RNA-directed DNA methylation and demethylation in plants

CHINNUSAMY Viswanathan & ZHU Jian-Kang†

Institute for Integrative Genome Biology and Department of Botany and Plant Sciences, University of California, Riverside, California 92521, USA

RNA-directed DNA methylation (RdDM) is a nuclear process in which small interfering RNAs (siRNAs) direct the cytosine methylation of DNA sequences that are complementary to the siRNAs. In plants, double stranded-RNAs (dsRNAs) generated by RNA-dependent RNA polymerase 2 (RDR2) serve as precursors for Dicer-like 3 dependent biogenesis of 24-nt siRNAs. Plant specific RNA polymerase IV (Pol IV) is presumed to generate the initial RNA transcripts that are substrates for RDR2. siRNAs are loaded onto an argonaute4-containing RISC (RNA-induced silencing complex) that targets the de novo DNA methyltransferase DRM2 to RdDM target loci. Nascent RNA transcripts from the target loci are generated by another plant-specific RNA polymerase, Pol V, and these transcripts help recruit complementary siRNAs and the associated RdDM effector complex to the target loci in a transcription-coupled DNA methylation process. Small RNA binding proteins such as ROS3 may direct target-specific DNA demethylation by the ROS1 family of DNA demethylases. Chromatin remodeling enzymes and histone modifying enzymes also participate in DNA methylation and possibly demethylation. One of the well studied functions of RdDM is transposon silencing and genome stability. In addition, RdDM is important for paramutation, imprinting, gene regulation, and plant development. Locus-specific DNA methylation and demethylation, and transposon activation under abiotic stresses suggest that RdDM is also important in stress responses of plants. Further studies will help illuminate the functions of RdDM in the dynamic control of epigenomes during development and environmental stress responses.

1 Introduction

Histone variants, histone N-terminal tail modifications, and DNA methylation determine the chromatin structure. Histone N-tail acetylation, phosphorylation, and ubiquitination enhance gene expression, while dimethylation of H3K9 and H3K27, biotinylation, and sumoylation result in repressive chromatin that inhibits gene expression. DNA methylation also induces transcriptionally repressive chromatin structure. In mammals, cytosine is methylated primarily at CpG sequences, while in plants cytosine in all sequence contexts (CpG, CpHpG, and CpHpH, where H is adenine, cytosine, or thymine) can be methylated. In the Arabidopsis genome, about 24% of CG, 6.7% of CHG, and 1.7% of CHH are methylated\(^1\). In bacteria, DNA methylation protects the host genome from its own restriction enzymes that cleave invasive phage DNA, whereas DNA methylation in vertebrates and angiosperms has several functions such as transposon silencing, transcriptional gene silencing (TGS), imprinting, and paramutation\(^2\). Hersatile histone-N-tail-modifications and DNA methylation are epigenetic marks. In response to developmental and environmental cues, genome sequence information superimposed by epigenetic marks can give rise to several

Received March 5, 2009; accepted March 15, 2009
doi: 10.1007/s11427-009-0052-1

\(^1\)Corresponding author (email: jian-kang.zhu@ucr.edu)
epigenomes from a single genome\textsuperscript{[6]}. Small interfering RNAs (siRNAs) mediate both TGS and post-transcriptional gene silencing (PTGS). DNA methylation requires specific histone modifications (for example, H3K9 dimethylation) and siRNAs. siRNAs appear to control at least one-third of methylated loci in the \textit{Arabidopsis} genome\textsuperscript{[7]}. This review focuses on RNA-directed DNA methylation and its functions in plants.

2 RNA-directed DNA methylation (RdDM)

RdDM was discovered in viroid-infected tobacco plants in 1994. Transgenic tobacco expressing the potato spindle tuber viroid (PSTVd) exhibited viroid methylation after autonomous viroid RNA-RNA replication\textsuperscript{[8]}. TGS due to DNA methylation was found to be induced by sequence homology of the transgene promoter\textsuperscript{[9]}. Later, it was shown that heritable promoter methylation is directed by double-stranded RNA (dsRNA), which is cleaved into small RNAs of 23 nt in tobacco\textsuperscript{[10]}. This finding of the involvement of small RNAs in TGS was similar to that of siRNA-induced PTGS in plants\textsuperscript{[11]}. Hence, a link between the RNA interference (RNAi) machinery and induction of DNA methylation was suggested\textsuperscript{[10]}. Forward and reverse genetics analyses led to the identification of components involved in siRNA biogenesis and RdDM.

2.1 siRNA biogenesis

Flowering plants produce three classes of endogenous siRNAs, namely heterochromatic siRNAs (hc-siRNAs), \textit{trans}-acting siRNAs (tasiRNAs), and natural antisense transcripts-derived siRNAs (nat-siRNAs). hc-siRNAs are also called chromatin-associated siRNAs (casiRNAs) or repeat-associated siRNAs (rasiRNAs), tasiRNAs and nat-siRNAs mainly induce PTGS, while rasiRNAs trigger TGS. A prerequisite for siRNA biogenesis is the presence of dsRNA precursors. Transcripts from inverted repeats, pseudogenes, and natural antisense transcripts can form dsRNA by sequence complementarity, while transcripts from heterochromatic loci are converted into dsRNA by RNA-dependent RNA polymerases (RDRs). dsRNAs are cleaved by a Dicer-like family of ribonucleases III (DCL2, DCL3, and DCL4 in \textit{Arabidopsis}) to yield single-stranded \textasciitilde{} 20 to 24-nt siRNAs. These siRNAs are loaded onto argonaute (AGO) protein-containing RISC (RNA-induced silencing complex)\textsuperscript{[12-14]}.

Analysis of endogenous siRNA production from the SINE retroelement \textit{AtSN1} (\textit{Arabidopsis thaliana} short interspersed element 1) in rdr mutants of \textit{Arabidopsis} revealed that siRNA production requires DCL4. The \textit{rdr2-1} mutant exhibited a significant reduction in CHG and CHH methylation in \textit{AtSN1}. Further, endogenous rasiRNA production depends on DCL3 but not on DCL1 or DCL2. Thus, RDR2 and DCL3 are required for production of endogenous siRNAs involved in RdDM\textsuperscript{[15]}.

\textit{HUA ENHANCER-1} (HEN1), a nuclear methyltransferase, methylates the 2’-OH group of 3-terminal nucleotides in the small-RNA duplex, which is then separated into single-stranded siRNA. The \textit{hen1} mutation impairs the siRNA-dependent TGS pathway in \textit{Arabidopsis}\textsuperscript{[16]}.

hc-siRNAs are \textit{cis}-acting and induce RdDM-mediated heterochromatin formation\textsuperscript{[17]}. Transcription of highly methylated heterochromatin to generate the precursor RNA transcripts required for siRNA biogenesis remained a mystery until the two plant-specific DNA-dependent RNA polymerases, Pol IV (=Pol IVa) and Pol V (=Pol IVb), were identified in \textit{Arabidopsis}. These polymerases consist of 12 subunits that are paralogous or identical to the 12 subunits of Pol II. NRPD1 (=NRPD1a) and NRPE1 (=NRPDb) are the largest subunits of Pol IV and Pol V, respectively. Both of these RNA polymerases share the same second-largest subunit NRPD2/NRPE2\textsuperscript{[18,19]}. A forward genetic screen for mutants of transgene silencing led to the identification of the silencing defective 4 \textit{(sde4)} mutant with reduction in siRNA production and DNA methylation at the \textit{AtSN1} locus. \textit{SDE4/NRPD1A} encodes the largest subunit of Pol IV\textsuperscript{[20]}. With the use of a two-component \textit{a’} promoter::GFP reporter silencing system in seeds, mutants defective in RNA-directed DNA methylation (\textit{dmd} mutants) were identified in \textit{Arabidopsis}\textsuperscript{[21]}. The \textit{dmd2} (\textit{nrpd2/nrpe2}) mutation reduced siRNA production as well as DNA methylation at \textit{AtSN1} and \textit{SS} loci\textsuperscript{[22]}. Further, mutation in \textit{NRPD2} also impaired siRNA production. Thus, NRPD1a and NRPD2 were found to be necessary for siRNA production and the silencing of certain transposons and repetitive DNA\textsuperscript{[23]}. \textit{nrpd1a nrpd1b} double-mutant plants showed loss of cytosine DNA methylation at pericentromeric \textit{SS}-gene clusters and \textit{AtSN1}, and impaired heterochromatin formation\textsuperscript{[24]}. Further analysis revealed that both NRPD1a and NRPD1b are required for efficient silencing of transposons and other highly repetitive sequences, while silencing of low-repetitive
DNA depends only on NRPD1a\textsuperscript{[25]}. Pol IV may transcribe methylated heterochromatin, transposons, and repetitive DNA to produce the precursor transcripts, which are converted into dsRNA by RDR2\textsuperscript{[26,27]}

A genetic screen for suppressors of repressor of silencing 1 (ros1, a DNA demethylase mutant) led to the identification of NRPD1, NRPE1, and a shared fourth subunit of Pol IV and Pol V in RdDM. The rdm2 (RNA-directed DNA methylation2) mutant showed substantial reduction in both siRNA accumulation and DNA methylation in several Pol IV- and Pol V-dependent loci (5S rDNA, MEA-ISR, AtSN1, AtGP1, and AtMU1). RDM2 encodes a protein similar to the RPB4 subunit of Pol II. Coimmunoprecipitation and coimmunolocalization studies revealed that RDM2 is associated with Pol II. Coimmunoprecipitation and coimmunolocalization studies revealed that RDM2 is associated with Pol II. Immunoprecipitated NRPE1-containing high-molecular-weight complex (~500 kDa) consisted of eight subunits and 10 associated proteins in cauliflower (Brassica oleracea). Functions of two of these protein homologs were examined in Arabidopsis. T-DNA knockout mutants of the RPB5b (NRPE5A) subunit and KTF1 (a KOW-domain transcription factor) exhibited reduction in both NRPE1-dependent siRNAs and DNA methylation at the AtSN1 locus but only a reduction in the DNA methylation level at the 5S-rDNA locus and two LTR loci. KTF1 is similar to NRPE1 in that it also contains WG/GW motifs, a binding site for AGO proteins\textsuperscript{[29]}. Cloning and sequencing of siRNAs from wild-type and RNA Pol IV-mutant Arabidopsis plants showed that Pol IV is required for the production of >90% of siRNAs\textsuperscript{[30]}. Thus, Pol IV is necessary for production of siRNAs from heterochromatin, transposons, and repetitive sequences.

An Arabidopsis SNF2 domain-containing protein (CLASSY1) is also required for Pol IV- and RDR2-dependent biogenesis of dsRNAs, which are substrates for DCL3\textsuperscript{[31]}. siRNA production is reduced drastically in nrpd1a and rdr2 mutants but is reduced only slightly in nrpd1b and ago4 mutants. Immunolocalization studies revealed that Pol IVa and Defective in RNA-directed DNA methylation 1 (DRD1, a SWI2/SNF2-like putative chromatin-remodeling protein) are localized external to the nucleolus and colocalize with endogenous repeat loci, while RDR2, DCL3, AGO4, and NRPD1b colocalize with siRNAs within the nucleolus. Maintenance of the heterochromatic state involves locus-specific Pol IV transcription followed by the synthesis of RDR2-dependent dsRNAs. dsRNAs are cleaved by DCL3 to produce 24-nt hc-siRNAs\textsuperscript{[27]}. Site-directed mutagenesis revealed that the magnesium-binding metal A and metal B motifs of Pol IV and V are essential for siRNA production, RdDM, retrotransposon silencing, and the specific nuclear localization of both polymerases\textsuperscript{[32]}

### 2.2 DNA methyltransferases

Cytosine methyltransferases catalyze the transfer of an activated methyl group from S-adenosyl methionine to the 5’ carbon of the cytosine ring. New cytosine methylation is catalyzed by de novo DNA methyltransferases, while maintenance DNA methyltransferases propagate the symmetric methylation marks after each cycle of DNA replication. Maintenance methyltransferases recognize the methylcytosine on the parental DNA strand and catalyze the methylation of cytosine on the daughter DNA strand. The Arabidopsis genome encodes 10 DNA methyltransferase-like proteins that can be grouped into the following three families: methyltransferase (MET, ortholog of mammalian DNMT1), chromomethylase (CMT, chromodomain methyltransferase, plant specific), and DOMAINS REARRANGED METHYLTRANSFERASE (DRM, a relative of mammalian Dnmt3). De novo methyltransferases DRM1 and DRM2 catalyze new cytosine methylation. Propagation of symmetric CG and CHG methylation through cell division is catalyzed by the MET1 and the plantspecific enzyme CMT3 (Chromomethylase 3), respectively\textsuperscript{[14]}. Non-CG methylation is maintained by DRM2 and CMT3\textsuperscript{[33]}

RdDM requires both de novo methyltransferases and maintenance methyltransferases in plants\textsuperscript{[17,34]}. A two-component genetic screen consisting of a 35S-GFP transgene and modified RNA virus-driven expression of the partial 35S-promoter sequence showed RdDM of 35S promoter and thus GFP transgene expression is silenced. In this system, suppression of MET1 did not affect RdDM. However, in the absence of the RNA trigger, maintenance of methylation and TGS in the subsequent generations required MET1\textsuperscript{[35]}. Another two-component transgene system, which consisted of the 35S promoter driven expression of an inverted repeat of the NOPALINE SYNTHASE promoter (NOSpro) and an unlinked target transgene NOSpro::NPTII, showed
heavy methylation of CG, CHG, and CHH sites on NOSpro. Studies on these plants crossed with drm and cmt3 mutants revealed that maintenance of pre-established RNA-directed CpG methylation does not depend on DRM2 and CMT3. DRM2 is required, however, for the initial establishment of RdDM in CG, CHG, and CHH sites. Non-CG methylation at the MEA-ISR (MEDEA INTERSTITIAL SUBTELOMERIC REPEATS) locus of Arabidopsis is maintained by DRM2[36,37]. Thus, DRM2 is required for both establishment of DNA methylation and maintenance of CHG- and CHH-DNA methylation[37]. Epigenome sequencing in wild type and mutants of DNA methyltransferases revealed that the functional boundaries of DNA methyltransferases overlap, because DRM1 and DRM2 are also required for maintenance of symmetric methylation, and MET1 and CMT3 may also participate in de novo methylation[7].

2.3 DNA methylation complex

Small RNA-binding AGO proteins are the first identified component of the RdDM complex. AGO proteins contain PAZ, MID, and PIWI domains. The PAZ domain recognizes the 3' end of small RNAs, the MID domain binds to the 5' phosphate of small RNAs, and the PIWI domain has an RNaseH-like activity that cleaves single-stranded target RNA[38]. The Arabidopsis genome encodes 10 AGO proteins while the rice genome encodes 17. The role of AGO4 in RdDM was evident from the impairment in siRNA accumulation and the loss of both non-CG DNA methylation and H3K9 methylation at the SUPERMAN and AtSN1 loci in ago4 mutants[39]. siRNAs are loaded onto RNA-induced gene silencing complex consisting of AGO4 and other effector proteins for RdDM[38,40,41]. A mutation in the Asp-Asp-His catalytic motif of AGO4 abolished only its catalytic activity without affecting its siRNA-binding activity. Experiments with this mutant form of AGO4 revealed that, besides being required for DNA methylation, AGO4 catalytic activity is necessary for the generation of secondary siRNAs[42]. Forward genetic screens for suppressors of ros1 revealed that both AGO4[16] and AGO6[43] are required for RdDM. AGO6 is required for specific heterochromatin-related siRNAs biogenesis, DNA methylation, and TGS in Arabidopsis. The roles of AGO6 were partially redundant with those of AGO4[43]. AGO4 functions in both siRNA-dependent and -independent pathways of TGS[46].

Some of the proteins important for histone modifications and chromatin remodeling are also involved in RdDM and/or maintenance of CG methylation. The histone deacetylase HDA6[45] and histone H3K9 methyltransferase KRYPTONITE/SUVH4[45] are required for maintenance of CG methylation established by RdDM. HDA6 is necessary for both RdDM- and siRNA-independent DNA methylation[16]. KYP/SUVH4-mediated methylation of H3K9 provides a binding site for CMT3 through its chromodomain[46]. In addition to SUVH4, other SRA-SET proteins (SUVH2, SUVH5, SUVH6, and SUVH9) also contribute to methylation of H3K9 and maintenance of DNA methylation[45,47,48]. SUVH2 and SUVH9 are also essential for both DRM2-mediated de novo DNA methylation and maintenance DNA methylation. SUVH2 and SUVH9 do not participate in siRNA biogenesis but are involved in the latter methylation stage of RdDM, because suvh2 and suvh9 mutations do not eliminate siRNAs in Arabidopsis. SUVH2 and SUVH9 preferentially bind to methylated CG sites and asymmetric sites, respectively. This suggests that the SRA-domain proteins help target or retain DRM2 and CMT3 at the target loci[49]. In addition to deacetylation and methylation, histone de-ubiquitination also contributes to RdDM-dependent TGS. The Arabidopsis sup32-1 mutation, which partially suppressed the TGS caused by ros1, caused a reduction in CHpG and CHpHp methylation at RdDM target loci. The SUP32 encodes an ubiquitin-specific protease (UBP26), which catalyzes deubiquitination of H2B. De-ubiquitination of H2B is necessary for H3K9 dimethylation[49].

Two SWI2/SNF2-like putative chromatin-remodeling proteins are important for DNA methylation. The Arabidopsis ddm1 (decrease dna methylation1) mutant exhibited >70% reduction in DNA methylation. DDM1 encodes a putative SWI2/SNF2 class chromatin-remodeling ATPase protein with a vital role in the maintenance of CG and non-CG methylation[50]. ddm1 mutants show a loss of siRNA accumulation and DNA methylation in transposable elements[51]. A screen for mutants defective in RdDM identified the DRD1 gene encoding another SWI2/SNF2-like putative chromatin-remodeling protein. drd1 mutants had a complete loss of siRNA-induced non-CG methylation at a target promoter, while methylation of centromeric and rDNA repeats was unaffected. DRD1 may mediate a sequence homology search to locate siRNA-targeted DNA sequences for methylation[51]. Recently, DRD1 was also identified from the ros1 suppressor screen and was found to function spe-
cifically in siRNA-dependent TGS\[16\]. A DNA helicase was recently implicated to function in TGS. In rice, Os-RecQ1, a homolog of Neurospora QUELLING DEFECTIVE 3, is involved in RNAi-induced transgene silencing, and is probably required for unwinding inverted-repeat double-stranded DNA\[52\].

A viral enhancer-based two-component transgene system was used to identify the dms3 mutant defective in transgene silencing. DMS3 encodes a protein similar to the hinge-domain region of structural-maintenance-of-chromosomes (SMC) proteins\[53\]. DMS3 was proposed to bind and stabilize siRNA-DNA and/or siRNA-RNA complexes and to enable the spread of DNA methylation in cis beyond the region targeted by the primary siRNAs\[51,54\].

Because de novo DNA methylases have little sequence specificity beyond their target nucleotides (CG, CHG, CHH), targeting of de novo DNA methylases to specific sequences depends on small RNAs and specific histone modifications. Small RNA-binding proteins can directly or indirectly through adapter proteins recruit de novo DNA methylases to the target sequence. Pol V is one of the proteins implicated in the methylation step of RdDM, because drd3 (nrpe1) mutants showed loss of non-CG methylation at AtSN1 and 5S rDNA loci without affecting siRNA production\[22\]. Several studies showed the requirement for Pol V mainly in the methylation step of RdDM\[18,25,27,30\], although in some cases Pol V was implicated in both siRNA production and DNA methylation\[25,55\].

The mechanism of action of Pol V in the methylation step was revealed only recently. The C-terminal domain (CTD) of NRPE1 contains a conserved WG/GW-motif through which it interacts with AGO4\[56\]. NRPE1 and AGO4 colocalize in Cajol bodies\[26\] and AB-bodies along with DRM2\[57\]. Pol V transcribes noncoding regions in both heterochromatin and euchromatin. The Pol V-dependent noncoding transcripts are ~ 200 nt in size, with 5' triphosphates or 7-methylguanosine caps but without polyA tails. Pol V transcription requires the chromatin-remodeling protein DRD1. Because both Pol V-generated non-coding RNAs and siRNA production were required for RdDM, it was proposed that siRNAs in the AGO4-RISC complex interact with nascent Pol V transcripts, thereby recruiting chromatin-modifying complexes, including DRM2 and histone-modifying enzymes, to the target loci. The AGO4-RISC complex containing siRNAs that guide DRM2 for cytosine methylation of the DNA sequence complementary to the siRNAs (Figure 1)\[58\]. Once primary RdDM is established, unidirectional spreading of methylation from an upstream enhancer element into downstream sequences was in some cases induced by secondary siRNAs in Arabidopsis. Primary RdDM attracts Pol IV, RDR2, and DCL3. This siRNA machinery generates secondary siRNAs, which in turn direct methylation in the downstream region\[59\].

3 DNA demethylation

Studies on the ros1 mutant of Arabidopsis unraveled the role of DNA glycosylases in DNA demethylation. ros1 mutants showed locus-specific DNA hypermethylation and enhanced TGS. ROS1 encodes a bifunctional DNA glycosylase/lyase that actively removes DNA methylation by a base excision repair mechanism in Arabidopsis\[60\]. Transgenic plants overexpressing ROS1 showed erasure of cytosine methylation in CpG, CpHpG, and CpHpHp sequences, and an increased level of target gene expression\[61\]. Arabidopsis DEMETER (DME), a protein similar to ROS1, is necessary for endosperm gene imprinting and seed viability in Arabidopsis\[62\]. Further, DNA methylation profiling of ros1, dml2 (DEMETER-LIKE-2), and dml3 mutants and wild-type Arabidopsis using genome-tiling microarray analysis and whole genome bisulfite sequencing and other approaches revealed that these demethylases target hundreds of loci including many transposons\[7,63,64\]. During germination, decondensation of 5S rDNA heterochromatin is accompanied by ROS1-mediated active demethylation of 5S rDNA. Following this decondensation, Pol IV-dependent recondensation occurs at 5S rDNA chromatin to attain the mature nuclei stage\[65\]. Thus, DNA demethylation regulates epigenome dynamics. Recently, it was found that targeting of the demethylases to specific loci requires the small RNA-binding protein ROS3. ROS3, an RNA-recognition motif (RRM) containing protein, may enable small RNAs to direct demethylation at specific loci. ROS3 colocalizes with ROS1 in discrete loci in the nucleus. Small RNA-bound ROS3 may direct ROS1 for sequence-specific demethylation (Figure 1)\[66\].

4 Functions of RdDM

Heterochromatins are rich in transposable elements and repetitive sequences. siRNAs produced from hetero-
Figure 1  RNA-directed DNA methylation and demethylation in plants. Heterochromatin region is decondensed by the chromatin remodeling proteins CLY, DRD1 and DDM1. Pol IV may transcribe methylated heterochromatin at transposons and other repetitive DNA to produce the precursor transcripts for siRNAs. During de novo DNA methylation, aberrant transcripts from the locus may serve as template for Pol IV amplification. These precursor transcripts are converted into dsRNA by RDR2. DCL3 cleaves the dsRNAs to produce 24-nt hc-siRNAs. These siRNAs are loaded onto AGO4-containing RISC. siRNAs in the AGO4-RISC complex interact with nascent Pol V transcripts, thereby recruiting chromatin-modifying complexes, including DRM2 and histone-modifying enzymes, to the target loci. The AGO4-RISC complex containing siRNAs guide DRM2 for cytosine methylation of the DNA sequence complementary to the siRNAs. Chromatin remodeling proteins, histone modifying proteins and DNA methyltransferases are all important for maintaining DNA methylation. Small RNAs bound to ROS3 may direct ROS1 for sequence-specific DNA demethylation (H3K9 indicates repressive histone methylation marks on the nucleosome; "c", cytosine methylation; AGO4, Argonaute 4; AGO6, Argonaute 6; CLY, CLASSY1, a SNF2 domain-containing chromatin remodeling protein; CMT3, Chromomethylase 3; CTD, C-terminal domain of the largest subunit of Pol V; DCL3, DICER-Like 3, a type III ribonuclease; DDM1, DECREASED DNA METHYLATION1, a SWI2/SNF2 class chromatin-remodeling ATPase; DME, DEMETER, a DNA demethylase; DRD1, DEFECTIVE IN RNA-DIRECTED DNA METHYLATION1, a SWI2/SNF2-like putative chromatin-remodeling protein; DRM2, DOMAINS REARRANGED METHYLTRANSFERASE2, a de novo DNA methyltransferase; HDA6, HISTONE DEACETYLASE 6; HEN1, HUA ENHANCER-1, a small RNA methyltransferase; KYP, KRYPTONITE/SUVH4, a histone H3K9 methyltransferase; MET1, methyltransferase 1; Pol IV, plant-specific RNA polymerase IV; Pol V, plant-specific RNA polymerase V; RDR2, RNA-dependent RNA polymerase 2; RISC, RNA-induced silencing complex; ROS1, REPRESSOR OF SILENCING 1, a DNA demethylase; ROS3, REPRESSOR OF SILENCING 3, a small RNA-binding protein).
chromatic regions guide the DDM1 chromatin-remodeling ATPase, which mediates heterochromatin formation\cite{51}. Genome-wide DNA-methylation mapping by microarray\cite{67,68} and shotgun bisulfite-sequencing\cite{1,7} of Arabidopsis revealed that DNA methylation is most abundant in the heterochromatic regions. Comparison of genome-wide methylation patterns and small RNAs in Arabidopsis indicated that about 37% of the methylated loci were associated with siRNA clusters. Loss of DNA methylation in met1 and ddm1 mutants resulted in large-scale transcriptional reactivation of transposons and pseudogenes\cite{67,69,70}. Thus, transposon silencing is one of the most important functions of DNA methylation\cite{2,3,51}. Nucleolar dominance is a phenomenon in which rRNA genes from one progenitor of interspecific hybrids of plants are expressed, while the rRNA genes from the other parent are clustered at nucleolus organizer regions (NORs) and silenced. DNA methylation and repressive histone modifications silence rRNA genes. Genetic analysis showed that 24-nt siRNAs produced by the RDR2- and DCL3-dependent pathway guides DRM2-catalyzed rDNA methylation. Methylcytosine-binding-domain proteins MBD6 and MBD10 bind to methylated cytosines and recruit histone deacetylases to deacetylate histones. Thus DNA methylation and histone deacetylation lead to silencing of rDNA loci\cite{71}. Besides transposon silencing and nucleolar dominance, RdDM is also involved in gene regulation, stress responses and plant development\cite{2,3,51}.

4.1 Gene expression and epialleles

Heterochromatin is found near centromeres and telomeres but are often dispersed in heterochromatic knobs. Heterochromatic knobs consist of retrotransposons and tandem repeats, with very few transcribed genes. Transposons/repeats-derived siRNAs can mediate RdDM of nearby genes and thus affect gene expression. Natural variation in flowering time in Arabidopsis thaliana was found to be due to an epiallele of FLC generated by RdDM. The Landsberg erecta (Ler) ecotype contains an MPF (Methylated region near Promoter of FLC) element from which siRNAs are generated. These siRNAs direct methylation and heterochromatization of MPF and thus lead to repression of FLC. Further, RdDM-induced epigenetic differences at several loci were found between the Ler and Col ecotypes of Arabidopsis\cite{72}. In ddr1 and pol I vb mutants, solo LTR (long terminal repeat) retrotransposon in the euchromatin region is derepressed because of a decrease in cytosine methylation. These intergenic retrotransposon LTRs were found to generate siRNAs and induce epigenetic repression of adjacent genes\cite{73}.

In petunia, pMADS3, an ortholog of Arabidopsis AGAMOUS, is specifically expressed in the stamen and carpels of developing flowers. Transgenic plants ectopically expressing a part of the pMADS3 genomic sequence showed ectopic expression and upregulation of endogenous pMADS3 due to CG methylation in a negative cis-element in intron 2 of pMADS3. This methylation was established by siRNAs produced from the inverted repeat region of intron 2 of the transgene. Once established, CG methylation in this epiallele is maintained in subsequent generations even in the absence of siRNA triggers. Thus, RdDM can also induce transcriptional activation\cite{74}.

Paramutation is a phenomenon in which interaction between alleles of genes leads to heritable changes in gene expression. In an example from maize, the booster1 (bl) locus encodes a bHLH transcription factor, which activates genes involved in anthocyanin biosynthesis in maize. The bl locus has two alleles, namely B-I (paramutable allele, high expression) and B’ (paramutagenic allele, weak expression). B-I and B’ alleles do not differ in DNA sequence but do differ in DNA methylation pattern. When B-I and B’ plants are crossed, the B’ allele silences the B-I allele, and B-I acquires paramutagenic ability. A mediator of paramutation, MOP1, which encodes an RDR is necessary for paramutation and transposon silencing\cite{75}. Transcripts from bidirectional transcription of tandem repeats located 100 kb upstream of the transcription start site of bl1 appears to be processed by DCL and MOP1 to produce siRNAs, which in turn mediate paramutation\cite{76}. Paramutation in the maize purple plant 1 (p1l) locus changes its p1l-Rhoades (P1l-Rh) allele to P1l-mahogany (P1l’). The p1l-Rh allele shows high expression of the p1l MYB-like transcription factor and strong pigmentation in aerial parts, while P1l’ shows low expression and reduction in pigmentation. A CACTA-like DNA transposon related to the doppia subfamily is present immediately 5’ to the p1l-Rh coding region. siRNAs (24-nt) derived from the doppia direct RdDM at the p1l-Rh locus. MOP1 (=RMR1, Required to Maintain Repression 1) is required for accumulation of siRNAs and paramutation. In addition, rmr6 mutants also show hypomethylation of the 5’ region
of pl1-Rh and enhanced expression of pl1-Rh. The RMR6 (ZmRPD1) gene encodes the largest subunit of Pol IV. Nuclear run-on transcription assays found no detectable transcription by Pol IV. The rmr6 mutant has about 82% reduction in 24-nt siRNAs in maize cobs, suggesting that maize Pol IV is involved in the biogenesis of siRNAs. These evidences support that RdDM at the pl1-Rh locus in maize requires Pol IV[77].

The role of RdDM in regulation of gene expression was studied by tiling microarray analysis of gene expression in rdr2, drmldrm2cmt3, and wild-type Arabidopsis seedlings (15 days old) grown on MS medium. Eighteen transcripts with AGI codes and 19 predicted transcriptional units with siRNAs were upregulated more than two-fold in both rdr2-1 and drmldrm2cmt3 mutants[78]. siRNAs generated from repetitive sequences regulate expression of nearby genes. The rice genome has a significantly higher fraction of repetitive sequences proximal to genes than the Arabidopsis genome. siRNA-dependent epigenetic silencing might be more important in plants with larger and more complex genomes[4].

Many of the genes with promoter methylation showed tissue-specific expression in Arabidopsis[67]. Compared to the Arabidopsis genome, the rice genome is enriched in promoter DNA methylation, which is correlated with transcriptional repression[79]. Although the rdr2 mutant of Arabidopsis does not show discernible developmental phenotypes, the mop1 (mediator of paramutation1) mutant of maize had delayed flowering in all environments. Other phenotypes (such as reduction in height, feminized tassels, and spindly and barren stalks) were displayed stochastically depending upon the environmental conditions[80]. In larger and more complex genomes, transposons and repeats are abundant in euchromatic regions. Hence, RdDM may be essential for the regulation of genes involved in plant development and plant responses to environmental factors.

Widespread gene-body methylation (in 33% of genes) was found in the euchromatin region of Arabidopsis[4,7]. Arabidopsis ecotypes Col and Ler differed in gene-body methylation status[81]. Gene-body methylation may prevent cryptic transcription, recombination, or transposon insertion within the genes[6,68]. The role of siRNAs in regulating gene-body methylation is not known.

### 4.2 Plant development

Each stage of plant development is mediated by extensive transcriptional reprogramming. Analyses of mutants defective in DNA methyltransferases revealed the functions of DNA methylation in plant development. Arabidopsis MET1-antisense transgenic plants showed reduced CG methylation accompanied by developmental defects such as reduced height, reduced apical dominance, low fertility, and altered leaf size and flowering time[82]. The dmr1 drm2 cmt3 (ddc) triple mutant impaired in non-CG DNA methylation had curled leaves, retarded plant growth, reduction in biomass, and partial sterility[33,83]. Recently, a suppressor of dmr1drm2cmt3 (sdc) mutant was isolated. Overexpression of SDC resulted in a curled-leaf phenotype, while the sdcdrm1drm2cmt3 quadruple mutant had normal leaves. In wild-type Arabidopsis plants, the SDC gene is silenced because of promoter methylation by DRM2 in an RdDM-dependent pathway. Once tandem repeats of SDC promoter are methylated, spreading of DNA methylation into flanking nonrepetitive sequences is mediated by siRNAs and MET1. RdDM-mediated silencing of SDC is necessary for normal leaf development[84]. The Arabidopsis ddm1 mutant showed defects in leaf structure, flowering time, and flower structure[85]. A ddm1 mutation-induced developmental defect called bonsai (bns) was caused by hypermethylation of the BONSAI gene encoding a putative Anaphase-Promoting Complex (APC) 13 protein. A long interspersed nuclear element (LINE) retrotransposon insertion within the 3' non-coding region of the BONSAI gene was found to generate siRNAs required for hypermethylation[86].

Mutants defective in Pol IV (nrpd1a, nrpd2a) and Pol V (nrpd1b, nrpd2b, nrpe5-1) had more rosette leaves and delayed flowering compared to the wild type Arabidopsis, and the effect was more pronounced under short-day conditions[19,22,24,25]. In larger and more complex genomes like that of maize, Pol IV is crucial for plant development. Studies on the maize rmr6 mutant showed that Pol IV is involved in sex determination and abaxial leaf fates. The rmr6 mutant showed drastic reduction in 24-nt siRNAs[77].

Flowering only when environmental conditions are appropriate is crucial for the successful reproduction of plants. Plants employ an epigenetic mechanism to program genes controlling flowering time. Cold temperature treatment known as vernalization represses the FLOWERING LOCUS C (FLC, a MADS-box protein) gene by chromatin modification[87]. Pol IVa/RDR2/DCL3-dependent siRNAs-directed chromatin silencing may contribute to FLC regulation[88], but the role of
RdDM in this regulation is unclear. In a suppressor screen for mutants impaired in transgene silencing, mutants of fca and fpa were recovered. These mutants were late flowering and showed transposon (AtSN1 and AtMu1) reactivation and partial suppression of transgene silencing. Similar to rdr2 and nrpd1a mutants, the fca-11fpa-8 double mutant also had reduced transgene siRNA levels. FCA and FPA proteins contain RRM-domain and regulate DNA methylation at some loci probably through binding of aberrant RNAs and recruiting or stabilizing Pol IV. Besides flowering time, female gametophyte development and early embryonic development were also affected by fca and fpa mutations\cite{89}. Significant differences in siRNA populations were found between seedlings and inflorescence, and floral tissues contained twice as many unique siRNAs compared to seedlings\cite{89}. Further studies are needed to clarify the role of siRNAs in reproductive development of plants.

Imprinting of FLOWERING WAGENINGEN (FWA), MEDEA (MEA), and FERTILIZATION-INDEPENDENT SEED2 (FIS2) genes is controlled by DNA methylation in Arabidopsis. Active demethylation of these genes from the female genome in the endosperm results in their expression. The DME DNA demethylase, expressed in the central cell before fertilization, removes methylation of FWA, MEDEA, and FIS2\cite{91}. The FWA promoter contains tandem repeats, which generate siRNAs that in turn direct promoter methylation\cite{51,83,92}. The variant in methylation vim1vim2vim3 triple mutant, which is defective in CG methylation\cite{83}, and RdDM mutants (rdr2, dcl3, ago4, nrpd1a, nrpd1b, drm, and met1) showed release of FWA silencing and delayed flowering\cite{25,37,92}. These results suggest that RdDM mediates de novo methylation of the FWA gene. Transposable elements are active in the vegetative nucleus of pollen because of downregulation of DDM1 in Arabidopsis. The terminally differentiated vegetative nucleus accompanies sperm cells in pollen. Transposon reactivation in the vegetative nucleus may help suppress transposons in the sperm via 21-nt Athila siRNAs and thus protect the genome of gametes\cite{94}.

### 4.3 Abiotic stress responses

Environmental factors regulate plant growth and development. Above-optimum or below-optimum levels of water, light, nutrients, temperature, and other environmental factors limit plant growth and survival. Hence, plants must be able to sense the environment and reprogram their growth and development to survive and reproduce. Reprogramming of gene expression in response to environmental stresses depends at least in part on nucleosome histone post-translational modifications and DNA methylation in plants. Epigenetic stress memory may provide adaptive advantages to plants\cite{95}. As discussed in an earlier section, RdDM is a key defense against transposon activation. Abiotic stresses may activate transposons through DNA demethylation, and thus affect genome stability. Cold stress-induced downregulation of MET1 expression resulted in hypomethylation of the Ac/Ds transposon region in maize root tips\cite{90}. Cold stress-induced hypomethylation resulted in transposition of the Tam-3 transposon in Antirrhinum majus\cite{97}. Cold stress-induced hypomethylation was associated with induction of ZmMII expression in maize\cite{98}. The Arabidopsis genome consists of about 200 Copia-like retrotransposons. One of them, AtCopeg1, showed tissue-specific expression, activation by cytokinin, and repression by ABA\cite{99}. In newly synthesized wheat amphloids, activation of Wis 2-IA LTR retrotransposons provided promoter activity to proximal genes. This LTR promoter activity resulted in either sense-strand transcription, which increased the expression of adjacent gene, or antisense-strand transcription, which resulted in gene silencing\cite{100}. The rice genome contains about 1000 copies of Dasheng LTR Retrotransposons. Transposon methylation display of leaves versus roots revealed a direct correlation between tissue specific LTR methylation and tissue-specific expression of the flanking rice genes. Further, the rice subspecies japonica and indica differed in LTR methylation and LTR-mediated gene expression. LTR methylation appears to regulate several adjacent genes in rice\cite{101}. Thus, stress-induced transposon activation can profoundly influence gene expression and genome stability.

Abiotic stresses can induce hypo- or hypermethylation of DNA, which in turn can change gene expression. Abiotic stress-induced gene-body hypomethylation enhanced the expression of the Nigpdl (a glycophosphodiesterase-like protein) gene in tobacco\cite{102}. Abiotic stresses induced DNA methylation at some loci in tobacco\cite{103}, pea\cite{104}, and Mesembryanthemum crystallinum\cite{105}. Abiotic stress-induced reduction in methylation and/or increase in demethylation can cause hypomethylation, while RdDM and heterochromatic histone modifications can result in hypermethylation of DNA. In maize, stress-induced hypomethylation was correlated with re-
duction in the expression of MET1\textsuperscript{106}. Because abiotic stress-induced methylation changes are not global changes, sequence-specific change in DNA methylation/demethylation may be mediated by small RNAs. The small RNAs induced by abiotic stress have been identified in Arabidopsis\textsuperscript{106}. As discussed earlier, ROS3 and other RRM proteins may help DNA demethylases to target specific DNA sequences\textsuperscript{66}, siRNAs can also regulate gene expression by PTGS under abiotic stresses. In Arabidopsis, the antisense overlapping gene pair consisting of constitutively expressed Δ\textsuperscript{1}-pyrroline-5-carboxylate dehydrogenase (P5CDH) and salt-stress induced SROS5 produce 3\textsuperscript{′} complementary transcripts. These dsRNAs generate nat-siRNAs through a pathway involving DCL2, RDR6, SGS3, and NRPD1A. The 24-nt SRO5-P5CDH nat-siRNA targets P5CDH mRNAs for cleavage, resulting in a decrease in P5CDH transcript levels. This leads to enhanced proline accumulation and salt stress tolerance\textsuperscript{107}, NRPD1A is required for the biogenesis of both nat-siRNAs and hc-siRNAs that mediate RdDM. The nrpd1a mutant was less tolerant than wild type to salt stress\textsuperscript{107}. Abnormal phenotypes of the mop1 mutant were influenced by the environment\textsuperscript{80}. These results suggest that the siRNA machineries are necessary for environmental stress responses. Small RNAs matching the complementary region of several cis-NAT genes have been identified in Arabidopsis. Many of these cis-NAT gene pairs are regulated by abiotic stresses\textsuperscript{108}. Further studies on these cis-NAT pairs will clarify the role of siRNAs in PTGS and TGS under abiotic stress and stress recovery.

5 Conclusions and perspectives

DNA methylation is a dynamic epigenetic mark. DNA methylation at specific loci is induced by heterochromatic histone modifications and siRNAs. RdDM depends on the components of siRNA biogenesis and methylation complex. The substrates for siRNA biogenesis include natural antisense transcript pairs or transcripts that are converted into dsRNA by RDR. Plant specific RNA polymerases Pol IV and Pol V are involved in siRNA biogenesis and siRNA-induced methylation, respectively. 24-nt siRNAs can direct DNA methylation at complementary loci through RdDM complex containing AGO4/6 and DNA methyltransferases. RdDM induces methylation at CG, CHG and CHH sequences and requires both de novo- and maintenance methyltransferases. Chromatin remodeling enzymes and histone modifying enzymes help accessibility of RdDM machinery to the target loci. Pol IV and Pol V transcribe methylated loci but the mechanism of recognition of these loci needs further studies. In the case of new DNA methylation marks induced by environmental and developmental cues, whether Pol IV and Pol V are involved is not known. RdDM protects genome by silencing transposons. Other important functions of RdDM include paramutation, imprinting and gene regulation during development and stress response. Small RNAs are also involved in target-specific demethylation. ROS3 and other small RNA-binding proteins may direct ROS1 and other demethylases to specific loci for demethylation.

Abiotic stresses induce methylation or demethylation at specific loci. Small RNAs probably direct methylation or demethylation enzymes to these loci. Systematic analysis of siRNAs, DNA methylation, and stress response of mutants defective in siRNA biogenesis, DNA methylation, and demethylation will help unravel the role of RdDM in development and stress responses. The DME DNA demethylase, expressed in the central cell before fertilization, removes methylation of imprinted FWA, MEDEA, and FIS2 genes. It is unclear whether target specific demethylation by DME may be directed by small RNAs.

Gene-body methylation is widespread in the Arabidopsis genome. The role of RdDM and the function of gene body methylation in gene regulation need further study. Transposon silencing and activation are important in crop plants, as these processes can activate or silence proximal genes. Better understanding of RdDM and its functions in plant development and environmental stress responses will help lay the foundation for rational strategies for engineering economically important transgenic plants by promoting/inhibiting transcriptional silencing of transgenes and endogenous genes.

Work in J.-K.Z lab was supported by National Institutes of Health Grants R01GM070795 and R01GM059138. V.C. is thankful to ICAR. New Delhi for providing study leave.

2 Huettel B, Kanno T, Daxinger L, et al. RNA-directed DNA methyla-


44 Aufsatz W, Mette M F, van der Winden J, et al. HDA6, a putative histone deacetylase needed to enhance DNA methylation induced by double-stranded RNA. EMBO J, 2002, 21: 6832—6841
45 Ebbs M L, Bender J. Locus-specific control of DNA methylation by the Arabidopsis SUVH5 histone methyltransferase. Plant Cell, 2006, 18: 1166—1176
61 Erhardt K F Jr, Stonaker J L, Parkinson S E, et al. RNA polymerase IV


82 Finnegan E J, Peacock W J, Dennis E S. Reduced DNA methylation in Arabidopsis thaliana results in abnormal plant development. Proc Natl Acad Sci USA, 1996, 93: 8449—8454


92 Steward N, Kusano T, Sano H. Expression of ZmMET1, a gene encoding a DNA methyltransferase from maize, is associated not only with DNA replication in actively proliferating cells, but also with altered DNA methylation status in cold-stressed quiescent cells. Nucleic Acids Res, 2000, 28: 3250—3259


