 151:831–38
82. Finnegan EJ, Peacock WJ, Dennis ES. 1996. *Proc. Natl. Acad. Sci. USA* 93: 8449–54

83. Ronemus MJ, Galbiati M, Ticknor C, Chen J, Dellaporta SL. 1996. *Science* 273: 654–57
84. Xie S, Wang Z, Okano M, Nogami M, Li Y, et al. 1999. *Gene* 236:87–95
85. Ramsahoye BH, Biniszkiewicz D, Lyko F, Clark V, Bird AP, Jaenisch R. 2000. *Proc. Natl. Acad. Sci. USA* 97:5237–42
86. Okano M, Bell DW, Haber DA, Li E. 1999. *Cell* 99:247–57
87. Kaneda M, Okano M, Hata K, Sado T, Tsujimoto N, et al. 2004. *Nature* 429:900–3
88. Chen TP, Ueda Y, Xie SP, Li E. 2002. *J. Biol. Chem.* 277:38746–54
89. Xu GL, Bestor TH, Bourc'his D, Hsieh CL, Tommerup N, et al. 1999. *Nature* 402:187–91
90. Kondo T, Bobek MP, Kuick R, Lamb B, Zhu XX, et al. 2000. *Hum. Mol. Genet.* 9:597–604
91. Lorincz MC, Schubeler D, Hutchinson SR, Dickerson DR, Groudine M. 2002. *Mol. Cell. Biol.* 22:7572–80
92. Chen TP, Ueda Y, Dodge JE, Wang ZJ, Li E. 2003. *Mol. Cell. Biol.* 23:5594–605
93. Aapola U, Shibuya K, Scott HS, Ollila J, Vihinen M, et al. 2000. *Genomics* 65:293–98
94. Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH. 2001. *Science* 294:2536–39
95. Chedin F, Lieber MR, Hsieh CL. 2002. *Proc. Natl. Acad. Sci. USA* 99:16916–21
96. Judson H, Hayward BE, Sheridan E, Bonthron DT. 2002. *Nature* 416:539–42
97. Hayward BE, De Vos M, Judson H, Hodge D, Huntriss J, et al. 2003. *BMC Genet.* 4:2
98. Moglabey YB, Kircheisen R, Seoud M, El Mogharbel N, Van den Veyver I, Slim R. 1999. *Hum. Mol. Genet.* 8:667–71
99. Patel CV, Gopinathan KP. 1987. *Anal. Biochem.* 164:164–69
100. Lyko F, Ramsahoye BH, Jaenisch R. 2000. *Nature* 408:538–40
101. Gowher H, Leismann O, Jeltsch A. 2000. *EMBO J.* 19:6918–23
102. Xu S, Xiao J, Posfai J, Maunus R, Benner J 2nd. 1997. *Nucleic Acids Res.* 25:3991–94
103. Cao X, Jacobsen SE. 2002. *Proc. Natl. Acad. Sci. USA* 99(Suppl. 4):16491–98
104. Cao X, Aufsatz W, Zilberman D, Mette MF, Huang MS, et al. 2003. *Curr. Biol.* 13:2212–17
105. Kouzminova E, Selker EU. 2001. *EMBO J.* 20:4309–23
106. Miao VP, Freitag M, Selker EU. 2000. *J. Mol. Biol.* 300:249–73
107. Tamaru H, Selker EU. 2003. *Mol. Cell. Biol.* 23:2379–94
108. Selker EU, Tountas NA, Cross SH, Margolin BS, Murphy JG, et al. 2003. *Nature* 422:893–97
109. Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, et al. 2003. *Nature* 422:859–68
110. Galagan JE, Selker EU. 2004. *Trends Genet.* 20:417–23
111. Shen JC, Rideout WM 3rd, Jones PA. 1992. *Cell* 71:1073–80
112. Zolan ME, Pukkila PJ. 1986. *Mol. Cell. Biol.* 6:195–200
113. Lindroth AM, Cao X, Jackson JP, Zilberman D, McCallum CM, et al. 2001. *Science* 292:2077–80
114. Henikoff S, Comai L. 1998. *Genetics* 149: 307–18
115. Jones DO, Cowell IG, Singh PB. 2000. *BioEssays* 22:124–37
116. Rose TM, Schultz ER, Henikoff JG, Pietrokovski S, McCallum CM, Henikoff S. 1998. *Nucleic Acids Res.* 26:1628–35
117. Finnegan EJ, Kovac KA. 2000. *Plant Mol. Biol.* 43:189–201
118. Bartee L, Malagnac F, Bender J. 2001. *Genes Dev.* 15:1753–58
119. Pinarbasi E, Elliott J, Hornby DP. 1996. *J. Mol. Biol.* 257:804–13
120. Okano M, Xie S, Li E. 1998. *Nucleic Acids Res.* 26:2536–40
121. Hermann A, Schmitt S, Jeltsch A. 2003. *J. Biol. Chem.* 278:31717–21
122. Tang LY, Reddy MN, Rasheva V, Lee TL, Lin MJ, et al. 2003. *J. Biol. Chem.* 278:33613–16

123. Van den Wyngaert I, Sprengel J, Kass SU, Luyten WH. 1998. *FEBS Lett.* 426:283–89
124. Lee YF, Tawfik DS, Griffiths AD. 2002. *Nucleic Acids Res.* 30:4937–44
125. Nan X, Campoy FJ, Bird A. 1997. *Cell* 88:471–81
126. Hendrich B, Bird A. 1998. *Mol. Cell. Biol.* 18:6538–47
127. Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, et al. 1998. *Nature* 393:386–89
128. Klose RJ, Bird AP. 2004. *J. Biol. Chem.* 279:46490–96
129. Balmer D, Arredondo J, Samaco RC, LaSalle JM. 2002. *Hum. Genet.* 110:545–52
130. Hendrich B, Guy J, Ramsahoye B, Wilson VA, Bird A. 2001. *Genes Dev.* 15:710–23
131. Guy J, Hendrich B, Holmes M, Martin JE, Bird A. 2001. *Nat. Genet.* 27:322–26
132. Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, Zoghbi HY. 1999. *Nat. Genet.* 23:185–88
133. Bestor TH, Tycko B. 1996. *Nat. Genet.* 12:363–67
134. Tamaru H, Selker EU. 2001. *Nature* 414:277–83
135. Freitag M, Hickey PC, Khlafallah TK, Read ND, Selker EU. 2004. *Mol. Cell* 13: 427–34
136. Jackson JP, Lindroth AM, Cao X, Jacobsen SE. 2002. *Nature* 416:556–60
137. Malagnac F, Bartee L, Bender J. 2002. *EMBO J.* 21:6842–52
138. Lehnertz B, Ueda Y, Derijck AA, Braunschweig U, Perez-Burgos L, et al. 2003. *Curr. Biol.* 13:1192–200
139. Vongs A, Kakutani T, Martienssen RA, Richards EJ. 1993. *Science* 260:1926–28
140. Jacobsen SE, Sakai H, Finnegan EJ, Cao X, Meyerowitz EM. 2000. *Curr. Biol.* 10:179–86
141. Geiman TM, Tessarollo L, Anver MR, Kopp JB, Ward JM, Muegge K. 2001. *Biochim. Biophys. Acta* 1526:211–20
142. Dennis K, Fan T, Geiman T, Yan QS, Muegge K. 2001. *Genes Dev.* 15:2940–44
143. Gibbons RJ, Picketts DJ, Villard L, Higgs DR. 1995. *Cell* 80:837–45
144. Gibbons RJ, Higgs DR. 2000. *Am. J. Med. Genet.* 97:204–12
145. Kanno T, Mette MF, Kreil DP, Aufsatz W, Matzke M, Matzke AJ. 2004. *Curr. Biol.* 14:801–5
146. Wassenegger M, Heimes S, Riedel L, Sanger HL. 1994. *Cell* 76:567–76
147. Cao X, Jacobsen SE. 2002. *Curr. Biol.* 12:1138–44
148. Chan SW, Zilberman D, Xie Z, Johansen LK, Carrington JC, Jacobsen SE. 2004. *Science* 303:1336
149. Wassenegger M. 2000. *Plant Mol. Biol.* 43:203–20
150. Papp I, Mette MF, Aufsatz W, Daxinger L, Schauer SE, et al. 2003. *Plant Physiol.* 132:1382–90
151. Xie Z, Johansen LK, Gustafson AM, Kasschau KD, Lellis AD, et al. 2004. *PLoS Biol.* 2:E104
152. Morris KV, Chan SW, Jacobsen SE, Looney DJ. 2004. *Science* 305:1289–92
153. Kawasaki H, Taira K. 2004. *Nature* 431:211–17

Annual Review of Biochemistry
Volume 74, 2005

CONTENTS