RNA-directed DNA methylation
Mechanisms and functions
Magdy M. Mahfouz
Center for Plant Stress Genomics & Technology; King Abdullah University of Science & Technology (KAUST); Thuwal, Kingdom of Saudi Arabia; and Agricultural Genetic Engineering Research Institute (AGERI); Agricultural Research Center; Giza, Egypt

Key words: epigenetics, RdDM, DNA methylation, DNA demethylation, epialleles, abiotic stress, transgenerational inheritance

Epigenetic RNA based gene silencing mechanisms play a major role in genome stability and control of gene expression. Transcriptional gene silencing via RNA-directed DNA methylation (RdDM) guides the epigenetic regulation of the genome in response to disease states, growth, developmental and stress signals. RdDM machinery is composed of proteins that produce and modify 24-nt-long siRNAs, recruit the RdDM complex to genomic targets, methylate DNA and remodel chromatin. The final DNA methylation pattern is determined by either DNA methyltransferase alone or by the combined action of DNA methyltransferases and demethylases. The dynamic interaction between RdDM and demethylases may render the plant epigenome plastic to growth, developmental and environmental cues. The epigenome plasticity may allow the plant genome to assume many epigenomes and to have the right epigenome at the right time in response to intracellular or extracellular stimuli. This review discusses recent advances in RdDM research and considers future perspectives.

Most of the DNA sequences in the mammalian and plant genomes are transcribed into non-coding RNAs (ncRNAs). The recently discovered roles of the ncRNAs in gene regulation processes have revolutionized our understanding of molecular, cellular, developmental and evolutionary biology. Most of these ncRNAs are of different sizes, sequences, genomic loci and biogenesis and they all facilitate sequence specific gene silencing. An expanded and diverse ncRNAs-based gene silencing machineries operate in plants. The ncRNAs and their silencing machinery may function in a cell or tissue specific manner, at certain developmental stage and in response to a disease state or adverse environmental cues. Four RNA silencing pathways have been characterized in Arabidopsis namely microRNA (miRNA), transacting short interfering RNA (ta-siRNA), natural antisense transcript derived siRNA (nat-siRNA) and heterochromatic interfering RNA (siRNA). Unlike miRNA, ta-siRNA and nat-siRNA which function at the post-transcriptional level through mRNA degradation or translational repression, heterochromatic siRNA mediates the gene silencing at transcriptional level by directing DNA methylation.

dsRNAs can be produced from overlapping and inverted repeats transcription by RNA polymerase II (RNA Pol II). dsRNA formation can also be initiated by the plant specific RNA Pol IV. These dsRNAs are then processed by the gene silencing machinery to generate short interfering RNAs (siRNAs) which guide DNA and chromatin modifications on homologous sequences in the genome. RNA-directed DNA methylation (RdDM) pathway which may be unique to plants cells, similar and related process takes place in mammals, functions to mediate epigenetic modifications. The RdDM process was first discovered in plants infected with viroids. Plants use methylation epigenetic modifications to control gene expression possibly as an adaptation mechanism to ensure survival under unfavorable conditions. Methylation dependent silencing at transposons sequences protects the genome. Methylation in gene body regulatory sequences regulate gene expression and ultimately various cellular processes important to growth, development and survival. The epigenetic methylation marks occur at the cytosine nucleotide of the DNA and may be inherited over many generations. Three cytosine methylation systems operate in the Arabidopsis genome in three different sequence contexts: CpG, CpHpG and CpHpH (H is adenine A, thymine T or cytosine C). The methylation can be directed by siRNAs to cause gene silencing. The methylation at some target loci can be reversed by the action of cytosine demethylases. The methylation/demethylation systems provide a dynamic control of gene expression patterns and hence genome plasticity in response to various developmental, growth and stress signals. Throughout this review we will highlight recent advances in RdDM mechanisms and functions pertinent to development, stress tolerance and genome defense.

The Key Players of the RdDM Silencing Machinery

DNA methylation of cytosine residues is required by all multicellular organisms to maintain normal development and proper responses to disease states and environmental cues. Targeted disruption of various DNA methyltransferases result in developmental abnormalities. DNA methylation is known to be the major modification of the plant genome but information on how the methylation marks is directed to certain parts of the genome is unclear. RdDM provided the first evidence that RNA molecules can feedback on the genome and induce epigenetic modification.

Correspondence to: Magdy M. Mahfouz; Email: magdy.mahfouz@kaust.edu.sa
Submitted: 03/02/10; Accepted: 03/03/10
Previously published online: www.landesbioscience.com/journals/psb/article/11695
to their cognate DNA sequences. In the cell nucleus, siRNAs of sequences homologous to DNA in the genome are used to guide a specific and often reversible DNA methylation.

The RdDM methylation machinery is composed of several proteins that process and produce the siRNAs, modify the histones, remodel the chromatin and methylate the cytosine in all sequence contexts. The siRNAs are produced from dsRNAs. The dsRNA can be generated from inverted repeats or overlapping transcription by RNA PolII. Pol IV transcription of methylated DNA can produce single stranded RNA (ssRNA) which is converted into double stranded RNA (dsRNA) by the RNA dependent RNA polymerase 2 (RDR2). The dsRNAs are then processed by DCL3 followed by the HEN1 methylation and loaded on AGO4 which interacts with WG/GW CTD of the largest subunit of Pol V, NRPE1. The siRNAs guided AGO4 and its associated proteins to the homologous DNA sequences and facilitate target methylation by the domains rearranged 2 (DRM2) de novo methyltransferase. AGO4/siRNA may interact with the nascent Pol V transcripts or DNA to guide the methylation process. The Pol IV generates precursor transcripts that produce and or amplify the siRNA trigger. In conclusion, The RdDM machinery is composed of DNA methyl transferases, histone modifying enzymes, chromatin remodelling proteins, plant specific RNA polymerases (Pol IV and Pol V) and RNAi machinery proteins (RDR2, DCL3, HEN1 and AGO4/6). I will address in more detail the molecular structure and function of these key players of the RdDM machinery in the following sections.

**Plant specific Pol IV and Pol V polymerases.** Eukaryotic organisms have three multisubunit DNA-dependent RNA polymerases with distinct functions namely RNA Pol I, II and III. RNA Pol I transcribes the 45S ribosomal DNA genes to generate 45S ribosomal RNAs, RNA Pol II transcribes the DNA coding sequences to generate mRNAs and RNA Pol III transcribes the 5S ribosomal RNA (rRNA) and transfer RNA (tRNA) genes. Higher plants have two extra RNA polymerases, RNA Pol IV and Pol V that function mainly in gene silencing and their catalytic subunits, unlike Pol II, are not required for viability. The largest and the second largest catalytic subunits of these two polymerases have been identified and characterized. The other components of the multisubunit complexes of the two polymerases have been recently identified and are yet to be molecularly characterized.

The largest and the second largest subunits for these two polymerases have been named as NRPA1 and NRPA2 for RNA Pol IV and NRPA1 and NRPA2 for RNA Pol V.

Forward genetics screens and reverse genetics analyses have shown that RNA Pol IV and Pol V are involved in RdDM. Cytosine methylation and siRNA accumulation have been studied in Pol IV and Pol V mutants. siRNA accumulation is abolished in Pol IV mutants and not in Pol V mutants indicating that Pol IV acting upstream of Pol V and function at different steps in the RdDM. 94% of the loci that produce siRNAs in Arabidopsis are dependent on Pol IV and only one third of these loci require also Pol V function. The components of RdDM machinery including Pol IV and Pol V appeared to localize at the chromosomal loci of the sources of siRNA production and targets of RdDM. 16

The siRNAs produced by RdDM colocalize with AGO4, DCL3, NRPA1 and RDR2 in Cajal bodies. It is worth mentioning that Pol IV is also involved in the production of the natural antisense siRNAs (nat-siRNA) using the dsRNA formed by the overlapping 3’ ends of gene pairs. Along with other components of the RdDM, Pol IV was shown to be involved in the spreading of short and long range silencing signals. The substrates and the products of these two polymerases are still not fully known. Possible molecular functions may involve methylated DNA, single stranded RNA or dsRNA as template to produce siRNAs. However, previous studies suggested that the template of Pol IV is RNA as its function is disrupted by RNase A.

Two important recent studies have shown that Pol IV and Pol V evolved from Pol RNA Pol II. The first study led to the identification of the components of Pol IV and Pol V subunits and suggests that these polymerases are Pol II-like polymerases that evolved a specialized role in siRNA production and silencing of transposons, endogenous repeats and transgenes. In the second study, elegant forward genetic screen has identified a new component referred to as RDM2 (for RNA directed DNA methylation 2). RDM2 shares sequence similarity with the fourth subunit of RNA Pol II (RPB4), but is not an orthologue of RPB4, and evolved to assume different function in Pol IV and Pol V complexes. DRM2 is named as NRPA1/ NRPA2 for the Pol IV and Pol V complexes respectively. NRPA4 is required for high level accumulation of siRNA, DNA methylation and transcriptional silencing at transposons and repetitive sequences. This second study confirmed that Pol IV and Pol V exist in a Pol II-like multisubunit complexes and provided a strong evidence that the functionality of Pol IV and Pol V require more subunits in addition to their catalytic subunits.

Current model of RdDM pathway suggests that Pol IV acts first to transcribe methylated DNA to produce the initial transcripts that move into the nucleolus and convert into dsRNA by RDR2. Then, dsRNA is diced by DCL3 and methylated by HEN1 to generate the siRNAs. The siRNAs are loaded onto AGO4 which can bind to NRPA1 and exit the nucleolus together to find and base pair with the Pol V transcripts at the target loci. DRM2 may bind to the siRNA/AGO4/NRPA1 complex in the nucleoplasm as it exits from the nucleolus to methylate the target sequences.

**AGO proteins.** Argonaute genes are found in bacteria, archaea and eukaryotes and their number varies among different species. Argonaute proteins have been discovered in plants more than a decade ago. Argonautes are the major players in RNA-based gene silencing pathways. Through the involvement of Argonautes in many RNAi-based silencing mechanisms, they contribute to maintain the genome, to produce small noncoding RNAs, to form the heterochromatins and to control RNA stability and protein synthesis. Argonaute proteins are classified into three groups based on their phylogenetic relationships and ability to bind small RNAs. Group 1 members are referred to as AGO proteins and they are capable of binding miRNA and siRNAs. Group 2 members are referred to as PIWI, that bind interacting RNAs (piRNAs) and group 3 members are found only in worms and bind to secondary siRNAs. Arabidopsis and rice genomes contain 10 and 18 group 1 argonaute-like genes respectively. Biochemically,
Argonaute proteins are similar to RNase H endonucleases but they use RNA instead of DNA to target the RNA molecules. Argonautes contain four distinct functional domains namely the N-terminal, PAZ, MID and PIWI domains. The PAZ domain recognizes the 3' end of small RNAs and the MID domain binds to the 5' phosphate. The PIWI domain exhibits an endonuclease activity that is similar to that of RNase H enzymes.

Plant AGOs are grouped in three clades: in Arabidopsis, clade 1 includes AGO1, AGO5 and AGO10; clade 2 includes AGO2, AGO3 and AGO7; and clade 3 includes AGO4, AGO6, AGO8 and AGO9. AGO1 was identified more than a decade ago from Arabidopsis mutants that exhibit pleiotropic developmental defects and later it was shown to be involved in post-transcriptional gene silencing. AGO4 was identified first in a forward genetic screen for mutants defective in transcriptional gene silencing of SUPERMAN (SUP) gene locus. This study established AGO4 as the AGO that facilitates the transcriptional gene silencing. AGO4 was shown to control locus-specific siRNA accumulation and DNA methylation. Moreover, AGO4 was shown to be localized in cajal bodies with other members of the RdDM pathway including 24-nt siRNA, Pol IV, RDR2, DCL3 and NRPE1. The proteins of the RdDM pathway have to work in order (Pol IV, RDR2, DCL3 and AGO4) or otherwise will be mislocalized. In Arabidopsis, AGO4 was shown to interact with NRPE1. The NRPE1 C-terminal domain forms a complex with AGO4 and siRNA. AGO4, siRNA and NRPE1 complex binds to target sequence through the siRNA:target base pairing. AGO4 may recruit other RdDM components to cause methylation or use its catalytic activity to generate secondary siRNAs to reinforce silencing. AGO6 was identified in a forward genetic screen for second site suppressors of the transcriptional gene silencing (TGS) in rsl1-1 mutant. Repressor of gene silencing 1 (ROS1) prevents hypermethylation and rsl1-1 mutants exhibit hypermethylation and transcriptional gene silencing of the luciferase reporter gene. ago6iros1 double mutant exhibits a reactivation of the transcriptionally silent luciferase gene. However, transgene reactivation was stronger in ago4-iros1-1 double mutant. AGO6 was shown to be important for the accumulation of specific heterochromatin siRNAs and its functions are partially redundant with that of AGO4.

DNA cytosine methyl transferases. Up to 50% of cytosines are methylated in the Arabidopsis genome. DNA methylation is an important epigenetic modification. In addition to the primary DNA sequence, the chromatin organization plays a major role in determining the gene transcription. The genome is composed of euchromatic and heterochromatic regions. Euchromatin is less compacted and accessible for the transcription machinery and leads to gene expression. However, heterochromatin is highly compacted and renders the DNA not accessible to the transcription machinery and hence leads in general to gene repression. The heterochromatin regions are determined in part by the methylation of cytosine nucleotides. Hence, DNA methylation leads to gene repression and demethylation leads to gene expression. For example, drastic loss of methylation results in massive reactivation of transposons. Moreover, the control of gene expression by DNA methylation depends on the position of the methylation marks relative to the gene. Methylation in the promoter sequences tends to repress gene expression. However, body methylated genes are expressed in moderate to high levels and may lose their tissue specificity. The loss of gene body methylation does not lead to significantly higher gene expression and it might help fine tune the expression in response to developmental or environmental stimuli. However there are exceptions to this rule for example SUPERMAN (SUP) and AGAMOUS gene body methylation causes transcriptional repression because important regulatory regions reside in gene body.

The methylation of cytosine nucleotides in the genome is catalyzed by cytosine methyltransferases. The methylation of cytosine nucleotides determines the extent of heterochromatin and hence the level of gene expression. Cytosine nucleotides are methylated at the 5' position of the pyrimidine ring. The methylation reaction is catalyzed by cytosine methyltransferases by transferring the methyl group from S-adenosyl methionine (SAM) onto the 5' position of the pyrimidine ring generating 5-methyl cytosine (5-me C). The methylation mark on the DNA can serve to attract methyl binding proteins (MBP) that may function as a platform to recruit other chromatin modifying and remodeling complexes. These complexes may function in heterochromatin formation which is repressive for gene expression due to the inaccessibility of the gene regulatory sequences to the transcriptional machinery. Several cytosine methyltransferases are present in the genomes of prokaryotes and eukaryotes. Cytosines can be found in three sequence contexts; 5' CpG 3', 5' CpHpG 3' and 5' CpHpH 3'. Cytosine in CpG and CpHpG are called symmetric cytosines and in CpHpH is called asymmetric cytosine. After each round of DNA replication during cell division, each daughter cell has a hemimethylated DNA (one methylated parental strand and one newly synthesized and unmethylated strand). For the CpG and CpHpG symmetric methylation sequences, the methylation can be established on the unmethylated strand by maintenance methyltransferases based on the information from the old methylated strand. This methylation can occur in the absence of the original methylation signal. General observations on methylation patterns have been made from methylation profiling studies on the whole genome scale. First, methylation is targeted to transposable elements and repeat rich sequences like centromeric repeats and ribosomal DNA sequences (rDNA). This explains that genomes that are rich in repeat sequences exhibit higher methylation relative to genomes with less repeat sequences (25% of cytosines are methylated in maize compared to 6% in Arabidopsis). Second, methylation at CpG loci is highly abundant followed by CpHpG and CpHpH loci respectively. It should be noted that cytosines occur primarily at CpG dinucleotides in the mammalian genome. However, in plant genomes, cytosine methylation occurs at all sequence contexts indicating a rich repertoire of regulatory methylation machineries and complex regulatory mechanisms.
A typical cytosine methyltransferase contains four important domains; a binding domain for SAM, a binding site for the DNA target, a catalytic domain that catalyzes the methyl transfer reaction and a genome targeting domain.39,40 Three major cytosine methyltransferase classes have been characterized in plants.41 Methyltransferase 1 (MET1), a plant homologue of the mammalian DNMT1, a chromomethyltransferase (CMT) and domains rearranged methyltransferase (DRM). MET1 is the major methylation maintenance enzyme at the CpG dinucleotides. CpG methylation is strongly reduced in met1 mutants.3 met1 mutant exhibits morphological defects such as delayed flowering and reduced size.4,42,43 These morphological defects could be traced back to specific loci with reduced methylation like FWA (flow- ering wageningen) gene sequences. Several forward and reverse genetic studies have concluded that MET1 is required to maintain CpG methylation but not non-CpG methylation.43 The residual methylation observed in met1 mutants is facilitated by other cytosine methyltransferases. The CMT chromomethylase, ubiquitous and specific to plants, is characterized by the presence of a chromodomain motif embedded in the C-terminal domain.44 Forward genetic screens have shown that CMTases are involved in non-CpG methylation. A loss-of-function allele of cmt3 exhibited a strong reduction in CpHpG methylation. cmt3 mutants do not exhibit severe morphological defects like met1 indicating that the CpG methylation is the basal and primary pattern of methylation.45,46 These observations suggest that non-CpG patterns could provide a secondary level of regulation. Domains rearranged methyltransferase (DRM) is the third class of cytosine methyltransferases in the plant genome. DRM is the plant homologue of the mammalian DNMT3.47 DRM contains all of the domains found in the mammalian DNMT3 but in a different arrangement. drm1 and drm2 double mutant exhibits no morphological defects and subtle changes in the methyla- tion patterning.48,49 This mutant was found to establish a new methylation imprints on FWA and SUP genes respectively. These observations suggest that DRM1 and DRM2 play a role in establishing a new methylation imprint in response to RNA trigger. In conclusion, the RdDM machinery requires different methyltransferases to methylate cytosine in different sequence contexts.50,51 MET1 was shown to be dispensable to the initiation of the RdDM but indispensable to maintain the methylation in the absence of RNA trigger. RdDM requires the activity of DRM to establish methylation in all symmetric and asymmetric sequence contexts. drm2 mutant shows a complete loss of asym- metric methylation and partial loss of CpHpG symmetric methyla- tion the remaining of which is catalyzed by CMT3.52

As RdDM establishes DNA methylation on the non-methyl- ated target sequence, DRM methyltransferase proteins are very important component of RdDM. In Arabidopsis, drm1drm2 double mutant lacks all types of de novo DNA methylation including transformed tandem repeats and transcribed inverted repeats.48,52,53 In the current model of the DRM targeting to specific sequences, transcription of a given gene target produces nascent RNA transcripts that base pair with the 24-nt siRNAs tethering a complex of chromatin modifying and remodeling enzymes and DRM methyltransferase. This model is supported by the fact that Pol IV/V are important for siRNA directed DNA methylation. In conclusion, DNA methyltransferases play the key role in RdDM and the final methylation patterns and levels are shaped by the activity of both DNA methylases and demethylases.54,55 For some genes methyltransferases may be sufficient to generate the exact level and pattern of methylation but some other genes may need the action of demethylases to adjust the level and pattern of methylation. Indeed, DNA demethylases are important in the control of gene activity and to ensure the reversibility and the plasticity of the epigenetic modifications.55 It should be noted that the methylation patterns of the genome may not be final throughout the plant life cycle. Methylation pat- terns can change substantially in response to developmental and environmental signals.56

**Histone modification and chromatin remodeling proteins.** Chromatin structure has emerged not only to provide a packaging solution of DNA in the nucleus but also as a source of passing a relatively stable genetic information.57 As mentioned earlier, highly condensed chromatin, hetero- chromatin, is repressive for gene expression and open and relaxed chromatin, euchromatin, is permissive for active gene expression.58 Most coding sequences are located in euchromatin regions whereas repetitive and transposable elements are located in heterochromatin regions. Usually heterochromatin is subjected to tight and somewhat permanent silencing but euchromatin is sub- jected to both silencing and activation machineries. Chromatin is composed of nucleosome units which in turn are composed of the DNA helix wrapped around the histone octamers (two copies of each H2A, H2B, H3 and H4). The nucleosomal units exhibit the beads on the string structure and can be connected by histone H1 to generate higher order and compacted chromatin.

Histone proteins can be modified by methylation, acetylation, deacetylation, phosphorylation and ubiquitination to regulate the chromatin structure. Another layer of complexity is added by the fact that lysines in histones can be monomethylated, dimethylated and trimethylated. Also arginines can be monomethylated or dimethylated. It should be noted that histone acetylation is associated with active gene transcription and histone deacetylation is associated with gene repression. However, histone methyla- tion can be associated with either gene expression or repression.59 For example, the di or trimethylation of of lysine 4 of histone H3 (H3K4me2/3) and the di or trimethylation of of lysine 36 of histone H3 (H3K36me2/3) are histone modifications associated with gene expression. However, trimethylation of lysine 27 of histone H3 (H3K27me3) is a histone modification that associates with gene repression. These specific modifications of the histones serve as a code to favor certain chromatin states that contribute to gene expression control.

There is a crosstalk between cytosine methylation and histone modifications on the molecular level. Silent genes were shown to be DNA methylated and deacetylated at histones H3 and H4.60 Cytosine methylation can be used as a nucleating signal for his- tone modifications. The histones of the promoters silenced by RdDM was shown to be deacetylated by AthHDA6.61 Forward genetics screens have identified AthHDA6 as an important fac- tor to maintain the silenced state and as an essential component...
of RdDM that trigger the covalent modifications of histones in a sequence specific manner.\textsuperscript{65} RdDM silenced promoters are reactivated in AThDA6 loss of function mutants. Silencing of promoters by RdDM involves a cooperation between cytosine methylation and histone deacetylation. Histone methyltransferases including SUVH4/KRYPTONITE (KYP) histone 3 lysine 9 (H3K9) methyltransferase, SUVH5 and SUVH6 are required to maintain the cytosine methylation at CHG and non-CG specific loci respectively.\textsuperscript{63-65}

Chromatin remodeling factors use the energy from ATP to move, destabilize or slide the nucleosomes. Chromatin remodelers play an important regulatory role in gene transcription and DNA replication, repair and recombination. Some remodelers promote dense nucleosome packaging while other remodelers act to slide or eject nucleosomes to help the access of the transcription machinery. There are three main ATP-dependent chromatin remodeling complexes in all eukaryotes including the switching defective/sucrose nonfermenting (SWI/SNF) ATPases, the initiation switch (ISWI) ATPases and the chromodomain and helicase like domain (CHD) helicases. One of the important components of RNA-directed DNA methylation is the chromatin remodeling factors. A screen for mutants defective in RdDM has identified defective in RNA-directed DNA methylation 1 (DRD1) that belongs to a subfamily of SWI2/SNF2 chromatin remodelers. Arabidopsis genome contains 41 members of the SWI2/SNF2 like chromatin remodeling proteins. DRD1 subfamily of SNF2-like chromatin remodeling proteins is found only in plants and has similarities to Rad54, ATRX and JBP2 subfamilies. DRD1 subfamily of SNF2-like chromatin remodelers. One of the SWI2/SNF2 subfamily members DDM1 disrupts a remethylation machinery. It appears that constitutive heterochromatin regions remain repressed and silent and chromatin regions that can change between euchromatin and heterochromatin states can either be active or repressed.\textsuperscript{71} This indicates that the methylation/demethylation machineries act on facultative heterochromatin whereas methylation machineries of DNA and chromatin act on constitutive heterochromatin. As RdDM is known to interact with DNA glycosylases and disturbs the DNA methylation it may be acting preferentially on facultative heterochromatin.

**Sources and targets of RdDM.** The 24-nt-long siRNAs can be generated from dsRNAs from various sources. These sources include RNA Pol II transcription of inverted repeats or overlapping transcription of a single transcript, replication of ssRNA viruses, transposons, aberrant RNAs and ssRNAs generated by Pol V from methylated DNA sequences. Theoretically, all of the sources can serve as targets for the RdDM. Moreover, all the DNA sequences complementary to the isolated siRNAs could be targeted for methylation. Genome wide methylation and transcript profiling have been used to identify the endogenous targets of the RdDM pathway in wild type and mutant backgrounds. Intergenic regions, transposons, repeats and plant genes in euchromatin were identified as major targets of RdDM.\textsuperscript{11,13,14}

In a screen of *drd1* and *Pol V* mutants to identify targets, transposons were found to be upregulated. The reactivation of soloLTR of *Copia*-type LTRCO family was associated with reduced cytosine methylation and increased expression of neighboring sequences. The chromatin state affects the reactivation of targets. Targets that are strongly upregulated were found to reside in euchromatin where targets with weak reactivation reside in heterochromatin. The transposons soloLTR, LTR1 and LTR3 have similar DNA sequence but their reactivation is different according to their chromatin location. soloLTR is highly derepressed because it resides between heterochromatin and euchromatin and carries negligible H3K9me2. However LTR1 and LTR3 reside in heterochromatin and are modified by H3K9me2. It should be noted, however, that not all transposons are regulated by RdDM and alternate mechanisms may exist to control their expression.\textsuperscript{23-27} As the RdDM modification is reversible and the potential targets reside in close proximity to genes in the euchromatin it is likely that RdDM plays a much more significant role, than thought before, in gene expression control in response to developmental or environmental stimuli.

**Demethylation Machinery**

The final DNA methylation pattern of the genome, in a given tissue or at developmental stage, is determined by the action of DNA methyltransferases and DNA demethylases. Specific de novo DNA methylation that produces a desirable methylation pattern may not require the action of DNA demethylases. However, promiscuous de novo DNA methylation requires the activity of DNA demethylases to produce a desirable pattern.\textsuperscript{55} Passive and active methylation can occur in the genome. DNA is hemimethylated after DNA replication and the newly synthesized DNA strand requires the action of the maintenance DNA methyltransferases to complete the new DNA strand methylation process. The final DNA methylation pattern of the genome is not necessarily stable and DNA demethylases are involved in adjusting DNA methylation to fit the cellular environment. DNA demethylation can occur in a variety of circumstances including (1) DNA hypomethylation in the vicinity of CpG islands in generally unmethylated DNA, (2) active DNA demethylation in retrotransposons and other repetitive elements and (3) DNA replication instability.

DNA demethylation is an ATP dependent process that involves two distinct mechanisms: passive and active DNA demethylation. DNA demethylases of diverse origin are involved in passive and active DNA demethylation. Several classes of DNA demethylases are known; however, the DNA demethylases of plants are not as well studied. DNA demethylases in plants are divided into two distinct classes: (1) type I DNA demethylases that preferentially demethylate 5mC and (2) type II DNA demethylases that preferentially demethylate 5hmC. The type I DNA demethylases include Arabidopsis DNA demethylase 1 (DDM1) and DDM2, whereas the type II DNA demethylases include Arabidopsis DNA demethylase 3 (DDM3) and DDM4. DDM1 is a chromatin remodeling helicase that was shown to maintain the cytosine methylation at CHG and non-CG specific loci respectively. DDM2 and DDM4 are homologs of DDM1 and are involved in DNA demethylation. The active DNA demethylation is catalyzed by a family of Arabidopsis DNA demethylases that act on methylated DNA sequences. Theoretically, all of the DNA sequences complementary to the isolated siRNAs could be targeted for methylation. Genome wide methylation and transcript profiling have been used to identify the endogenous targets of the RdDM pathway in wild type and mutant backgrounds. Intergenic regions, transposons, repeats and plant genes in euchromatin were identified as major targets of RdDM.\textsuperscript{11,13,14}
ases play a central role in keeping some of the transposons in active DNA demethylation. Methylation of the coding regions is perspective of DNA replication which indicate the involvement of DEMETER (DME) is a DNA demethylase that controls the independent seed (2). DEMETER like 2 (DML2) and DEMETER like 3 (DML3). The demethylases may use one of several proposed mechanisms to achieve the DNA demethylation. The most well-studied and confirmed mechanism is a base excision repair that is initiated by 5-mC DNA glycosylases. The glycosidic bond between 5-mC and the deoxyribose is cleaved by the DNA glycosylase creating abasic AP site. The deoxyribose at the AP site is removed by the action of AP endonuclease creating a gap that is filled by the DNA polymerase and sealed by DNA ligase. These reactions result in the replacement of the 5-mC by unmethylated cytosine. Other proposed mechanisms of DNA demethylation that awaits more confirmation and replication include nucleotide excision repair and hydrolysis, 5-mC deamination coupled with G/T mismatch repair and oxidation demethylation.

DNA demethylases in mammals have been implicated in early development, memory formation, immune responses and tumorigenesis. In plants, DNA demethylases function to prevent the RdDM of transgenes and endogenous genes, regulate imprinting and transposons and involve in 5S rDNA chromatin decondensation. Repressor of silencing 1 (ROS1), a DNA glycosylase/lyase, functions to maintain the expression of a transgene and its homologous endogenous gene by preventing the promoter methylation through the RdDM pathway. Two ROS1 homologues, DML2 and DML3, have been identified in Arabidopsis. A genomic tiling microarray approach was applied to wild type and demethylases triple Arabidopsis mutant (ros1, dml2, dml3). 180 loci were shown to be demethylated and more than 80% of which lie in genic regions indicating that demethylation protects the genes from deleterious methylation patterns. The expression of several maternally imprinted genes require the action of demethylases. DEMETER (DME) is a DNA demethylase that controls the expression of the maternally imprinted MEA polycomb group gene in the endosperm tissue. The disruption of the demethylase activity leads to the silencing of the MEA gene in the endosperm and consequently to impaired seed development. Maternally imprinted genes such as FWA and FIS2 (fertilization independent seed 2) are controlled by demethylases by similar mechanisms. Several studies have indicated that demethylases play a central role in keeping some of the transposons in a dynamic state and maintaining a basal level of expression. It is quite important to keep basal levels of expression of some transposons especially in large plant genomes because they play central roles in maintaining genome structure and architecture and in inducing genetic variation.

Plants subjected to stress show less methylation profiles irrespective of DNA replication which indicate the involvement of active DNA demethylation. Methylation of the coding region of tobacco glycerophosphodiesterase gene were shown to be greatly reduced when plants are subjected to salt, drought and oxidative stresses. DNA demethylases function in chromatin decondensation to permit gene expression at specific loci. The 5S rDNA repeats in the centromeric heterochromatin are silenced by RdDM and chromatin remodeling proteins. ROS1 mediates active DNA demethylation and decondenses the 5S rDNA chromatin to permit gene expression during the periods of high protein synthesis demand. The decondensations is reversed by the RdDM machinery. This condensation and decondensation of 5S rDNA allows plants to adjust to growth signals.

In plants, it has been shown that the levels and activities of the demethyllases is tightly regulated and coupled to the levels and activities of methyltransferases. In met1 mutant genome methylation is significantly reduced and the level of ROS1 expression is undetectable. Also, in mutants of the key components of RdDM (npr1, rdr2, dcl3, dsm2) the expression of ROS1 is very low. It is likely that sensors of the methylation activities exist in the genome and relay signals to control the expression of ROS1, DML2 and DML3. ROS3, an RNA-binding protein and a component of ROS1 demethylation machinery at some target loci was shown to be regulated also by the methylation levels. ROS3 expression is enhanced in ros1 mutant and ROS1 expression is enhanced in ros3 mutant because of increasing methylation at some target loci in ros1 and ros3 mutants. The interplay between methylation and demethylation machineries is used to up or downregulate specific genes or to fine tune the epigenetic states in response to developmental or environmental stimuli. The plasticity of the plant epigenome is maintained by this dynamic control of methylation and demethylation machineries.

**Functions of The RdDM**

Methylation is very important to plant life. In fact, without the methylation epigenetic modification plant life may not be possible. Methylation is very important in many aspects of plant growth, development, variation, responses to biotic and abiotic stresses and genome stability. As plants rely heavily on the epigenetic changes for their development and responses to the environment and due to their sessile nature, plants may need to assume several epigenomes to control the global expression pattern at any given point in their life cycle. DNA methylation, histone modifications and chromatin remodeling factors contribute to produce several epigenomes with different expression patterns in response to growth, developmental and environmental stimuli. This notion is supported by the fact that plant genomes contain the largest numbers of chromatin regulatory proteins (e.g., more than 500 in Arabidopsis), DNA and histone methyltransferases. These several epigenomes could produce a long and/or short cellular memories to adjust the plant responses to different internal or external stimuli.

**Development.** Mitotic epigenetic inheritance through DNA methylation, chromatin modification and remodeling is of utmost importance to the development of eukaryotes. As cells differentiate to new cell types or tissues and respond differently to different intercellular and environmental signals,
the information in the almost static genome sequence may not be enough. Cellular memory comes into play to activate or silence genes. This is of particular importance to plant cells which, unlike mammalian cells, do not terminally differentiate and remain totipotent throughout their life. I will discuss the involvement of several RdDM key players in the development. The morphological phenotypes of mutants of these key players have established the connection of their roles to development. For example, FLOWERING LOCUS C (FLC) is the key gene that mediates the vernalization responses. Vernalization is a cold treatment to seeds that induces epigenetic modification that can be maintained through mitotic cell divisions but not from one generation to the next. Vernalization treatment leads to the repression of FLC by histone methylation.96,97 The repression of FLC was shown to be mediated by a 24-nt siRNA.98 The 24-nt siRNA is absent from Pol IV, RDR2 and DCL3 mutants. These observations suggest, yet to be rigorously tested, that the repression of FLC might be mediated through RdDM pathway.

Pol IV and Pol V subunits may not be essential for the viability of plants. However, they seem to play important RdDM dependent roles in the development. nrpd1 and nrpd2 mutants exhibit a late flowering phenotype when grown under short day conditions as observed in rdr2, dcl3 and ago4 mutants.13,14 The interaction between two alleles that gives rise to a meiotically heritable change in gene expression of one allele is called paramutation. MOP1 (mediator of paramutation 1) and RMR1 (required to maintain repression 1) have been identified in maize by forward genetic screens.99,100 MOP1 and RMR1 were found to be orthologues of RDR2 and SNF2 subfamily member DRD1 respectively. mop1 mutant of maize exhibits delayed flowering, reduced size and spindely and barren stalks.101

Several forward genetic screens and reverse genetic analysis studies have shown that plant methyltransferases play a major role in plant development. met1 and met1dm1 mutants exhibit a much stronger developmental phenotype than that of the met1 mutant alone. These observations indicate a functional redundancy between DNA methyltransferases in development. Differential DNA methylation controls the activities of MEA, FIS2 and FWA imprinted genes in paternal and maternal alleles. The methylation of these genes is mediated by MET1.84,85,102 These genes are methylated and silenced as a default state. However, the maternal alleles are hypomethylated by the action of demethylases. The imprinting of these genes is catalyzed by the DME DNA demethylase.

DECREASE in DNA METHYLATION 1 (DDM1), a chromatin remodeling ATPase, is involved in CG and non-CG methylation and heterochromatin silencing.103 The ddm1 mutant exhibits several developmental abnormalities. The ddm1 induced loss-of-function of BONASI (bns) phenotype was due to the hypermethylation and accumulation of small RNAs. The methylation was shown to be dependent on the presence of a retrotransposon in the 3’ region. The presence of this retrotransposon could generate the trigger for the epigenetic modification pathway. The putative histone demethylase increase in BONSAI methylation 1 prevents hypermethylation by protecting genes from CHG methylation.104-106 In these studies the IBM was shown to be required for the proper development and may have thousands of gene targets.

**Stress responses.** Epigenetic stress memory provides the necessary information for the cellular machineries to assume the right epigenome at the right time under stress conditions.107 The reversibility of the epigenetic changes is crucial for the plant adaptation under different stress conditions. Plants tend to slow their growth to conserve energy as a response to stress so that they do not run the risk of dying.108 Ribosome biogenesis, a massive energy consuming process, and the rRNA transcription increase in response to increasing growth to support the protein synthesis demand. However, under slowed growth conditions the rate of rRNA transcription and ribosome biogenesis decreases as a cellular economic regulatory mechanism to conserve energy. The rRNA transcription is controlled by a DNA methylation and histone modification epigenetic switch.109 This rRNA transcription epigenetic switch may be controlled by stress signals.

The reactive oxygen species (ROS) generated by salt stress induces the expression of SRO5 and P5CDH genes which leads to the production of a 24-nt natural small interfering RNAs (nat-siRNA) from their complementary sequences. The formation of 24-nt nat-siRNAs target the P5SDH messages for cleavage and this process is dependent on the NRPD.110 Accordingly, the production of nat-siRNAs downregulates the expression of the P5CDH leading to the accumulation of proline which plays an important role in plant stress tolerance.

Differential expression levels of maintenance and de novo DNA methyltransferases was observed in rice plants under cold and salt stress.110 Analysis of the methylation patterns in the pea root tips subjected to water stress has shown that hypermethylation is induced by water deficit and that the hypermethylation at the CCGG sequence occurs in specific DNA targets.111 Mild osmotic stress induces reversible hypermethylation of the CpCpG dinucleotides in tobacco cell suspension cultures.112 In this system, there was no changes in the CpG methylation within the CCGG motif. The CG dinucleotides appeared to be fully methylated under normal physiological conditions. The genome of the tobacco cells grown in suspension culture seems to adjust its responses to osmotic stress through de novo methylation of DNA. Aluminum stress induces the transcription of a glycosphosphodiesterase-like protein (NtGPDL). Methylation and demethylation have been observed only in the coding region of NtGPDL. In response to stress, the CG dinucleotide is selectively methylated. Salt, cold and oxidative stresses induce the same demethylation patterns indicating that the gene body methylation may affect the chromatin structure and gene expression of NtGPDL gene.89

The activity of the transposable elements is known to be regulated by reversible methylation. The regulation of rice retrotransposon can affect the expression patterns of the flanking gene sequences. Changes of the methylation patterns of the retrotransposon are induced by stress and show tissue specificity.113 The regulation of the transposable elements methylation patterns could provide another layer of controlling gene expression in response to environmental stress. AtCopeg1 is a member of the Copia-like retrotransposons that shows another example of the involvement of retrotransposons in stress responses. AtCopeg1
is expressed in a tissue specific manner and its expression level is
modulated by hormones and stress signals.114

Genome stability. Transposons have been discovered because
of their propensity to cause damage to the chromosomes. Most
of the eukaryotic genomes have a high number of transposons
that, if activated, can cause high levels of mutations. However,
most of the transposons are quiescent or silenced through epi-
genetic mechanisms. One of the important functions of RdDM
is to target transposons to protect the genome from their del-
eterious effects. Indeed, siRNAs from transposons have been
detected in Arabidopsis and tobacco indicating that transpo-
sons are major sources and targets of RdDM. Transposons can
also affect the transcription of host genes. For example, FWA
gene transcription is inhibited due to the presence of short
interspersed nuclear element (SINE) transposon in the pro-
moter of the gene.115 But how the RdDM machinery recognizes
these transposons sequences provides recognition features. Active
CACTA transposon in a mutant background can remain active
after being introduced in wild type background for several gen-
erations.116 The recognition, however, could be due to the fact
that transposons produce aberrant RNAs that can be detected
by surveillance system. The aberrant RNAs produce dsRNAs
which act as precursors to produce the siRNAs and target the
RdDM machinery to the transposons sequences in the genome.
Indeed, the aberrant RNAs produced from transposons provide
a self-reinforcing silencing mechanism.

In Arabidopsis, the CACTA class transposon is methylated at
CG and non-CG sequences and almost silent in wild type plants.
meth1 and cmnt3 single mutations can lead to partial activation
with no mobility of the transposon. However, meth1cmnt3 double
mutant leads to full activity and mobility of the transposon.117
This indicates that CG and non-CG methylation is required for
full silencing of the transposon. It is worth emphasizing, that
transposons are methylated at the ends as well as at the middle of
their sequence. These methylations features suggest that suppres-
sion of transposons activities could be achieved not only through
the transcriptional silencing but also by blocking the access of
transposases to the ends.117

Transgenerational Inheritance in Plants

Epigenetic silent information is maintained throughout mitosis
and meiosis over generations. The heritability of these silencing
information led to the concept of epigenetic alleles. Epigenetic
alleles or epialleles share the same DNA sequence but exhibit dif-
ferential expression due to the epigenetic marks. Several intrigu-
ing examples of epialleles have been characterized in Arabidopsis,
the epiallele of SUPERMAN is involved in floral development.118
Epialleles can also interact in cis or trans to cause heritable changes.
These epiallelic interactions give rise to concepts like paramutation
and imprinting. In paramutation, the interaction of two alleles
give rise to a meiotically heritable change of one allele. Maize B1
locus, encodes a transcription factor and controls the anthocyanin
pigment, provides an excellent example of paramutation. The two
alleles of a paramutagenic B’ and paramutable B-1 interact. In a
heterozygous background, B’ epiallele converts B-I epiallele into
B’ epiallele. Extensive studies have characterized a 7 repeat region
100 kb upstream of the B locus as a control region for this para-
mutagenic interaction. A genetic suppressor analysis of paramuta-
tion has identified a mediator of paramutation (Mop1) as an RNA
dependent RNA polymerase 2. This clearly links the paramuta-
tion to siRNA machinery. In imprinting, as discussed earlier, the
two alleles have differential expression depending on the parent of
origin. For example, MEDEA (MEA) is a maternally imprinted
gene in the seed endosperm, the paternal allele of MEA is silent
while the maternal allele is expressed.

The alleles of epigenetic states and their associated phenotypes
were shown to be inherited through meiotic cell division over sev-
eral generations. This transgenerational heritability was shown to
be dependent on the accurate propagation of DNA methylation
at the CpG dinucleotides.122 Mutants defective in the mainten-
tance of CG methylation show severe developmental abnor-
malities.123 Loss of function mutation of MET1 suppresses the
activity of DNA demethylases and leads to altered RdDM and
redistribution of methylation marks. It is likely that the CpG
methylation coordinates and stabilizes the epigenetic memory
of transgenerational inheritance in Arabidopsis.

An efficient remethylation mechanism that corrects defects of
transgenerational DNA methylation has been shown to operate in
the Arabidopsis genome.123 The remethylation has been shown to
be facilitated by RNAi. When the RNAi machinery is com-
promised the selective and progressive remethylation is impaired.
Some repeat elements are efficiently targeted by RNAi-dependent
de novo methylation. The elucidation of this RNAi-based mech-
anism in protecting the genome against the loss of transgenera-
tional marks points to key controls of epialleles in variation and
adaptation.

Some plants are very amenable to epigenetic inductions
where others are resistant. For example, rapid and heritable
changes were induced in the inbred flax variety Sormont Cirrus
(Pl) by growth media and stress conditions.124 DNA meth-
ylation and histone modifications were suggested as control
mechanisms that cause these epigenetic modifications.125 Plants
subjected to stress and poor growth conditions exhibit progres-
sive epigenetic modification over generations, which leads to
stress tolerant phenotype. Moreover, plants that grow in very
rich growth medium and optimum conditions also exhibit pro-
gressive epigenetic modifications over generations that lead to
increased plant size and or yield. However, these plants may
lose some parental stress tolerance epigenetic modifications and
become more susceptible once exposed again to adverse envi-
ronmental conditions.

The extensive genotypic and phenotypic changes that occur
in flax in one generation provide a rich resource to study the
epigenetic mechanisms and the points of control of such epigenetic modifications. Using this system, important questions like why certain varieties are sensitive to epigenetic inductions and others are resistant? what are the key controls that can unleash the batteries of epigenetic modifications? and how to engineer plants to be sensitive to epigenetic modifications for desirable traits? Such questions might be answered using the information from this system or similar system.

Concluding Remarks

For plants to thrive and adapt in a dynamic and changing environment, Mendelian inheritance may not be sufficient. Epigenetics adaptations to fluctuations in growth conditions and environmental cues help fine tune the epigenome. This epigenome fine tuning helps plants endure for many generations under limited nutrition or stressful conditions. However, plant epigenetic mechanisms appear to be remarkably complex when compared to mammals. DNA methylation is considered to be the most important epigenetic mark. The RdDM serves multiple functions including genome defense, imprinting, paramutation and gene expression control in response to growth, developmental or stress signals. The RdDM involves the production of 24-nt long siRNAs that can be produced from heterochromatin repetitive sequences, inverted repeats or overlapping transcription. The siRNAs biogenesis takes place in the nucleolus and siRNAs are loaded onto AGO4 which interacts with NRP1. siRNA/AGO4/NRP1 complex exits the nucleolus to the nucleoplasm to recruit the DRM1/2 de novo methyltransferases to the target DNA sequences. The hallmark of the RdDM is that cytosine methylation occurs in all sequence contexts (CpG, CpHpG and CpHpH). The sources and targets of RdDM appears to reside with high density in heterochromatic regions of transposons and repetitive sequence and with much wider and pronounced roles in complex genomes. For example unlike the methylation patterns in Arabidopsis, the promoters of the rice genome are enriched in methylation. The epigenetic regulation is affected by the genome architecture and complex and larger genomes may warrant broader roles of epigenetic silencing. There are still a lot to be learned from comparative studies especially from the analysis of dicots vs. monocots and small vs. large genomes. It is expected that RdDM plays a much wider and pronounced roles in complex genomes.

As the technology for whole genome microarrays and high-throughput sequencing and profiling of DNA methylation is available, comparative epigenomic approaches ought to be utilized to identify more RdDM key players and targets across plant genomes and to link the methylation pattern of a given epigenome to growth, developmental and stress cues. This information will help us to learn more about the molecular mechanisms of epigenetics and its contribution to natural phenotypic variation.

Acknowledgements

I am grateful to Prof. Jian-Kang Zhu, the director of the Center for Plant Stress Genomics & Technology at KAUST, for his helpful discussions and valuable comments on the manuscript. I also thank Prof. Ray Bressan for reading the manuscript and insightful discussions.

References


