Active DNA Demethylation in Plants and Animals

H. ZHANG1 AND J.-K. ZHU1,2

1Department of Horticulture and Landscape Architecture, Purdue University, West Lafayette, Indiana 47907
2Shanghai Center for Plant Stress Biology and Shanghai Institute of Plant Physiology and Ecology, Shanghai Institutes of Biological Sciences, Chinese Academy of Sciences, Tianjin, China

Correspondence: jkzhu@purdue.edu

Active DNA demethylation regulates many vital biological processes, including early development and locus-specific gene expression in plants and animals. In Arabidopsis, bifunctional DNA glycosylases directly excise the 5-methylcytosine base and then cleave the DNA backbone at the abasic site. Recent evidence suggests that mammals utilize DNA glycosylases after 5-methylcytosine is oxidized and/or deaminated. In both cases, the resultant single-nucleotide gap is subsequently filled with an unmodified cytosine through the DNA base excision repair pathway. The enzymatic removal of 5-methylcytosine is tightly integrated with histone modifications and possibly noncoding RNAs. Future research will increase our understanding of the mechanisms and critical roles of active DNA demethylation in various cellular processes as well as inspire novel genetic and chemical therapies for epigenetic disorders.

In combination with histones and nonhistone architectural proteins, eukaryotic genomic DNA is hierarchically packed into complex chromatin structures. Chromatin can be adjusted from a compact status to an easy-to-access condition, and vice versa, through individual or combined epigenetic alterations in DNA methylation patterns and histone modifications as well as by actions of chromatin-remodeling factors and noncoding RNAs.

DNA methylation at the fifth position of cytosine is an important epigenetic modification. Via its effects on chromatin structures, DNA methylation regulates many critical biological processes such as gene imprinting, silencing of transposable elements (TEs), and X-chromosome inactivation. In plants, methylation occurs in all DNA sequence contexts including CG, CHG, and CHH (H represents either A, T, or G), with transposon-rich heterochromatic regions being the major targets (Zhang et al. 2006; Henderson and Jacobsen 2007). In mammals, cytosine in CG context throughout the genome is the predominant methylation target, with the exception of those densely clustered into CG islands near gene promoters (Ehrlich et al. 1982; Suzuki and Bird 2008; Cedar and Bergman 2009). Mammalian DNA can also be methylated in non-CG contexts, as has been observed in embryonic stem cells (Ram-sahoye et al. 2000; Lister et al. 2009). On a whole-genome scale, both Arabidopsis and human cytosine methylation shows a preference for nucleosome-bound DNA over flanking sequences (Chodavarapu et al. 2010), indicating an interplay between nucleosomes and DNA methyltransferases or demethylases in determining genomic DNA methylation patterns. Cytosine methylation can be faithfully inherited during DNA replication and is thus often considered as a stable epigenetic mark. Nevertheless, DNA methylation is subject to dynamic regulation of establishment, maintenance, and removal in response to developmental and environmental cues.

Establishment and maintenance of cytosine methylation are mediated by DNA methyltransferases (DNMTs) that use S-adenosyl-L-methionine (SAM) as the methyl donor. Mammalian DNMT3 family members and their plant homolog DOMAIN REARRANGED METHYLTRANSFERASE 2 (DRM2) catalyze de novo DNA methylation (Cao and Jacobsen 2002; Goll and Bestor 2005). The targeting of DRM2 for de novo methylation can be mediated by the RNA-directed DNA methylation (RdDM) pathway, which requires 24-nucleotide small interfering RNAs (siRNAs) (Law and Jacobsen 2010; Haag and Pikaard 2011; Zhang and Zhu 2011). Similarly, mammalian piRNAs, which are small RNAs (25 to 30 nucleotides) that mediate posttranscriptional silencing of transposons, are thought to pair with nascent transcripts and directly or indirectly recruit de novo methyltransferases (Law and Jacobsen 2010; Watanabe et al. 2011). In addition, evidence has been published in support of a model in which DNMT3A is activated or recruited to target loci via protein interaction with DNMT3L, a noncatalytic DNMT3 paralog that binds unmethylated histone 3 lysine 4 (H3K4) tails (Ooi et al. 2007; Cheng and Blumenthal 2008). Once established, mammalian DNA methylation is maintained by DNMT1, which restores the hemi-methylated DNA to a fully methylated status during DNA replication (Goll and Bestor 2005). In plants, DNA methylation in CG and CHG contexts is maintained by MET1 (DNA METHYLTRANSFERASE 1) and CMT3 (CHROMOMETHYLASE 3), respectively (Law and Jacobsen 2010). Cytosine methylation in the asymmetric CHH context cannot be maintained and thus must occur de novo during every cell cycle.
Preexisting DNA methylation can be lost as a consequence of passive or active demethylation processes. The former refers to the failure in maintaining DNA methylation owing to DNA methyltransferase dysfunction or shortage of methyl group supplies, whereas the latter is an outcome of enzymatic action resulting in the replacement of 5-methylcytosine (5-mC) with cytosine. This review focuses on active DNA demethylation in plants and animals. We start by reviewing active DNA demethylation processes and their functional significance. We then consider the molecular players involved in removing methylated cytosines as well as the established and emerging models of enzymatic demethylation. Finally, we highlight the interplay between DNA demethylases and other epigenetic regulators in carrying out a fully functional demethylation pathway.

PRUNING DNA METHYLATION PATTERNS

An extensive literature has documented strong correlations between various diseases and aberrant DNA methylation patterns. Active demethylation is critical for the pruning of DNA methylation patterns that is required for normal developmental reprogramming and proper transcriptional activities. In both plants and animals, enzyme-mediated removal of 5-mC has been observed in specific early developmental stages and in somatic cells as well.

Genome-Wide Demethylation during Development

In some mammals including humans, two waves of global DNA demethylation have been documented at specific times during early development. The first genome-wide demethylation occurs in paternal pronuclei in the zygote (Mayer et al. 2000; Oswald et al. 2000). Detection of 5-mC by immunostaining suggested a global loss of genomic DNA methylation that was observed before the completion of the first round of DNA replication. Inconsistence, such loss of DNA methylation was not abolished by treatment of zygotes with the replication inhibitor aphidicolin (Mayer et al. 2000). Closer examination revealed that such paternal demethylation bypassed certain genomic loci, including centromeric and pericentromeric heterochromatin, some transposons, and imprint regions (Olek and Walter 1997; Rougier et al. 1998; Santos et al. 2002; Lane et al. 2003). In addition, the maternal genome maintains its methylation patterns despite the ongoing paternal demethylation in the same zygote. Although how such a wave of active demethylation locates its target regions remains to be explored, some proteins that protect DNA methylation from demethylation at specific imprinted loci have been identified (Nakamura et al. 2007; Reese et al. 2007; Li et al. 2008). Several recent studies (Gü et al. 2011; Inoue and Zhang 2011; Iqbal et al. 2011) have suggested that passive demethylation probably accounts, at least in part, for this global demethylation, although this replication-dependent demethylation process is indeed initiated by enzymatic catalysis (discussed later). Nevertheless, active DNA demethylation does occur in the paternal pronucleus before the zygote enters into the two-cell stage, as indicated by the rapid reduction in methylation levels at the two TEs, Lin1 and Etn (Okada et al. 2010).

Zygotic paternal demethylation is followed by a replication-dependent passive DNA demethylation in the maternal genome before the formation of the blastocyst, at which stage the de novo DNA methyltransferases reestablish genomic DNA methylation in both parental origins (Wu and Zhang 2010). Most of these newly created DNA methylation patterns are erased during the second wave of global demethylation, which occurs in primordial germ cells (PGCs). Genome-wide loss of 5-mC in PGCs is considered as active demethylation, because this process undergoes several cell cycles in the presence of the maintenance methyltransferase DNMT1 (Morgan et al. 2005). In support of the view that this active process is replication independent, chronological investigation revealed that the profound epigenetic changes occur while the PGCs are in the G2 phase of the cell cycle (Hajkova et al. 2008). Unlike zygotic paternal DNA demethylation, imprinted methylation marks are targets for demethylation in the PGC genome, although some repetitive elements still showed incomplete demethylation (Hajkova et al. 2002).

Following the onset of DNA demethylation in PGCs, extensive chromatin rearrangement occurs (Hajkova et al. 2008), consistent with the notion that this global erasure of DNA methylation allows for epigenetic reprogramming toward totipotency. Similarly, epigenetic reprogramming involving active DNA demethylation is required for the differentiated somatic cell nucleus to become totipotent during nuclear transfer or cloning (Dean et al. 2001; Zhu 2009). Indeed, researchers have postulated that incomplete and aberrant epigenetic reprogramming is the cause of the high frequency of failed cloning and of developmental abnormalities in cloned animals (Morgan et al. 2005; Yang et al. 2007; Zhu 2009). In contrast to its significance in PGCs, the significance of paternal DNA demethylation in zygotes is less clear. Given that TEs and other repeat sequences account for half of the genome in mammals, the question has been raised as to whether global demethylation generally releases silencing of such types of sequences, even though some of them escape paternal demethylation (Wu and Zhang 2010). In fact, such a scenario has been observed in plants.

In Arabidopsis, evidence exists for genome-wide DNA demethylation concurrent with transposon reactivation during gametogenesis. Like most seed plants, Arabidopsis uses a double-fertilization strategy that depends on the multicellular nature of male and female gametophytes (Fig. 1). Mature pollen grains are tricellular, containing two sperm cells and one vegetative nucleus, whereas the female gametophyte contains one egg cell, one diploid central cell, two synergid cells, and three antipodal cells. The two sperm cells fertilize the egg cell and the central cell, thereby producing the embryo and the endosperm, respectively. Genome-wide hypomethylation was observed in the Arabidopsis endosperm (Gehring et al.
Mutation of DEMETER (DME), one of the bifunctional DNA glycosylases known for active DNA demethylation in Arabidopsis, partially prevents the global loss of DNA methylation (Hsieh et al. 2009). Gene expression of DME is detectable in the female gametophyte but quickly disappears after fertilization of the central cell with a sperm cell (Choi et al. 2002), suggesting that the endosperm hypomethylation may result from demethylation in the central cell nucleus. At the same time, MET1 is transcriptionally repressed during female gametogenesis (Jullien et al. 2008). Therefore, endosperm hypomethylation is likely to be a combined outcome of active and passive DNA demethylation.

In addition to the global hypomethylation patterns in the endosperm, increased CHH methylation relative to adult tissues at siRNA target loci in both the endosperm and the embryo has also been reported (Gehring et al. 2009). Such local hypermethylation was hypothesized to result from augmented RdDM, an enhancement owing to elevated siRNA production that is attributed to transcriptional activation of TEs and other siRNA target loci by the genome-wide demethylation in the endosperm. Although the cell-non-autonomous effects of small RNAs have been reported in other parts of the plant, including male gametocytes (Slotkin et al. 2009; Molnar et al. 2010), transport of siRNAs from the central cell (or endosperm) to the egg (or embryo) has yet to be demonstrated. Nevertheless, this enhanced-silencing hypothesis seems appealing. Arabidopsis endosperm is a transient nutrient reservoir that is lost by the time the seed germinates, whereas the embryo eventually develops into a mature plant. Therefore, genome-wide DNA demethylation in the endosperm and the resultant siRNA burst would not threaten transgenerational genome integrity; instead, it could serve as a “phishing” attempt with the purpose of identifying and silencing potentially mobile DNA elements in the embryo (Fig. 1C).

Demethylation-induced activation of TEs was also observed in the vegetative nucleus of Arabidopsis pollen (Slotkin et al. 2009). Consistent with DNA hypomethylation, the vegetative cell exhibits significant down-regulation of DDM1 (DECREASED DNA METHYLATION 1), a chromatin-remodeling factor necessary for maintaining DNA methylation (Slotkin et al. 2009). Dysfunction of DDM1 reduces global 5-mC levels by more than 70% in Arabidopsis (Vongs et al. 1993). Additionally, DME is expressed in the vegetative cell but not in the sperm cells (Scholt et al. 2011). Therefore, both passive and active demethylation seems to contribute to DNA hypomethylation in the vegetative nucleus. Whereas activation of TEs
seems to be restricted to the vegetative cell, 21-nucleotide small RNAs from the activated TEs accumulate in sperm cells (Slotkin et al. 2009). In support of the hypothesis that these 21-nucleotide small RNAs traveled from the vegetative cell to sperm cells, an artificial miRNA from the vegetative cell successfully silenced a reporter gene specifically expressed in sperm cells (Slotkin et al. 2009). In the sperm cells, DDM1 is expressed and TEs are silenced as in adult tissues (Martienssen et al. 2008; Slotkin et al. 2009). Therefore, these mobile 21-nucleotide small RNAs may reinforce TE silencing in sperm cells by posttranscriptional cleavage of the primary TE transcript. In addition, these small RNAs may be carried over during fertilization and thereby contribute to TE silencing in the zygote, for instance, by cleaving retrotransposon transcripts to prevent reverse transcription in the endosperm (Fig. 1C).

Locus-Specific Active DNA Demethylation

In addition to genome-wide regulation, active DNA demethylation also functions extensively in epigenetic regulation at discrete genetic loci. The Arabidopsis REPRESSOR OF SILENCING 1 (ROS1) is a well-established DNA demethylase that prunes DNA methylation for dynamic transcriptional regulation. Unlike DME, which appears to have little preference for specific DNA sequences during global demethylation (Huh et al. 2008; Gehring et al. 2009; Hsieh et al. 2009), ROS1 seems to preferentially counteract rigorous DNA methylation established by the RdDM pathway (Gong et al. 2002; He et al. 2009; Zhu 2009; Gao et al. 2010). The RdDM pathway effectively silences TEs, repetitive sequences, and foreign DNA (Matzke et al. 2007). However, this de novo methylation process may overexert its function and cause unwanted methylation for several reasons. First, RdDM establishes methylation in a target DNA sequence that matches the 24-nucleotide siRNA(s), but the methylation may spread to neighboring sequences. If an RdDM target locus is near a transcriptional unit that needs to be expressed, prevention of spreading of DNA methylation may become important. Second, RdDM blocks transposition of TEs and thereby protects genome integrity under normal conditions. However, TE transposition when an organism is exposed to harsh environmental stress may be evolutionarily beneficial (Ito et al. 2011a). In addition, transcripts from TE or other DNA repeats may carry regulatory functions and are needed under certain circumstances, as exemplified by the mammalian SINE (short interspersed nuclear element) transcripts that control mRNA production at multiple levels (Ponicsan et al. 2010). Third, siRNAs guide RdDM to their homologous loci. Unwanted silencing may thus occur if an otherwise transcriptionally active locus shows homology to newly generated siRNAs, such as those from foreign DNA. Therefore, active DNA demethylation is required for the proper control of epigenetic regulation (Fig. 2).

Transcription of RD29A (RESPONSIVE TO DESSICATION 29A) is highly induced by some abiotic stress conditions in wild-type Arabidopsis. In the transgenic line containing the transgene P_{RD29A}\_LUC (luciferase under the promoter of RD29A), RD29A is silenced along with the transgene when ROS1 is mutated (Gong et al. 2002). This transcriptional silencing in the ros1 mutant is attributable to DNA hypermethylation in the endogenous and transgenic RD29A promoters, which are located in different chromosomes (Gong et al. 2002). This is evidence that active DNA demethylation can protect genes from being incorrectly silenced. Bisulfite sequencing of the endogenous RD29A promoter in this transgenic plant revealed that methylated cytosines closer to the transcriptional start site are more prone to ROS1-mediated demethylation (Duan and Zhu, unpubl.), supporting the view that active DNA demethylation may restrict over-spreading of DNA methylation. Genetic screening for ros1 suppressors using this transgenic line has identified many known RdDM components (Zhang and Zhu 2011), strongly supporting an antagonistic relationship between active DNA demethylation and RdDM. Intriguingly, the ROS1 transcript level is significantly repressed in RdDM mutants and in met1 as well as by treatment with the DNA methyltransferase inhibitor 5-aza (Huettel et al. 2006; Mathieu et al. 2007; Martinez-Macias et al. 2012), indicating a feedback regulation of ROS1-dependent active DNA demethylation. Recently, an example of DNA demethylation-regulated TE transposition was reported. Active DNA demethylation by DNG701, a rice homolog of AtROS1, promotes transposition of the retrotransposon Tos17 in rice calli (La et al. 2011). Although ROS1 does not change global DNA methylation levels (Penterman et al. 2007), a recent characterization of the DNA methylome has identified nearly 5000 regions in the ros1 mutant that are hypermethylated relative to the wild type (Qian et al. 2012). Among these ROS1 targets, 44% are TEs, 26% are intergenic regions, and 30% are genic loci.

Mammalian cells also use active DNA demethylation for locus-specific gene regulation. One recent example is the transcriptional regulation of a human tumor suppressor gene, p15\^{ink4b} (Thillainadesan et al. 2012). In human breast cancer MCF7 cells, p15\^{ink4b} is silenced by promoter hypermethylation, which involves the oncogene ZNF217 and a corepressor complex CoREST that collectively recruit the DNA methyltransferase DNMT3A. Treatment with transforming growth factor β (TGF-β) repelled the DNMT3A/ZNF217/CoREST silencing complex and reduced DNA methylation in the p15\^{ink4b} promoter by recruiting a demethylation complex containing thymine DNA glycosylase (TDG), one of the mammalian enzymes involved in active DNA demethylation. On the other hand, knockdown of TDG or MBD4 prevents TGF-β-dependent demethylation of p15\^{ink4b} (Thillainadesan et al. 2012). Also in the MCF7 cells, the pS2/TFI1 gene promoter undergoes periodic methylation and active demethylation that coincides with cyclical estrogen receptor binding and gene expression of pS2 (Métévier et al. 2008). Studies in mammals have also revealed important roles of active DNA demethylation in neurogenesis and immune responses. The promoter of brain-derived neurotrophic factor (BDNF), an important gene for adult neural plasticity, is methylated under unstimulated con-
ditions. In response to electroconvulsive treatment, the BDNF promoter exhibits active DNA demethylation that involves Gadd45b (growth arrest and DNA-damage-inducible 45 b) (Ma et al. 2009). In T lymphocytes, active DNA demethylation occurs in a small region of the promoter-enhancer of the interleukin-2 gene shortly after stimulation and independently of DNA replication (Bruniquel and Schwartz 2003).

ENZYMATIC REMOVAL OF 5-METHYLCYTOSINES

The importance of active DNA demethylation in epigenetic regulation has motivated extensive research to identify DNA demethylases as well as the underlying mechanisms. Studies in Arabidopsis have established the role of the bifunctional DNA glycosylases in direct removal of 5-mC through a base excision repair (BER) pathway. Although orthologous DNA glycosylases have not been identified in mammals, research to date has suggested several mechanisms involving some DNA glycosylases that mediate mammalian active DNA demethylation.

Direct Excision of 5-mC

DNA demethylation can be achieved by a process that starts with direct excision of 5-mC (Fig. 3). In Arabidopsis, removal of 5-mC is catalyzed by a subfamily of bifunctional DNA glycosylases that include ROS1, DME, DML2 (DMETER like 2), and DML3 (Zhu 2009). Each of these four enzymes excises 5-mC from all sequence contexts in vitro, although discrepancies regarding substrate preference exist (Agius et al. 2006; Gehring et al. 2006; Morales-Ruiz et al. 2006; Penterman et al. 2007). In addition to catalyzing the hydrolysis of a glycosylic bond between the base and the deoxyribose, these DNA demethylases also possess apurinic/apyrimidinic (AP) lyase activity that nicks the DNA backbone at the abasic site (Zhu 2009). During demethylation, these bifunctional enzymes consecutively exert DNA glycosylase and AP lyase activities, generating a single-nucleotide gap that is eventually refilled with an unmethylated cytosine through the DNA damage repair pathway. On the basis of genetic interaction analyses of maternal effects on seed development, Arabidopsis DNA LIGASE 1 (LIG1) has been suggested to participate in DME-mediated demethylation (Andreuzza et al. 2010), whereas the DNA polymerase responsible for repair remains unknown. Recently, a zinc finger DNA 3' phosphoesterase (ZDP) was shown to function in ROS1-mediated DNA demethylation (Martinez-Macias et al. 2012). Biochemical results demonstrated that ZDP removes the 3' phosphate from ROS1-nicked DNA, facilitating subsequent repair. ZDP mutation consistently reduces DNA demethylation in vitro and causes DNA hypermethylation at hundreds of loci in vivo. In addition, ZDP interacts with ROS1 in vitro and its transcript level is down-regulated in RdDM pathway mutants. Although ZDP is undoubtedly involved in active DNA demethylation, there appears to be genetic redundancy for ZDP function, given that ZDP target loci represent fewer than one-tenth of ROS1 targets, as determined in a parallel study (Qian et al. 2012).
Whereas DME functions during gametogenesis, the other three *Arabidopsis* demethylases are expressed in vegetative tissues (Penterman et al. 2007; Zhu et al. 2007; Huh et al. 2008). Triple mutation of ROS1, DML2, and DML3 in the *rdd* mutant does not alter total DNA methylation levels compared with the wild-type background (Penterman et al. 2007), yet these enzymes clearly monitor a broad range of target loci, as revealed by the characterization of the *Arabidopsis* methylome (Penterman et al. 2007; Lister et al. 2008; Qian et al. 2012). The most recent results showed \( \sim 5000 \) hypermethylated regions in *ros1* mutants and \( \sim 9000 \) in *rdd* mutants (Qian et al. 2012). More than half of the hypermethylated regions in *rdd* do not appear to be targeted by ROS1 (Qian et al. 2012), indicating that DML2 and DML3 possess unique functions compared with ROS1, although partial redundancy of these three glycosylases has been observed (Penterman et al. 2007). Meanwhile, DME is distinctive for its well-known function in transcriptional activation of the maternal allele of imprinted genes (Huh et al. 2008).

In vertebrates including mammals, TDG and the methyl-CpG-binding protein 4 (MBD4) are reported to mediate active DNA demethylation by functioning as DNA glycosylases (Rai et al. 2008; Wu and Zhang 2010; Cortellino et al. 2011; Thillainadesan et al. 2012). However, the glycosylase activities of TDG and MBD4 against 5-mC are very weak compared with their glycosylase activities in removing thymine in T/G mismatch (Zhu et al. 2000a,b). Excision of 5-mC by TDG and MBD4 can be enhanced by the presence of other cellular components (Frémont et al. 1997; Jost et al. 1999; Rai et al. 2008), suggesting that these two enzymes may have strong 5-mC glycosylase activities in vivo. Alternatively, the role of TDG and MBD4 in active DNA demethylation may be attributed to their T/G mismatch repair activities, if 5-mC is deaminated into T. In addition, these DNA glycosylases may recognize other derivatives of 5-mC for the purpose of demethylation.

**Excision after Derivatization of 5-mC**

In contrast to plants that directly excise 5-mC, mammals apparently utilize DNA glycosylases after 5-mC is oxidized and/or deaminated (Fig. 3). In both cases, DNA base excision repair follows. Recent advances highlight several key enzyme families in regulating mammalian active demethylation, including the Tet (ten-eleven translocation) family that catalyzes iterative oxidation of 5-mC, the AID/APOBECs (activation-induced cytidine deaminase/apolipoprotein B mRNA-editing catalytic polypeptides) family that deaminates 5-mC and its oxidative derivative 5-hydroxymethylcytosine (5-hmC), and DNA glycosylases (especially TDG) that excise 5-mC derivatives and initiate subsequent DNA repair.

**Figure 3.** Active DNA demethylation pathways in plants and animals. *Arabidopsis* bifunctional DNA glycosylases, such as ROS1 and DME, can directly excise 5-mC and then cleave the DNA backbone at the abasic site, resulting in a gap that is filled with an unmodified cytosine through the BER pathway. In mammals where no ROS1 ortholog has been identified to date, 5-mC can undergo deamination and/or hydroxylation before being excised by DNA mismatch repair glycosylases, which triggers the BER pathway to complete the active demethylation process. See text for details.
In mammals, 5-mC can be hydroxylated to 5-hmC by the Tet family of dioxygenases, which include TET1, 2, and 3 in humans and mice (Bhutani et al. 2011; Williams et al. 2011). In mouse embryonic stem cells, 5-hmC shows strong enrichment within exons and near transcriptional start sites (Pastor et al. 2011), yet the conversion of 5-mC to 5-hmC alone does not activate genes silenced by DNA methylation (Bhutani et al. 2011; Williams et al. 2011). However, 5-hmC can be either deaminated by AID/APOBEC family proteins to yield 5-hydroxymethyluracil (5-hmU) (Cortellino et al. 2011) or further oxidized by Tet proteins to consecutively generate 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-caC) (He et al. 2011; Ito et al. 2011b). Mammalian DNA glycosylase TDG recognizes and efficiently excises 5-fC and 5-caC, while leaving 5hmC intact (He et al. 2011; Maiti and Drohat 2011; Zhang et al. 2012). In fact, TDG is the only known mammalian DNA glycosylase that cleaves 5-fC and 5-caC, because other tested DNA glycosylases including MBD4 and SMUG1 (single-strand-selective monofunctional U DNA glycosylase 1) showed no significant activity in removing 5-fC and 5-caC (He et al. 2011; Maiti and Drohat 2011). It is now clear that TDG-initiated DNA repair completes the process of active DNA demethylation (Cortellino et al. 2011; He et al. 2011). However, it remains to be determined whether a putative decarboxylase may remove the carboxyl group from 5-caC to regenerate unmodified cytosine (Wu and Zhang 2010; Ito et al. 2011b), because depletion of TDG can only partially restore the 5-caC level to that under normal conditions in mouse embryonic stem cells, in which Tet2 catalyzes 5-caC production with high efficiency (He et al. 2011; Ito et al. 2011b).

The importance of 5-mC oxidation is also indicated by recent findings that active demethylation in the zygotic paternal pronucleus is dependent on TET3-dependent conversion of 5-mC to 5-hmC (Gu et al. 2011; Iqbal et al. 2011). Deficiency of TET3 was also shown to impede demethylation of two embryonic stem cell marker genes and to cause developmental defects (Gu et al. 2011), although it is unknown whether 5-hmC in the paternal pronucleus undergoes deamination or further oxidation before being excised. Although new 5-hmC may be generated while the preexisting 5-hmC is being processed for excision, 5-hmC seems to persist throughout the early embryonic stages (Gu et al. 2011), suggesting that the global loss of 5-mC observed during the one-cell stage (Mayer et al. 2000) may not completely yield in replacement with unmodified cytosines as quickly as previously thought. Because several cycles of cell division occur after the one-cell embryonic stage, the presence of 5-hmC probably contributes to passive DNA demethylation, given that 5-hmC is not recognized by Dnmt1 for methylation maintenance during DNA replication (Valinluck and Sowers 2007). Indeed, a recent study reported that, in the two-cell embryonic stage, only one of the two sister chromatids of sperm-derived chromosomes is enriched for 5-hmC (Inoue and Zhang 2011). Moreover, blastomeres at the four- and eight-cell stages showed a gradual loss of 5-hmC, suggesting a replication-dependent demethylation during preimplantation development (Inoue and Zhang 2011).

Besides 5-fC and 5-caC, T and 5-hmU can be produced from deamination of 5-mC or 5-hmC, respectively (Fig. 3). The AID/APOBEC family of proteins catalyzes these deamination reactions and thereby initiates mismatch repair for active DNA demethylation (Cortellino et al. 2011; Guo et al. 2011). Mammalian AID is originally known for its function in generating antibody diversity in B lymphocytes (Muramatsu et al. 2000). AID belongs to the APOBEC family, which is so named because members other than AID were originally identified as RNA editors (Conticello et al. 2007). The ability of AID and APOBEC1 to mediate 5-mC deamination has been documented in vitro and in Escherichia coli (Morgan et al. 2004), and the first substantiation of AID’s role in mediating vertebrate active DNA demethylation came from zebrafish embryos (Rai et al. 2008). Overexpression of AID and MBD4 together in zebrafish embryos caused demethylation of the bulk genome and of the injected methylated DNA fragments. Gadd45a was then shown to enhance the cooperative demethylation by AID and MBD4. On the other hand, depletion of AID or Gadd45a attenuated demethylation of the injected methylated DNA, which was linearized and nonreplicating. It is therefore postulated that MBD4 excision of thymine follows AID deamination of 5-mC, a process that is promoted by Gadd45 (Rai et al. 2008). Indeed, increase in T/G mismatch was observed when AID was expressed with a noncatalytic MBD4 (Rai et al. 2008).

Similar to the proposed deamination → excision → repair demethylation process in zebrafish, a process involving AID, TDG, and Gadd45a was identified in a mouse study that focused on TDG (Cortellino et al. 2011). RNAi targeting TDG in embryonic cells compromised DNA demethylation in the promoter of an in vitro-methylated reporter gene. Contribution of passive demethylation was excluded because DNA demethylation occurred within 12 h of transfection and because the reporter plasmid contained no replication origin. TDG was also shown to control several development- and hormone-regulated genes through DNA demethylation of their corresponding promoters, which is consistent with the fact that knockout or catalytic dysfunction of TDG leads to embryonic lethality (Cortellino et al. 2011). In contrast, mice that lack AID are viable and fertile, although their PGCs exhibit genome-wide hypermethylation that is still low (~20%) relative to that in embryonic stem cells or somatic cells (70%–80%) (Popp et al. 2010), pointing to possible functional redundancy in the AID/APOBEC family. Alternatively, the lack of dependency on AID may indicate that the deamination step, on a whole-genome scale, can be largely bypassed through, for instance, iterative oxidation of 5-hmC followed by enzymatic removal of 5-fC and/or 5-caC (Fig. 3). In addition, because this study used bisulfite sequencing, which cannot distinguish between 5-mC and 5-hmC (Huang et al. 2010; Pastor et al. 2011), it remains to be explored to what extent 5-mC or 5-hmC contributes as AID deamination substrates to demethylation. Nevertheless, AID was demonstrated to
be necessary for promoter demethylation and concomitant gene induction of *OCT4* and *NANOG*, two human pluripotency genes whose promoters are rapidly demethylated during nuclear reprogramming (Bhutani et al. 2010). AID binds to methylated *OCT4* and *NANOG* promoters in fibroblasts, but not to the demethylated promoters in embryonic stem cells, supporting the direct involvement of AID in active DNA demethylation. In addition to TDG, SMUG1 also efficiently cleaves 5-hmU in double-stranded DNA, yet depletion of TDG alone is lethal in mice (Cortellino et al. 2011).

BER is a well-established pathway that begins with DNA glycosylase activity for nucleotide mismatch repair (Norbury and Hickson 2001). In both plants and mammals, BER can accomplish active DNA demethylation (Zhu 2009; Wu and Zhang 2010; Bhutani et al. 2011). In addition to DNA glycosylases, the proteins in the DNA damage-responsive Gadd45 family have been implicated in active DNA demethylation (reviewed in Niehrs and Schäfer 2012). Moreover, single-stranded DNA breaks and BER were shown to be mechanistically linked to global DNA demethylation in mouse PGCs and in the paternal pronucleus (Hajkova et al. 2010).

**INTERPLAY BETWEEN 5-mC REMOVAL AND OTHER EPIGENETIC MODIFICATIONS**

Active DNA demethylation is a coordinated process. In addition to proteins that directly mediate enzymatic replacement of 5-mC with unmodified cytosine, other molecular components can also be involved by recruiting the 5-mC removal machinery to its target loci, by establishing a chromatic environment that allows for excision of 5-mC, and conversely, by repelling the demethylation enzymes in order to protect the 5-mC mark.

**Recruitment for 5-mC Removal**

Although enzymes that excise 5-mC can be targeted to specific loci, the molecular mechanisms underlying the recruitment remain elusive. In theory, target specificity can originate from distinct characteristics of the chromatin or the individual chromatin components including DNA sequences and histones. Enzymes involved in 5-mC excision may carry functional domains that facilitate the recognition of their targets. DME and ROS1 each contain a characterized domain that mediates methylation-independent DNA binding (Mok et al. 2010; Ponferrada-Marin et al. 2010). Such domains may either facilitate target searching and/or the enzymatic removal of 5-mC. Similarly, the mammalian DNA glycosylase MBD4 contains a methy1 DNA-binding domain that may help MBD4 identify methylated substrates (Ballestar and Wolffe 2001). Nevertheless, other factors would also be required to precisely target active DNA demethylation to specific loci.

In *Arabidopsis*, an RNA-binding protein named ROS3 has been implicated in targeting ROS1 to certain loci (Zheng et al. 2008). ROS3 is required for demethylation at a number of loci, some of which overlap with ROS1 targets. ROS3 and ROS1 proteins colocalize in discrete foci dispersed throughout the nucleus in wild-type plants, whereas ROS1 localization was severely disrupted in the ros3 mutant, suggesting that the two proteins may function in a demethylation complex. ROS3 binds single-stranded small RNAs of specific sequences in vivo; therefore, it was proposed that ROS3 may use small RNAs to identify homologous target loci for demethylation (Zheng et al. 2008). However, the general sequence features of ROS3-bound small RNAs are currently obscure except that they appear to be rich in guanine (Zheng et al. 2008). It is also unknown whether ROS3 can bind long, noncoding RNAs (lncRNAs), such as those generated by RNA polymerase II (Pol II) or the plant-specific Pol V. In the RdDM pathway, Pol II- or Pol V-dependent, nascent lncRNAs can pair with their homologous 24-nucleotide siRNAs to recruit DNA methylases (Law and Jacobsen 2010; Zhang and Zhu 2011). Although Pol V mutation usually results in desilencing, it also causes hypermethylation at certain loci where DNA methylation patterns are independent of the siRNAs (Mosher et al. 2008), indicating a possibility that lncRNA might also facilitate the recruitment of DNA demethylases. In fact, the initial hint that RNA might contribute to DNA demethylation was the observation of RNA-dependent 5-mC glycosylase activity against hemimethylated DNA in chicken embryo extracts (Frémont et al. 1997). 5-mC DNA glycosylase activity from chicken embryos or mice G8 myoblasts was blocked by RNase treatment and was restored by the addition of synthetic RNA complementary to the methylated strand of the substrate DNA (Jost et al. 1999; Zhu et al. 2000a). Although these in vitro results may reflect direct assistance during enzymatic removal of 5-mC, a role of RNA in targeting active DNA demethylation is plausible.

DNA demethylation can be induced by hormones and other signals. Recruitment of the demethylases in these cases may be mediated by the signaling components or their downstream effectors. TDG can directly interact with the estrogen receptor α in activating transcription of estrogen-regulated genes (Chen et al. 2003). In addition, TDG interacts with SRC1, NCoA-3, CBP, and p300 (Tini et al. 2002; Lucey et al. 2005; Chiang et al. 2010), all of which are transcriptional coactivators that activate transcription at genes downstream from various signaling pathways (Goodman and Smolik 2000); whether these protein interactions result in active DNA demethylation in the gene promoters remains to be determined. Similarly, the retinoid receptor physically associates with human TDG (Um et al. 1998), which causes demethylation of a retinoic acid-responsive promoter linked to a reporter gene (Zhu et al. 2001). These results suggest that DNA demethylases may be guided to their target loci through physical interaction with certain transcriptional regulators.

**Interplay with Histone Modifications**

Removal of 5-mC may occur directly in the established chromatic environment. Alternatively, a call for demethylation may also lead to other modifications to the chromatin before or during the removal of 5-mC. Recently, a...
histone acetyltransferase named IDM1 (INCREASED DNA METHYLATION 1) was identified as a regulator of DNA demethylation in Arabidopsis (Qian et al. 2012). IDM1 recognizes chromatin containing CG methylation and low histone 3 lysine 4 (H3K4) and arginine 2 (H3R2) methylations that are characteristic of IDM1’s target loci and subsequently catalyzes H3K18 and H3K23 acetylation. Analysis of genetic interactions revealed that IDM1 and ROS1 function in the same pathway for DNA demethylation, although IDM1 controls only a portion of the ROS1 targets. It is still unclear how ROS1 is recruited to the modified chromatin, but the discovery of IDM1 underscores the contribution of histone modifications to active DNA demethylation. Similarly, maintenance of 5-mC in the CHG context can be impaired by IBM1 (INCREASE IN BONSAI METHYLATION 1), which is a histone demethylase that prevents H3K9 methylation (Saiz et al. 2008; Miura et al. 2009). Arabidopsis SSRP1 (STRUCTURE SPECIFIC RECOGNITION PROTEIN 1) is required for DNA demethylation and regulation of many paternal imprinted genes in the central cell (Ikeda et al. 2011). This non-histone chromosomal protein is a component of a heterodimeric histone chaperone complex, which has a conserved function in chromatin remodeling in Arabidopsis, Drosophila, and humans (Orphaneides et al. 1999; Shimojima et al. 2003; Duroux et al. 2004). Interestingly, hypomethylation by MET1 dysfunction in the ssrp1–1 met1–3 double mutant only partially activates the maternal allele of the imprinted FWA-GFP reporter in the central cell; in contrast, FWA-GFP is fully active in dme-1 met1–3. Therefore, mutation of SSRP1 apparently leads to additional repressive chromatin modification besides DNA hypermethylation.

In Arabidopsis and humans, nucleosome-bound DNA is more densely methylated than linker DNA sequences (Chodavarapu et al. 2010). Such a bias may be attributable to preferential methylation of nucleosomal DNA, more frequent demethylation in linker DNA, or both. Each member of the bifunctional DNA glycosylase family in Arabidopsis possesses an amino-terminal domain similar to the linker histone H1 (Zhu 2009; Law and Jacobsen 2010). Although it is unknown whether these H1-like domains are functional in binding to linker DNA, it might be interesting to determine whether nucleosome positioning contributes to genome-wide DNA demethylation patterning. In fact, researchers have speculated that the obvious differences in histone compositions and histone marks between the two parental pronuclei could underlie the paternal-specific demethylation in zygotes (Wu and Zhang 2010). In mouse embryonic stem cells, genome-wide profiling revealed a strong correlation between 5-hmC enrichment and the dual histone mark H3K27 trimethylation plus H3K4 trimethylation in gene promoter regions (Pastor et al. 2011).

**Preventing Demethylation**

During the genome-wide loss of 5-mC in the zygotic paternal pronucleus, DNA methylation in the maternal genome and in certain paternal loci remains unchanged. Although the docking of demethylation enzymes to DNA substrates can potentially be blocked by repressive histone modifications, several nonhistone proteins have been shown to protect DNA methylation regardless of the global demethylation. Mutation of stella/PGC7, which binds to both unmethylated and methylated DNA in vitro, leads to DNA demethylation in both paternal and maternal pronuclei (Nakamura et al. 2007). MBD3, which is an essential component of a chromatin-remodeling complex, also may block demethylation because demethylation in the promoter of the paternally imprinted H19 gene was observed in MBD3-depleted mouse embryos (Reese et al. 2007). Chromatin immunoprecipitation showed that MBD3 binds to the methylated region in the H19 promoter (Reese et al. 2007). However, it is unknown whether stella and MBD3 repel demethylation enzymes to protect 5-mC or whether they maintain DNA methylation imprint through interactions with methyltransferases as does the zinc finger protein ZFP57 (Li et al. 2008). Another way to prevent active DNA demethylation (in addition to counteracting the activities of DNA demethylases) is via transcriptional repression of DNA demethylases. For instance, DME is nondetectable in the germ cells but is expressed in the central cell (Choi et al. 2002; Schott et al. 2011), thereby allowing the establishment of gene imprinting in the endosperm.

**CONCLUDING REMARKS AND PERSPECTIVES**

In both plants and animals, genome-wide active DNA demethylation mediates developmental reprogramming, whereas locus-specific removal of 5-mC ensures proper gene regulation. As exemplified in Arabidopsis, identification of active DNA demethylation targets on a whole-genome scale has revealed thousands of such loci (Qian et al. 2012), indicating that dynamic editing of DNA methylation patterns is a common epigenetic feature. The biochemical mechanism for 5-mC cleavage is well established in plants, but convincing evidence of active DNA demethylation pathways in mammals has been obtained only recently. In particular, the discovery of Tet-dependent 5-hmC has triggered a flood of studies, resulting in significant advances in understanding of mammalian active DNA demethylation processes. Nevertheless, more efforts are needed to understand fully this critical epigenetic regulation. Gene families exist for enzymes known to be involved in 5-mC excision. Do all members of each family mediate active DNA demethylation? How much redundancy is there among their functions? How may they cooperate in shaping the cellular methylome? For instance, array-based data show that ROS1 transcript abundance increases during early embryo development in Arabidopsis seeds (Schmid et al. 2005). This might indicate a temporal coediting of the embryonic methylome by DME and ROS1. In fact, ROS1-dependent demethylation mediates postembryonic remodeling of the pericentromeric 5S rDNA chromatin (Douet et al. 2008). Another interesting question is how...
the demethylation enzymes might be regulated in accordance with the need of demethylation, as exemplified by the repression of Arabidopsis ROS1 transcript (Huettel et al. 2006; Mathieu et al. 2007; Martinez-Macias et al. 2012).

The whole spectrum of biological processes under regulation by active DNA demethylation has yet to be completely revealed, even though some enzymes have long been known as DNA demethylases. A recent study reported that telomeres were longer in ros1 mutants than in wild-type Arabidopsis, but the underlying mechanism is unknown (Liu et al. 2010). Mechanistic investigations have been focusing on 5-mC excision, and molecular players in the follow-up repairing process have received less attention. Research is needed to determine which BER proteins are involved in active DNA demethylation and how active DNA demethylation interacts with the canonical DNA damage repair pathways. In addition, although recent advances have provided novel insights into mammalian genome-wide demethylation, how such global demethylation is processed remains to be determined. At the same time, the mechanisms underlying locus-specific recruitment or protection from active DNA demethylation are still unclear. As a major means of epigenetic regulation, active DNA demethylation clearly interacts with other epigenetic processes, such as histone modifications, and thereby needs to be evaluated in the context of active DNA demethylation and how active DNA demethylation interacts with the follow-up repairing process.

REFERENCES


Guo J, Cao Y, Zhong C, Ming GL, Song H, Shi L. 2002. ROS1 transcript (Huettel et al., but the underlying mechanism remains to be determined. Future research on this topic will greatly increase our understanding of epigenetics and should also lead to the development of novel genetic and chemical therapies for epigenetic disorders.
ACTIVE DNA DEMETHYLATION


Active DNA Demethylation in Plants and Animals

H. Zhang and J.-K. Zhu

Access the most recent version at doi:10.1101/sqb.2012.77.014936

References

This article cites 112 articles, 41 of which can be accessed free at:
http://symposium.cshlp.org/content/77/161.full.html#ref-list-1

Article cited in:
http://symposium.cshlp.org/content/77/161.full.html#related-urls

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here

To subscribe to Cold Spring Harbor Symposia on Quantitative Biology go to:
http://symposium.cshlp.org/subscriptions

© 2012 Cold Spring Harbor Laboratory Press; all rights reserved