Review

Genetic analysis of RNA-mediated transcriptional gene silencing

Marjori Matzke*, Werner Aufsatz, Tatsuo Kanno, Lucia Daxinger, Istvan Papp, M. Florian Mette, Antonius J.M. Matzke

Institute of Molecular Biology, Austrian Academy of Sciences, Billrothstrasse 11, A-5020, Salzburg, Austria

Received 25 July 2003; received in revised form 21 October 2003; accepted 21 October 2003

Abstract

The ‘nuclear side’ of RNA interference (RNAi) is increasingly recognized as an integral part of RNA-mediated gene silencing networks. Current data are consistent with the idea that epigenetic changes, such as DNA (cytosine-5) methylation and histone modifications, can be targeted to identical DNA sequences by short RNAs derived via Dicer cleavage of double-stranded RNA (dsRNA). To determine the relationships among RNA signals, DNA methylation and chromatin structure, we are carrying out a genetic analysis of RNA-mediated transcriptional gene silencing (TGS) in Arabidopsis. Results obtained so far indicate that in response to RNA signals, different site-specific DNA methyltransferases (DMTases) cooperate with each other and eventually with histone-modifying enzymes to establish and maintain a transcriptionally inactive state at a homologous target promoter. Processing of dsRNA in Arabidopsis occurs in the nucleus and in the cytoplasm, where distinct Dicer-like (DCL) activities are thought to generate functionally distinct classes of short RNAs. RNA silencing pathways thus operate throughout the cell to defend against invasive nucleic acids and to regulate genome structure and function.

D2004 Elsevier B.V. All rights reserved.

Keywords: DNA methylation; Dicer; Histone deacetylase; MicroRNA; Short interfering RNA; Transcriptional gene silencing

1. Introduction

During the last several years, a major advance has occurred in our understanding of eukaryotic gene regulation. To an extent that was impossible to foresee, regulatory RNAs have been linked to various gene silencing phenomena that have essential roles in plants, animals and fungi. Different types of regulatory RNA act in distinct ways to induce gene silencing. Long noncoding RNAs, exemplified by mammalian Xist and some RNAs involved in imprinting, appear to serve as scaffolds for the assembly of large heterochromatic domains [1]. This predominantly structural role contrasts to the informational role played by regulatory short RNAs derived via cleavage of double-stranded RNA (dsRNA) precursors. These short RNAs, which are ~21–26 nucleotides (nt) in length, serve as guides for enzyme complexes that degrade, modify or inhibit the function of homologous nucleic acids. They therefore furnish specificity determinants for silencing. Gene silencing phenomena that are induced by nucleotide sequence-specific interactions mediated by RNA are termed collectively ‘RNA silencing’ [2]. RNA silencing, which has been described in diverse organisms, provides a new paradigm for studying eukaryotic gene regulation.

2. RNA silencing and short RNAs

2.1. RNAi and siRNAs

The most familiar form of RNA silencing occurs in the cytoplasm and has been termed posttranscriptional gene silencing (PTGS) in plants, RNA interference (RNAi) in animals, and quelling in the filamentous fungus Neurospora crassa. This evolutionarily conserved process involves a perfectly dsRNA that is processed throughout its length by an RNaseIII activity termed Dicer into ~21-nt short interfering RNAs (siRNAs) of both polarities. Following ATP-dependent unwinding of the siRNA duplex, the antisense siRNA guides a ribonuclease complex, RISC (RNA-induced silencing complex), to the cognate mRNA and targets it for degradation (Fig. 1, middle) [3,4]. RNAi and related phenomena were initially
observed following microinjection of dsRNA into cells or when transcription of transgenes, transposons or viruses yielded dsRNA. RNAi/PTGS/quelling thus play a major role in defending organisms against foreign or invasive sequences. Host defense is not the only function of RNA silencing, however; examination of native short RNA populations is revealing their pervasive involvement in the regulation of endogenous genes that are important for development.

2.2. miRNAs and development

Efforts to clone size-fractionated RNAs from normal cells have recovered various classes of natural short RNAs including microRNAs (miRNAs) [5], antisense siRNAs, tiny noncoding RNAs (tnRNAs) [6], heterochromatic siRNAs [7], and repeat associated small RNAs [8]. These short RNAs are distinguished by various features including the nature of the dsRNA precursor and the type of silencing they elicit.

MiRNAs are derived via Dicer cleavage of imperfect RNA duplexes encoded in intergenic regions of animal and plant genomes. They usually accumulate from one region and one side only of a ~70–130-nt hairpin precursor (Fig. 1, right). In animals, most miRNAs are thought to base pair imprecisely with 3' -untranslated regions (UTRs) of mature mRNAs and block translation. In plants, miRNAs can induce either mRNA degradation [9,10] or translational arrest [11]. Although initial reports suggested that full complementarity of a miRNA was required to induce degradation of target mRNAs (Fig. 1, middle) [9], recent results indicate that partial complementarity is in some cases sufficient to effect target cleavage [10]. The factors that determine the mode of silencing induced by different miRNAs are not yet clear. Association with accessory proteins, such as different members of the Argonaute family (Section 4.4), might facilitate miRNA interaction with different effector complexes.

In Arabidopsis thaliana, 19 unique miRNAs have been cataloged to date [12]. Because some Arabidopsis miRNAs are encoded by multiple loci, these 19 miRNAs correspond to 41 miRNA genes. This probably represents only a fraction of the total number of miRNAs in Arabidopsis. Computational methods, which have not yet been adapted for plants, have estimated the number of miRNA genes in vertebrates and Caenorhabditis elegans to be ca. 230 ± 30 and 105 ± 15, respectively [12]. The number of experimentally verified miRNA genes in C. elegans has risen to 96 [6]. About 30% of the C. elegans miRNA genes are conserved in vertebrates. MiRNAs genes thus comprise similar fractions (0.2–0.5%) of the total gene number for both C. elegans and vertebrates [6].
comprehensive cloning effort has identified a set of 62 nonredundant miRNAs [8].

The expression of many miRNAs in plants and animals is temporally and spatially regulated, consistent with a key role for miRNA silencing pathways in development of these organisms [6,8,12]. Notably, of 61 predicted mRNA targets of Arabidopsis miRNAs, 40 are known or putative transcription factors that modulate cell division and cell fate decisions throughout the plant [12]. Several of the nontranscription factors that modulate cell division and cell fate are known to be directly regulated by miRNAs, such as the floral gene APETALA2, probably by inhibiting translation of the cognate mRNA [11].

The targets of miRNAs in animals are more difficult to predict since their degree of complementarity to target mRNAs is less than for many Arabidopsis miRNAs. Nevertheless, the targets of several animal miRNAs have been identified recently including Hes1, a helix-loop-helix transcription factor miRNA targets are involved in RNA or DNA metabolism. Most strikingly, the predicted targets of Arabidopsis miR162 and miR168, respectively, are mRNAs for DCL1, a Dicer-like protein (Section 3.2.1), and ARGONAUTE1 (AGO1), a protein required for PTGS (Section 4.4) [12]. Specific miRNAs that target genes involved in controlling in leaf morphogenesis and floral development have been reported recently. MiR-JAW regulates the TCP family of leaf transcription factors by inducing TCP mRNA cleavage [10]. MiR-172 regulates the homeotic floral gene APETALA2, probably by inhibiting translation of the cognate mRNA [11].

The targets of miRNAs in animals are more difficult to predict since their degree of complementarity to target mRNAs is less than for many Arabidopsis miRNAs. Nevertheless, the targets of several animal miRNAs have been identified recently including Hes1, a helix-loop-helix transcriptional repressor, which is regulated by miR-23 in humans [13] and the Drosophila pro-apoptotic hid gene, which is regulated by the bantam miRNA [14]. These examples can be added to the founding miRNAs, lin-4 and let-7 of C. elegans. Originally termed small temporal RNAs because of their heterochronic effects on development when mutated [15], lin-4 and let-7 target, respectively, a novel nuclear factor as well as a probable RNA binding protein (lin-4), and a RING-B-Box-Coiled-coil protein (let-7) [16].

2.3. Endogenous antisense siRNAs

Initial experiments to clone short RNAs from C. elegans [17,18] and Drosophila [19] did not reveal an obvious role for RNAi in regulating normal gene expression, as evidenced by the failure to recover any short RNAs that were fully complementary to protein coding regions of endogenous genes. The lack of conventional siRNAs in cloned populations prompted the suggestion that Dicer and its cofactors are ordinarily used to generate miRNAs important for development, and are only recruited for siRNA production and RNAi upon viral or other dsRNA induction [20]. Recently, however, modified cloning strategies have revealed the existence in C. elegans of ~750 endogenous siRNAs that perfectly match genomic sequences that are antisense to predicted protein coding exons [6]. Many short RNAs cloned from Arabidopsis that are not authentic miRNAs also might be antisense siRNAs that can target endogenous genes [9]. It is thus too soon to conclude that RNAi is solely a host defense to viruses, transposons and transgenes, and possible roles in development remain to be investigated (Fig. 1, middle).

2.4. TncRNAs

A new class of short RNA comprising 33 distinct members has been identified recently in C. elegans [6]. TncRNAs are similar in length to miRNAs (~20–21 nt) but they are not processed from typical miRNA hairpin precursors. Some sort of duplex structure is involved in the generation of tncRNAs, however, as their accumulation often requires Dicer activity [6]. Similarly to miRNAs, some tncRNA genes are developmentally regulated. Unlike miRNAs, however, tncRNAs are not evolutionarily conserved and they do not comprise families of related genomic sequences [6]. The function and mode of silencing (if any) triggered by tncRNAs is not yet clear. Since they are derived from non-protein-coding regions, they could conceivably target chromatin modifications to genetic regulatory sequences. Roles in heterochromatin formation have been proposed for heterochromatic siRNAs and repeat associated small RNAs (Section 5).

3. RNA-guided genome modifications

SiRNAs and miRNAs induce silencing at the post-transcriptional level with great specificity because they rely on RNA–RNA sequence recognition. RNA can also base pair with DNA, however, raising the possibility that RNA may guide genome modifications and induce silencing at the transcriptional level. Indeed, during the last several years, the idea that short RNAs can target epigenetic alterations, such as DNA (cytosine-5) methylation and histone modifications, to specific regions of the genome has gained tremendous momentum (Fig. 1, left).

3.1. RNA-directed DNA methylation (RdDM)

The first example of an RNA-guided epigenetic modification of the genome was discovered in a plant system. RdDM was detected in 1994 by Wasseneeger et al. [21] in viroid infected plants. Viroids are plant pathogens consisting solely of non-protein coding, highly base paired, rod-shaped RNAs several hundred base pairs in length. During viroid replication in infected plants, cDNA copies of the viroid that had been integrated into the plant genome became methylated de novo. This suggested that the replicating viroid was somehow initiating methylation of the homologous DNA copies. RdDM has subsequently been shown to require a dsRNA that is processed to short RNAs 21–26 nt in length [22,23]. There is evidence implicating a longer class of short RNA ~24–26 nt in length in RdDM [24,25]. However, a requirement for short RNAs and not dsRNA directly remains to be demonstrated conclusively by showing that
Dicer mutants block RdDM. Many cases of PTGS have been associated with methylation in protein coding regions [26], the significance of which is not yet clear. DsRNAs that contain sequences identical to promoter regions can induce methylation of unlinked homologous promoters and transcriptional gene silencing (TGS) [22,23,27–29].

RdDM is an extraordinarily specific process. Methylation is primarily limited to the region of RNA–DNA sequence identity; there is little spreading of methylation into adjacent DNA sequences [26,27]. This notable specificity suggests that RNA–DNA base pairing provides the substrate for cytosine (C) methylation. RdDM is also a highly refined process: the minimum length of DNA that can be targeted for methylation by RNA signals is ~ 30 bp [26]. Finally, RdDM leads to an unusual pattern of methylation: not only conventional CG dinucleotides but also Cs in other sequence contexts become modified [26,27].

3.2. RNA-mediated TGS

Although the basic features of RdDM have been defined, much remains to be learned about the mechanism of this process. To identify components of the RNA-mediated TGS pathway, we are carrying out genetic screens using different transgene systems in Arabidopsis. The basic system consists of two unlinked transgene loci: a silencer locus and a target locus (Fig. 2). The silencer locus comprises an inverted DNA repeat of target promoter sequences, which is transcribed by a second unrelated promoter. Transcription through this region produces an RNA that is self-complementary and can fold back on itself, thus providing the source of promoter dsRNA. Processing of the promoter dsRNA into short RNAs 21–24 nt in length presumably occurs as a result of Dicer activity. The short RNAs—in conjunction with one or more DNA methyltransferases (DMTases) and possibly chromatin-modifying factors—are thought to trigger de novo methylation and transcriptional silencing of the homologous promoter at the target locus [27]. These well-defined transgene systems are being used to understand the relationships among RNA signals, DNA methylation and chromatin modifications.

3.2.1. DsRNA processing: heterogeneous siRNAs and multiple DCL enzymes

With respect to RNA signals in the RNA-mediated TGS pathway, we would like to know: (a) how and where in the cell the promoter dsRNA—which is synthesized in the nucleus and ultimately exerts its effects in that compartment—is processed to short RNAs and whether these indeed induce RdDM; (b) how RNA signals gain access to DNA; and (c) whether processing of transgene short RNAs and endogenous miRNAs requires similar components. These questions are complicated in plants because of the existence of heterogeneous siRNA populations and multiple Dicer-like (DCL) activities.

Northern blot analyses of short RNAs in plants have consistently revealed the existence of two size classes: a shorter class 21–22 nt in length, which corresponds to short RNAs in other organisms, and a longer class 24–26 nt in length, which is possibly unique to plants [22,24,27]. Evidence that these two size classes of short RNA are functionally distinct comes from studies using viral proteins that suppress PTGS. These proteins, which are derived from different RNA and DNA viruses [30], provide unique tools for studying RNA silencing in plants. Different viral proteins can prevent the accumulation of one or both size classes of short RNA. In one set of experiments, blocking the accumulation of the 21–22 nt class suppressed the mRNA degradation step of PTGS, whereas repressing accumulation of the larger 24–26 nt class prevented RdDM and movement of a mobile silencing signal [24]. Whether these findings are generally applicable remains to be seen, but they provide a precedent for grouping short RNAs into functionally discrete populations and implicate specifically the larger size class in RdDM.

The size and functional diversity of short RNAs in plants is reflected in the multiplicity of DCL enzymes. Whereas

![Fig. 2. Transgene system for analyzing RNA-mediated TGS and promoter methylation. A silencing locus comprising an inverted repeat of target promoter sequences (red) is transcribed by a second unrelated promoter (pro). This produces promoter dsRNA, which is presumably processed by Dicer into short RNAs 21–24 nt in length. Together with one or more DMTases and chromatin factors, RNA signals target methylation (filled circles) to the homologous promoter at an unlinked target locus. Although short RNAs are probably required for this step [24,25], the involvement of dsRNA cannot yet be ruled out (dotted arrow). In response to RNA signals, cytosines (Cs) in all sequence contexts acquire methylation, which spreads little beyond the region of RNA–DNA sequence identity.](image-url)
mammals, *C. elegans* and fission yeast (*Schizosaccharomyces pombe*) have one Dicer protein and *Drosophila* has two, the *Arabidopsis* genome encodes four DCL proteins [31]. Although all DCL proteins contain RNase III and DExH-box RNA helicase domains, which are presumably the minimal requirements for Dicer function, they vary in the presence and number of dsRNA binding domains, PAZ domains (Section 4.4), and nuclear localization signals (NLS) [31]. Two DCL proteins, DCL1 and DCL4, have one or more predicted NLSs, suggesting the existence of both nuclear and cytoplasmic pathways for processing dsRNA in plants. The NLS of DCL1 has been shown to localize a GFP fusion protein to nuclei of plant cells and thus appears to be a functional NLS [32]. Similar experiments have not yet been carried out for DCL4.

Several previously identified mutants that have dramatic effects on plant development ([abnormal suspensor (sus1); short integuments1 (sin1); carpel factory1 (caf-1)] have subsequently been found to be defective in DCL1 [31]. The fact that the *sus1*, *sin1*, and *caf-1* mutations were identified solely from their aberrant phenotypes prior to the discovery of RNA silencing underscores the importance of dsRNA processing for plant development. Consistent with this, DCL1 has been shown to be required for processing various substrates (for example, perfect vs. imperfect RNA duplexes). Conceivably, DCL1 and DCL4 process dsRNAs in the nucleus and the putative cytoplasmic DCL activities, DCL2 and DCL3, process dsRNAs in the cytoplasm [32]. To determine DCL requirements in an RNA-mediated TGS system, we are using reverse genetics approaches to down-regulate DCL1 and DCL4 (L. Daxinger, W. Aufsatz and A. Matzke, unpublished data).

The extent to which nuclear processing of dsRNA occurs in organisms other than plants is not known. Although the Dicers from *S. pombe* and *C. elegans* lack predicted NLSs, one of the two *Drosophila* Dicers (Dicer-1) and the mouse and human Dicer each contain one predicted NLS [31]. Whether these NLSs are functional remains to be established. Data so far indicate that the human Dicer is located in the cytoplasm and acts in the RNAi pathway in that compartment [36,37]. This is consistent with the subcellular compartmentalization of processing of miRNA precursors in human cells, which takes place in two steps catalyzed by distinct RNase III proteins: a nuclear step, catalyzed by Drosha [38], which releases single 70-nt miRNA precursors from primary polycistronic transcripts, and a cytoplasmic step catalyzed by Dicer to process miRNA hairpin precursors into mature miRNAs [39]. Restriction of Dicer to the cytoplasm does not rule out the possibility that siRNAs or miRNAs produced in this compartment could be transported to the nucleus and induce genome modifications [28,40,41]. Drosha homologs are found in *C. elegans*, *Drosophila*, mice and humans [38], but apparently not in plants, where nuclear DCLs can possibly compensate for their absence. In addition to having pathways for processing dsRNA to short RNAs by DCL activities in the nucleus, plants also differ from mammals in lacking ADARs (adenosine deaminases acting on RNA), nuclear enzymes that edit long perfect dsRNAs [42]. Plants and animals thus appear to deal differently with long extended duplex RNAs in the nucleus.

### 3.2.2. De novo and maintenance methylation

With respect to the cytosine methylation step of the RdDM pathway, we are interested in identifying the DMTases responsible for establishing and maintaining the unusual pattern of methylation induced by RNA. Establishing and maintaining DNA methylation are two separate steps and they are thought to require distinct enzymes. Recent results, however, are challenging this strict division of activities.

Establishment involves the de novo methylation of a previously unmodified DNA sequence. The signals for de novo methylation are still mysterious. One of the best known is the RNA signals discussed here, but their ability to guide homologous DNA methylation has been confirmed so far only in plants. Additional signals for de novo methylation in higher organisms possibly include pairing of DNA repeats [43] or other unusual DNA structures [44,45]. DMTases that are considered de novo enzymes include Dnmt3a and Dnmt3b in mammals [46,47] and their plant homologs, the domains rearranged methyltransferases, DRM1 and DRM2 [48].

Once a particular sequence acquires methylation, it is possible to maintain this modification at symmetrical CGs (and in plants, CNGs, where N is not G) through subsequent rounds of DNA replication through the activity of so-called maintenance DMTases. These enzymes can recognize methylated Cs in the template strand and catalyze methylation of the opposite C in the newly synthesized strand. The traditional maintenance activity for CG dinucleotides in mammals is Dnmt1 [49]; the plant homolog is MET1 [50,51]. In plants, a special type of DMTase, chromomethylase3 (CMT3), can maintain methylation in CNG trinucleotides [52]. There is no known maintenance activity for asymmetrical CNN (where N is not G) nucleotide groups. Therefore, if methylation is observed in asymmetrical Cs, this can be taken as a measure of ongoing de novo methylation [27,53]. Once de novo methylation stops for any reason, CNN methylation is rapidly lost and the only methylation that persists is that which can be maintained. This includes CG methylation in mammals, and CG and CNG methylation in plants.

There is a third, enigmatic class of putative DMTases in many eukaryotes—the Dnmt2 family—whose function is
not fully understood. Members of this family are present in mammals, plants and two organisms not normally thought

to methylate their DNA: D. melanogaster and S. pombe [54]. The Drosophila dDnmt2 is active during early develop-

ment and catalyzes low levels of non-CG methylation [55]. The gene encoding the Dnmt2 homolog in S. pombe is

mutated and does not encode a protein with catalytic ability [56]. Nevertheless, the gene is expressed [57] for reasons

that are not clear.

All of the DMTases discussed above are potential can-

didates for establishing and/or maintaining RdDM.

3.2.3. RdDM and chromatin modifications

The third aspect of the RNA-mediated TGS pathway

considered here concerns the identification of chromatin

modifications that might be needed to initiate or retain

RdDM. Given the close relationships between DNA meth-
ylation and chromatin remodeling [58–60], and DNA methylation and histone modifications, such as deacetyl-
yation [61–64] and methylation [65–67], it is likely that the

RdDM pathway will require at one or more steps histone-

modifying activities and chromatin remodeling factors.

4. Genetic analysis of RNA-mediated transcriptional silencing

The basic question under investigation is how RNA signals are produced and used to assemble a transcription-

ally inactive state at a homologous target promoter. Differ-

ent promoters are likely to attract and maintain methylation
to varying extents and to exhibit individual sensitivities to

cG or non-cG methylation. Therefore, we are analyzing

several target promoters that differ in their strength, tissue-

specificity, and density of CG dinucleotides. Two discussed

here are the nopaline synthase promoter (NOSpro), a moderately active, constitutive plant promoter ~ 300 bp in

length, and the α′ promoter (α′ pro), a strong, seed-

specific promoter that is ~ 250 bp in length [68]. Although

both promoters have similar lengths and GC contents (~ 45%), the NOSpro contains 19 CG dinucleotides in

contrast to only nine in the α′ pro. A priori, one might hypothesize that the NOSpro is more sensitive to CG methylation
than the α′ pro.

The silencer locus in each case contains a transcribed inverted repeat of the target promoter (Fig. 2). The respec-
tive target genes are NOSpro-NPTII (nopaline phospho-

transferase II) and α′ pro-GFP (green fluorescent protein)

Plant lines that are doubly homozygous for the target and

silencing loci are being used in mutant screens to identify

components of RNA-mediated silencing pathways.

4.1. NOSpro RNA-mediated TGS

Screens for suppressors of RNA-mediated silencing of

the NOSpro-NPTII target gene recovered rts mutants (RNA-

mediated transcriptional gene silencing) that could be placed in three complementation groups: rts1, 2, and 3. The rts1

mutation was mapped and found to correspond to an RPD3-
lke histone deacetylase, HDA6 [53], which is one of 18

histone deacetylases encoded in the Arabidopsis genome

[69]. Examination of DNA methylation patterns revealed

that the primary defect in hda6 mutants is a failure to

reinforce CG methylation above the ‘de novo’ level, which

was set at ~ 30–50% based on the degree of CNN

methylation, which cannot be maintained by any known

DMTase [53].

Associations between CG methylation and histone deace-
tylation are known from animal systems [61–64]. In our

experiments, however, the signal for de novo methylation

is known (RNA) and the resulting pattern of de novo methyl-

ation is distinctive by including Cs in all sequence contexts,

including CNN. This information allowed us to dissect the

effects of the rts1 mutation on de novo vs. maintenance

methylation. Accordingly, HDA6 was placed in a pathway

between the de novo methylation step catalyzed in the

presence of RNA signals, in which all Cs in the region of

RNA–DNA sequence identity become modified to an

intermediate degree, and subsequent reinforcement and/or

maintenance methylation of preferentially CG dinucleotides

[53].

The DMTase maintaining CG methylation in plants is

MET1 (Section 3.2.2). Interestingly, the rts2 mutant turned out to be defective in this enzyme [W. Aufsatz, M.F. Mette, A. Matzke, unpublished results]. The rts2 mutant revealed, however, that MET1 is more than a CG main-

tenance activity. This was demonstrated by a bisulfite

sequence analysis, which showed that in the met1 mutant, nearly half the CG dinucleotides in the target NOSpro

sequence lack methylation on both DNA strands in the

presence of the RNA signal. This suggests that in response
to RNA signals, MET1 can act as a de novo DMTase for

CG dinucleotides [W. Aufsatz, M.F. Mette, M. Matzke, unpublished results]. The CG ‘de novo’ activity of MET1

probably occurs in cooperation with the traditional de novo

DMTases (DRM1 and DRM2) because all de novo methy-
lilation of the target NOSpro is abolished in drm1 drm2 double mutants [70].

The proposal that MET1 is a de novo methyltransferase

for a subset of CGs in cooperation with traditional de novo

activities DRM1/DRM2 is consistent with data from mam-
malian systems which indicate that: (1) Dnmt1, normally

considered a CG maintenance DMTase, also has de novo

activity [71–74]; (2) Dnmt1 cooperates with Dnmt3a/3b—

the conventional de novo enzymes—to establish and/or

maintain methylation [74–77]; and (3) preexisting methy-
lation stimulates de novo activity of Dnmt1 [78]. These

findings all indicate a degree of redundancy among

DMTases, blurring the distinction between traditional ‘de

novo’ and ‘maintenance’ functions [77]. CMT3 is a

possible exception, serving a bona fide maintenance func-
tion for CNG trinucleotides in plants [52,79]. Classifying
the other DMTases by site-specificity is probably still valid: CG dinucleotides are methylated primarily by MET1/Dnmt1 and non-CGs predominantly by DRM1,2/Dnmt3a,3b.

A model based on the findings from the rts1 and rts2 mutations depicts MET1 and DRM1/DRM2 acting together in response to RNA signals to catalyze de novo methylation at CGs and nonCGs, respectively, in the target NOSpro (Fig. 3). The resulting de novo methylation reaches a level of ~30–50% at most Cs within the region of RNA–DNA sequence identity. According to the model, the relatively uniform, intermediate methylation leads to partial silencing and recruits HDA6, which deacetylates histones and attracts more MET1 to reinforce specifically CG methylation. This presumably locks in silencing, perhaps in conjunction with other histone modifications.

This model is similar to one proposed for the establishment of de novo methylation during mammalian embryonic development. In this model, Dnmt3a/3b initially methylate non-CGs, resulting in the recruitment of Dnmt1 to catalyze de novo methylation of CG dinucleotides [75]. Missing from the mammalian model is a description of the signals specifying the initial pattern of de novo methylation catalyzed by Dnmt3a/3b. In our experiments, de novo methylation is cued by RNA. Similarly, RNA signals could fulfill this role during mammalian development (Section 7), although this suggestion remains to be verified experimentally.

4.2. Sequence composition of promoters determines dependence on C methylation

The NOSpro is clearly sensitive to CG methylation: both of the rts mutants identified so far have a defect in some aspect of CG methylation. Independent genetic analyses in other laboratories have indicated that other promoters might be silenced by non-CG methylation. For example, screens for suppressors of silencing and hypermethylation of the SUPERMAN and PAI loci in Arabidopsis recovered mutants defective in the chromomethylase CMT3 [52,79], which maintains CNG methylation, and SUVH4 [67,80], a histone H3 Lys9 methyltransferase needed to maintain CNG methylation (Fig. 3, top). The ‘CNG methylation-sensitive’ pathway has been linked indirectly to ~24-nt short RNAs by the identification of a mutant deficient in an Argonaute protein, AGO4 [25], which is thought to play a role in generating short RNAs or helping them carry out their function (Section 4.4). Finally, certain promoters appear to be unaffected by C methylation, as indicated by recovery of

Fig. 3. Genetic analysis of RNA-mediated TGS and promoter methylation. The experimental system is described in Fig. 2. Forward genetic screens with the NOSpro system have recovered rts mutants defective in the DMTase MET1 (rts2), and the histone deacetylase HDA6 (rts1). MET1 appears to act as both a de novo and reinforcement/maintenance DMTase for CG dinucleotides. Maintenance occurs in the absence of RNA signals; reinforcement in their presence. The ability of MET1 to act as a de novo DMTase is probably stimulated by RNA signals and requires cooperation with the traditional de novo enzymes DRM1,2 which methylate non-CGs in response to RNA signals. HDA6 appears to act in between the de novo and reinforcement steps to enhance preferentially CG methylation. The identity of rts3 is not yet known, but preliminary data suggest that it might affect short RNA production or accumulation. Molecular characterization of mutants recovered in a screen for abolishing seed-specific RNA silencing of the α-pro (asr1, 2, 3) suggests they are involved in facilitating RNA–DNA interactions and/or de novo methylation of non-CGs. The nuclear enzyme Dicer-like1 (DCL1) is not responsible for producing promoter short RNAs; other candidates are DCL2, DCL3 and DCL4, the last of which—like DCL1—has a predicted nuclear localization signal. A silencing pathway that has been described by others in Arabidopsis [25,52,67,79,80] relies on CNG methylation (light gray ovals) that is maintained by chromomethylase (CMT3), histone H3K9 methylation (KYP), and a member of the Argonaute protein family (AGO4). The position of AGO4 illustrated here is consistent with its postulated role in short RNA production [25].
the mom mutation, which releases TGS of a methylated transgene [81].

4.3. α’ promoter RNA-mediated TGS

The seed-specific α’ pro offers an opportunity to identify tissue-specific silencing factors and to analyze a promoter that may be less sensitive than the NOSpro to RNA-mediated CG methylation. The α’ pro, which originates from the α’ subunit of the soybean seed storage protein β-conglycinin [68], has been shown to maintain its seed-specificity in transgenic Arabidopsis (T. Kanno, unpublished results). A genetic screen for suppressors of RNA-mediated silencing and methylation of an α’ pro-GFP target gene in Arabidopsis has recovered asr mutants (abolish seed-specific RNA silencing) that could be placed in three complementation groups: asr1, 2 and 3 [T. Kanno, A. Matzke, unpublished results]. None of the asr mutations tested so far appear to be allelic to the rts mutations, suggesting that the NOSpro and α’ pro are silenced by distinct silencing pathways. In contrast to the rts mutants, which are deficient in CG methylation of the target NOSpro, the asr mutants are impaired in de novo methylation of non-CGs in the target α’ pro. The sharp reduction of non-CG methylation occurs despite the continued presence of α’ pro dsRNA and short RNAs in the asr mutants (T. Kanno, M.F. Mette, M. Matzke, unpublished results). These findings suggest that the asr mutations are defective in either: (1) de novo DMTase activities targeting non-CGs (i.e. DRM1 or DRM2); (2) proteins that allow RNA signals to access DNA, such as chromatin remodeling factors; or (3) proteins that interface between RNA signals and epigenetic effector complexes. Potential candidates for the last category are members of the Argonaute protein family.

4.4. Proteins functionally linking short RNAs and DMTases/chromatin modifications

So far our genetic screen for suppressors of RNA-mediated silencing of the NOSpro, which is sensitive to CG methylation, has revealed proteins that act at the genome level: a DMTase, MET1, and a histone-modifying enzyme, HDA6. As mentioned previously, we are using reverse genetics to identify the DCL activity required to produce promoter short RNAs and to confirm their role in RdDM (Section 3.2.1). Still missing from the RdDM pathway, however, are proteins that associate with short RNAs (or other RNA signals) and facilitate their interaction with homologous DNA regions and/or epigenetic modifiers. Given the defects in de novo methylation in asr mutants despite the continued production of α’ pro short RNAs, one or more of these mutants might be deficient in such an accessory protein. Possible candidates are members of the Argonaute family, which have been recovered repeatedly in screens to identify mutants impaired in various types of RNA-mediated silencing.

Argonaute proteins are ~ 100-kDa highly basic proteins comprising PAZ and PIWI domains [82]. The PAZ domain, which is also found in many Dicer enzymes, might mediate protein–protein interactions [83]. Indeed, two Drosophila Argonaute proteins have been found to communoprecipitate with Dicer [83]. The function of the conserved PIWI domain is unknown. The basic character of Argonaute proteins might give them RNA binding ability. Although the precise roles of these proteins in RNA-mediated silencing are not yet clear, they probably participate in producing and/or ‘shepherding’ different short RNAs into appropriate effector complexes [83]. In addition to roles in silencing, Argonaute proteins have been implicated in development and stem cell fate determination.

Argonaute proteins are encoded by multigene families comprising ten members in Arabidopsis, four in Drosophila, three in C. elegans, seven in humans, and eight in the mouse [83]. Screens for mutants defective in RNAi/PTGS/quelling identified the related Argonaute proteins rde1 in C. elegans [84], AGO1 in Arabidopsis [85] and QDE2 in Neurospora [86]. With respect to genome-level silencing, mutations in the Drosophila piwi gene disrupt TGS of transgene [87]. The ago1 mutant in S. pombe is impaired in RNA-mediated heterochromatin formation [40,41,88–90]. Arabidopsis AGO4 is needed for silencing, CNG methylation and histone H3 Lysine 9 (H3K9) methylation of SUPERMAN and the retroelement AtSN1 [25].

Different Argonaute proteins probably participate to varying extents in different types of RNA-mediated silencing and/or in development. For example, Arabidopsis AGO1 is required for PTGS and RdDM triggered by a sense transgene construct [91] and can have pleiotropic effects on plant architecture when mutated [92,93]. In contrast, PINHEAD/ZWILLE, a second member of the Argonaute family in Arabidopsis that has been relatively well studied, has a role in maintaining undifferentiated stem cells in the shoot apical meristem [94] but is dispensable for PTGS [91]. Although Arabidopsis AGO4 has been implicated in maintaining CNG methylation and H3K9 methylation, it does not induce obvious morphological defects when mutated [25].

5. RNAi-dependent heterochromatin assembly in S. pombe

Although we have focused on RdDM in plants, recent work on S. pombe has revealed an extensive role for components of the RNAi machinery in gene silencing and chromatin modifications that occur in the absence of detectable DNA methylation. The RNAi proteins Dicer, Argonaute and RNA-dependent RNA polymerase are required in conjunction with the histone methyltransferase Clr4 to establish heterochromatin at S. pombe centromeres [41,88,89] and the silent mating type locus [40]. This pathway can also target normally euchromatic genes, provided the appropriate complementary short RNAs are pro-
duced, and long terminal repeats (LTRs) of retrotransposons [90]. In contrast to RdDM, which is largely confined to the region of RNA–DNA sequence identity, RNAi-dependent heterochromatin can spread several kilobases from the RNA-targeted nucleation site in a manner that depends on Swi6, the S. pombe ortholog of heterochromatin protein1 [40,41,88–90]. As a consequence of the spreading process, heterochromatin nucleated on LTRs can infiltrate into and silence neighboring genes, thus contributing to cell differentiation [90]. Collectively, these studies suggest a general role for RNAi-mediated heterochromatin assembly in gene silencing [40,90], chromosome structure [41], and segregation [88,89] in S. pombe. Whether RNAi-dependent heterochromatin assembly in S. pombe is totally independent of proteins of the DNA methylation machinery is discussed in Section 7.

The targets of the RNAi-dependent heterochromatin pathway appear to be primarily transposable elements and their degenerate remains: in addition to retrotransposon LTRs, the centromere outer repeats [41,88] and mat locus [40] contain transposon relics. This pathway thus illustrates how an ancient defense pathway acting on invasive sequences can be recruited to regulate host genes and other regions of the genome that contain transposon-derived sequences. Thus, from an evolutionary perspective, the host defense and developmental roles of RNA silencing (Fig. 1) are likely to be related [95,96].

6. Natural short RNAs inducing promoter methylation or heterochromatin formation

Natural sources of short RNAs that target DNA methylation or histone modifications to homologous regions of the genome can potentially originate from hairpin precursors encoded in intergenic regions or by hybridization of overlapping sense-antisense transcripts. In S. pombe, heterochromatic siRNAs derived from an outer centromere repeat [7] are needed for centromeric heterochromatin [41,88]. Several miRNAs identified in Arabidopsis might target endogenous promoter sequences [34]. Some of the C. elegans ncRNAs or endogenous siRNAs could initiate chromatin modifications [6]. In Drosophila, repeat associated small RNAs, which are derived from every known type of transposable element in this organism, may contribute to heterochromatin formation of cognate DNA regions [8]. As more short RNAs are identified and functionally characterized, a full accounting of those capable of guiding genome-level modifications should eventually emerge.

7. General model: short RNAs interact with proteins of the DNA methylation machinery

It is noteworthy that RNA silencing and DNA methylation pathways frequently co-exist [96]. Plants, vertebrates and Neurospora, all of which methylate their DNA, also carry out RNAi. Similarly, Drosophila and S. pombe perform RNAi and express Dnmt2-like genes, although only the Drosophila enzyme can catalyze C methylation. By contrast, Saccharomyces cerevisiae lacks the machineries for both RNAi and DNA methylation. An exception is C. elegans, which carries out RNAi but whose genome does not encode recognizable DMTases. The phylogenetic distribution of RNAi and DNA methylation is striking and suggests links between these two processes in the establishment of epigenetic modifications in diverse organisms.

For the establishment of epigenetic modifications during plant and animal development, we favor a general model in which proteins of the DNA methylation machinery initially recognize an RNA-DNA hybrid (Fig. 4). This step will usually lead to de novo C methylation that is co-extensive with the region of RNA–DNA sequence identity and distributed among Cs depending on the site-specificity of available DMTases (Fig. 4, left). For plants, Cs in all sequence contexts would become modified, because both MET1 (targeting CGs) and DRM1 DRM2 (targeting non-CGs) are able to catalyze de novo methylation induced by RNA signals. Although methylation in animals genomes has been assumed to be restricted to CG dinucleotides, recent work has detected substantial non-CG methylation early in mammalian [97] and Drosophila [55] development. It is conceivable that this non-CG methylation is guided by RNA. Consistent with this, mammalian Dnmt3a/Dnmt3b [97] and Drosophila dDnmt2 [55] are able to catalyze non-CG methylation; moreover, the activity of these enzymes is high in embryos but declines thereafter, making them unavailable for catalyzing non-CG methylation in the genomes of adult cells.

According to the original pattern of de novo methylation that is laid down in response to RNA signals, chromatin-modifying enzymes would be recruited to establish a repressive chromatin structure. This step may or may not be accompanied by DNA methylation depending on the continued availability of RNA signals, and DMTases with de novo and/or maintenance activities (Fig. 4, right). Plants are unique in having a maintenance activity for CNG trinucleotides (CMT3) and they continue to carry out de novo methylation throughout development, which probably accounts for their highly methylated genomes. At the opposite extreme, Drosophila—which has no active maintenance DMTase in adult flies—would passively lose all methylation established in the embryo and rely on chromatin factors recruited previously to maintain the silent state. The idea that established DNA methylation patterns induce histone modifications during mammalian development has recently gained experimental support [98]. DMTases lacking catalytic ability, such as mammalian Dnmt3L [99] and S. pombe Dnmt2 [56,57], might also respond to RNA signals and recruit chromatin factors without directly catalyzing C methylation.
The proposal that proteins of the DNA methylation machinery recognize an RNA–DNA hybrid can account for the extraordinary specificity of RdDM. The resolution of RdDM can be regulated virtually down to the nucleotide level over a DNA target sequence as short as 30 bp. This contrasts to histone modifications, which take place in a chromatin context comprising 146 bp of DNA per nucleosome. The fact that the entire DNA sequence spanned by the triggering RNAs is equally accessible to DMTases, and that methylation fails to spread significantly beyond the region of RNA–DNA sequence identity, supports the notion that RdDM requires direct RNA–DNA base pairing that is not restricted by a complex chromatin environment. One possibility is that RdDM occurs during or directly after DNA replication, before histones are deposited and nucleosomes form. Indeed, a structural resemblance between short RNA–DNA hybrids and a DNA replication fork could explain the ability of MET1 to be both a CG de novo DMTase in the presence of RNA signals, and a CG maintenance DMTase during DNA replication [96].

Conceivably, RNA-guided chromatin modifications could occur within a nucleosomal context, but this might be mechanistically distinct from RdDM and occur independently of DMTases. For example, this could be achieved through an ‘RNA recognition’ mechanism, in which short RNAs do not base pair with DNA but with nascent transcripts synthesized from target genomic sequences [100]. This model might be applicable to cases where the primary epigenetic effect involves covalent modifications of histones, leading to the formation of heterochromatin that can spread relatively long distances from the nucleation site. This process might be fundamentally different from RdDM, in which the primary epigenetic mark is C methylation that does not spread significantly beyond the genomic region targeted by RNA. These features are more compatible with the ‘DNA recognition model’ invoking RNA–DNA base pairing [100]. An important question for future research is to determine the way(s) in which short RNAs or other RNA signals elicit DNA cytosine methylation and/or histone modifications.

Fig. 4. Hypothetical model for the establishment and maintenance of transcriptional silencing in different organisms via recognition of RNA–DNA hybrids by proteins of the DNA methylation machinery. Known or putative de novo DMTases and the type of cytosine methylation (filled circles) they catalyze (CG, CNG, CNN) in response to short RNAs (red wavy lines) are depicted at the left. Following establishment of the de novo methylation pattern, maintenance DMTases in plants and mammals, and chromatin factors (black and gray ovals) such as histone-modifying enzymes and polycomb proteins, are recruited in all organisms to stabilize the silent state (right). In plants, the maintenance DMTases, MET1 and CMT3, can preserve methylation in CGs and CNGs, respectively, even if the RNA signals are removed. If RNA signals remain available, de novo methylation can continue owing to the activity of de novo DMTases in adult plants. In mammals, Dnmt1 will maintain CG methylation even if RNA signals are removed (parentheses), but de novo non-CG methylation will cease even if RNA signals persist because the required DMTases, Dnmt3a/3b, decline in activity as development proceeds. In Drosophila, non-CG methylation is catalyzed early in development by dDnmt2. Owing to the loss of dDnmt2 activity during development, methylation will be passively lost during cell divisions, leaving chromatin modifications to maintain the silenced state in adult flies. In S. pombe the catalytically inactive Dnmt2 protein could initially respond to short RNA signals, contributing to a structure that recruits chromatin factors to stabilize silencing in the region of RNA–DNA sequence identity.
8. Summary and outlook

RNA-directed genome modifications are taking their place alongside siRNA-guided mRNA degradation and miRNA-guided translational repression as a fundamental means for achieving sequence-specific gene silencing. It is becoming evident that RNA-mediated silencing is important for development, genome defense and chromosome architecture in diverse eukaryotes. RdDM, the first example of an RNA-guided genome modification, provides the most compelling evidence that RNA–DNA sequence interactions can directly elicit epigenetic modifications in a highly specific manner.

Genetic analysis of the RdDM pathway in Arabidopsis lends support to the following scenario: (1) in the presence of RNA signals, site-specific DMTases cooperate to establish intermediate levels of de novo methylation at CG and non-CG nucleotide groups within a region of RNA–DNA sequence identity; (2) the RNA-directed pattern of de novo methylation promotes the recruitment of histone-modifying activities; (3) histone modifications lead to reinforcement of C(N)G methylation, which can also be maintained in the absence of the RNA trigger. This sequence of events implies that DNA methylation can be both a cause and a consequence of silencing. This dual role might be attributable to the structural resemblance between short RNA–DNA hybrids, which provide a substrate for de novo methylation, and DNA replication forks, where preexisting epigenetic modifications must be preserved. Depending on their sequence composition, individual promoters appear to vary in their sensitivity to different types of C methylation and rely on different DMTases and histone-modifying enzymes to maintain silencing.

Recent work showing that RNA can target homologous DNA sequences for elimination during macronuclear development in Tetrahymena [101] illustrates another possible outcome of RNA–DNA sequence recognition in the nucleus. Determining the full scope of RNA-directed genome modifications, as well as the evolutionary origins and transcriptional regulation of the guide RNAs themselves, offers exciting prospects for future research.

Acknowledgements

Work in our lab is supported by the Austrian Fonds zur Förderung der wissenschaftlichen Forschung (Grant nrs: Z21-MED and P15611) and the European Union (contract no. QLK3-CT-2000-00078).

References


[40] M. Okano, D. Bell, D. Haber, E. Li, DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development, Cell 99 (1999) 247–257.


