5-7-2014

The Mechanism of Small RNA Biogenesis, Degradation, and Function in *Arabidopsis*

Meng Xie
University of Nebraska-Lincoln, meng.xie2008@gmail.com

THE MECHANISM OF SMALL RNA BIOGENESIS, DEGRADATION, AND FUNCTION IN ARABIDOPSIS

by

Meng Xie

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Biological Sciences

Under the Supervision of Professor Bin Yu

Lincoln, Nebraska

May, 2014

THE MECHANISM OF SMALL RNA BIOGENESIS, DEGRADATION, AND FUNCTION IN ARABIDOPSIS

Meng Xie, Ph.D.

University of Nebraska, 2014

Advisor: Bin Yu

Eukaryotic small RNAs play important roles in many biological processes through sequence-specific RNA silencing. In plants, there are mainly two small RNAs triggering gene silencing: microRNAs (miRNAs) and small interfering RNAs (siRNAs). The biogenesis and precise regulation of small RNA abundance are crucial for plant growth, development, genomic stability, and the resistance to both abiotic and biotic stresses. In this study, we used *Arabidopsis thaliana*, the model plant, to study the mechanism of RNA-directed DNA methylation (RdDM), in which siRNAs can trigger DNA methylation and gene silencing. In addition, we investigated the mechanism of miRNA biogenesis and degradation. For RdDM, we demonstrated that two SUPPRESSOR OF GENE SILENCING3 (SGS3)-like homologs named FACTOR of DNA METHYLATION 1 (FDM1) and 2 (FDM2) are required for *de novo* methylation established by RdDM. DNA methylation level and siRNA level are reduced significantly in *fdm*1 and 2 mutants. FDM1 and 2 are potential RNA-binding proteins with four domains: zinc-finger, XS, Coil-coiled, and XH domains. By studying the function of each domain, we propose that FDM1/2 can form a complex with other SGS3-like proteins and acts as a scaffold to stabilize the AGO4-siRNA-POL V transcripts, which is the essential structure to trigger *de novo* methylation. For miRNA biogenesis, we studied two protein factors: TOUGH and CDC5. We observed that TOUGH and CDC5 are required for

proper function of DCL1 and miRNA processing. Moreover, TOUGH and CDC5 associate with key components in DCL1 processing complex, such as HYL1 and DCL1, which generate miRNAs. In addition, CDC5 interacts with DNA-dependent RNA polymerase II and is a positive transcription factor of genes encoding miRNAs. For miRNA degradation, we studied the mechanism underling miRNA uridylation catalyzed by one nucleotidyl transferase termed HESO1. Our data suggest that AGO1, the effector protein of miRNAs in *Arabidopsis*, associates with HESO1, which is necessary for the uridylation and degradation of unmethylated miRNAs by HESO1.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

First, I would like to express my appreciation to my advisor, Dr. Bin Yu, for his great mentorship and support, and his creative and constructive guidance to me help me through my whole Ph.D. study. Great thanks to all my committee members, Dr. Heriberto D. Cerutti, Dr. Audrey L. Aktin, Dr. David R. Holding and Dr. Chi Zhang for their advice and encouragement. Thanks to Dr. Thomas E. Elthon for helping me with high-perfermance liquid chromatography and sharing machines with us, otherwise I would not get the gel filtration data of FDM1, which is the key result for my conclusion of FDM1 function in RdDM.

I am also deeply appreciated to all of my lab members and my friends. Thanks to all the lab members, both present and past, Dr. Guodong Ren, Dr. Shuxin Zhang, Dr. Zuyu Chen, Dr. Xiang Liu, for their support and help in my experiments. Specifically thanks to Dr. Guodong Ren and Dr. Shuxin Zhang for their collaborations in CDC5, TOUGH, and HESO1 projects. I learned a lot from them. With their help, I learned several critical techniques used in my project. Thanks to members in Dr. Chi Zhang lab, Dr. Yongchao Dou and Dr. Tang Lu, for their data analysis, which gave me helpful ideas on my project. Thanks to Dr. Yong Ding, the past member of Dr. Zoya Avramova, for being generous to discuss with me on technical problems and give me advice. I would like to extend my gratitude to Dr. Joe Zhou and Christian Elowsky from the UNL Microscopy Core Facility for their patience and invaluable feed back in assisting me with confocal microscopy.

Finally, I want to express my great thanks to my family members, to my father, my mother, and my sister, for believing in me continuously, for loving me unconditionally, for supporting me at all times.

LIST OF FIGURES

CHAPTER 1

Literature Review

1. Overview

In eukaryotes, small RNAs play important roles in many biological processes through sequence-specific RNA silencing [1]. Due to differences in precursors and biogenesis processes, small RNAs can be classified into three major types: small interfering RNAs (siRNAs), microRNAs (miRNAs), and Piwi-interaction RNAs (piRNAs) [1]. siRNAs are usually 21-24 nucleotides (nt) in size and arise from long double-stranded RNAs (dsRNAs), which are often derived from inverted repeats, transposable elements (TEs) and viral replication. [2,3]. In contrast, the majority primary transcripts of miRNAs (pri-miRNAs) are generated by the DNA-dependent RNA polymerase II (Pol II) from miRNA coding genes (*MIR*). Pri-miRNAs contain stem-loop structures, where miRNAs reside, and are cleaved by RNase III enzymes (Drosha and Dicer in animals; DICER-LIKE in plants) into 21-24 nt mature miRNAs [2,4]. Different from siRNAs and miRNAs, piRNAs are specific to animals [5,6]. The length of piRNAs is usually 24-32 nt, longer than siRNAs and miRNAs [5,6]. Precursors of piRNAs are believed to be ssRNAs, which are independent of Dicer for processing [6]. Similar to siRNAs, sources of piRNAs are TEs, intergenic regions, and certain genes [7-9].

Upon production, miRNAs and siRNAs are loaded onto members of the ARGONAUTE (AGO) protein family to form an RNA-induced silencing complex (RISC). miRNAs and siRNAs then guide AGO to repress the expression of genes at post-transcriptional levels through target miRNA cleavage or translational inhibition, or at transcriptional levels through directing DNA methylation or histone modifications [1-4]. Recent studies have

established the framework of the mechanism governing miRNA- and siRNA-mediated gene silencing. They are summarized below.

2. RNA-dependent DNA methylation (RdDM)

2.1 Introduction

Epigenetics is the study of gene expression regulation caused by molecular modifications of chromatin such as DNA methylation (5-Methylcytosine) and histone modification rather than genetic information changes like DNA sequence alteration [10]. It is a very active topic of contemporary biology because many diseases such as cancer are related to abnormal chromatin modifications [11,12]. In plants, over 50% of the genome, including centromeric region and repetitive sequences, is methylated. This suggests that plants are excellent resources to study mechanisms controlling DNA methylation [13,14]. Consequently *Arabidopsis thaliana* (*Arabidopsis*), a model flowering plant, has been used for studying DNA methylation mechanism for decades.

Different from mammals in which methylation mainly occurs on cytosine in CG sites, cytosine methylation in plants commonly occurs in three sequence contexts: CG, CHG, and CHH, where H represents any nucleotide other than guanine [15]. During plant cell division, DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2), a DNA methyltransferase creates new methylation marks on DNA via *de novo* methylation [16]. On the other hand, more methyltransferases are involved in the maintenance of DNA methylation by adding methylation marks to daughter strands after DNA replication according to methylation patterns of parental strands [17-19]. CG methylation is

preserved during cell division and DNA replication by maintenance pathway. In contrast, CHG and CHH have been found to require *de novo* methylation for their maintenance in DNA replication cycles [15]. A conserved *de novo* DNA methylation mechanism is RNA-directed DNA methylation (RdDM) by which small interference RNAs (siRNAs) trigger DNA methylation. RdDM was firstly found in transgene silencing in potato [20]. Later, RdDM was recognized as a transcriptional gene silencing mechanism existing in plants and some animals [21,22]. RdDM is involved in diverse epigenetic processes such as transgene silencing, transposon suppression, gene imprinting [11, 23-26].

2.2 Current Model of RdDM in *Arabidopsis*

Recent studies from *Arabidopsis* have greatly increased our understanding of the mechanism of RdDM. Many components critical for RdDM have been identified by genetic and proteomic approaches. Studies on these genes have established the framework of RdDM. As shown in Figure 1-1, the RNAse III enzyme DICER-LIKE 3 (DCL3) produces ra-siRNAs from dsRNAs synthesized by RNA-dependent RNA polymerase 2 (RdR2) from single-stranded RNAs [27], which are thought to be produced by plant-specific DNA-directed RNA polymerase IV (Pol IV) from RdDM target loci [28-31]. ARGONAUTE 4 (AGO4) binds ra-siRNA to form an AGO4–ra-siRNA complex [32-34], which is recruited to chromatin by interaction of AGO4 and plant specific DNA-directed RNA polymerase V (Pol V) [35] and/or base pairing between siRNA and Pol V-dependent transcripts [36,37]. Recruitment of AGO4 to some low-copy-number loci also requires DNA-directed RNA polymerase II (Pol II) [38]. After loading onto chromatin, AGO4 is thought to recruit the protein DRM2, which then

catalyzes *de novo* cytosine DNA methylation at symmetric CG or CHG sites and asymmetric CHH sites [16,35,37]. The KOW-CONTAINING TRANSCRIPTION FACTOR 1/ SPT5-LIKE protein (KTF1/SPT5L) is required for RdDM. Its interaction with chromatin, AGO4 and Pol V-dependent transcripts is thought to assist the recruitment of DRM2 to chromatin [39,40]. Recruitment of SPT5L to Pol V-dependent transcripts and chromatin is AGO4- independent [41]. CLASSY 1 (CLSY1), a chromatin-remodeling protein, and SAWADEE HOMEODOMAIN HOMOLOG 1 (SHH1)/DNA-BINDING TRANSCRIPTION FACTOR 1 (DTF1) are essential for ra-siRNA accumulation and DNA methylation [42-44]. These three proteins are co-purified with Pol IV, indicating that they form a complex [43]. DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1 (DRD1; a chromatin-remodeling protein), DEFECTIVE IN MERISTEM SILENCING 3 [DMS3; a protein containing a hinge domain of structural maintenance of chromosome (SMC) protein], and RNA-DIRECTED DNA METHYLATION 1 (RDM1; a methylated DNA-binding protein) are required for generation of Pol V-dependent transcripts and RdDM [10,45-47]. It has been shown that DRD1, DMS3 and RDM1 function as a complex in RdDM [10]. RDM1 also interacts with AGO4 and DRM2, and may help recruit the silencing complex to chromatin [47].

2.3 Two plant specific DNA-dependent RNA polymerases (POL IV and POL V) are essential for RdDM in *Arabidopsis*.

2.3.1 Overview

Besides POL I, II, and III, plants also contain another two DNA-dependent RNA polymerases, which are named as DNA-dependent RNA polymerases IV (POL IV) and

DNA-dependent RNA polymerases V (POL V). POL IV and POL V are plant-specific and conserved among different plant species, such as *Arabidopsis*, rice, and maize [28,48]. Although POL IV and POL V are Pol II-like enzymes, they have evolved specialized roles in the production of noncoding transcripts for siRNA biogenesis and genomic DNA methylation [10].

2.3.2 The role of POL IV and POL V in RdDM

RdDM pathway and proper DNA methylation require POL IV and POL V function. POL IV and POL V loss-of-function mutants show significant deficiencies on siRNA accumulation and DNA methylation levels in many RdDM target loci [28,49]. However, roles of POL IV and POL V in RdDM pathway are different. Deep sequencing analysis of siRNA population in wild type and mutants of POL IV or POL V shed light on different roles of POL IV and POL V [50]. The majority of detected siRNAs (>94%) are dependent on POL IV for accumulation. In contrast, not all the POL IV-dependent siRNAs require POL V for their accumulation even thought methylation levels of their corresponding targets are almost eliminated in POL V mutants. These results suggest that POL IV is key for siRNA production in RdDM but the role of POL V in siRNA biogenesis may be separated from its role in DNA methylation [50]. Based on the requirement of POL IV or POL V for accumulation, siRNAs can be classified into to types [38,50,51]. Type I siRNAs require both POL IV and POL V for their production and they are usually high-copy-number repeats or transposons, such as AtSN1, siR1003, and Copia2 [28,30,31]. Type II siRNAs only depend on POL IV for accumulation and the majority of them are low-copy-number repeats and intergenic sequences, for example siR02, Cluster2, and soloLTR [31,52].

The study on low-abundance intergenic noncoding (IGN) transcripts that are produced from flanking regions of RdDM loci in *Arabidopsis* revealed the role of POL V in DNA methylation [36]. The largest subunit of POL V (NRPE1) associates with the IGN region and is required for the accumulation of IGN transcripts, suggesting Pol V may be responsible for the transcription of IGN RNAs. The POL V-dependent transcripts from IGN regions are independent of POL IV, DCL3 or RDR2. In *nrpe1*, the association of AGO4 with target DNA loci and DNA methylation are eliminated [37], suggesting these POL V-dependent transcripts may act as scaffolds to recruit the downstream silencing machinery, such as AGO4 [37]. Similar to *Arabidopsis*, gene silencing in fission yeast also requires transcripts (POL II-dependent) for the establishment of DNA methylation and heterochromatin [53].

2.3.3 Structural features of POL IV and POL V

Similar to POL II, POL IV and POL V are also large protein complexes, with a molecular mass close to 1 Megadalton (MDa), containing multiple subunits [31]. The largest subunits of POL IV and POL V are NRPD1 and NRPE1 respectively, which share similarities with the largest subunits of POL I, II, and III [31,54]. The N-terminal regions of NRPD1, NRPE1, and NRPB1 (largest subunit of POL II) are highly conserved. All the three polymerases contain evolutionary conserved regions A to H [31]. However, the C-terminal shows variations among these polymerases, which are proposed to cause differences in polymerase activities of POL II, IV, and V. Different from the C-terminal domain (CTD) of NRPB1, the C-terminal of NRPD1 shares similarity with the C-terminal half of a nuclear-encoded protein named DEFECTIVE CHLOROPLAST AND LEAVES (DCL), which regulates rRNA processing in chloroplasts [31]. Compared with

NRPB1 and NRPD1, NRPE1 shows additional features in its C-terminal: firstly, NRPE1 has a long CTD that extends beyond the DCL-like motif. Secondly, NRPE1 has multiple potential phosphorylation sites in a highly hydrophilic domain composed by ten complete repeats of a 16-amino-acid consensus sequence [31].

The other reason for functional variations of POL II, IV, and V is the subunit differences. Although most of their subunits are paralogous or identical to the 12 subunits of POL II, POL IV and POL V have their own specific subunits [54]. In the POL IV complex, there are four subunits distinct from their POL II paralogs, while POL V has six distinct subunits from POL II. Even between POL IV and POL V, there are four subunits with different features [54]. More interestingly, the subunit differences occur in key positions relative to the template channel and RNA exit paths, which determine specific polymerase activity and specific target recognition [54]. In summary, variations in the largest subunit and other complex components cause the functional divergence among Pol II, POL IV and POL V.

2.3.4 Functions of POL IV and POL V subunits

Besides NRPD1 and NRPE1, the largest subunit of POL IV and POL V respectively, as well as several smaller subunits of POL IV and POL V, are studied by mutation analysis. NRPD2/NRPE2 is the second largest subunit, which is shared by POL IV and POL V. Without NRPD2/NRPE2, both siRNAs produced from RdDM loci and DNA methylation are almost undetectable, suggesting NRPD2/NRPE2 is essential for the function of both Pol IV and Pol V [29]. The other well-studied subunit required for RdDM is NRPD4/NRPE4, which is also a common subunit for POL IV and POL V.

NRPD4/NRPE4 shares sequence similarity with NRPB4, a subunit of POL II, but has unique functions different from NRPB4 and POL II. NRPD4/NRPE4 was found physically associated with NRPD1 and NRPE1 [55]. NRPD4/NRPE4 mutants showed reduction in DNA methylation levels and siRNA accumulation at DNA loci regulated by RdDM, such as 5S rDNA and AtSN1, which illustrate that NRPD4/NRPE4 is involved in RdDM [55]. In addition, NRPE5 was found to function exclusively in POL V [56]. However, mutation analysis does not reveal the function of other subunits of POL IV and POL V in RdDM, suggesting that they are redundant or that they are non-essential for the function of POL IV and POL V in RdDM.

2.3.5 Protein factors related with POL IV and POL V function

In eukaryotes, POL II requires transcription factors for proper activity. Indeed, POL IV and POL V also require protein partners. Several proteins associated with POL IV and POL V have been identified.

CLSY1, a putative chromatin-remodeling factor, was identified involved in the production of 24 nt siRNAs and the spreading of transcriptional gene silencing (TGS) signals [42]. CLSY1 contains one SNF2 domain and one helicase domain. Studies by Smith et al [42] using immunolocalization technique showed that in wild-type cells, signals of NRPD1 display punctate distribution in the nucleoplasm (small foci) and near the chromocenter periphery for NRPD1. However in CLSY1 loss-of-function mutants, NRPD1 localizes in only one to three large foci or is diffuse without detectable foci. In addition, RDR2 can be detected in nucleoplasmic foci, nucleolar dots, and nucleolar perimeter ring in wild-type plants. In contrast, when CLSY1 is knocked out, the majority

of RDR2 signals can only be detected in nucleoplasmic foci [42]. These results suggest that CLSY1 affects the nuclear localization of NRPD1 and RDR2 and may regulate siRNA production at the level of POL IV or RDR2 activity.

SUPPRESSOR OF TY INSERTION5-LIKE (SPT5L, also known as KTF1), the homolog of the yeast transcription elongation factor Spt5, was identified as a potential transcription factor associated with NRPE1 [39,40]. Rowley et al [41] found that the chromatin association of POL V is independent on SPT5L. However SPT5L requires POL V for its proper chromatin association at RdDM target loci illustrating that SPT5L acts downstream of POL V, which is consistent with the transcription elongation function of the yeast homolog [41].

Besides SPT5L, a homolog of yeast transcription factor IWR1 termed RDM4/DMS4 was identified by forward genetic screening and determined to affect the accumulation of 24 nt siRNAs [57,58]. RDM4/DMS4 loss-of-function mutants showed significant reduction in POL V-dependent transcripts, suggesting that POL V requires the assistance of RDM4/DMS4 to produce scaffold transcripts [57,58]. Consistently, RDM4/DMS4 has been found to physically interact with NRPE1, the largest subunit of POL V, which further supports the role of RDM4/DMS4 in POL V transcription [58].

Two protein factors aiding in POL V chromatin association were also identified through genetic screens: DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1 (DRD1), a putative chromatin-remodeling factor with SNF2 domain, and DEFECTIVE IN MERISTEM SILENCING 3 (DMS3), a protein with a domain that is similar to the hinge region of structural maintenance of chromosome proteins (SMC) [36,37,45,46,59-61].

Chromatin immunoprecipitation (ChIP) analysis showed that without DRD1 or DMS3, the association of NRPE1 with chromatin regions generating POL V-dependent IGN transcripts is impaired [36,37]. In addition, the production of POL V-dependent IGN transcripts was undetectable in DRD1 and DMS3 null alleles [45,46]. Furthermore, affinity purification studies showed that DRD1 and DMS3 can be co-purified with POL V subunits suggesting that DRD1 and DMS3 function together with POL V and act downstream of RdDM [61].

2.4 RDR2 produces a long double-strand siRNA precursor.

RNA dependent RNA polymerases (RDRs), with a conserved RNA-dependent RNA polymerase catalytic domain, can use ssRNA molecule as a template to synthesize dsRNA [62]. These proteins have been identified in plants, fungi, *C. elegans*, but not in mice or human [62]. Based on phylogenic analysis, eukaryotic RDRs can be classified into three major types: RDRα, RDRβ, and RDRγ [63]. RDRα exists in both plants and lower animals. RDRβ is specific in lower animals, while RDRγ is specific in plants [63].

There are six identifiable RDRs in *Arabidopsis*, RDR1-6 [64]. Among six RDRs, RDR1, 2, and 6 share the C-terminal canonical catalytic DLDGD motif of eukaryotic RDRs [64] and are well studied. Initially, the three RDRs were thought to be involved in plant anti-virus mechanism, such as posttranscriptional gene silencing (PTGS). However, further studies made it apparent that they have unique molecular functions, even though all of them are belong to the RDRα subfamily [65]. Unlike RDR1, RDR2, and RDR6, the other three RDRγ proteins in *Arabidopsis* have not been assigned functions.

RDR1, 2, and 6 function in synthesizing double-strand RNA (dsRNA) molecules using

single-stranded RNAs as templates. The resulting dsRNAs are cleaved into different types of siRNAs targeting specific endogenous loci [65]. RDR1 is involved in the production and amplification of virus-derived siRNAs and may protect plants from virus infection [66-68]. RDR1 is critical for the production of the majority of virus-derived siRNAs based on the analysis of small RNA library [68]. In addition, studies by Wang et al [68] showed that RDR1 preferentially amplified siRNAs derived from 5' terminal of viral RNAs. RDR6 has multiple defined functions, including pathogen defense, abiotic stress response, and plant development [62]. Together with AGO1 and DCL1, elements of miRNA pathway, RDR6 acts to amplify siRNAs. With the assistance of SUPRESSOR OF GENE SILENCING3 (SGS3), a dsRNA-binding protein which prefers 5'-overhang-containing dsRNAs, RDR6 converts partially AGO-cleaved transcripts into dsRNAs, which will be processed into 21-24 nt siRNAs by DCL4/DCL1 to down-regulate the expression of targets [69-72]. At least two types of siRNAs are dependent on RDR6 for amplification: TRANS-ACTING siRNA (ta-siRNAs) generated from non-protein-coding precursors targeted by miRNAs and nat-siRNAs processed from overlapped double-stranded regions formed by sense-antisense transcripts generated from opposite coding strands [73-75]. As miRNAs, ta-siRNAs and nat-siRNAs silence genes by cleaving target RNAs [73,75]. Both ta-siRNAs and nat-siRNAs are involved in regulating development and biotic and abiotic response of plants [74-76].

Among RDR1, RDR2, and RDR6, RDR2 is the only one participating in 24 nt siRNA-mediated DNA methylation. RDR2 T-DNA insertion mutants lack the 24 nt siRNAs for RdDM pathway, such as siRNA02, AtSN1, Cluster2, and siRNA 1003 [27]. However miRNA and ta-siRNA production is unaffected in *rdr2*, suggesting RDR2 functions

specifically in RdDM pathway. Consistent with the loss of siRNAs production, DNA methylation levels in *rdr2* mutants are reduced significantly [27]. Different from RDR6, RDR2 usually acts together with RdDM components like POL IV, POL V, DCL3 and AGO4 [28-31,33,79,80]. In fact, RDR2 co-localizes with NRPD1, NRPE1, DCL3, and AGO4 in nucleolar dots [60]. The fact that 98.5% of POL IV-dependent siRNAs are lost in *rdr2* mutant and that POL IV and RDR2 are physically associated *in vivo* suggests that RDR2 functions together with POL IV to synthesize double-strand siRNA precursors [51]. Consistent with this notion, *in vitro* biochemical studies show that RDR2's polymerase activity is dependent on POL IV. In the absence of POL IV, RDR2 does not synthesize RNA fragments using DNA-RNA bipartite templates [80].

Two biochemical activities of RDR2 have been proposed based on the studies of a *Neurospora* RDR gene termed QUELLING DEFECTIVE1 (QDE1), which acts in RNA silencing and DNA repair pathways [81,82]. The observation that QDE1 can use RNA template to synthesize a RNA ladder with RNA products of all sizes demonstrates that QDE1 is able to initiate 3' to 5' transcription in the middle of mRNAs, which is independent of template. On the other hand, QDE1 also has the activity to start the synthesis from the free 3' terminal of mRNA templates [83]. According to the dual role of QDE1, firstly RDR2 may move together with POL IV along DNA and synthesize a series of discontinuous second strands from the internal of POL IV-dependent transcripts before the termination of POL IV transcription, which is analogous to lagging-strand Okazaki fragment generated during DNA duplication. The second possibility is that RDR2 may use complete transcripts of POL IV as templates. In this way, RDR2 can initiate transcription from the free 3' end and generate the full-length fragment [84].

2.5 Dicer proteins involved in RdDM.

Dicer proteins are multi-domain ribonucleases that process dsRNAs to release a ~21-24 bp RNA duplexes, which have a 5' phosphate and a 2nt 3' overhang at each strand [85]. Six domains are included in Dicer proteins: DEAD box, helicase-C, DUF283, PAZ, RNase III, and dsRBD [86]. PAZ, RNaseIII and dsRBD are thought to be responsible for dsRNA cleavage and binding, respectively [87]. PAZ domain is connected with RNaseIII domain by a long α helix, binds the 3' terminal nucleotide of a dsRNA with a 2nt 3' overhang, and therefore, is critical for substrate recognition [87]. Structural and biochemical analysis suggest that Dicer functions by forming an intermolecular dimer with two RNaseIII domains [88], each of which hydrolyzes one strand of the substrate.

Vertebrates encode one Dicer to generate both miRNAs and siRNAs. In contrast, plants posses several Dicer-like genes (DCL) to meet the requirement of multiple small RNA pathways. Four Dicer-like genes exist in *Arabidopsis*, DCL1-DCL4 [85]. All of them have RNaseIII activity and can cleave double-strand RNAs into short double-strand RNA fragments. They show distinct roles in small RNA biogenesis. DCL1 is primarily responsible for miRNA generation [27]. DCL2 and DCL4 are mainly related with the generation of viral siRNAs, such as cucumber mosaic virus and cauliflower mosaic virus [89-90]. However, DCL3 is responsible for the production of 24 nt siRNAs used in the RdDM pathway. Long double-stranded siRNA precursors are cleaved by DCL3 into short siRNA duplexes [91]. Without proper DCL3 function, most 24 nt siRNAs involved in RdDM will be eliminated [91]. On the other hand, DCLs also have partially overlapped functions. For example, DCL1 has also been found to be involved in the siRNA pathway for specific loci [92]. DCL2, 3, and 4 also have overlapping functions in 24 nt siRNA

production [70].

Similar to Dicer proteins in animals, plant Dicers have been found to be associated with double-stranded RNA-binding proteins (dsRBPs) [92,93]. *Arabidopsis* contains five potential dsRBPs, termed DSRNA-BINDING PROTEIN1-5 (DRB1-5). Interestingly, the four Dicers have preference for dsRBPs. DCL1 exclusively couples with DRB1/HYL1 for its function. In contrast, DCL4 operates exclusively with DRB4 [93]. DCL2 and DCL3 do not need dsRBPs to produce siRNAs [93].

In conclusion, four Dicers in *Arabidopsis* act redundantly and hierarchically. The associated dsRBPs may determine specific substrate recognition of DCLs and cause distinct functions of the four Dicers for plants.

2.6 AGO4 acts downstream of RdDM

ARGONAUTE (AGO) proteins are the effector proteins in small RNA-induced gene silencing pathways. They exist in most eukaryotes and bind the three major small RNAs, miRNA, siRNA, and piRNA to form RISC in order to cleave mRNAs or trigger DNA modifications [94,95]. AGO usually contains four major domains: N-terminal domain, PAZ, MID and PIWI domains [96]. Crystal structure and biochemical analysis revealed that the PAZ domain binds to the 3’ end of small RNA and the MID domain binds to the 5’ end of small RNA [96]. The PIWI domain shows similarity to ribonuclease-H enzyme with conserved Asp-Asp-Asp/Glu/His/Lys motif and is responsible for the cleavage of target mRNAs [97]. However, not all AGOs have such slicer activities to cleavage mRNAs.

Animals and plants encode multiple AGOs, which often have specific function in various small RNA pathways. There are ten AGO proteins (AGO1-AGO10) in *Arabidopsis,* which can be classified into three groups based on sequence similarities: Group1, AGO1, AGO5, and AGO10; Group 2, AGO2, AGO3, and AGO7; Group 3, AGO4, AGO6, AGO8, and AGO9 [98]. Among ten AGOs, AGO1 is the effector protein for most miRNAs while AGO10 and AGO7 bind to specific miRNAs. AGO4, AGO6 and AGO9 have been shown to act in 24 nt siRNA-mediated DNA methylation [98]. In addition, the AGOs of *Arabidopsis* show preference on 5' nucleotides. For instance, AGO2 and AGO4 preferentially recruit small RNAs with 5' terminal adenosine, while AGO5 prefer to bind small RNAs with 5' terminal cytosine [99].

The function of AGO4 in RdDM has been extensively studied. AGO4 has slicer activity. However, its function in RdDM is independent of its slicer activity. AGO4 binds RdDM loci and lack of AGO4 significantly reduces DNA methylation and siRNA amplification. *In vivo* immunolocalization analysis demonstrates that AGO4 either co-localizes with NRPE1 in Cajal bodies, which are a dynamic compartments for siRNA processing, or with NRPE1, NRPE2 and DRM2 at a separate discrete nuclear body termed as the AGO4-NRPE1 (AB) body, which is a potential active site for RdDM [100,101]. Further studies show that AGO4 physically interacts with NRPE1, the largest subunit of POL V, through GW/WG repeats in CTD region of NRPE1 [35]. In addition, AGO4 is associated with POL V transcripts and is dependent on POL V transcripts for its chromatin association, AGO4/siRNA complex is proposed to interact with POL V transcripts by the base pairing between siRNA and POL V transcripts [36,37]. In *nrpe1*, the association of AGO1 with chromatin is disrupted, suggesting that the Pol V-AGO4 interaction and the

association of AGO4-siRNA with Pol V-dependent transcripts may recruit AGO4 to the RdDM target regions to trigger DNA methylation [37].

Previously, RdDM was thought to be solely nuclear process because both biogenesis and functioning of 24 nt siRNAs take place in nucleus. However by separately deep sequencing siRNA populations in cytoplasm and nucleus, recent studies discovered that the abundance of individual 24 nt siRNAs is about ten times higher in cytoplasm compared with the nucleus [102]. The majority of cytoplasmic 24 nt siRNAs are duplexes while 24 nt siRNAs in nucleus are single-stranded. Furthermore, a small fraction of AGO4 can be detected in the cytoplasm and associated with only single-stranded cytoplasmic 24 nt siRNAs but not duplexes. This suggests that in cytoplasm the passenger strand of siRNA duplex is removed by AGO4 slicer activity in order to form mature AGO4/siRNA complex and RISC. These results reveals that the loading of siRNAs into AGO4 seems to occur in cytoplasm and that the formation of mature AGO4/siRNA complex is critical for their selective nuclear import, which may be another regulatory pathway of RdDM.

Besides the transcription elongation function associated with POL V, SPT5L/KTF1 is also an adapter of AGO4 and aids in the recruitment of AGO4 to POL V-dependent transcripts. Similar to NRPE1, SPT5L physically interacts with AGO4 through its GW/WG repeats motif [39,40]. *In vivo* RNA-immunoprecipitation experiments showed that similar to AGO4, SPT5L binds POL V-dependent transcripts indicating the adapter role of SPT5L for AGO4 and POL V-dependent transcripts. Besides the adapter function, SPT5L is also involved in POL V-dependent transcription or the production of POL V-dependent transcripts [40].

2.7 Methyltransferases involved in *de novo* methylation by RdDM

Three methyltransferases have been identified that are involved in plant DNA methylation: DRM2, DNA METHYLTRANSFERASE 1 (MET1), and CHROMOMETHYLASE 3 (CMT3) [10]. MET1 is the plant homolog of DNMT1, which is the methyltransferase responsible for maintaining DNA methylation patterns during cell division in mammals [18,77]. DRM2 also has a mammalian homolog termed DNMT3, the *de novo* methyltransferase setting up DNA methylation patterns in the early stage of development [16,77]. However, CMT3 is a plant specific methyltransferase without any mammalian homolog, which has been found to be involved in the maintenance of CHG methylation [79,103].

The three methyltransferases have different functions in the establishment and maintenance of DNA methylation during cell division. The establishment of DNA methylation is mainly catalyzed by DRM2 via *de novo* methylation [16]. Considering the fact that DRM2 can be detected in RdDM downstream complex and is associated with AGO4, it is proposed that DRM2 is recruited to chromatin by AGO4/siRNA/POL V-dependent transcripts complex [61]. In contrast, the maintenance of DNA methylation is dependent on DRM2, MET1, and CMT3. However different sequence contexts require different enzymes for the maintenance: CG methylation by MET1, CHG methylation by CMT3, and CHH methylation by DRM2 [77].

2.8 Challenges in understanding RdDM mechanism

In order to comprehensively study DNA methylation regulation in *Arabidopsis*, Hume et al [104] analyzed the methylome of 86 *Arabidopsis* gene silencing mutants by whole-

genome bisulfite sequencing (BS-seq). Their findings suggest that the mechanism of establishment and maintenance of plant DNA methylation is much more complicated than previously thought. The current RdDM model cannot cover all the loci regulated by DNA methylation and DNA methylation is regulated in a site-specific manner involving interplays between different pathways and different protein factors [104]. For example, some specific sites are regulated by RNAi factors but not *de novo* methylation pathway [104]. There is a POL II related pathway for DNA methylation, which is independent of POL IV and POL V [104]. In addition, new protein factors controlling DNA methylation have been identified, such as SUVH5/6 and CAF-1 complex. They are involved in different DNA methylation pathway from RdDM and *de novo* methylation [104]. Studies of Dominique et al [105] have defined 21 nt siRNA-dependent chromatin-based pathway in *Arabidopsis* for the methylation of psORF and AT1TE93275 loci. Apart from 24 nt siRNA-dependent RdDM, this pathway requires PTGS factors, such as SILENCING DEFECTIVE 3 (SDE3), RDR6 and AGO2, and NEEDED FOR RDR2-INDEPENDENT DNA METHYLATION (NERD), an unmethylated H3K4 binding protein.

In conclusion, for the thousands of RdDM target loci in *Arabidopsis*, not all of them follow the model to establish, maintain, and modify their DNA methylation patterns. In order to understand how plants accurately target, maintain and even modify DNA methylation patterns of specific loci in plants, more protein factors involved in plant DNA methylation and more methylated loci need to be studied in detail.

3. MiRNA biogenesis and degradation

3.1 Introduction

MicroRNAs (miRNAs) are ~22 nt noncoding RNAs, which are indispensable for various biological processes in plants and animals, such as development, physiology, and stress response [106-108]. The first miRNA discovered by scientists is *lin-4* miRNA in *C. elegans*, which is generated from *lin-4* gene and repress on the translation of *lin-14* mRNA to LIN-14 protein [109]. Later, numerous miRNAs were discovered in various organisms, such as human and plants. In the human genome, there are over 1000 miRNAs identified, which are predicted to target about 60% of all protein-coding genes [110,111]. Expression and functional studies demonstrate that miRNAs exist in various cell types and tissues and participate in the regulation of many cellular processes [106-108]. In addition, various human pathologies are correlated with dysregulation of miRNAs. For example, dysfunction of miR-96 can cause hereditary progressive hearing loss [112]. MiR-21 is involved in several types of cancer, such as glioblastoma and astrocytoma [113]. In *Arabidopsis*, more than 100 miRNAs have been identified by both genetic and bioinformatics approaches. Aberrant reduction or elevation in miRNA levels can cause many developmental and physiological defects. For instance, miR172 loss-of-function mutants show late flowering, supernumerary petals and stamens, while overexpression of miR172 can induce early flowering, lack of petals, and transformation of sepals to carpels [114,115]. Thus, the accumulation of miRNAs needs tight control for correct function in plants.

Studies on miRNA biogenesis, and functional mechanism illustrate that plant and animal miRNA pathway share many similarities [116]. For example, both plant and animal miRNAs depend on dicer proteins for their production. miRNAs need to bind to AGO

proteins to form RISC to repress target gene expression through target cleavage and/or translational inhibition. However, miRNA pathways in plants and animals are not exactly the same. For instance, the biogenesis of miRNAs only occurs in the nucleus in plants; while in animals generation of miRNAs need both cytoplasm and nuclear processes [116]. Studies in the past decades have established a general model for miRNA pathway in plants.

3.2 Overview of miRNA pathway in *Arabidopsis*

The model for miRNA pathway in plants is shown in Figure 1-2. In *Arabidopsis*, the majority of miRNA genes are located in intergenic regions and encoded as independent transcriptional units. DNA-dependent RNA polymerase II (POL II) generates primary transcripts of miRNAs (pri-miRNAs) from miRNA loci [117,118]. After transcription, nuclear pri-miRNAs are then initially processed by DCL1, an RNase III endonuclease, into miRNA precursors (pre-miRNAs), which are stem-loops with a 2 nt 3' overhang and harbor the miRNA/miRNA* [86]. Then pre-miRNAs are cleaved by DCL1 again to produce miRNA/miRNA* with 2 nt 3' overhangs [119]. In *Arabidopsis*, the zinc finger protein SERRATE (SE) and the dsRNA binding protein HYPONASTIC LEAVES1 (HLY1/DRB1) work together with DCL1 for pri-miRNA processing efficiency and accuracy. After DCL1 processing, the small RNA methyltransferase HUA ENHANCER1 (HEN1) adds a methyl group to the 3' end of miRNA/miRNA* duplex to stabilize them [120]. Most miRNA molecules exit the nucleus and enter the cytoplasm with assistances of HASTY (HST), the plant homolog of EXPORTIN 5, which is responsible for the nucleocytoplasmic transport of miRNAs in animals [121]. However, not all plant

miRNAs require HST for nuclear export [121]. The major effector of miRNAs in *Arabidopsis* is AGO1, which majorly represses the expression of target through cleavage or translational inhibition [122].

3.3 Regulation of miRNA abundance

Because the proper abundance of miRNAs is crucial for growth and development, plants have evolved multi-tiered and sophisticated regulative systems to precisely control miRNA levels in an acceptable range. Such regulations mainly affect miRNA biogenesis and turnover.

3.3.1 Regulation of miRNA biogenesis

3.3.1.1 Transcriptional Regulation

Two general transcription factors of Pol II have been shown to regulate transcription of pri-miRNAs: Mediator and NOT2. Mediator is a multi-subunit complex, which exists in yeast, plants, and mammals [123,124]. The mediator complex is essential for activator-dependent transcription in eukaryotes [124]. With a large surface area and the potential of protein-protein interaction, the mediator complex acts as a bridge between POL II and transcription factors. In *Arabidopsis*, the mediator complex has been found to interact with transcriptional activators and facilitate POL II recruitment to *MIR* genes [118]. NOT2 is a negative transcriptional regulator and is highly conserved in eukaryotes [125]. Studies in yeast showed that NOT2 is the core component of CARBON CATABOLITE REPRESSION4 (CCR4)-NOT complex, which is involved in mRNA transcription,

mRNA decay and miRNA-directed mRNA degradation [126-128]. Recently, Wang et al [129] revealed that two homolog proteins NOT2a and NOT2b in *Arabidopsis*, which contain conserved NOT2_3_5 domain, are required for the transcription of miRNA genes. In loss-of-function mutants for both NOT2a and NOT2b, the abundance of pri-miRNA and mature miRNA is reduced. However, NOT2a and NOT2b may act as general transcription factors since they also regulate the transcription of protein-coding genes and NOT2b physically interacts with POL II.

Transcription factors specific for some miRNAs family have also been identified. POWERDRESS (PWR), a SANT-domain-containing protein with putative transcription factor and chromatin remodeling activity, has been found to regulate POL II recruitment to some miR172 family members loci and be required for the accumulation of miR172 [130]. The accumulation of some MIR156 family members requires the proper function of transcription factor FUSCA3 [129]. APETALA2 (AP2), a transcription factor involved in seed development, stem cell maintenance, and floral organ identity, is associated with the miR156 and miR172 loci and seems to act oppositely for miR156 and miR172. Impairment of APETALA2 represses miR156 expression and promotes miR172 expression [131]. miRNA gene expression can also be regulated by various stresses via specific transcription factors [132]. For example, the expression of miR398b and c is induced in response to copper deficiency via SQUAMOSA PROMOTER BINDING PROTEIN-LIKE7 (SPL7) [133] while the expression of MYB2 (a transcription factor), which binds to the promoter of miR399f gene, is induced to activate miR399f transcription under phosphate starvation [134].

3.3.1.2 Regulation of processing of miRNA precursors

Processing of miRNA precursors by DCL1 is regulated to ensure the proper levels of miRNAs. Several protein factors have been shown to regulate DCL1 function.

HYL1 and SE are critical for efficient and accurate miRNA processing by DCL1. Although DCL1 alone is able to process pri-miRNAs and pre-miRNAs, its cleavage efficiency and accuracy requires HYL1 and SE, which are RNA binding proteins that physically associate with DCL1 [135,136]. In loss-of-function alleles of HYL1 and SE, misplaced cleavages of several pri-miRNAs were detected by RNA-seq [135]. Actually, DCL1, HYL1, and SE were shown to form small nuclear bodies called Dicer-body (D-body) *in vivo* [137,138]. The fact that pri-miRNAs also localize in D-bodies suggests that miRNA processing may occur in them [138].

Studies on the crystal structure of HYL1 RNA binding domain revealed that HYL1 probably binds to the miRNA/miRNA* duplex region of miRNA precursors as a dimer [139]. On the other hand, the crystal structure of SE showed that the appearance of the SE core is similar to a walking man, in which N-terminal a helices, C-terminal non-canonical zinc-finger domain and novel middle domain resemble the leading leg, the lagging leg and the body, respectively [140]. This scaffold-like structure together with protein and RNA binding capability of SE suggest that SE may act to position miRNA precursor toward the DCL1 catalytic site within miRNA processing machinery [140]. In addition, SICKLE (SIC), a proline-rich protein, co-localizes with HYL1 and is required for the accumulation of a subset of miRNAs, suggesting that it may act as a partner of HYL1 to regulate the biogenesis of some miRNAs [141].

Besides protein-protein interaction, phosphorylation of HYL1 and DCL1 also affects precursor processing. Manavella et al [142] reported the effect of HYL1 phosphorylation status on miRNA processing and identified a new player in miRNA biogenesis termed C-TERMINAL DOMAIN PHOSPHATASE-LIKE1 (CPL1), which was previously found to be able to dephosphorylate a serine motif in CTD of POL II [143]. CPL1 was found to be critical for DCL1 activity and required for accurate precursor cleavage [142]. CPL1 is required to maintain the hypophosphorylated state of HYL1, which is a phosphorylated protein and needs to be dephosphorylated for optimal activity. [142]. In the absence of CPL1, the dephosphorylation of HYL1 and accurate processing and strand selection from miRNA duplexes are compromised [142]. SE is also required for the dephosphorylation of HYL1 [142]. CPL1 physically interacts with SE and lack of SE disrupts the CPL1-HYL1 interaction, suggesting that SE functions as a scaffold to mediate CPL1 interaction with HYL1 [142].

DCL1 is also phosphorylated in vivo, which may be essential for DCL1 function [144]. The forkhead-associated domain (FHA)-containing protein DAWDLE (DDL) was shown to be involved in miRNA biogenesis. The *ddl* mutants are growth delayed, produce defective roots, shoots, and flowers, have reduced seed set and show reduced levels of pri-miRNAs as well as mature miRNAs [144]. DDL binds RNA and physically associates with DCL1, suggesting that DDL is involved in DCL1 function [144]. The crystal structure of DDL FHA domain shows that DDL contains a conserved phosphothreonine binding cleft, which can recognize and bind to the phospho-threonine of DCL1 [145]. Co-immunoprecipitation experiments (Co-IP) showed that the phosphothreonine binding cleft is important for the direct interaction between DDL and

the DCL1 fragments targeted for phosphorylation, suggesting that DCL1 phosphorylation *in vivo* may guide the association between DDL and DCL1 [145].

The transcription of DCL1, HYL1, and SE are regulated to control miRNA processing. Several transcription factors have been shown to regulate their proper expression. STABILIZED1 (STA1), an *Arabidopsis* pre-mRNA processing factor 6 homolog, is required for DCL1 expression. Disruption of STA1 shows decreased DCL1 transcript levels [146]. Histone acetyltransferase GCN5 shows a general repressive effect on miRNA production through inhibiting the transcription of HYL1 and SE [147].

Recently, MODIFIER OF SNC2 (MOS2), an RNA-binding protein, was determined to be involved in the assembling of nuclear dicing body [148]. MOS2 interacts with pri-miRNAs *in vivo* [148]. Although MOS2 does not interact with DCL1, HYL1, or SE, it is required for the recruitment of pri-miRNAs to HYL1 and HYL1 localization in the nuclear dicing body [148]. NOT2s directly interact with DCL1, which is conserved between rice and *Arabidopsis* [129]. Impairment of NOT2s results in the disruption of DCL1 contained D- bodies, suggesting that it affects DCL1 subcellular localization [129].

3.3.1.3 Splicing machinery in miRNA processing

The cap-binding complex (CBC), composed by CAP BINDING PROTEINs CBP20 and CBP80, is required for the correct splicing of the first intron in plants [149] and in animals [150]. Lack of CBP80 and CBP20 reduces the accumulation of miRNAs and increases the abundance of pri-miRNAs, suggesting that they both may be involved in

pri-miRNA processing. However, the function of CBP80/20 in pri-miRNA processing may be independent of their roles in mRNA splicing since the accumulation of both pri-miRNAs with and without introns is increased in *cbc20* and *cbc80* mutants [151]. However, whether or not CBP80/20 affects processing accuracy remains to be determined. Although several splicing factors are involved in miRNA processing, the relationship between splicing machinery and miRNA process is still unclear. In animals, it is proposed that splicing factors regulate miRNA process via the modulation of pri-miRNA structures [152]. In *Arabidopsis*, alternative splicing of pri-miRNAs has been revealed to affect miRNA processing. Studies by Schwab et al [153] showed that introns following the 3' end of the stem-loop of some pri-miRNAs could promote the accumulation of mature miRNAs. Accompanied with reduced mature miRNA level, introns in the 3' end of the stem-loop are spliced efficiently in *dcl1* mutants [153]. However, the underlying mechanism of splicing-regulated miRNA biogenesis requires further studies.

3.3.1.4 Regulation of RISC Assembly

In plants, RISC assembly is monitored to regulate miRNA function. Plant cells need to ensure that the miRNA strand of miRNA/miRNA* duplex is loaded into AGO1 to form RISC. Several AGO1-associated protein factors are critical for this process. HEAT SHOCK PROTEIN90 (HSP90) directly interacts with AGO1 during its association with the guide/passenger duplex [154]. Biochemical studies showed that the disassociation of HSP90 triggered by ATP hydrolysis of HSP90, could promote RISC assembly and

passenger strand removal [155]. In addition, another AGO1-associated protein termed SQUINT (SQN) has the similar function as HSP90 in passenger strand removal [155]. HSP90 is proposed to trigger AGO1 conformational changes by its chaperone activity. In this way, the association of HSP90 to AGO1 can determine whether the passenger strand is removed or not [155]. A similar animal mode has been established based on biochemical data [156,157]. HYL1 and CPL1 are necessary for correct strand selection during RISC loading [142,158]. In the absence of CPL1, HYL1 is phosphorylated and the strand selection from miRNA/miRNA* duplex is compromised [158].

3.3.2 miRNA stability control

In contrast to miRNA biogenesis and processing, decay of miRNAs has received limited attention. Originally, miRNAs were generally thought to be relatively stable because they are too short to be the substrates of RNases [116]. However, recent studies on both plants and animals unveiled the regulative role of miRNA turnover on miRNA accumulation. Actually, the stability of miRNAs has been found to be regulated by 3' methylation and uridylation of miRNAs, which act oppositely in miRNA degradation process.

3.3.2.1 Degradation of miRNAs by exonuclease

Enzymes responsible for miRNA turnover have been identified in various organisms. In *C. elegans*, the 5' to 3' exonuclease XRN-2 has been found to catalyze the degradation of mature miRNAs [159]. The degradation triggered by XRN-2 requires the release of miRNA from RISC, which is proposed to facilitate the enzyme to access miRNA 5' end [159]. Consequently, in *C. elegans* it is believed that miRNAs can be specifically released from RISC and degraded in the absence of its complementary targets in order to

make AGO proteins available for loading new miRNAs [159]. In animals, 3' to 5' trimming of miRNAs is catalyzed by Nibbler, a putative 3' to 5' exoribonuclease [160]. In *Arabidopsis*, a family of 3' to 5' exoribonucleases named SMALL RNA DEGRADING NUCLEASE 1, 2, and 3 (SDN1, SDN2, and SDN3) were found to be involved in mature miRNA turnover [161]. Inactivation of SDN proteins results in stabilization of several miRNAs [161].

3.3.2.2 Methylation protects miRNAs from degradation and uridylation (3' untemplated uridine addition)

In *Arabidopsis*, HUA1 ENHANCER1 (HEN1), an Mg^{2+}-dependent methyltransferase (MTase), was identified to catalyze 2'-O-methylation in the 3' ends of miRNA/miRNA* duplexes [120,162]. This methylation probably occurs before the disassociation of guide and passenger strands because HEN1 prefer 21-24 nt double-stranded RNAs (dsRNA) [162]. HEN1 recognizes substrates with the 2 nt overhang of miRNA/miRNA* duplex, and the 2' and 3' OH of the 3' end [162]. Later, studies on the crystal structure unveiled the mechanism of substrate recognition of HEN1 [163]. HEN1 functions as monomer to bind the duplex substrate [163]. The two dsRNA-binding domains (dsRBDs) were found to be critical for substrate recognition [163]. In addition, the distance between MTase domain and La-motif-containing domain (LCD) determines the substrate length specificity [163]. HEN1 homologs, which also induce 2'-O-methyl modification, have been identified in animals and flies [164-166]. However, the animal HEN1 lacks the dsRNA-binding domain and acts on miRNAs after RISC loading [167,168].

Studies of *hen1* mutants in *Arabidopsis* reveal that methylation affects the stability of

miRNAs. In *hen1* mutants, the abundance of miRNAs is reduced. In addition, miRNA size heterogeneity can be detected in *hen1* mutants by northern blotting, which is reflected by a ladder of bands [120,169]. Small RNA sequencing of *hen1* mutants revealed that the heterogeneous species are composed by both tailing and trimming miRNAs [169]. The size heterogeneity arises form the 3' end of miRNAs. miRNAs tend to have an oligonucleotide U tail at the 3' end in *hen1* mutants [169]. Besides U tailing, miRNAs display truncation from 3' ends [169]. Considering the fact that HEN1 adds methyl group to 3' end of miRNAs, it is proposed that methylation protects miRNAs from 3' uridine addition and truncation.

3.4 Uridylation of miRNAs

Uridylation of miRNAs is the addition of non-templated uridine to the 3' terminal, which is catalyzed by terminal nucleotidyl transferases. Uridylation is a critical regulatory mechanism for small RNAs functions in both plants and animals.

The characterization of *C. reinhardtii* gene MUT68 suggests that 3' uridylation may trigger miRNA trimming and degradation [170,171]. MUT68 is a terminal nucleotidyl transferase, which is involved in the degradation of both 5' RNA cleavage products generated by RISC and small RNAs. MUT68 adds U-tails to 3' termini small RNAs [170]. Cooperating with RRP6, which is the peripheral exosome subunit and degrades RNAs from 3'-to-5', MUT68 stimulates the efficient decay of small RNAs [171]. Consequently, the abundance of miRNAs is elevated in *mut68* mutants [171]. The function of MUT68 and RRP6 in miRNA uridylation and truncation has been proved by *in vitro* biochemical experiments [171]. MUT68 and RRP6 together, but not RRP6 alone,

can trigger the degradation of unmethylated RNA substrate. However, if a 2'-O-methylated miRNA is used as the substrate, MUT68 and RRP6 can not trigger uridylation and degradation, demonstrating that 3' methylation of miRNAs can block 3' uridylation and protect miRNAs from degradation [171].

In *Arabidopsis*, a nucleotidyl transferase responsible for this miRNA uridylation and degradation has been identified termed HEN1 SUPPRESSOR1 (HESO1) [172,173]. *In vitro* biochemical studies reveal that HESO1 preferentially adds untemplated U to the 3' terminal of unmethylated miRNAs, and this is blocked by 3' terminal methylation [172,173]. In the *hen1* background, *heso1* increases the abundance of normal sized miRNAs and reduces miRNA tails, demonstrating that HESO1 is the enzyme to catalyze 3' uridylation of miRNAs in the absence of methylation [172,173]. Furthermore, overexpression of HESO1 in *hen1* reduces the abundance of small RNAs in *hen1*, confirming that uridylation triggers degradation of miRNAs in *Arabidopsis* [172]. However, *heso1* increases the abundance of 3' truncated miRNAs in *hen1* while overexpression of HESO1 decreases the levels of 3' truncated miRNAs in *hen1* [172]. These results suggest that unlike in the green alga, uridylation may trigger miRNA degradation through a mechanism other than 3'-to-5' truncation.

A similar phenomenon has been observed in animals. In zebrafish, in the absence of a HEN1 homolog, piRNAs are uridylated and adenylated, which are accompanied with the reduction of piRNA levels [165]. In flies, without the protection from 3' methylation, AGO2-bound siRNAs are uridylated or adenylated, which can induce 3' trimming [174]. Besides triggering degradation, uridylation seems to also effect small RNA activity. In human cells, the uridylation of miR-26 has been reported to only reduce its repressive

activity against targets but not its accumulation [175].

In humans, multiple nucleotidyl transferases including MTPAP, PAPD4, PAPD5, ZCCHC6, ZCCHC11, and TUT1 were shown to add nucleotide to 3' terminal of miRNAs in a miRNA sequence specific manner. For instance PAPD5 is responsible for the adenylation of 4 miRNAs, while TUT1 is associated with 3' uridylation [176]. These enzymes are responsible not only for uridylation but also for adenylation [176]. However it seems that functions of uridylation and adenylation are different. Uridylation usually triggers miRNA decay while adenylation usually has no effect on miRNAs stability, or increases their stability [177,178]. An enzyme responsible for siRNA uridylation in *C. elegans* is CDE-1, which destabilizes siRNAs [179].

3.5 Unsolved problems in the mechanisms of miRNA biogenesis and degradation

Studies on the regulation of miRNA biogenesis and degradation revealed a sophisticated regulative network in miRNA accumulation. The involvement of phosphorylation, transcriptional regulation, and intron splicing in pri-miRNA processing mechanism demonstrates that DCL1-mediated miRNA processing is far more complicated than previously thought. Consequently, further studies on miRNA processing, especially discovering and functional characterization of new genes involved in this process, is necessary to reveal the regulatory network of DCL-mediated processing.

Although HESO1-mediated uridylation has been found to trigger miRNA degradation, how HESO1 recognizes miRNAs *in vivo* and how uridylation triggers miRNA degradation in *Arabidopsis* is still unknown. Furthermore, understanding of the uridylation process and degradation mechanisms will aid in our use of RNAi technology.

The uridylation and degradation mechanisms will help in recovering gene expression after RNAi, which will make RNAi technology more flexible.

Figures

Figure 1-1. Model for RNA-directed DNA methylation in *Arabidopsis* (reproduced from [10]). Single-stranded RNA (ssRNA) transcripts from transposons and repeated elements are generated by POL IV. CLSY1 is thought to be involved in POL IV transcription. Then RDR2 is proposed to use ssRNAs as the template to synthesize long double-stranded RNAs (dsRNAs), which will be processed by DCL3 into 24 nt siRNAs. To trigger DNA methylation and gene silencing, siRNAs are bound by AGO4 to form AGO4-siRNA complex, which is recruited to RdDM target by base pairing with transcripts generated by POL V. After associating with the target chromatin, AGO4 is proposed to recruit methyltransferases, such as DRM2, SUVH2/9, to catalyze DNA methylation. AGO4 colocalizes with NRPE1 in Cajal bodies, which are thought to be siRNA processing bodies. Several protein factors have been found to be involved in AGO4 and POL V functions. DRD1 and DMS3 are critical for POL V transcription. SPT5L is related with AGO4 function.

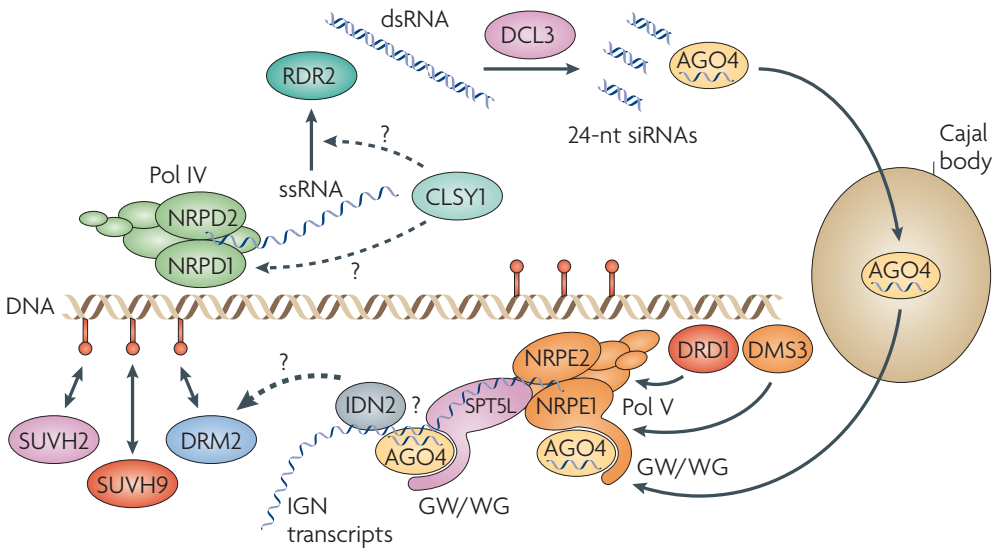
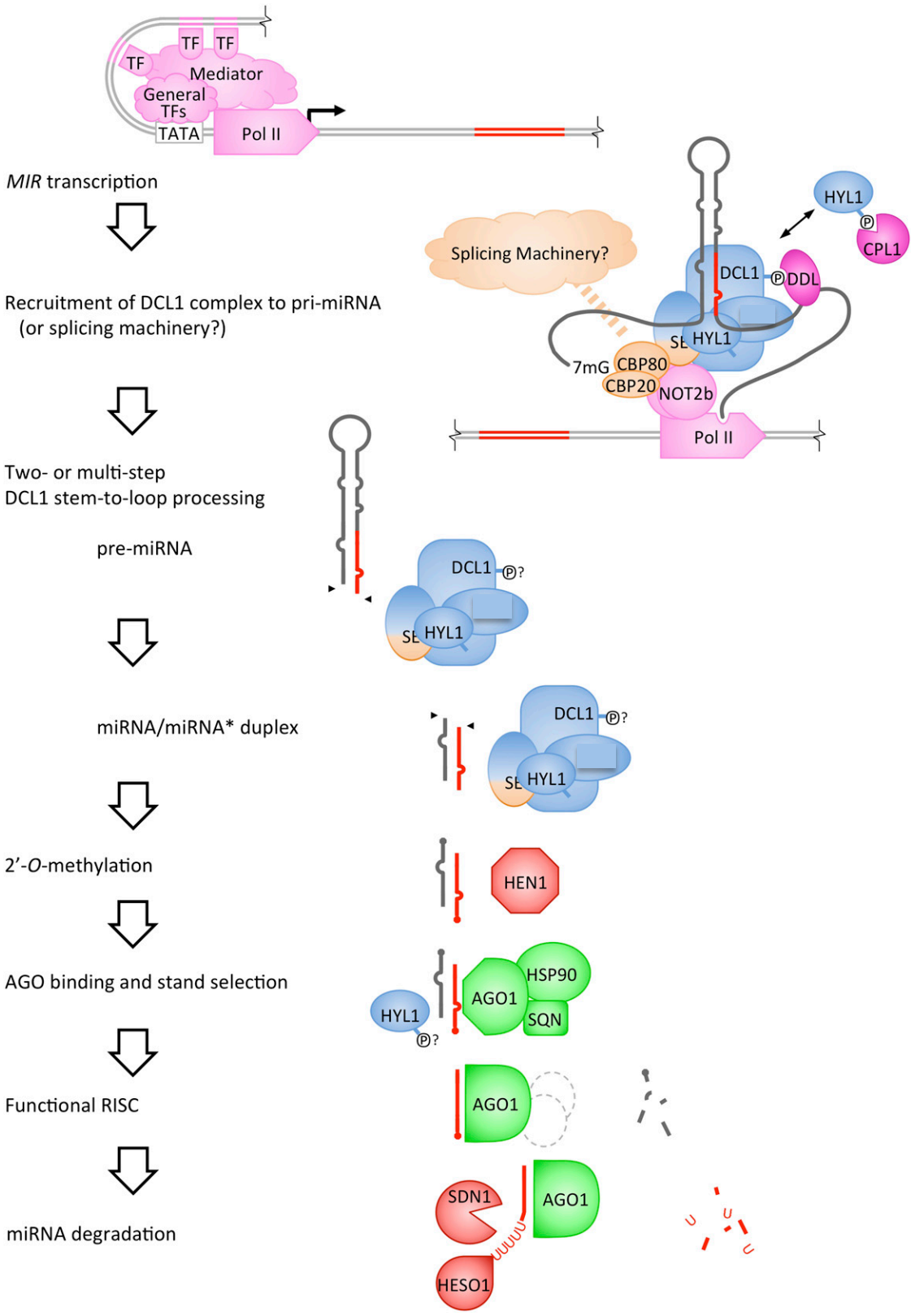


Figure 1-2. Model for miRNA biogenesis and turnover in *Arabidopsis* (modified from [132]). The transcription of MIR genes by POL II is controlled by several general transcription factors, such as Mediator complex. After transcription pri-miRNAs are processed by DCL1 into mature miRNA/miRNA* duplexes. Several protein factors are involved to promote the efficiency and accuracy of DCL1 processing, including HYL1, SE, CPL1, CBP20/80, and NOT2. The 3' terminal of miRNA/miRNA* duplex is methylated by HEN1, which is proposed to protect miRNAs from turnover. Then the sense strand of the duplex is loaded into AGO1 to form RISC, which triggers gene silencing by cleaving target mRNAs or repressing translation. HSP90 and SQN are critical for the separation of sense and anti-sense strands and the formation of RISC. For miRNA turnover, HESO1 is proposed to induce the uridylation and degradation of miRNAs. SDN1 is thought to degrade the uridylated miRNAs.


TF
TF
TF
Mediator
General
TFs
TATA
Pol II
MIR transcription
Recruitment of DCL1 complex to pri-miRNA
(or splicing machinery?)
Splicing Machinery?
HYL1
P
CPL1
DCL1
P
DDL
SE
HYL1
7mG
CBP80
CBP20
NOT2b
Pol II
Two- or multi-step
DCL1 stem-to-loop processing
pre-miRNA
DCL1
P?
SE
HYL1
miRNA/miRNA* duplex
DCL1
P?
SE
HYL1
2'-O-methylation
HEN1
AGO binding and stand selection
HSP90
AGO1
SQN
HYL1
P?
Functional RISC
AGO1
miRNA degradation
SDN1
AGO1
UUUU
HESO1
U
U
U

References

1. Chen, X. (2009). Small RNAs and their roles in plant development. Annual Review of Cell and Developmental, 25, 21-44.

2. Kim, V. N., Han, J., & Siomi, M. C. (2009). Biogenesis of small RNAs in animals. Nature reviews Molecular cell biology, 10(2), 126-139.

3. Zhang, H., & Zhu, J.-K. (2011). RNA-directed DNA methylation. Current opinion in plant biology, 14(2), 142-147.

4. Xie, Z., Khanna, K., & Ruan, S. (2010). Expression of microRNAs and its regulation in plants. Paper presented at the Seminars in cell & developmental biology.

5. Vagin, V. V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., & Zamore, P. D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. Science, 313(5785), 320-324.

6. Siomi, M. C., Sato, K., Pezic, D., & Aravin, A. A. (2011). PIWI-interacting small RNAs: the vanguard of genome defence. Nature reviews Molecular cell biology, 12(4), 246-258.

7. Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., & Hannon, G. J. (2007). Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila. Cell, 128(6), 1089-1103.

8. Gunawardane, L. S., Saito, K., Nishida, K. M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, M. C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5'end formation in Drosophila. Science, 315(5818), 1587-1590.

9. Aravin, A. A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K. F., Hannon, G. J. (2008). A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. Molecular cell, 31(6), 785-799.

10. Law, J. A., & Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nature Reviews Genetics, 11(3), 204-220.

11. Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes & development, 16(1), 6-21.

12. Suzuki, M. M., & Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. Nature Reviews Genetics, 9(6), 465-476.

13. Henderson, I. R., & Jacobsen, S. E. (2007). Epigenetic inheritance in plants. Nature, 447(7143), 418-424.

14. Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S. W.-L., Chen, H., Jacobsen, S. E. (2006). Genome-wide High-Resolution Mapping and Functional Analysis of DNA Methylation in Arabidopsis. Cell, 126(6), 1189-1201.

15. Cokus, S. J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C. D., Jacobsen, S. E. (2008). Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature, 452(7184), 215-219.

16. Cao, X., & Jacobsen, S. E. (2002). Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. Proceedings of the National Academy of Sciences, 99(suppl 4), 16491-16498.

17. Saze, H., Scheid, O. M., & Paszkowski, J. (2003). Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. Nature genetics, 34(1), 65-69.

18. Kankel, M. W., Ramsey, D. E., Stokes, T. L., Flowers, S. K., Haag, J. R., Jeddeloh, J. A., Richards, E. J. (2003). Arabidopsis MET1 cytosine methyltransferase mutants. Genetics, 163(3), 1109-1122.

19. Malagnac, F., Bartee, L., & Bender, J. (2002). An Arabidopsis SET domain protein required for maintenance but not establishment of DNA methylation. The EMBO journal, 21(24), 6842-6852.

20. Wassenegger, M., Heimes, S., Riedel, L., & Sänger, H. L. (1994). RNA-directed de novo methylation of genomic sequences in plants. Cell, 76(3), 567-576.

21. Yoder, J. A., Walsh, C. P., & Bestor, T. H. (1997). Cytosine methylation and the ecology of intragenomic parasites. Trends in genetics, 13(8), 335-340.

22. Martienssen, R. A., & Colot, V. (2001). DNA methylation and epigenetic inheritance in plants and filamentous fungi. Science, 293(5532), 1070-1074.

23. Bender, J., & Fink, G. R. (1995). Epigenetic control of an endogenous gene family is revealed by a novel blue fluorescent mutant of Arabidopsis. Cell, 83(5), 725-734.

24. Kinoshita, T., Miura, A., Choi, Y., Kinoshita, Y., Cao, X., Jacobsen, S. E., Kakutani, T. (2004). One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation. Science, 303(5657), 521-523.

25. Lawrence, R. J., Earley, K., Pontes, O., Silva, M., Chen, Z. J., Neves, N., Pikaard, C. S. (2004). A concerted DNA methylation/histone methylation switch regulates

rRNA gene dosage control and nucleolar dominance. Molecular cell, 13(4), 599-609.

26. Ramsahoye, B. H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A. P., & Jaenisch, R. (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proceedings of the National Academy of Sciences, 97(10), 5237-5242.

27. Xie, Z., Johansen, L. K., Gustafson, A. M., Kasschau, K. D., Lellis, A. D., Zilberman, D., Carrington, J. C. (2004). Genetic and functional diversification of small RNA pathways in plants. PLoS biology, 2(5), e104.

28. Herr, A., Jensen, M., Dalmay, T., & Baulcombe, D. (2005). RNA polymerase IV directs silencing of endogenous DNA. Science, 308(5718), 118-120.

29. Kanno, T., Huettel, B., Mette, M. F., Aufsatz, W., Jaligot, E., Daxinger, L., Matzke, A. J. (2005). Atypical RNA polymerase subunits required for RNA-directed DNA methylation. Nature genetics, 37(7), 761-765.

30. Onodera, Y., Haag, J. R., Ream, T., Nunes, P. C., Pontes, O., & Pikaard, C. S. (2005). Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. Cell, 120(5), 613-622.

31. Pontier, D., Yahubyan, G., Vega, D., Bulski, A., Saez-Vasquez, J., Hakimi, M.-A., Lagrange, T. (2005). Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in Arabidopsis. Genes & development, 19(17), 2030-2040.

32. Zilberman, D., Cao, X., & Jacobsen, S. E. (2003). ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. Science, 299(5607), 716-719.

33. Zheng, X., Zhu, J., Kapoor, A., & Zhu, J. K. (2007). Role of Arabidopsis AGO6 in siRNA accumulation, DNA methylation and transcriptional gene silencing. The EMBO journal, 26(6), 1691-1701.

34. Havecker, E. R., Wallbridge, L. M., Hardcastle, T. J., Bush, M. S., Kelly, K. A., Dunn, R. M., Baulcombe, D. C. (2010). The Arabidopsis RNA-directed DNA methylation argonautes functionally diverge based on their expression and interaction with target loci. The Plant Cell Online, 22(2), 321-334.

35. El-Shami, M., Pontier, D., Lahmy, S., Braun, L., Picart, C., Vega, D., Lagrange, T. (2007). Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. Genes & development, 21(20), 2539-2544.

36. Wierzbicki, A. T., Haag, J. R., & Pikaard, C. S. (2008). Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. Cell, 135(4), 635-648.

37. Wierzbicki, A. T., Ream, T. S., Haag, J. R., & Pikaard, C. S. (2009). RNA polymerase V transcription guides ARGONAUTE4 to chromatin. Nature genetics, 41(5), 630-634.

38. Zheng, B., Wang, Z., Li, S., Yu, B., Liu, J.-Y., & Chen, X. (2009). Intergenic transcription by RNA polymerase II coordinates Pol IV and Pol V in siRNA-directed transcriptional gene silencing in Arabidopsis. Genes & development, 23(24), 2850-2860.

39. Bies‐Etheve, N., Pontier, D., Lahmy, S., Picart, C., Vega, D., Cooke, R., & Lagrange, T. (2009). RNA‐directed DNA methylation requires an AGO4‐interacting member of the SPT5 elongation factor family. EMBO reports, 10(6), 649-654.

40. He, X.-J., Hsu, Y.-F., Zhu, S., Wierzbicki, A. T., Pontes, O., Pikaard, C. S., Zhu, J.-K. (2009). An Effector of RNA-Directed DNA Methylation in Arabidopsis Is an ARGONAUTE 4-and RNA-Binding Protein. Cell, 137(3), 498-508.

41. Rowley, M. J., Avrutsky, M. I., Sifuentes, C. J., Pereira, L., & Wierzbicki, A. T. (2011). Independent chromatin binding of ARGONAUTE4 and SPT5L/KTF1 mediates transcriptional gene silencing. PLoS genetics, 7(6), e1002120.

42. Smith, L. M., Pontes, O., Searle, I., Yelina, N., Yousafzai, F. K., Herr, A. J., Baulcombe, D. C. (2007). An SNF2 protein associated with nuclear RNA silencing and the spread of a silencing signal between cells in Arabidopsis. The Plant Cell Online, 19(5), 1507-1521.

43. Law, J. A., Vashisht, A. A., Wohlschlegel, J. A., & Jacobsen, S. E. (2011). SHH1, a homeodomain protein required for DNA methylation, as well as RDR2, RDM4, and chromatin remodeling factors, associate with RNA polymerase IV. PLoS genetics, 7(7), e1002195.

44. Liu, J., Bai, G., Zhang, C., Chen, W., Zhou, J., Zhang, S., Zhu, J.-K. (2011). An atypical component of RNA-directed DNA methylation machinery has both DNA methylation-dependent and-independent roles in locus-specific transcriptional gene silencing. Cell research, 21(12), 1691-1700.

45. Kanno, T., Mette, M. F., Kreil, D. P., Aufsatz, W., Matzke, M., & Matzke, A. J. (2004). Involvement of putative SNF2 chromatin remodeling protein DRD1 in RNA-directed DNA methylation. Current Biology, 14(9), 801-805.

46. Kanno, T., Bucher, E., Daxinger, L., Huettel, B., Böhmdorfer, G., Gregor, W., Matzke, A. J. (2008). A structural-maintenance-of-chromosomes hinge domain–containing protein is required for RNA-directed DNA methylation. Nature genetics, 40(5), 670-675.

47. Gao, Z., Liu, H.-L., Daxinger, L., Pontes, O., He, X., Qian, W., Zhang, S. (2010). An RNA polymerase II-and AGO4-associated protein acts in RNA-directed DNA methylation. Nature, 465(7294), 106-109.

48. Erhard, K. F., Parkinson, S. E., Gross, S. M., Barbour, J.-E. R., Lim, J. P., & Hollick, J. B. (2013). Maize RNA Polymerase IV defines trans-generational epigenetic variation. The Plant Cell Online, 25(3), 808-819.

49. Hamilton, A., Voinnet, O., Chappell, L., & Baulcombe, D. (2002). Two classes of short interfering RNA in RNA silencing. The EMBO journal, 21(17), 4671-4679.

50. Mosher, R. A., Schwach, F., Studholme, D., & Baulcombe, D. C. (2008). PolIVb influences RNA-directed DNA methylation independently of its role in siRNA biogenesis. Proceedings of the National Academy of Sciences, 105(8), 3145-3150.

51. Zhang, X., Henderson, I. R., Lu, C., Green, P. J., & Jacobsen, S. E. (2007). Role of RNA polymerase IV in plant small RNA metabolism. Proceedings of the National Academy of Sciences, 104(11), 4536-4541.

52. Huettel, B., Kanno, T., Daxinger, L., Aufsatz, W., Matzke, A. J., & Matzke, M. (2006). Endogenous targets of RNA-directed DNA methylation and Pol IV in Arabidopsis. The EMBO journal, 25(12), 2828-2836.

53. Grewal, S. I., & Elgin, S. C. (2007). Transcription and RNA interference in the formation of heterochromatin. Nature, 447(7143), 399-406.

54. Ream, T. S., Haag, J. R., Wierzbicki, A. T., Nicora, C. D., Norbeck, A. D., Zhu, J.-K., . . . Pikaard, C. S. (2009). Subunit compositions of the RNA-silencing enzymes Pol IV and Pol V reveal their origins as specialized forms of RNA polymerase II. Molecular cell, 33(2), 192-203.

55. He, X.-J., Hsu, Y.-F., Pontes, O., Zhu, J., Lu, J., Bressan, R. A., Zhu, J.-K. (2009). NRPD4, a protein related to the RPB4 subunit of RNA polymerase II, is a component of RNA polymerases IV and V and is required for RNA-directed DNA methylation. Genes & development, 23(3), 318-330.

56. Lahmy, S., Pontier, D., Cavel, E., Vega, D., El-Shami, M., Kanno, T., & Lagrange, T. (2009). PolV (PolIVb) function in RNA-directed DNA methylation requires the conserved active site and an additional plant-specific subunit. Proceedings of the National Academy of Sciences, 106(3), 941-946.

57. Kanno, T., Bucher, E., Daxinger, L., Huettel, B., Kreil, D. P., Breinig, F., Gurazada, S. G. R. (2010). RNA‐directed DNA methylation and plant development require an IWR1‐type transcription factor. EMBO reports, 11(1), 65-71.

58. He, X.-J., Hsu, Y.-F., Zhu, S., Liu, H.-L., Pontes, O., Zhu, J., Zhu, J.-K. (2009). A conserved transcriptional regulator is required for RNA-directed DNA methylation and plant development. Genes & development, 23(23), 2717-2722.

59. Ausin, I., Mockler, T. C., Chory, J., & Jacobsen, S. E. (2009). IDN1 and IDN2 are required for de novo DNA methylation in Arabidopsis thaliana. Nature structural & molecular biology, 16(12), 1325-1327.

60. Pontes, O., Li, C. F., Nunes, P. C., Haag, J., Ream, T., Vitins, A., Pikaard, C. S. (2006). The Arabidopsis Chromatin-Modifying Nuclear siRNA Pathway Involves a Nucleolar RNA Processing Center. Cell, 126(1), 79-92.

61. Law, J. A., Ausin, I., Johnson, L. M., Vashisht, A. A., Zhu, J.-K., Wohlschlegel, J. A., & Jacobsen, S. E. (2010). A Protein Complex Required for Polymerase V Transcripts and RNA-Directed DNA Methylation in Arabidopsis. Current Biology, 20(10), 951-956.

62. Willmann, M. R., Endres, M. W., Cook, R. T., & Gregory, B. D. (2011). The functions of RNA-dependent RNA polymerases in Arabidopsis. The Arabidopsis book/American Society of Plant Biologists, 9.

63. Zong, J., Yao, X., Yin, J., Zhang, D., & Ma, H. (2009). Evolution of the RNA-dependent RNA polymerase (RdRP) genes: duplications and possible losses before and after the divergence of major eukaryotic groups. Gene, 447(1), 29-39.

64. Wassenegger, M., & Krczal, G. (2006). Nomenclature and functions of RNA-directed RNA polymerases. Trends in plant science, 11(3), 142-151.

65. Voinnet, O. (2008). Use, tolerance and avoidance of amplified RNA silencing by plants. Trends in plant science, 13(7), 317-328.

66. Kasschau, K. D., Fahlgren, N., Chapman, E. J., Sullivan, C. M., Cumbie, J. S., Givan, S. A., & Carrington, J. C. (2007). Genome-wide profiling and analysis of Arabidopsis siRNAs. PLoS biology, 5(3), e57.

67. Donaire, L., Barajas, D., Martínez-García, B., Martínez-Priego, L., Pagán, I., & Llave, C. (2008). Structural and genetic requirements for the biogenesis of tobacco rattle virus-derived small interfering RNAs. Journal of virology, 82(11), 5167-5177.

68. Wang, X.-B., Wu, Q., Ito, T., Cillo, F., Li, W.-X., Chen, X., Ding, S.-W. (2010). RNAi-mediated viral immunity requires amplification of virus-derived siRNAs in Arabidopsis thaliana. Proceedings of the National Academy of Sciences, 107(1), 484-489.

69. Allen, E., Xie, Z., Gustafson, A. M., & Carrington, J. C. (2005). microRNA-Directed Phasing during Trans-Acting siRNA Biogenesis in Plants. Cell, 121(2), 207-221.

70. Gasciolli, V., Mallory, A. C., Bartel, D. P., & Vaucheret, H. (2005). Partially Redundant Functions of Arabidopsis DICER-like Enzymes and a Role for DCL4 in Producing trans-Acting siRNAs. Current Biology, 15(16), 1494-1500.

71. Yoshikawa, M., Peragine, A., Park, M. Y., & Poethig, R. S. (2005). A pathway for the biogenesis of trans-acting siRNAs in Arabidopsis. Genes & development, 19(18), 2164-2175.

72. Garcia, D., Collier, S. A., Byrne, M. E., & Martienssen, R. A. (2006). Specification of Leaf Polarity in Arabidopsis via the trans-Acting siRNA Pathway. Current Biology, 16(9), 933-938.

73. Peragine, A., Yoshikawa, M., Wu, G., Albrecht, H. L., & Poethig, R. S. (2004). SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in Arabidopsis. Genes & development, 18(19), 2368-2379.

74. Vazquez, F., Vaucheret, H., Rajagopalan, R., Lepers, C., Gasciolli, V., Mallory, A. C., Crété, P. (2004). Endogenous trans-Acting siRNAs Regulate the Accumulation of Arabidopsis mRNAs. Molecular cell, 16(1), 69-79.

75. Borsani, O., Zhu, J., Verslues, P. E., Sunkar, R., & Zhu, J.-K. (2005). Endogenous siRNAs Derived from a Pair of Natural cis-Antisense Transcripts Regulate Salt Tolerance in Arabidopsis. Cell, 123(7), 1279-1291.

76. Katiyar-Agarwal, S., Morgan, R., Dahlbeck, D., Borsani, O., Villegas, A., Zhu, J.-K., Jin, H. (2006). A pathogen-inducible endogenous siRNA in plant immunity. Proceedings of the National Academy of Sciences, 103(47), 18002-18007.

77. Chan, S. W.-L., Henderson, I. R., & Jacobsen, S. E. (2005). Gardening the genome: DNA methylation in Arabidopsis thaliana. Nature Reviews Genetics, 6(5), 351-360.

78. Pikaard, C. (2006). Cell biology of the Arabidopsis nuclear siRNA pathway for RNA-directed chromatin modification. Paper presented at the Cold Spring Harbor symposia on quantitative biology.

79. Qi, Y., He, X., Wang, X.-J., Kohany, O., Jurka, J., & Hannon, G. J. (2006). Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. Nature, 443(7114), 1008-1012.

80. Haag, J. R., Ream, T. S., Marasco, M., Nicora, C. D., Norbeck, A. D., Pasa-Tolic, L., & Pikaard, C. S. (2012). In vitro transcription activities of Pol IV, Pol V, and RDR2 reveal coupling of Pol IV and RDR2 for dsRNA synthesis in plant RNA silencing. Molecular cell, 48(5), 811-818.

81. Goldoni, M., Azzalin, G., Macino, G., & Cogoni, C. (2004). Efficient gene silencing by expression of double stranded RNA in Neurospora crassa. Fungal Genetics and Biology, 41(11), 1016-1024.

82. Maiti, M., Lee, H.-C., & Liu, Y. (2007). QIP, a putative exonuclease, interacts with the Neurospora Argonaute protein and facilitates conversion of duplex siRNA into single strands. Genes & development, 21(5), 590-600.

83. Aalto, A. P., Poranen, M. M., Grimes, J. M., Stuart, D. I., & Bamford, D. H. (2010). In vitro activities of the multifunctional RNA silencing polymerase QDE-1 of Neurospora crassa. Journal of biological chemistry, 285(38), 29367-29374.

84. Pikaard, C., Haag, J., Pontes, O., Blevins, T., & Cocklin, R. (2012). A Transcription Fork Model for Pol IV and Pol V–dependent RNA-Directed DNA Methylation. Paper presented at the Cold Spring Harbor symposia on quantitative biology.

85. Liu, Q., Feng, Y., & Zhu, Z. (2009). Dicer-like (DCL) proteins in plants. Functional & integrative genomics, 9(3), 277-286.

86. Margis, R., Fusaro, A. F., Smith, N. A., Curtin, S. J., Watson, J. M., Finnegan, E. J., & Waterhouse, P. M. (2006). The evolution and diversification of Dicers in plants. FEBS letters, 580(10), 2442-2450.

87. MacRae, I. J., Zhou, K., Li, F., Repic, A., Brooks, A. N., Cande, W. Z., Doudna, J. A. (2006). Structural basis for double-stranded RNA processing by Dicer. Science, 311(5758), 195-198.

88. Zhang, H., Kolb, F. A., Jaskiewicz, L., Westhof, E., & Filipowicz, W. (2004). Single processing center models for human Dicer and bacterial RNase III. Cell, 118(1), 57-68.

89. Moissiard, G., & Voinnet, O. (2006). RNA silencing of host transcripts by cauliflower mosaic virus requires coordinated action of the four Arabidopsis Dicer-like proteins. Proceedings of the National Academy of Sciences, 103(51), 19593-19598.

90. Moissiard, G., Parizotto, E. A., Himber, C., & Voinnet, O. (2007). Transitivity in Arabidopsis can be primed, requires the redundant action of the antiviral Dicer-like 4 and Dicer-like 2, and is compromised by viral-encoded suppressor proteins. Rna, 13(8), 1268-1278.

91. Henderson, I. R., Zhang, X., Lu, C., Johnson, L., Meyers, B. C., Green, P. J., & Jacobsen, S. E. (2006). Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning. Nature genetics, 38(6), 721-725.

92. Hiraguri, A., Itoh, R., Kondo, N., Nomura, Y., Aizawa, D., Murai, Y., Fukuhara, T. (2005). Specific interactions between Dicer-like proteins and HYL1/DRB-family dsRNA-binding proteins in Arabidopsis thaliana. Plant molecular biology, 57(2), 173-188.

93. Curtin, S. J., Watson, J. M., Smith, N. A., Eamens, A. L., Blanchard, C. L., & Waterhouse, P. M. (2008). The roles of plant dsRNA-binding proteins in RNAi-like pathways. FEBS letters, 582(18), 2753-2760.

94. Cenik, E. S., & Zamore, P. D. (2011). Argonaute proteins. Current Biology, 21(12), R446-R449.

95. Ghildiyal, M., & Zamore, P. D. (2009). Small silencing RNAs: an expanding universe. Nature Reviews Genetics, 10(2), 94-108.

96. Hutvagner, G., & Simard, M. J. (2008). Argonaute proteins: key players in RNA silencing. Nature reviews Molecular cell biology, 9(1), 22-32.

97. Rivas, F. V., Tolia, N. H., Song, J.-J., Aragon, J. P., Liu, J., Hannon, G. J., & Joshua-Tor, L. (2005). Purified Argonaute2 and an siRNA form recombinant human RISC. Nature structural & molecular biology, 12(4), 340-349.

98. Vaucheret, H. (2008). Plant argonautes. Trends in plant science, 13(7), 350-358.

99. Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., Long, C. (2008). Sorting of Small RNAs into Arabidopsis Argonaute Complexes Is Directed by the 5′ Terminal Nucleotide. Cell, 133(1), 116-127.

100. Li, C. F., Pontes, O., El-Shami, M., Henderson, I. R., Bernatavichute, Y. V., Chan, S. W.-L., Jacobsen, S. E. (2006). An ARGONAUTE4-Containing Nuclear Processing Center Colocalized with Cajal Bodies in Arabidopsis thaliana. Cell, 126(1), 93-106.

101. Li, C. F., Henderson, I. R., Song, L., Fedoroff, N., Lagrange, T., & Jacobsen, S. E. (2008). Dynamic regulation of ARGONAUTE4 within multiple nuclear bodies in Arabidopsis thaliana. PLoS genetics, 4(2), e27.

102. Ye, R., Wang, W., Iki, T., Liu, C., Wu, Y., Ishikawa, M., Qi, Y. (2012). Cytoplasmic Assembly and Selective Nuclear Import of Arabidopsis ARGONAUTE4/siRNA Complexes. Molecular cell, 46(6), 859-870.

103. Lindroth, A. M., Cao, X., Jackson, J. P., Zilberman, D., McCallum, C. M., Henikoff, S., & Jacobsen, S. E. (2001). Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. Science, 292(5524), 2077-2080.

104. Stroud, H., Greenberg, M. V., Feng, S., Bernatavichute, Y. V., & Jacobsen, S. E. (2013). Comprehensive Analysis of Silencing Mutants Reveals Complex Regulation of the Arabidopsis Methylome. Cell, 152(1), 352-364.

105. Pontier, D., Picart, C., Roudier, F., Garcia, D., Lahmy, S., Azevedo, J., Cooke, R. (2012). NERD, a Plant-Specific GW Protein, Defines an Additional RNAi-Dependent Chromatin-Based Pathway in Arabidopsis. Molecular cell, 48(1), 121-132.

106. Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. Cell, 116(2), 281-297.

107. Chen, X. (2005). MicroRNA biogenesis and function in plants. FEBS letters, 579(26), 5923-5931.

108. Voinnet, O. (2009). Origin, biogenesis, and activity of plant microRNAs. Cell, 136(4), 669-687.

109. Lee, R. C., Feinbaum, R. L., & Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell, 75(5), 843-854.

110. Lewis, B. P., Burge, C. B., & Bartel, D. P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell, 120(1), 15-20.

111. Friedman, R. C., Farh, K. K.-H., Burge, C. B., & Bartel, D. P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. Genome research, 19(1), 92-105.

112. Mencía, Á., Modamio-Høybjør, S., Redshaw, N., Morín, M., Mayo-Merino, F., Olavarrieta, L., Dalmay, T. (2009). Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss. Nature genetics, 41(5), 609-613.

113. Møller, H. G., Rasmussen, A. P., Andersen, H. H., Johnsen, K. B., Henriksen, M., & Duroux, M. (2013). A systematic review of microRNA in glioblastoma multiforme: micro-modulators in the mesenchymal mode of migration and invasion. Molecular neurobiology, 47(1), 131-144.

114. Aukerman, M. J., & Sakai, H. (2003). Regulation of flowering time and floral organ identity by a microRNA and its APETALA2-like target genes. The Plant Cell Online, 15(11), 2730-2741.

115. Chen, X. (2004). A microRNA as a translational repressor of APETALA2 in Arabidopsis flower development. Science, 303(5666), 2022-2025.

116. Krol, J., Loedige, I., & Filipowicz, W. (2010). The widespread regulation of microRNA biogenesis, function and decay. Nature Reviews Genetics, 11(9), 597-610.

117. Xie, Z., Allen, E., Fahlgren, N., Calamar, A., Givan, S. A., & Carrington, J. C. (2005). Expression of Arabidopsis MIRNA genes. Plant Physiology, 138(4), 2145-2154.

118. Kim, Y. J., Zheng, B., Yu, Y., Won, S. Y., Mo, B., & Chen, X. (2011). The role of Mediator in small and long noncoding RNA production in Arabidopsis thaliana. The EMBO journal, 30(5), 814-822.

119. Liu, C., Axtell, M. J., & Fedoroff, N. V. (2012). The helicase and RNaseIIIa domains of Arabidopsis Dicer-Like1 modulate catalytic parameters during microRNA biogenesis. Plant Physiology, 159(2), 748-758.

120. Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R. W., Chen, X. (2005). Methylation as a crucial step in plant microRNA biogenesis. Science, 307(5711), 932-935.

121. Park, M. Y., Wu, G., Gonzalez-Sulser, A., Vaucheret, H., & Poethig, R. S. (2005). Nuclear processing and export of microRNAs in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America, 102(10), 3691-3696.

122. Mourrain, P., Béclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J.-B., Picault, N. (2000). Arabidopsis SGS2 and SGS3 Genes Are Required for Posttranscriptional Gene Silencing and Natural Virus Resistance. Cell, 101(5), 533-542.

123. Biddick, R., & Young, E. T. (2005). Yeast mediator and its role in transcriptional regulation. Comptes rendus biologies, 328(9), 773-782.

124. Taatjes, D. J. (2010). The human Mediator complex: a versatile, genome-wide regulator of transcription. Trends in biochemical sciences, 35(6), 315-322.

125. Collart, M. A., & Struhl, K. (1994). NOT1 (CDC39), NOT2 (CDC36), NOT3, and NOT4 encode a global-negative regulator of transcription that differentially affects TATA-element utilization. Genes & development, 8(5), 525-537.

126. Denis, C. L., & Chen, J. (2003). The CCR4–NOT complex plays diverse roles in mRNA metabolism. Progress in nucleic acid research and molecular biology, 73, 221-250.

127. Collart, M. A., & Timmers, H. (2003). The eukaryotic Ccr4-not complex: a regulatory platform integrating mRNA metabolism with cellular signaling pathways? Progress in nucleic acid research and molecular biology, 77, 289-322.

128. Collart, M. A., & Panasenko, O. O. (2012). The Ccr4–not complex. Gene, 492(1), 42-53.

129. Wang, L., Song, X., Gu, L., Li, X., Cao, S., Chu, C., Cao, X. (2013). NOT2 Proteins Promote Polymerase II–Dependent Transcription and Interact with Multiple MicroRNA Biogenesis Factors in Arabidopsis. The Plant Cell Online, 25(2), 715-727.

130. Yumul, R. E., Kim, Y. J., Liu, X., Wang, R., Ding, J., Xiao, L., & Chen, X. (2013). POWERDRESS and diversified expression of the MIR172 gene family bolster the floral stem cell network. PLoS genetics, 9(1), e1003218.

131. Yant, L., Mathieu, J., Dinh, T. T., Ott, F., Lanz, C., Wollmann, H., Schmid, M. (2010). Orchestration of the floral transition and floral development in Arabidopsis by the bifunctional transcription factor APETALA2. The Plant Cell Online, 22(7), 2156-2170.

132. Rogers, K., & Chen, X. (2013). Biogenesis, turnover, and mode of action of plant microRNAs. The Plant Cell Online, 25(7), 2383-2399.

133. Yamasaki, H., Hayashi, M., Fukazawa, M., Kobayashi, Y., & Shikanai, T. (2009). SQUAMOSA promoter binding protein–like7 is a central regulator for copper homeostasis in Arabidopsis. The Plant Cell Online, 21(1), 347-361.

134. Baek, D., Park, H. C., Kim, M. C., & Yun, D.-J. (2013). The role of Arabidopsis MYB2 in miR399f-mediated phosphate-starvation response. Plant signaling & behavior, 8(3), 362-373.

135. Dong, Z., Han, M.-H., & Fedoroff, N. (2008). The RNA-binding proteins HYL1 and SE promote accurate in vitro processing of pri-miRNA by DCL1. Proceedings of the National Academy of Sciences, 105(29), 9970-9975.

136. Iwata, Y., Takahashi, M., Fedoroff, N. V., & Hamdan, S. M. (2013). Dissecting the interactions of SERRATE with RNA and DICER-LIKE 1 in Arabidopsis microRNA precursor processing. Nucleic acids research, 41(19), 9129-9140.

137. Song, L., Han, M.-H., Lesicka, J., & Fedoroff, N. (2007). Arabidopsis primary microRNA processing proteins HYL1 and DCL1 define a nuclear body distinct from the Cajal body. Proceedings of the National Academy of Sciences, 104(13), 5437-5442.

138. Fujioka, Y., Utsumi, M., Ohba, Y., & Watanabe, Y. (2007). Location of a possible miRNA processing site in SmD3/SmB nuclear bodies in Arabidopsis. Plant and cell physiology, 48(9), 1243-1253.

139. Yang, S. W., Chen, H.-Y., Yang, J., Machida, S., Chua, N.-H., & Yuan, Y. A. (2010). Structure of Arabidopsis HYPONASTIC LEAVES1 and Its Molecular Implications for miRNA Processing. Structure, 18(5), 594-605.

140. Machida, S., Chen, H.-Y., & Yuan, Y. A. (2011). Molecular insights into miRNA processing by Arabidopsis thaliana SERRATE. Nucleic acids research, 39(17), 7828-7836.

141. Zhan, X., Wang, B., Li, H., Liu, R., Kalia, R. K., Zhu, J.-K., & Chinnusamy, V. (2012). Arabidopsis proline-rich protein important for development and abiotic stress tolerance is involved in microRNA biogenesis. Proceedings of the National Academy of Sciences, 109(44), 18198-18203.

142. Manavella, P. A., Hagmann, J., Ott, F., Laubinger, S., Franz, M., Macek, B., & Weigel, D. (2012). Fast-forward genetics identifies plant CPL phosphatases as regulators of miRNA processing factor HYL1. Cell, 151(4), 859-870.

143. Engelsberger, W. R., & Schulze, W. X. (2012). Nitrate and ammonium lead to distinct global dynamic phosphorylation patterns when resupplied to nitrogen-starved Arabidopsis seedlings. The Plant Journal, 69(6), 978-995.

144. Yu, B., Bi, L., Zheng, B., Ji, L., Chevalier, D., Agarwal, M., Walker, J. C. (2008). The FHA domain proteins DAWDLE in Arabidopsis and SNIP1 in humans act in small RNA biogenesis. Proceedings of the National Academy of Sciences, 105(29), 10073-10078.

145. Machida, S., & Yuan, Y. A. (2013). Crystal structure of Arabidopsis thaliana Dawdle forkhead-associated domain reveals a conserved phospho-threonine recognition cleft for dicer-like 1 binding. Molecular plant, 6(4), 1290-1300.

146. Chaabane, S. B., Liu, R., Chinnusamy, V., Kwon, Y., Park, J.-h., Kim, S. Y., Lee, B.-h. (2013). STA1, an Arabidopsis pre-mRNA processing factor 6 homolog, is a

new player involved in miRNA biogenesis. Nucleic acids research, 41(3), 1984-1997.

147. Kim, W., Benhamed, M., Servet, C., Latrasse, D., Zhang, W., Delarue, M., & Zhou, D.-X. (2009). Histone acetyltransferase GCN5 interferes with the miRNA pathway in Arabidopsis. Cell research, 19(7), 899-909.

148. Wu, X., Shi, Y., Li, J., Xu, L., Fang, Y., Li, X., & Qi, Y. (2013). A role for the RNA-binding protein MOS2 in microRNA maturation in Arabidopsis. Cell research, 23(5), 645-657.

149. Raczynska, K. D., Simpson, C. G., Ciesiolka, A., Szewc, L., Lewandowska, D., McNicol, J., Jarmolowski, A. (2010). Involvement of the nuclear cap-binding protein complex in alternative splicing in Arabidopsis thaliana. Nucleic acids research, 38(1), 265-278.

150. Lewis, J. D., Izaurralde, E., Jarmolowski, A., McGuigan, C., & Mattaj, I. W. (1996). A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5'splice site. Genes & development, 10(13), 1683-1698.

151. Laubinger, S., Sachsenberg, T., Zeller, G., Busch, W., Lohmann, J. U., Rätsch, G., & Weigel, D. (2008). Dual roles of the nuclear cap-binding complex and SERRATE in pre-mRNA splicing and microRNA processing in Arabidopsis thaliana. Proceedings of the National Academy of Sciences, 105(25), 8795-8800.

152. Guil, S., & Cáceres, J. F. (2007). The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. Nature structural & molecular biology, 14(7), 591-596.

153. Schwab, R., Speth, C., Laubinger, S., & Voinnet, O. (2013). Enhanced microRNA accumulation through stemloop‐adjacent introns. EMBO reports, 14(7), 615-621.

154. Iki, T., Yoshikawa, M., Nishikiori, M., Jaudal, M. C., Matsumoto-Yokoyama, E., Mitsuhara, I., Ishikawa, M. (2010). In vitro assembly of plant RNA-induced silencing complexes facilitated by molecular chaperone HSP90. Molecular cell, 39(2), 282-291.

155. Iki, T., Yoshikawa, M., Meshi, T., & Ishikawa, M. (2012). Cyclophilin 40 facilitates HSP90‐mediated RISC assembly in plants. The EMBO journal, 31(2), 267-278.

156. Gu, S., Jin, L., Huang, Y., Zhang, F., & Kay, M. A. (2012). Slicing-independent RISC activation requires the argonaute PAZ domain. Current Biology, 22(16), 1536-1542.

157. Kwak, P. B., & Tomari, Y. (2012). The N domain of Argonaute drives duplex unwinding during RISC assembly. Nature structural & molecular biology, 19(2), 145-151.

158. Eamens, A. L., Smith, N. A., Curtin, S. J., Wang, M.-B., & Waterhouse, P. M. (2009). The Arabidopsis thaliana double-stranded RNA binding protein DRB1 directs guide strand selection from microRNA duplexes. Rna, 15(12), 2219-2235.

159. Chatterjee, S., & Großhans, H. (2009). Active turnover modulates mature microRNA activity in Caenorhabditis elegans. Nature, 461(7263), 546-549.

160. Liu, N., Abe, M., Sabin, L. R., Hendriks, G.-J., Naqvi, A. S., Yu, Z., Bonini, N. M. (2011). The Exoribonuclease Nibbler Controls 3′ End Processing of MicroRNAs in Drosophila. Current Biology, 21(22), 1888-1893.

161. Ramachandran, V., & Chen, X. (2008). Degradation of microRNAs by a family of exoribonucleases in Arabidopsis. Science, 321(5895), 1490-1492.

162. Yang, Z., Ebright, Y. W., Yu, B., & Chen, X. (2006). HEN1 recognizes 21–24 nt small RNA duplexes and deposits a methyl group onto the 2′ OH of the 3′ terminal nucleotide. Nucleic acids research, 34(2), 667-675.

163. Huang, Y., Ji, L., Huang, Q., Vassylyev, D. G., Chen, X., & Ma, J.-B. (2009). Structural insights into mechanisms of the small RNA methyltransferase HEN1. Nature, 461(7265), 823-827.

164. Horwich, M. D., Li, C., Matranga, C., Vagin, V., Farley, G., Wang, P., & Zamore, P. D. (2007). The Drosophila RNA Methyltransferase, DmHen1, Modifies Germline piRNAs and Single-Stranded siRNAs in RISC. Current Biology, 17(14), 1265-1272.

165. Kamminga, L. M., Luteijn, M. J., den Broeder, M. J., Redl, S., Kaaij, L. J., Roovers, E. F., Ketting, R. F. (2010). Hen1 is required for oocyte development and piRNA stability in zebrafish. The EMBO journal, 29(21), 3688-3700.

166. Kirino, Y., & Mourelatos, Z. (2007). The mouse homolog of HEN1 is a potential methylase for Piwi-interacting RNAs. Rna, 13(9), 1397-1401.

167. Saito, K., Sakaguchi, Y., Suzuki, T., Suzuki, T., Siomi, H., & Siomi, M. C. (2007). Pimet, the Drosophila homolog of HEN1, mediates 2′-O-methylation of Piwi-interacting RNAs at their 3′ ends. Genes & development, 21(13), 1603-1608.

168. Kurth, H. M., & Mochizuki, K. (2009). 2′-O-methylation stabilizes Piwi-associated small RNAs and ensures DNA elimination in Tetrahymena. Rna, 15(4), 675-685.

169. Li, J., Yang, Z., Yu, B., Liu, J., & Chen, X. (2005). Methylation Protects miRNAs and siRNAs from a 3′-End Uridylation Activity in Arabidopsis. Current Biology, 15(16), 1501-1507.

170. Ibrahim, F., Rohr, J., Jeong, W.-J., Hesson, J., & Cerutti, H. (2006). Untemplated oligoadenylation promotes degradation of RISC-cleaved transcripts. Science, 314(5807), 1893-1893.

171. Ibrahim, F., Rymarquis, L. A., Kim, E.-J., Becker, J., Balassa, E., Green, P. J., & Cerutti, H. (2010). Uridylation of mature miRNAs and siRNAs by the MUT68 nucleotidyltransferase promotes their degradation in Chlamydomonas. Proceedings of the National Academy of Sciences, 107(8), 3906-3911.

172. Ren, G., Chen, X., & Yu, B. (2012). Uridylation of miRNAs by HEN1 SUPPRESSOR1 in Arabidopsis. Current Biology, 22(8), 695-700.

173. Zhao, Y., Yu, Y., Zhai, J., Ramachandran, V., Dinh, T. T., Meyers, B. C., Chen, X. (2012). The Arabidopsis Nucleotidyl Transferase HESO1 Uridylates Unmethylated Small RNAs to Trigger Their Degradation. Current Biology, 22(8), 689-694.

174. Ameres, S. L., Horwich, M. D., Hung, J.-H., Xu, J., Ghildiyal, M., Weng, Z., & Zamore, P. D. (2010). Target RNA–directed trimming and tailing of small silencing RNAs. Science, 328(5985), 1534-1539.

175. Jones, M. R., Quinton, L. J., Blahna, M. T., Neilson, J. R., Fu, S., Ivanov, A. R., Mizgerd, J. P. (2009). Zcchc11-dependent uridylation of microRNA directs cytokine expression. Nature cell biology, 11(9), 1157-1163.

176. Wyman, S. K., Knouf, E. C., Parkin, R. K., Fritz, B. R., Lin, D. W., Dennis, L. M., Tewari, M. (2011). Post-transcriptional generation of miRNA variants by multiple nucleotidyl transferases contributes to miRNA transcriptome complexity. Genome research, 21(9), 1450-1461.

177. Kirino, Y., & Mourelatos, Z. (2007). 2′-O-methyl modification in mouse piRNAs and its methylase. Paper presented at the Nucleic acids symposium series.

178. Katoh, T., Sakaguchi, Y., Miyauchi, K., Suzuki, T., Kashiwabara, S.-i., Baba, T., & Suzuki, T. (2009). Selective stabilization of mammalian microRNAs by 3′ adenylation mediated by the cytoplasmic poly (A) polymerase GLD-2. Genes & development, 23(4), 433-438.

179. van Wolfswinkel, J. C., Claycomb, J. M., Batista, P. J., Mello, C. C., Berezikov, E., & Ketting, R. F. (2009). CDE-1 affects chromosome segregation through uridylation of CSR-1-bound siRNAs. Cell, 139(1), 135-148.

CHAPTER 2

A subgroup of SGS3-like proteins act reduncdantly in RNA-directed DNA methylation

Nucleic Acids Research. **(2012) Vol. 40(10):4422-4431**

Meng Xie, Guodong Ren, Pedro Costa-Nunes, Olga Pontes and Bin Yu

Abstract:

Plant specific SGS3-LIKE proteins are composed of various combinations of an RNA-binding XS domain, a zinc-finger zf-XS domain, a coil-coil domain and a domain of unknown function called XH. In addition to IDN2 and SGS3, the *Arabidopsis* genome encodes twelve uncharacterized SGS3-LIKE proteins. Here, we show that a group of SGS3-LIKE proteins act redundantly in RNA-directed DNA methylation (RdDM) pathway in *Arabidopsis*. Transcriptome co-expression analyses reveal significantly correlated expression of two SGS3-LIKE proteins, FACTOR of DNA METHYLATION 1 (FDM1) and FDM2 with known genes required for RdDM. The *fdm1* and *fdm2* double mutations but not the *fdm1* or *fdm2* single mutations significantly impair DNA methylation at RdDM loci, release transcriptional gene silencing and dramatically reduce the abundance of siRNAs originated from high-copy-number repeats or transposons. Like IDN2 and SGS3, FDM1 binds dsRNAs with 5' overhangs. Double mutant analyses also reveal that IDN2 and three uncharacterized SGS3-LIKE proteins FDM3, FDM4, and FDM5 have overlapping function with FDM1 in RdDM. Five FDM proteins and IDN2 define a group of SGS3-LIKE proteins that possess all four signature motifs in *Arabidopsis*. Thus, our results demonstrate that this group of SGS3-LIKE proteins is important component of RdDM. This study further enhances our understanding of the *SGS3* gene family and the RdDM pathway.

Introduction

In many eukaryotes, RNA-directed DNA methylation (RdDM) is often associated with transcriptional silencing (TGS) and is considered as an essential mechanism to maintain genome stability and to suppress the proliferation of transposable elements (1,2). A key component of RdDM is ~20-24-nucleotide (nt) small interfering RNA derived from transposon or repetitive sequences (rasiRNA) that associates with the ARGONAUTE (AGO) proteins to guide *de novo* cytosine methylation at its homolog loci (1,2). In *Arabidopsis thaliana*, the generation of 24-nt rasiRNAs depends on the RNA-dependent RNA polymerase 2 (RDR2), DICER-LIKE 3 (DCL3), the SNF2-like chromatin-remodeling factor CLASSY 1 (CLSY1) and the plant specific DNA-dependent RNA polymerase IV (Pol IV) (3-6). Pol IV associates with siRNA-generating loci and is thought to generate single-stranded RNAs (ssRNAs) from these loci, which are presumably converted into dsRNAs by RDR2 and subsequently processed by DCL3 into 24 nt rasiRNA duplex (3-7). CLSY1 is required for the correct localization of Pol IV and RDR2 (8).

After generation, one strand of siRNA duplexes is loaded into AGO4, AGO6 or AGO9 (9-11). Presumably through base-pairing between siRNA and Pol V-dependent transcripts and/or physical interaction with NRPE1, which is the largest subunit of Pol V, AGO4 is guided to targets to recruit Domains Rearranged Methyltransferase 2 (DRM2) to catalyze *de novo* cytosine DNA methylation at symmetric CG, CHG (H is adenine, thymine or cytosine) and asymmetric CHH context (12-14). It was recently shown that Pol II might recruit AGO4, Pol IV and Pol V to chromatin through its transcripts or transcription

activity at intergenic low-copy-number loci (7). Additional RdDM components include SUPPRESSOR OF TY INSERTION 5-LIKE (SPT5L, also known as KTF1), DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1 (DRD1), DEFECTIVE IN MERISTEM SILENCING 3 (DMS3) and RNA-DIRECTED DNA METHYLATION 1 (RDM1) (15-21). SPT5L interacts with both Pol V transcripts and AGO4 and is thought to act downstream of the RdDM pathway (16,18), whereas DRD1, DMS3 and RDM1 form a DDR complex that is required for the generation of Pol V-dependent transcripts (17,21).

The plant specific SGS3 gene family encodes proteins containing at least one of the following protein domains: XS, XH, and zf-XS that were named after *Arabidopsis* SGS3 and its rice homolog X1 (22,23). Among these protein domains, the XS domain is an RNA-binding domain, zf-XS domain is a C2H2 type zinc finger domain and XH domain refers to X-homolog domain with unknown function (22). In addition to these protein domains, some of SGS3-LIKE proteins also contain a coil-coil domain localized between the XS and XH domains (15,24). *Arabidopsis* encodes 14 SGS3-LIKE proteins including SGS3 and INVOLVED IN DE NOVO 2 (IDN2, also called RDM12) (15,23,24). While SGS3 is an essential component of post-transcriptional silencing (PTGS) required for the production of sense-transgene induced siRNAs and trans-acting siRNAs (23,25), IDN2/RDM12 is for RdDM and required for transcriptional silencing (TGS) (15,24). Both SGS3 and IDN2 bind dsRNAs with a 5’ overhang (15,26). However, the functions of remaining 12 SGS3-LIKE proteins are still unknown.

Here we identify five SGS3 homologs, FACTOR of DNA METHYLATION (FDM) 1, 2, 3, 4 and 5, as important components of RdDM. Using a combination of transcriptome co-expression analysis and reverse genetics, we found that *FDM1* and *FDM2* display a highly correlated expression pattern with known components of RdDM. Both FDM1 and FDM2 act redundantly in DNA methylation, accumulation of Pol V-dependent rasiRNAs and silencing of RdDM loci. However, FDM1 and FDM2 are not required for the accumulation of Pol V- and Pol II-dependent scaffold transcripts. Furthermore, we show that IDN2 and three uncharacterized SGS3-LIKE proteins FDM3, FDM4, and FDM5 have overlapping function with FDM1 in RdDM. FDM2 also have redundant function with IDN2 in RdDM. These findings broaden our knowledge of RdDM and the function of the SGS3 gene family.

Results:

AT1g15910 and At4g00380 co-expressed with genes in the RdDM pathway

Phylogenic analyses using full-length protein sequences assigned fourteen *Arabidopsis* SGS3 family members into three subfamilies (Figure 2-1A; 34). SGS3 from the first subgroup and IDN2 from the second subgroup have been shown to act in PTGS and TGS, respectively (15,23,24). However, no members from the third subgroup were studied. To extend our understanding of SGS3-LIKE proteins, we selected At1g15910 and At4g00380 from subgroup 3 for functional characterization as they contain the zf-XS, XS, XH and coil-coil domains (Figure 2-1B). The protein sequences of At1g15910 and At4g00380 are highly similar (93% identities and 96% similarities; Figure 2-7),

indicating that they might have redundant function. This was supported by the similar expression pattern between At1g15910 and At4g00380 in leaves, flowers, stem and roots (Figure 2-1C). However, they displayed altered expression levels in leaves, flowers, stem and roots, suggesting that their expression may be developmentally regulated (Figure 2-1C).

To infer the functions of At1g15910 and At4g00380, we searched for their co-expression genes within the ATTED-II developmental expression data set using a co-expression analysis program at the RIKEN PRIMe website (35,36). This search was based on the hypothesis that genes involved in a particular biological process often share regulatory systems thus having a similar expression pattern (37). Because of cross-hybridization between At1g15910 and At4g00380 in the microarray experiments, they were considered as a single gene in the analysis. The results showed that At1g15910/At4g00380 had a very strong correlation with *AGO4*, *NRPE1*, *DRD1*, *DMS3, IDN2* and *RDR2* (correlation coefficiency $r>0.83$; Figure 2-1D). These RdDM genes were coordinately expressed with At1g15910/At4g00380 in roots, embryos, siliques, leaves, stems, and flowers (Figure 2-8), according to the *Arabidopsis* eFP-Browser, which was developed to interpret gene expression data of *Arabidopsis* (38). At1g15910/ At4g00380 also had a considerably high correlation with *DCL3*, *NRPD1* and *DRM2* ($0.68<r<0.83$; table 2-1) as their expression was overlapped in various tissues and at different development stages (Figure 2-8). Altogether, these results showed the correlation between At1g15910/At4g00380 and known genes involved the RdDM pathway, and therefore, suggested their potential role in RdDM. We named At1g15910 and At4g00380 *FACTOR of DNA METHYLATION 1*

(*FDM1*) and *FACTOR of DNA METHYLATION 2* (*FDM2*), respectively, because we subsequently showed that they acted in RdDM (see below).

FDM1 and FDM2 have redundant and essential roles in RdDM

To examine the function of *FDM1* and *FDM2*, two T-DNA insertion lines, SALK_075378 for *FDM1* (39) and SAIL_291_F01 for *FDM2* (40) were obtained from the *Arabidopsis* stock center (http://www.arabidopsis.org) and further characterized. As a first step, plants homozygous for SALK_075378 (named *fdm1-1*) and SAIL_291_F01 (named *fdm2-1*) were identified by PCR genotyping (Figure 2-9). Sequence analysis of the flanking regions of the T-DNA revealed that *fdm1-1* contained a T-DNA insertion in the first intron (949 bp downstream from the ATG site) of *FDM1* and *fdm2-1* harbored a T-DNA insertion in the fifth intron (2252 bp downstream from the ATG site) of *FDM2* (Figure 2-1E). Using RT-PCR analysis, we failed to detect the transcripts of *FDM1* and *FDM2* in *fdm1-1* and *fdm2-1* (Figure 2-1F), respectively, indicating that they are potentially null alleles of *FDM1* and *FDM2*. As *FDM1* and *FDM2* might have redundant functions, we constructed a *fdm1-1 fdm2-1* double mutant by crossing the two respective single mutant lines. No obvious phenotypic abnormalities were observed in *fdm1-1*, *fdm2-1* and *fdm1-1 fdm2-1* (Figure 2-10).

To evaluate whether FDM1 and FDM2 have roles in the RdDM pathway, we examined DNA methylation status at known RdDM-regulated retrotransposon such as *AtSN1* and *ING5* in *fdm1-1*, *fdm2-1, fdm1-1 fdm2-1* and *Arabidopsis* ecotype Columbia (Wild type control; WT) plants by using methylation sensitive HaeIII restriction enzyme digestion

followed by PCR that identifies CHH methylation. HaeIII cannot cleave *ATSN1* and *ING5* DNAs from WT due to DNA methylation at its cleavage site (14,41). A reduction in DNA methylation will cause *AtSN1* and *ING5* DNAs to be less resistant to HaeIII cleavage, resulting in reduced or undetectable PCR products (14,41). As shown in Figure 2A, *fdm1-1* but not *fdm2-1* showed a moderate reduction of DNA methylation at *AtSN1* and *ING5* loci relative to WT. A reduction of DNA methylation at short interspersed repetitive elements upstream of *FWA* gene (*FWA SINE*) in *fdm1-1* but not in *fdm2-1* was also detected by methylation sensitive AvaII enzyme digestion analysis (Figure 2-2A) (27). The reduction of DNA methylation in *fdm1-1* but not *fdm2-1* may be correlated with the reduction of *FDM2* transcript abundance in *fdm1-1* and increased *FDM1* transcript levels in *FDM2-1* (Figure 2-1F). Introducing the wild-type *FDM1* genomic DNA into *fdm1-1* fully recovered the DNA methylation levels at the *AtSN*1 locus (Figure 2-11A), demonstrating that the reduction in DNA methylation in *fdm1-1* is due to *FDM1* loss-of-function. The restriction digestion patterns of *AtSN1*, *ING5 and FWA SINE DNAs* in *fdm1-1 fdm2-1* were similar to *nrpe1-1*, indicating a strong loss of DNA methylation at these loci (Figure 2-2A). The reduction of DNA methylation at *AtSN1* locus in *fdm1-1 fdm2-1* was further confirmed by McrBC enzyme digestion followed by PCR (Figure 2-2B). The McrBC enzyme cuts methylated but not unmethylated DNA. A reduction in DNA methylation will result in increased PCR products after McrBC treatment. This assay also revealed a reduction in DNA methylation at the *siR02* locus in *fdm1-1 fdm2-1* (Figure 2-2B). We further examined the DNA methylation status of *5S rDNA*, *AtMU1*, and *MEA-ISR* using the methylation-sensitive restriction enzyme HaeIII, HpaII (for CG and CHG methylation) and MspI (for CG methylation) followed by Southern blotting

(18,27,28). A strong reduction in DNA methylation at *5S rDNA*, *AtMU1* and *MEA-ISR* loci comparable to *nrpe1-1* was observed in *fdm1-1 fdm2-1* but not in *fdm1-1* and *fdm2-1* (Figure 2-2C, 2D and 2E). Next, we examined the methylation status of the highly repetitive 180-bp centromeric repeat that is not an RdDM target (11). The DNA methylation at this locus showed no obvious alteration in *fdm1-1 fdm2-1* and *nrpe1-1* compared with WT (Figure 2-2F). This indicated that the function of FDM1 and FDM2 in DNA methylation is rasiRNA-dependent. To confirm that the strong reduction of DNA methylation in *fdm1-1 fdm2-1* is due to lack of both *FDM1* and *FDM2*, we introduced the wild-type *FDM1* or *FDM2* genomic DNA into *fdm1-1 fdm2-1*. Two randomly chosen transgenic *fdm1-1 fdm2-1* lines harboring the *FDM1* transgene showed comparable DNA methylation levels at *AtSN1* and *ING5* with WT and *fdm2-1*, while two *fdm1-1 fdm2-1* lines containing the *FDM2* transgene have similar DNA methylation levels to *fdm1-1* (Figure 2-11B). These results demonstrated that FDM1 and FDM2 act redundantly in RdDM.

Next, we examined the expression levels of *AtSN1*, 5s rRNA spacer and siR02 in *fdm1-1*, *fdm2-1* and *fdm1-1 fdm2-1*, *nrpe1-1* and WT by RT-PCR. Their transcripts in *fdm1-1 fdm2-1* but not in *fdm1-1* and *fdm2-1* were significantly increased to levels comparable to *nrpe1-1* (Figure 2-3A and 2-3B). These results revealed that the reduction of DNA methylation in *fdm1-1 fdm2-1* is correlated with derepression of RdDM target loci.

The levels of Pol V-dependent rasiRNAs are reduced in *fdm1-1 fdm2-1*.

Based on their dependence on Pol V and Pol IV, rasiRNAs are classified into two types (27). The accumulation of type I rasiRNAs that are derived from highly repetitive DNA sequences, including AtSN1, siR1003 (from *5S rDNA*), AtREP2, SimpleHAT2, and AtCopia2, depends on both Pol V and Pol IV, whereas the levels of type II rasiRNAs generated from low-copy number DNA repeats, such as siR02, Cluster4, TR2558, Cluster2, and soloLTR, require Pol IV but not Pol V (27).

We examined the accumulation of both type I rasiRNAs and type II rasiRNAs in *fdm1-1*, *fdm2-1*, and *fdm1-1 fdm2-1* by Northern blotting. The accumulation of both type I rasiRNAs (AtSN1, siRNA 1003, Atcopia and SimpleHAT2) and type II rasiRNAs (siR02, Cluster4, TR2558) was reduced in *dcl3-1* but not in *fdm1-1* and *fdm2-1* relative to WT (Figure 2-3D and 2-3E). Like in *nrpe1-1*, the accumulation of type I but not type II rasiRNAs was significantly reduced in *fdm1-1 fdm2-1* compared with WT (Figure 2-3D and 2-3E). These results suggested that FDM1 and FDM2 act redundantly to promote the accumulation of type I rasiRNAs but not type II rasiRNAs. We next tested whether FDM1 and FDM2 were involved in the accumulation of microRNAs (miRNAs). However, the levels of DCL1-dependent miR172 and miR173 in *fdm1-1*, *fdm2-1* and *fdm1-1 fdm2-1* were similar to those in WT (Figure 2-12).

FDM1 and FDM2 are not required for the localization of NRPD1, RDR2, NRPE1 and AGO4 and for the accumulation of Pol V- or Pol II-dependent non-coding transcripts.

To explore the role of FDM1 and FDM2 in RdDM, we examined the nuclear localization of NRPD1, RDR2, NRPE1 and AGO4 in *fdm1-1 fdm2-1*. As shown in Figure 4, in both WT and *fdm1-1 fdm2-1* nuclei NRPD1 displayed punctate foci signals in the nucleoplasm. In contrast, as previously reported (31,42), RDR2, NRPE1 and AGO4 showed a round-shaped nucleolar signal in addition to puncta or diffuse signals outside the nucleolus both in WT and *fdm1-1 fdm2-1* (Figure 2-4). Thus, the *fdm1-1* and *fdm2-1* double mutations have no effects on the localization of the RdDM players NRPD1, NRPE1, RDR2 and AGO4.

Next, we tested the requirement of FDM1 and FDM2 for the accumulation of Pol V- or Pol II-dependent non-coding transcripts that serve as scaffolds to recruit AGO4-siRNA complex to chromatin (7,43). RT-PCR analyses showed that the Pol V-dependent transcripts at *AtSN1* locus (interval B) and Pol II-dependent transcripts at *siR02* locus (interval B) were not affected in *fdm1-1 fdm2-1* (Figure 2-3C).

FDM1 binds double-stranded RNAs (dsRNAs) with 5' overhangs

As the SGS3 and IDN2 have been shown to bind dsRNAs, we tested whether FDM1 is an RNA-binding protein using a GST-pull down assay. Because the truncated SGS3 and IDN2 proteins containing the XS and coil-coil domains are able to bind dsRNAs, we expressed a truncated version of FDM1 lacking the zinc finger and XH domain fused with GST tag at its N-terminus (GST-FDM1ΔZH) and a GST control protein in *E. coli*. The GST-FDM1ΔZH and GST proteins were purified with glutathione beads (Figure 2-5A). We prepared various radioactive-labeled RNA species including ssRNAs, dsRNAs

with 3’ overhangs and dsRNAs with 5’overhangs (Figure 2-5B and 2-5C). These probes were incubated with the glutathione beads containing GST-FDM1ΔZH or GST alone. GST-FDM1ΔZH retained radioactive 35 bp dsRNAs with 18 nt 5’ overhangs at each end but not 53 nt ssRNAs and a 36 bp dsRNAs with 17 nt 3’ overhang at each end, whereas GST alone did not bind any RNA species (Figure 2-5B and 2-5C). Furthermore, addition of unlabelled dsRNAs of the same sequence efficiently reduced the binding of radioactive probe by GST-FDM1ΔZH (Figure 2-5C). These results demonstrated that FDM1 binds dsRNAs with 5’ overhangs.

RNA-mediated in vitro AGO4-FDM1 interaction.

We next tested whether FDM1 interacts with AGO4 and RDR2 by *in vitro* protein pull-down assay in order to gain insight on the function of FDM1 in RdDM. A full-length FDM1 fused with a GST-tag at its N-termini was expressed in *E.coli* and purified with glutathione beads (Figure 2-13). The glutathione beads conjugated with GST-FDM1 were incubated proteins extracts containing HA-RDR2 or MYC-AGO4. Western blot detected the enrichments of MYC-AGO4 but not HA-RDR2 in the GST-FDM1 complex (Figure 2-13). In contrast, the control GST protein alone failed to pull down MYC-AGO4 (Figure 2-13). Because both AGO4 and FDM1 are RNA binding proteins, we tested whether the interaction is RNA-mediated. RNase A treatment abolished AGO4-FDM1 interaction (Figure 2-13).

FDM1 and FDM2 have overlapping functions with IDN2 in the RdDM pathway

Because FDM1 and FDM2 protein sequences share considerable similarities with that of

IDN2/RDM12 (~60%) and all of them are involved in RdDM, we asked whether they have overlapping functions. We obtained a T-DNA insertion line Salk_152144 for *IDN2/RDM12* from the *Arabidopsis* stock center and identified homozygous mutants by PCR genotyping (Figure 2-14). We named this line *idn2-3*. The transcript levels of *IDN2* were reduced in *idn2-3* (Figure 2-14C), resulting in a moderate reduction in DNA methylation at *AtSN1* and *ING5* loci (Figure 2-6A). We constructed two double mutants, *fdm1-1 idn2-3* and *fdm2-2 idn2-3* by crossing single mutants and analyzed DNA methylation status at *AtSN1* and *ING5* loci. Like *fdm1-1 fdm2-1*, *fdm1-1 idn2-3* and *fdm2-1 idn2-3* showed strong reduction in DNA methylation compared with each of single mutants (Figure 2-6A). It was noticed that the *fdm1-1 idn2-3* showed a stronger reduction in DNA methylation at *IGN5* locus than *fdm1-1 fdm2-1* and *fdm2-1 idn2-3*. This result may be related to the reduced expression of *FDM2* in the *fdm1-1* genetic background (Figure 2-1F). *fdm1-1 idn2-3* also displayed reduced DNA methylation at *5S rDNA* locus relative to *fdm1-1* and *idn2-3* (Figure 2-6B).

FDM3, FDM4 and FDM5 act redundantly with FDM1 in RdDM.

IDN2/RDM12, FDM1 and FDM2 have three additional homologs At3G12550 (subfamily 2), At1g13790 (subfamily 2) and At1g80790 (subfamily 3) that contain all four-signature motifs of SGS3 protein family in *Arabidopsis*. We named these proteins FDM3, FDM4, and FDM5 respectively and tested whether they have functions in RdDM. Homozygous T-DNA insertion lines Salk_020841 for At3G12550 (*fdm3-1*), Salk_008738 (*fdm4-1*) for At1G13790 and Salk_052192 (*fdm5-1)* for At1G80790 were obtained from *Arabidopsis* center (figure 2-14). No transcripts for *FDM3*, *FDM4* were detected in *fdm3-1* and *fdm4-*

1, respectively, whereas the abundance of FDM5 transcripts was reduced significantly in *fdm5-1* (Figure 2-14). The DNA methylation status of *ATSN1*, *ING5* and *5S rDNA* loci in *fdm3-1*, *fdm4-1* and *fdm5-1* showed no alteration relative to WT. We next tested whether FDM3, FDM4 and FDM5 have redundant functions with FDM1. In facts, the DNA methylation contents of *ATSN1*, *ING5* and *5S rDNA* loci are strongly reduced in *fdm1-1 fdm3-1, fdm1-1 fdm4-1*, *fdm1-1 fdm5-1* compared with each of single mutants and WT. In *fdm1-1 fdm3-1*, *fdm1-1 fdm4-1* and *fdm1-1 fdm5-1* expressing the *FDM3*, *FDM4* and *FDM5* transgenes under the control of their native promoters, respectively, the DNA methylation content of *ATSN1* and *ING5* is comparable with that in *fdm1-1*, indicating that lack of FDM3, FDM4 or FDM5 is responsible for the enhanced DNA methylation defects in the double mutants (data not shown).

Discussion

The *SGS3-LIKE* genes encode a large uncharacterized protein family. In this study, through a combination of transcriptome co-expression analysis, reverse genetics and biochemical assays, we show that two SGS3-LIKE proteins FDM1 and FDM2 from *Arabidopsis* are essential components of gene silencing triggered by small RNAs. FDM1 and FDM2 share high similarity and lack of both of them causes great reduction in DNA methylation levels and Pol V-dependent rasiRNA accumulation, resulting in release of transcriptional silencing. These results demonstrate that FDM1 and FDM2 have essential and redundant roles in the RdDM pathway.

Co-expression analysis revealed that *AGO4*, *NRPE1*, *DRD1*, *DMS3, IDN2/RDM12, FDM1/FDM2 DCL3,* and *RDR2* are highly correlated with each other (r>0.76; figure 2-1D and table 2-1). NRPD1 and DRM2 also display considerable correlation with these genes (r>0.6 and r>0.5; respectively; table 2-1). These results are supported by their coordinated high expression at DNA-replication active tissues such as inflorescence meristem, shoot meristem and developing embryo (figure 2-8), which agrees with their role in directing *de novo* DNA methylation (1,2). The correlation among genes involved in RdDM indicates that they may share a common regulatory system and tend to be co-expressed. Consequently, searching for co-expressed genes combined with reverse genetic analysis could be a powerful tool to identify novel genes that are involved in RdDM, especially those with functional redundancy.

How do FDM1 and FDM2 function in RdDM? They appear not to be required for the correct localization of NRPD1, RDR2, NRPE1 and AGO4, as these proteins have similar localizations in *fdm1-1 fdm2-1* as in WT (Figure 2-4). Like IDN2 and SGS3 (15,26), FDM1 binds dsRNAs with 5' overhangs (Figure 2-5). Given its sequence similarity and functional redundancy with FDM1, FDM2 most likely interacts with dsRNA with 5' overhangs too. These observations suggest at least two hypotheses for FDM1 and FDM2 function, as indicated for IDN2/RDM12 (15,24). The first is that FDM1 and FDM2 may bind dsRNA produced by RDR2 to stabilize it, which may be required for rasiRNA biogenesis (24). The second is that FDM1 may interact with AGO4-bound dsRNAs generated by base pairing between rasiRNAs and target transcripts produced by Pol II or Pol V to stabilize rasiRNA-target interaction or recruit downstream components such as

DRM2 to chromatin (15,24). *fdm1-1 fdm2-1* displayed reduced DNA methylation levels of both type I and type II rasiRNA generating loci (Figure 2-2) as well as reduced amount of type I rasiRNAs but not type II rasiRNAs (Figure 2-3). These molecular phenotypes of *fdm1-1 fdm2-1* resemble those of *nrpe1*, *ago4*, *rdm1* and *drd1*, indicating that like NRPE1, AGO4, DRD1 and RDM1, FDM1 and FDM2 may act downstream of ra-siRNA initiation in RdDM. In addition, FDM1 and FDM2 are not required for the accumulation of both Pol V-dependent and Pol II-dependent scaffold transcripts, indicating FDM1 and FDM2 may act downstream of Pol V and Pol II activities. Thus, we favor the suggestion that FDM1/FDM2 binds the rasiRNA-target duplex. In fact, an RNA-mediated AGO4-FDM1 association is observed, whereas an RDR2-FDM1 interaction is not detected (Figure 2-13).

The *Arabidopsis* genome encodes 14 SGS3-LIKE proteins (34) that can be assigned into three subfamilies. IDN2 and FDM1/FDM2 belong to subfamily 2 and 3, respectively (Figure 2-1A). However, their protein sequences are very similar (~ 60% similarity), indicating that they may have closely related functions. This notion is strongly supported by the facts that *fdm1-1 idn2-3* and *fdm2-1 idn2-3* show much stronger reduction in DNA methylation than each of single mutants (Figure 2-5). *Arabidopsis* encodes six SGS3-LIKE proteins from family 2 and family 3, including IDN2, FDM1 and FDM2, FDM3, FDM4 and FDM5, which contain all four-signature domains of SGS3-LIKE proteins. The double mutant analyses reveal that FDM3, FDM4 and FDM5 have redundant roles with FDM1 in RdDM (Figure 2-6). Thus our study defines a group of SGS3-LIKE proteins that play important roles in RdDM. Clearly, further work is required to

determine their molecular role in RdDM.

Materials and methods

Plant materials

The T-DNA insertional mutants, *fdm1-1* (SALK_075813) and *fdm2-1* (SAIL_291_F01) and *idn2-3* (Salk_152144) were obtained from the ABRC Stock Center (www.arabidopsis.org). The T-DNA insertions were identified through combination of gene specific primers and T-DNA left border primer (Primers FDM1RP, FDM1LP and LBa1 for *fdm1-1*; primers FDM2RP, FDM2LP and LB3 for *fdm2-1*; primers IDN2RP, IDN2LP and LBa1 for *idn2-3*). The *fdm1-1 fdm2-1*, *fdm1-1 idn2-3*, *fdm2-1 idn2-3* mutants were constructed by crossing single mutants. *nrpe1-1* (27), *dcl3-1* (6) and the *myc-AGO4* transgenic line were kindly gifts from Dr. Xuemei Chen. *Myc-AGO4* is in the Ler genetic background, whereas other mutants are in the Columbia genetic background.

Phylogenetic Analyses

Protein sequences for 14 *Arabidopsis* SGS3-LIKE proteins were obtained from the *Arabidopsis* website (http://www.arabidopsis.org). Full-length protein sequences of 14 SGS3-LIKE proteins were aligned using CLUSTALW at The Biology Work Bench (http://workbench.sdsc.edu/). Phylogenetic analysis was done by the unrooted neighbor-joining method. To assess the degree of reliability for each branch on the tree, bootstrap confidence values of each node were calculated with 1000 replicates using PAUP 4.0 (http://paup.csit.fsu.edu/).

DNA methylation assays

Genomic DNA was extracted from flowers and digested overnight with different methylation-sensitive restriction enzyme (HaeIII, AvaII, HpaII and MspI) or 1hr with McrBC. Approximate 5% of the digested DNA was subsequently used for PCR analysis of *AtSN1*, *IGN5, FWA SINE* and *siR02.* The undigested genomic DNA was amplified simultaneously as loading controls. PCR conditions were: 94°C for 30 seconds (s), 54 °C for 30 s, 72°C for 1 min, 32 cycles, and 72°C for 10 min. For Southern blotting, 5 μg of genomic DNA treated with HaeIII, HpaII, and MspI overnight was resolved in 1.2% agarose gel and transferred to Hybond-N^+ membranes. *5S rDNA*, *MEA-ISR* and *AtMU1* Southern blotting were carried out as described (18,27,28). The primer information was obtained from references (14,18,27,28).

RT-PCR analysis

Total RNA was extracted from flowers using Trizol reagent (Sigma). After DNase treatment, 2-5 μg of total RNA was used to synthesize cDNA with SuperScript III (Invitrogen) using oligo-dT or gene specific primers. The diluted cDNA reaction mixture was used for RT-PCR of *AtSN1, siR02* and *5s rRNA spacer* as previously described (7,29). The constitutively expressed *UBQ5* was used as an internal control. The cDNA reaction mixture without reverse transcriptase was used in PCR amplification to determine the absence of DNA contamination. Pol II- and Pol V- dependent transcripts were detected by RT-PCR according to (14).

SiRNA and miRNA detection

RNA isolation and hybridization were performed according to the method described by (30). siR1003, AtSN1, AtCopia 2, SimpleHAT2, siR02, Cluster4, and TR2558 were detected using 5′-End-labeled (^{32}P) antisense LNA oligonucleotides (7).

Immunolocalization

Leaves from 28-day-old plants were harvested and the immunolocalization experiments were performed as described (8,31).

RNA binding assay

The RNA and DNA binding assays were performed as previously described (32). GST and a truncated form of FDM1 (GST-FDM1ΔZH, amino acid 114-498) fused to GST were expressed in *E. coli* BL21 and purified as described by (33). The templates for RNA1, 2, and 3 were produced by PCR using primers RNA1F/1R, RNA2F/2R and RNA1F/3R, respectively. The template for RNA 1, 2 and 3 is the B region of the *AtSN1* locus. Primers RNA1F, RNA2R and RNA3R contain the T7 promoter. The RNAs were synthesized by *in vitro* transcription with T7 RNA polymerase at the presence or absence of [α^{32}P] UTP. RNA1 was used as ssRNAs in the binding assay. RNA1/RNA2 were annealed to generated dsRNAs with 5' overhangs at both ends. RNA3 is a dsRNA with 3' overhangs at both ends. Annealing was performed in the annealing buffer [10 mM Tris-HCl (pH 8.0), 20 mM NaCl, 1 mM EDTA (pH 8.0)] by incubating RNAs at 95 ^{o}C for 5 min and then gradually cooling to room temperature.

Figures

Figure 2-1. FDM1 and FDM2 are putative components of RNA-directed DNA methylation (RdDM) pathway. (A) Unrooted neighbor-joining phylogenies based on full-length amino acid sequences of 14 *Arabidopsis* SGS3 LIKE proteins. Bootstrap values were given for branch node. Dark grey: subfamily 1; Light Grey: subfamily 2; White: subfamily 3. (B) A scheme of protein structures of At1G15910 (FDM1) and At4G00380 (FDM2). Black box: the zf-XS domain; open box: the XS domain; Gray box: the coil-coil domain; hatched box: the XH domain. (C) RT-PCR analysis of At1G15910 and At4G00380 expression in root, leaf, flower, and stem. Amplification of *UBIQUITIN5* (At3g26650; *UBQ5*) with or without reverse transcription (-RT) is shown as a control. (D) Correlation among several genes involved in RNA-directed DNA methylation pathway and *FDM1/FDM2*. Black circle: *FDM1/FDM2*; Open circle: genes involved in RdDM. solid black line: r>0.9; dot line: 0.9>r>0.830. *: Because of cross hybridization of IDN2 andAt4g01780 in the microarray experiment, they were considered as a single gene during co-expression analysis. (E) Diagrams of T-DNA-insertion in *fdm1-1* and *fdm2-1*, respectively. Black box: coding region; open box: untranslated region; solid black line: intron; open triangle: T-DNA insertion site. Grey arrowheads: primer used for T-DNA genotyping; Black arrowheads: primer used for RT-PCR analysis. (F) RT-PCR analysis of *FDM1* and *FDM2* expression in *fdm1-1*, *fdm2-1* and Col (Wild type; WT). Amplification of *UBQ5* with or without RT (-RT) is shown as a control.

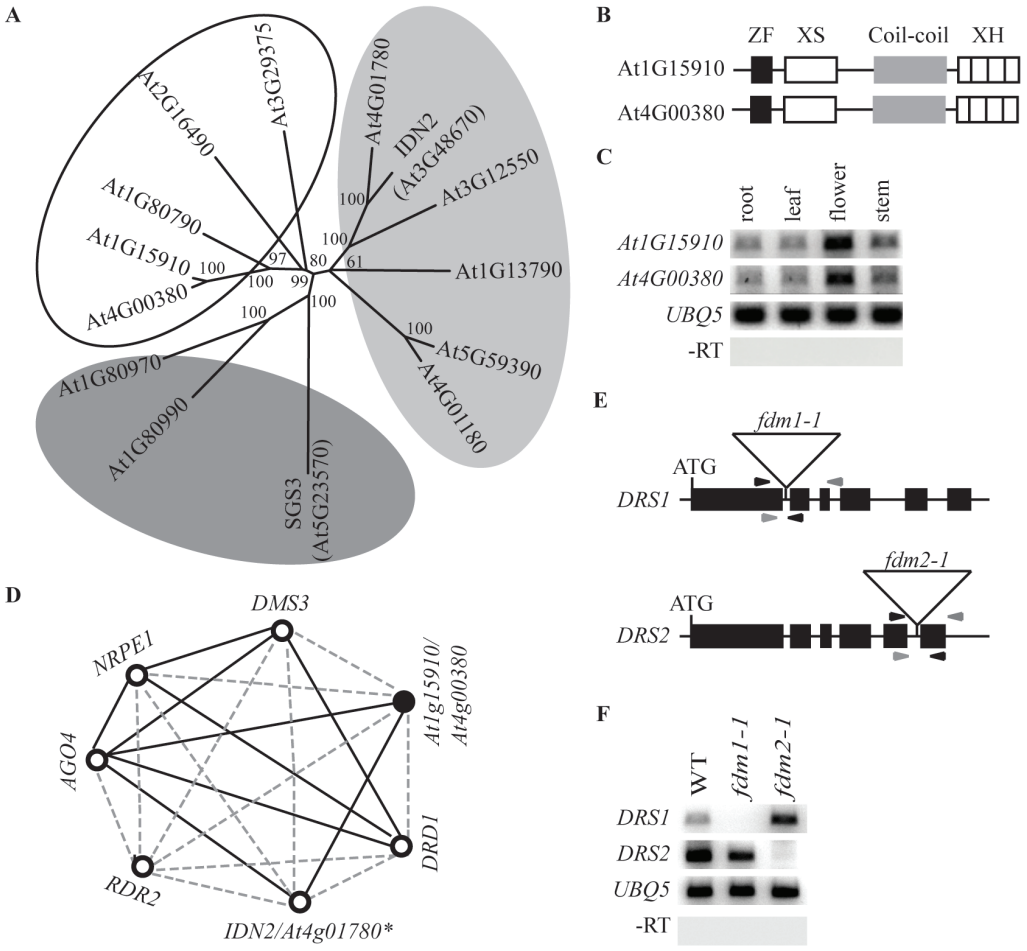


Figure 2-2. FDM1 and FDM2 play redundant and essential roles in RdDM.

(A) Reduced DNA methylation at *AtSN1, IGN5* and *FWA SINE* in *fdm1-1 fdm2-1*. HaeIII digested genomic DNAs from various genotypes were used for PCR amplification of

AtSN1 and *IGN5*, whereas AvaII treated genomic DNAs were used for the amplification of *FWA SINE*. Amplifications of undigested genomic DNA are used as loading controls. Col: wild-type plants. (B) Reduced DNA methylation at *AtSN1* and *siR02* loci in *fdm1-1 fdm2-1*. McrBC digested and undigested DNAs (control) were used for the amplification of *AtSN1* and *siR02* (C) Reduced DNA methylation at *MEA-ISR* in *fdm1-1 fdm2-1*. HpaII or MspI digested genomic DNAs from various genotypes were probed for *MEA-ISR*. Bands representing methylated (ME) or unmethylated (UM) DNA are indicated. (D) Reduced DNA methylation at *5S rDNA* locus in *fdm1-1 fdm2-1*. HaeIII, HpaII or MspI digested genomic DNAs from various genotypes were probed for *5S rDNA*. (E) Reduced DNA methylation at *AtMU1* locus in *fdm1-1 fdm2-1*. HaeIII digested genomic DNAs were probed for *AtMU1*. The three undigested bands presented in Col (WT) but not in *nrpe1-1* and *fdm1-1 fdm2-1* were indicated by arrows. (F) Unaffected DNA methylation at 180 bp centromeric repeats in *fdm1-1 fdm2-1.* Following HpaII, MspI or HaeIII treatment, genomic DNAs from various genotypes were probed for 180 bp centromeric repeats.

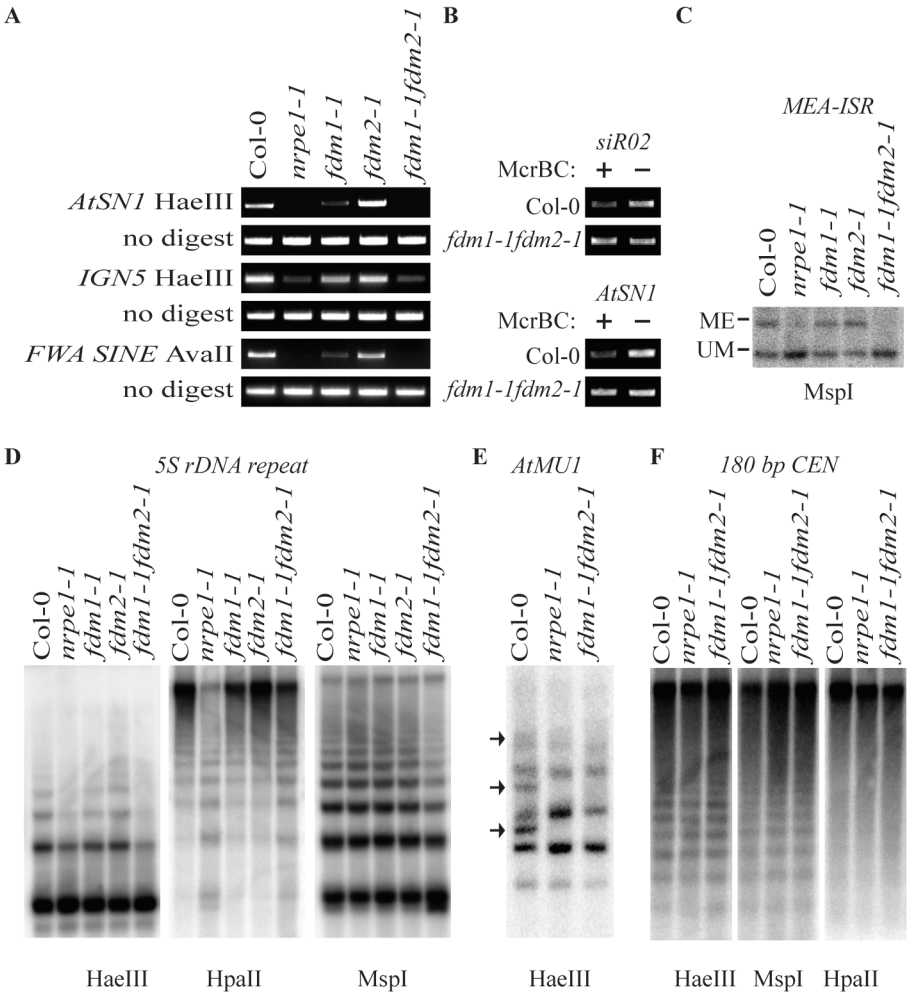

A
Col-0
nrpe1-1
fdm1-1
fdm2-1
fdm1-1fdm2-1
AtSN1 HaeIII
no digest
IGN5 HaeIII
no digest
FWA SINE AvaII
no digest
B
siR02
McrBC: + −
Col-0
fdm1-1fdm2-1
AtSN1
McrBC: + −
Col-0
fdm1-1fdm2-1
C
MEA-ISR
Col-0
nrpe1-1
fdm1-1
fdm2-1
fdm1-1fdm2-1
ME
UM
MspI
D
5S rDNA repeat
Col-0
nrpe1-1
fdm1-1
fdm2-1
fdm1-1fdm2-1
HaeIII
HpaII
MspI
E
AtMU1
Col-0
nrpe1-1
fdm1-1fdm2-1
HaeIII
F
180 bp CEN
Col-0
nrpe1-1
fdm1-1fdm2-1
HaeIII
MspI
HpaII

Figure 2-3. FDM1 and FDM2 prompt the accumulation of type I rasiRNAs and are required for silencing of RdDM loci.

(A-B) Enhanced transcription levels of *AtSN1*, *5S rRNA spacer,* and *siR02* in *fdm1-1 fdm2-1*. Transcripts of RdDM targets were detected by RT-PCR. For *5S rRNA* spacer transcripts, the band (~210 bp) indicated by an arrow corresponds to the silenced transcripts in Col (WT). (C) Unaffected Pol II- and Pol V-dependent noncoding transcripts at flanking region of *ATSN1* and s*iR02* in *fdm1-1 fdm2-1*. The transcripts were detected by strand-specific RT-PCR. The positions of amplified region by RT-PCR are indicated in the diagram on the right. Amplification of *UBQ5* with or without reverse transcription (-RT) is served as a control. (D) Reduced accumulation of type I rasiRNAs in *fdm1-1 fdm2-1.* (E) Unaffected accumulation of type II rasiRNAs in *fdm1-1 fdm2-1.* Various rasiRNAs were detected by northern blotting. The controls U6 rRNA blots and ethidium bromide-stained tRNAs were shown below the corresponding rasiRNA blots.

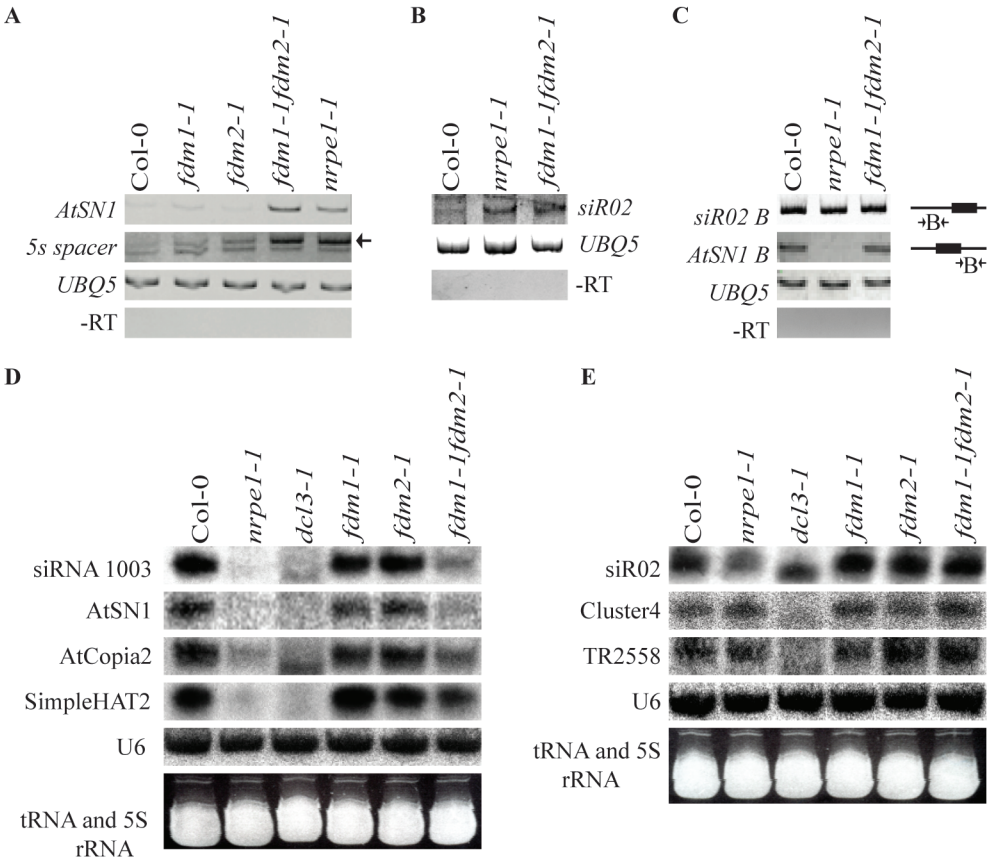

A
Col-0
fdm1-1
fdm2-1
fdm1-1fdm2-1
nrpe1-1
AtSN1
5s spacer
UBQ5
-RT
B
Col-0
nrpe1-1
fdm1-1fdm2-1
siR02
UBQ5
-RT
C
Col-0
nrpe1-1
fdm1-1fdm2-1
siR02 B
AtSN1 B
UBQ5
-RT
B
B
D
Col-0
nrpe1-1
dcl3-1
fdm1-1
fdm2-1
fdm1-1fdm2-1
siRNA 1003
AtSN1
AtCopia2
SimpleHAT2
U6
tRNA and 5S rRNA
E
Col-0
nrpe1-1
dcl3-1
fdm1-1
fdm2-1
fdm1-1fdm2-1
siR02
Cluster4
TR2558
U6
tRNA and 5S rRNA

Figure 2-4. The *fdm1-1* and *fdm2-1* double mutations have no effects on RdDM proteins nuclear localization. NRPD1, RDR2, NRPE1 and AGO4. Peptide antibodies specifically recognizing native NRPD1, RDR2, NRPE1 or AGO4 (in red) were used to perform immunolocalization experiments in *Arabidopsis* leaf nuclei from ecotype Columbia (WT) and *fdm1-1 fdm2-1* mutant line. DNA was counterstained with DAPI. Scale bar corresponds to 5μm.

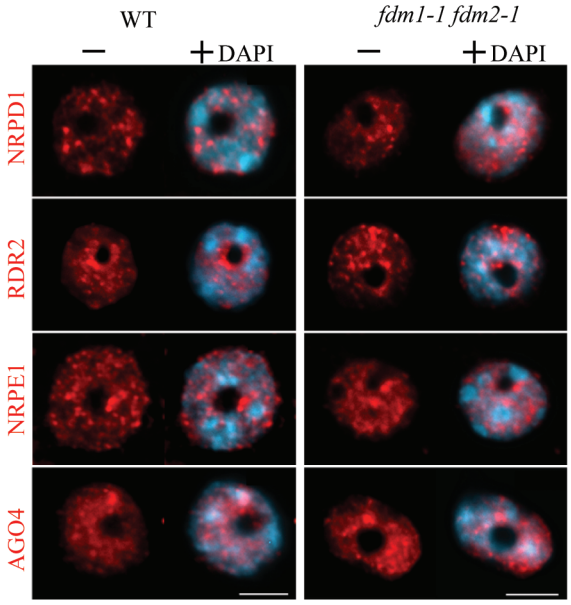


Figure 2-5. FDM1 binds double-stranded RNAs (dsRNAs) with 5' overhangs.

(A) The two purified proteins used in the binding assay, GST and GST-FDM1ΔZH (truncated FDM1 containing XS and SMC domain) were resolved in SDS-PAGE gel and stained with Coomassie Blue. The protein molecular weights are indicated on the right. (B-C) RNA- binding assays of FDM1 with various probes. The structure of various probes is shown on the right. *: radioactive labeled RNA strand. Approximately 50 μg of protein was used for the binding assay. For dsRNAs with 5' overhang, 1×, 10×, and 150× unlabeled RNAs of the same sequence were used for the competition assay.

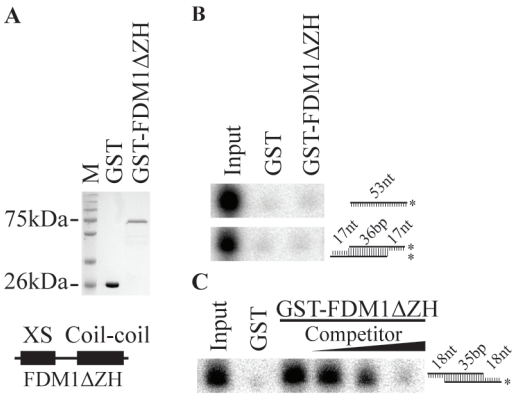


Figure 2-6. FDM1 has overlapping functions with IDN2, FDM3, FDM4 and FDM5. (A) DNA methylation levels at *AtSN1* and *IGN5* loci in various genotypes. HaeIII digested genomic DNAs were used for PCR-amplification of *ATSN1* and *ING5*. Amplification of undigested DNAs was used as loading controls. (B) DNA methylation at 5S rDNA locus in various genotypes.

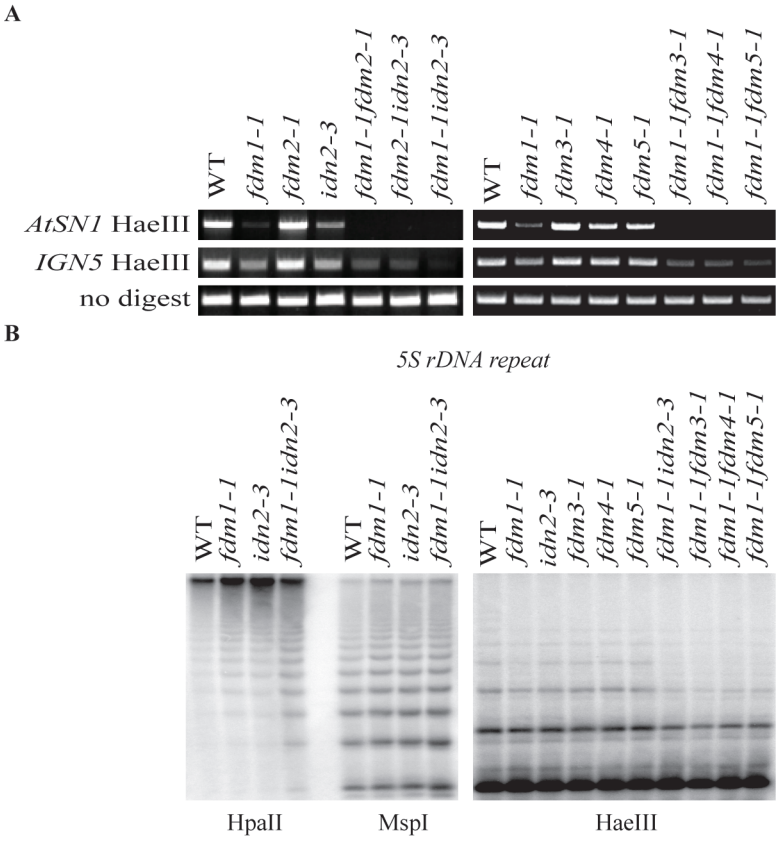


Figure 2-7. Protein sequence alignment between At1g15910 and At4g00380. Black boxes represent identical amino acids. Grey boxes stand for similar amino acids. The alignment is carried out using CLUSTALW software at the Biology Workbench (http://workbench.sdsc.edu/).

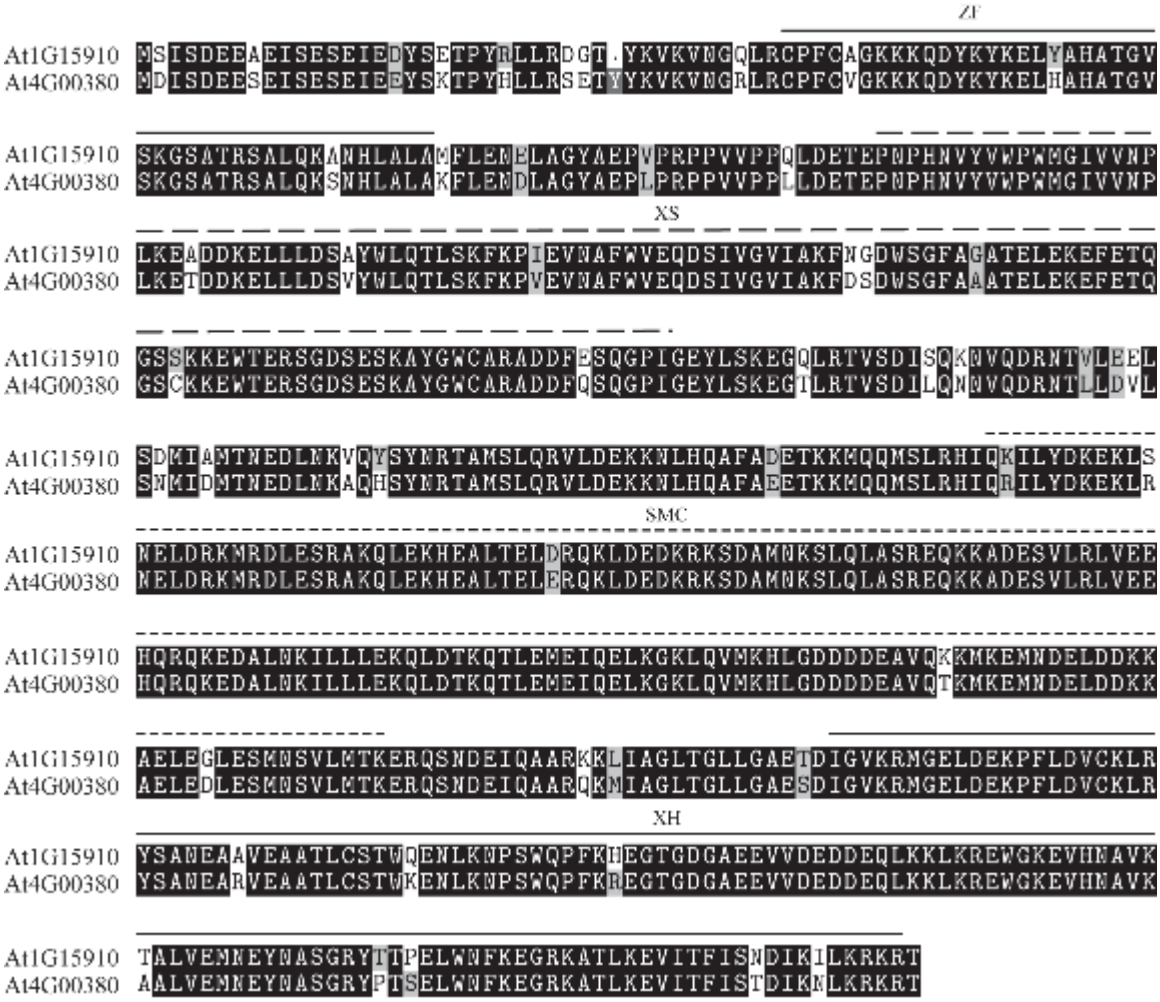
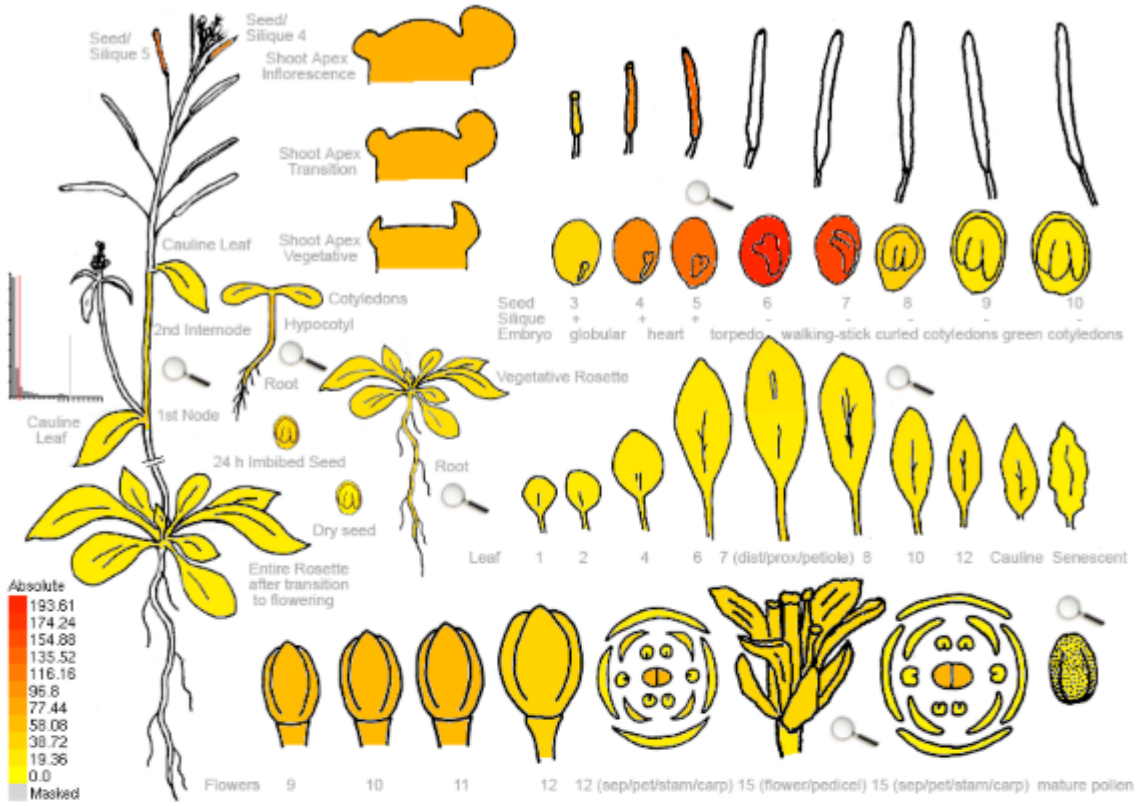


Figure 2-8. **Tissue-specific expression of *At1g15910/At4g00380 (FDM1 /FMD2)* and *AGO4.*** "Electronic fluorescent pictograph" of gene expression levels was generated by the *Arabidopsis* eFP-Browser. Absolute signal intensities were shown as a color scale with low levels of expression colored yellow and high levels colored red.

At1g15910/At4g30080 (*FDM1/FDM2*)



At2g27040 (*AGO4*)

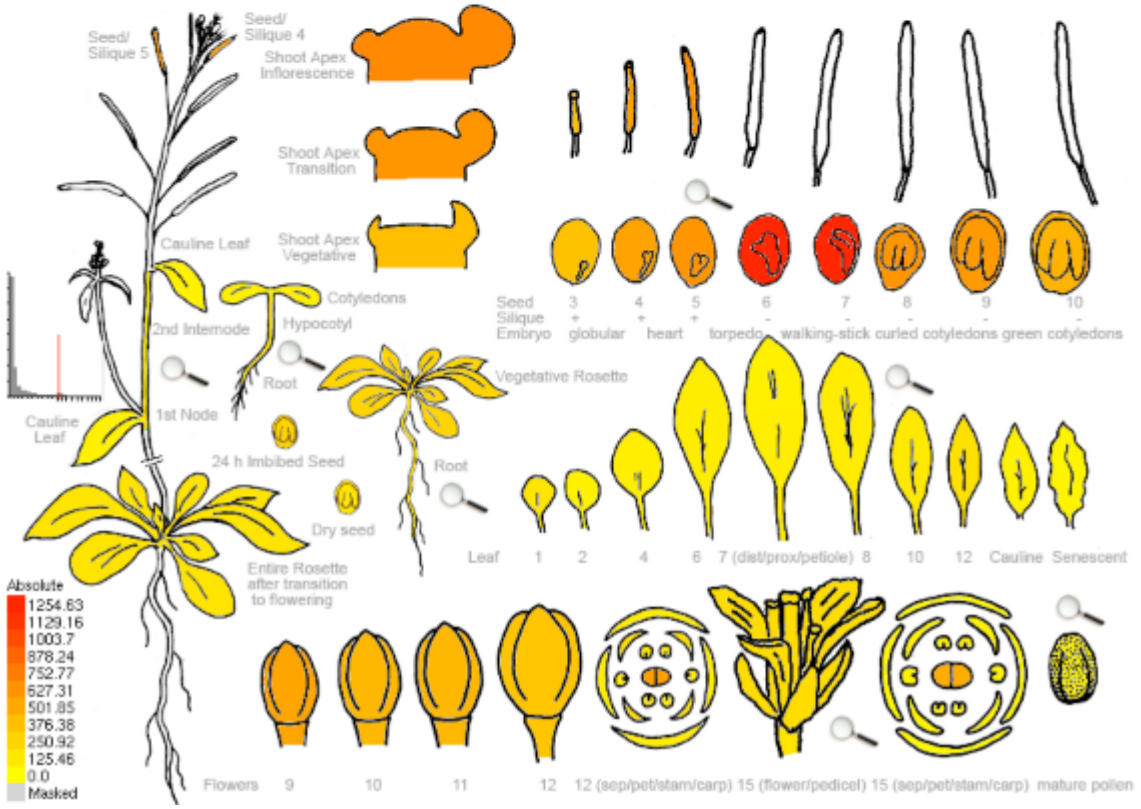


Figure 2-9. Identification of *fdm1-1* and *fdm2-1* by PCR analyses of T-DNA insertion in the *FDM1* and *FDM2* genes. Col: wild-type control (WT); *fdm1-1*: homozygous SALK_075813; *fdm2-1*: homozygous (SAIL_291_F01). FDM1LP/RP: Primer combinations used for FDM1 gene; FDM1RP/LBa1: primer combinations used for the T-DNA flanking genomic DNA of FDM1; FDM2LP/RP: Primer combinations used for *FDM2* gene; FDM2LP/LB3: primer combinations used for the T-DNA flanking genomic DNA of *FDM2*.

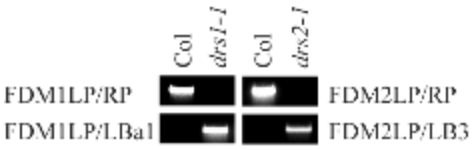


Figure 2-10. Growth of Col-0, *fdm1-1*, *fdm2-1*, and *fdm1-1 fdm2-1*. Picture of 14-day-old plants were taken.

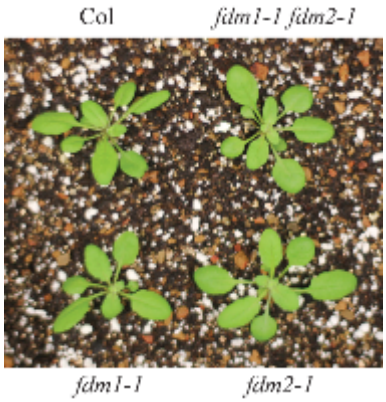


Figure 2-11. Complementation assay of DNA methylation defection in *fdm1-1* and *fdm1-1 fdm2-1*. (A) Complementation assay of *fdm1-1* by expression of *FDM1*. (B) Complementation assay of *fdm1-1 fdm2-1* by expression of *FDM1* or *FDM2*. HaeIII-treated or untreated (control) genomic DNAs from various genotypes were used for amplification of *ATSN1* or *ING5*.

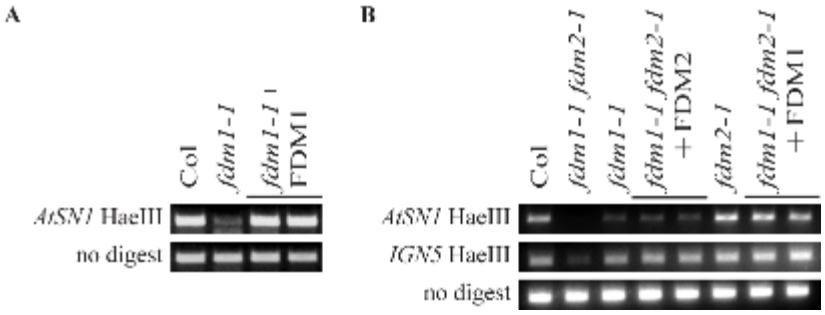


Figure 2-12. The accumulation of miR172 and miR173 in various genotypes. The accumulation of miR172 and miR173 in Col, *nrpe1-1*, *dcl3-1*, *fdm1-1*, *fdm2-1* and *fdm1-1 fdm2-1* was detected by northern blotting. The controls U6 RNA blot and the ethidium bromide-stained tRNAs were shown below the miRNA small blots.

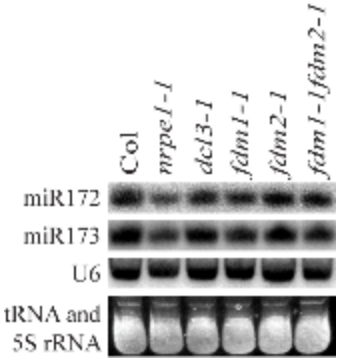


Figure 2-13. GST pull down assay of FDM1, AGO4 and RDR2. (A) The two proteins used for the GST pull down assay. GST and GST-FDM1 were resolved in SDS-PAGE gel to show that similar amount of the two proteins were used for protein pull down assay. The lower bands in the GST-FDM1 lane were truncated GST-FDM1 proteins as they were recognized by the anti-GST antibodies. (B-C) Pull down of AGO4 and RDR2 by GST and GST-FDM1. GST and GST-FDM1 conjugated to glutathione beads were used to perform pull down assay from protein extracts containing myc-AGO4 or HA-RDR2. After pull down, myc-AGO4 and HA-RDR2 were detected by western blotting with anti-Myc and anti-HA antibodies, respectively.

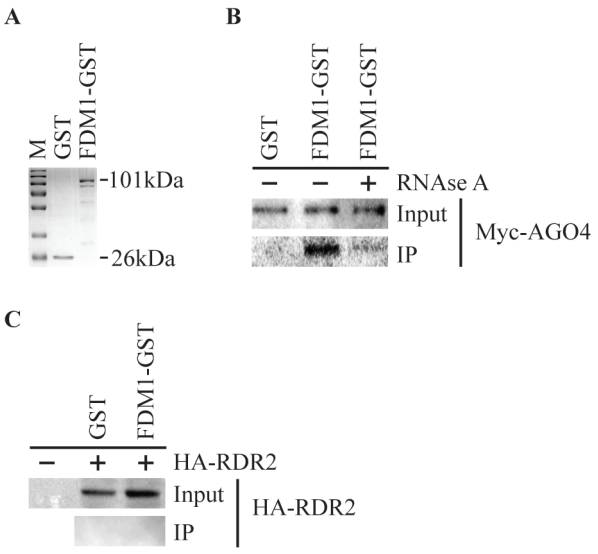


Table 2-1: Correlation coefficiency of genes required for RdDM.

	NRPE1	DRD1	AGO4	DMS3	RDR2	IDN2	SPTL5	DRM2	DCL3	NRPD1
FDM1	0.899	0.888	0.914	0.896	0.835	0.907	0.807	0.807	0.762	0.687
NRPE1		0.948	0.947	0.926	0.864	0.898	0.836	0.764	0.844	0.682
DRD1			0.925	0.901	0.862	0.887	0.839	0.779	0.815	0.668
AGO4				0.935	0.852	0.919	0.8	0.771	0.822	0.674
DMS3					0.833	0.885	0.778	0.706	0.789	0.603
RDR2						0.879	0.874	0.685	0.847	0.802
IDN2*							0.867	0.68	0.839	0.733
SPT5L								0.641	0.819	0.78
DRM2									0.51	0.651
DCL3										0.627

The correlation efficiency of genes used for Figure 1D was highlighted in yellow. *: IDN2/At4g01780; At4g01780 instead of IDN2 was used in the co-expression analysis, because these two genes cross hybridize in the microarray experiments and were counted for one gene by the co-expression analysis software.

References:

1. Feng, S., Jacobsen, S.E. and Reik, W. (2010) Epigenetic reprogramming in plant and animal development. *Science*, **330**, 622-627.
2. Moazed, D. (2009) Small RNAs in transcriptional gene silencing and genome defence. *Nature*, **457**, 413-420.
3. Herr, A.J., Jensen, M.B., Dalmay, T. and Baulcombe, D.C. (2005) RNA polymerase IV directs silencing of endogenous DNA. *Science*, **308**, 118-120.
4. Kanno, T., Huettel, B., Mette, M.F., Aufsatz, W., Jaligot, E., Daxinger, L., Kreil, D.P., Matzke, M. and Matzke, A.J. (2005) Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nat Genet*, **37**, 761-765.
5. Onodera, Y., Haag, J.R., Ream, T., Nunes, P.C., Pontes, O. and Pikaard, C.S. (2005) Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell*, **120**, 613-622.
6. Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E. and Carrington, J.C. (2004) Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol*, **2**, E104.
7. Zheng, B., Wang, Z., Li, S., Yu, B., Liu, J.Y. and Chen, X. (2009) Intergenic transcription by RNA polymerase II coordinates Pol IV and Pol V in siRNA-directed transcriptional gene silencing in Arabidopsis. *Genes Dev*, **23**, 2850-2860.
8. Smith, L.M., Pontes, O., Searle, I., Yelina, N., Yousafzai, F.K., Herr, A.J., Pikaard, C.S. and Baulcombe, D.C. (2007) An SNF2 protein associated with

nuclear RNA silencing and the spread of a silencing signal between cells in Arabidopsis. *Plant Cell*, **19**, 1507-1521.

9. Havecker, E.R., Wallbridge, L.M., Hardcastle, T.J., Bush, M.S., Kelly, K.A., Dunn, R.M., Schwach, F., Doonan, J.H. and Baulcombe, D.C. (2010) The Arabidopsis RNA-directed DNA methylation argonautes functionally diverge based on their expression and interaction with target loci. *Plant Cell*, **22**, 321-334.
10. Zheng, X., Zhu, J., Kapoor, A. and Zhu, J.K. (2007) Role of Arabidopsis AGO6 in siRNA accumulation, DNA methylation and transcriptional gene silencing. *EMBO J*, **26**, 1691-1701.
11. Zilberman, D., Cao, X. and Jacobsen, S.E. (2003) ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science*, **299**, 716-719.
12. Cao, X., Aufsatz, W., Zilberman, D., Mette, M.F., Huang, M.S., Matzke, M. and Jacobsen, S.E. (2003) Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation. *Curr Biol*, **13**, 2212-2217.
13. El-Shami, M., Pontier, D., Lahmy, S., Braun, L., Picart, C., Vega, D., Hakimi, M.A., Jacobsen, S.E., Cooke, R. and Lagrange, T. (2007) Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. *Genes Dev*, **21**, 2539-2544.
14. Wierzbicki, A.T., Haag, J.R. and Pikaard, C.S. (2008) Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell*, **135**, 635-648.

15. Ausin, I., Mockler, T.C., Chory, J. and Jacobsen, S.E. (2009) IDN1 and IDN2 are required for de novo DNA methylation in Arabidopsis thaliana. *Nat Struct Mol Biol*, **16**, 1325-1327.

16. Bies-Etheve, N., Pontier, D., Lahmy, S., Picart, C., Vega, D., Cooke, R. and Lagrange, T. (2009) RNA-directed DNA methylation requires an AGO4-interacting member of the SPT5 elongation factor family. *EMBO Rep*, **10**, 649-654.

17. Gao, Z., Liu, H.L., Daxinger, L., Pontes, O., He, X., Qian, W., Lin, H., Xie, M., Lorkovic, Z.J., Zhang, S. *et al.* (2010) An RNA polymerase II- and AGO4-associated protein acts in RNA-directed DNA methylation. *Nature*, **465**, 106-109.

18. He, X.J., Hsu, Y.F., Zhu, S., Wierzbicki, A.T., Pontes, O., Pikaard, C.S., Liu, H.L., Wang, C.S., Jin, H. and Zhu, J.K. (2009) An effector of RNA-directed DNA methylation in arabidopsis is an ARGONAUTE 4- and RNA-binding protein. *Cell*, **137**, 498-508.

19. Kanno, T., Mette, M.F., Kreil, D.P., Aufsatz, W., Matzke, M. and Matzke, A.J. (2004) Involvement of putative SNF2 chromatin remodeling protein DRD1 in RNA-directed DNA methylation. *Curr Biol*, **14**, 801-805.

20. Kanno, T., Bucher, E., Daxinger, L., Huettel, B., Bohmdorfer, G., Gregor, W., Kreil, D.P., Matzke, M. and Matzke, A.J. (2008) A structural-maintenance-of-chromosomes hinge domain-containing protein is required for RNA-directed DNA methylation. *Nat Genet*, **40**, 670-675.

21. Law, J.A., Ausin, I., Johnson, L.M., Vashisht, A.A., Zhu, J.K., Wohlschlegel, J.A. and Jacobsen, S.E. (2010) A protein complex required for polymerase V

transcripts and RNA- directed DNA methylation in Arabidopsis. *Curr Biol*, **20**, 951-956.

22. Bateman, A. (2002) The SGS3 protein involved in PTGS finds a family. *BMC Bioinformatics*, **3**, 21.

23. Mourrain, P., Beclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J.B., Jouette, D., Lacombe, A.M., Nikic, S., Picault, N. *et al.* (2000) Arabidopsis SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance. *Cell*, **101**, 533-542.

24. Zheng, Z., Xing, Y., He, X.J., Li, W., Hu, Y., Yadav, S.K., Oh, J. and Zhu, J.K. (2010) An SGS3-like protein functions in RNA-directed DNA methylation and transcriptional gene silencing in Arabidopsis. *Plant J*, **62**, 92-99.

25. Peragine, A., Yoshikawa, M., Wu, G., Albrecht, H.L. and Poethig, R.S. (2004) SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in Arabidopsis. *Genes Dev*, **18**, 2368-2379.

26. Fukunaga, R. and Doudna, J.A. (2009) dsRNA with 5' overhangs contributes to endogenous and antiviral RNA silencing pathways in plants. *EMBO J*, **28**, 545-555.

27. Pontier, D., Yahubyan, G., Vega, D., Bulski, A., Saez-Vasquez, J., Hakimi, M.A., Lerbs-Mache, S., Colot, V. and Lagrange, T. (2005) Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in Arabidopsis. *Genes Dev*, **19**, 2030-2040.

28. Cao, X. and Jacobsen, S.E. (2002) Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. *Proc Natl Acad Sci U S A*, **99 Suppl 4**, 16491-16498.

29. Vaillant, I., Schubert, I., Tourmente, S. and Mathieu, O. (2006) MOM1 mediates DNA-methylation-independent silencing of repetitive sequences in Arabidopsis. *EMBO Rep*, **7**, 1273-1278.

30. Park, W., Li, J., Song, R., Messing, J. and Chen, X. (2002) CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in Arabidopsis thaliana. *Curr Biol*, **12**, 1484-1495.

31. Pontes, O., Li, C.F., Nunes, P.C., Haag, J., Ream, T., Vitins, A., Jacobsen, S.E. and Pikaard, C.S. (2006) The Arabidopsis chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. *Cell*, **126**, 79-92.

32. Jiao, X., Trifillis, P. and Kiledjian, M. (2002) Identification of target messenger RNA substrates for the murine deleted in azoospermia-like RNA-binding protein. *Biol Reprod*, **66**, 475-485.

33. Yu, B., Bi, L., Zheng, B., Ji, L., Chevalier, D., Agarwal, M., Ramachandran, V., Li, W., Lagrange, T., Walker, J.C. *et al.* (2008) The FHA domain proteins DAWDLE in Arabidopsis and SNIP1 in humans act in small RNA biogenesis. *Proc Natl Acad Sci U S A*, **105**, 10073-10078.

34. Qin, Y., Ye, H., Tang, N. and Xiong, L. (2009) Systematic identification of X1-homologous genes reveals a family involved in stress responses in rice. *Plant Mol Biol*, **71**, 483-496.

35. Akiyama, K., Chikayama, E., Yuasa, H., Shimada, Y., Tohge, T., Shinozaki, K., Hirai, M.Y., Sakurai, T., Kikuchi, J. and Saito, K. (2008) PRIMe: a Web site that assembles tools for metabolomics and transcriptomics. *In Silico Biol*, **8**, 339-345.

36. Obayashi, T., Kinoshita, K., Nakai, K., Shibaoka, M., Hayashi, S., Saeki, M., Shibata, D., Saito, K. and Ohta, H. (2007) ATTED-II: a database of co-expressed genes and cis elements for identifying co-regulated gene groups in Arabidopsis. *Nucleic Acids Res*, **35**, D863-869.

37. Ihmels, J., Levy, R. and Barkai, N. (2004) Principles of transcriptional control in the metabolic network of Saccharomyces cerevisiae. *Nat Biotechnol*, **22**, 86-92.

38. Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G.V. and Provart, N.J. (2007) An "Electronic Fluorescent Pictograph" browser for exploring and analyzing large-scale biological data sets. *PLoS One*, **2**, e718.

39. Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R. *et al.* (2003) Genome-wide insertional mutagenesis of Arabidopsis thaliana. *Science*, **301**, 653-657.

40. McElver, J., Tzafrir, I., Aux, G., Rogers, R., Ashby, C., Smith, K., Thomas, C., Schetter, A., Zhou, Q., Cushman, M.A. *et al.* (2001) Insertional mutagenesis of genes required for seed development in Arabidopsis thaliana. *Genetics*, **159**, 1751-1763.

41. Hamilton, A., Voinnet, O., Chappell, L. and Baulcombe, D. (2002) Two classes of short interfering RNA in RNA silencing. *EMBO J*, **21**, 4671-4679.

42. Li, C.F., Pontes, O., El-Shami, M., Henderson, I.R., Bernatavichute, Y.V., Chan, S.W., Lagrange, T., Pikaard, C.S. and Jacobsen, S.E. (2006) An ARGONAUTE4-

containing nuclear processing center colocalized with Cajal bodies in Arabidopsis thaliana. *Cell*, **126**, 93-106.

43. Wierzbicki, A.T., Ream, T.S., Haag, J.R. and Pikaard, C.S. (2009) RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nat Genet*, **41**, 630-634.

CHAPTER 3

The DNA- and RNA-binding protein FACTOR of DNA METHYLATION 1 requires XH domain-mediated complex formation for its function in RNA-directed DNA methylation

The Plant Journal. (2012) Vol. 72(3):491-500

Meng Xie, Guodong Ren, Chi Zhang and Bin Yu

Abstract

Studies have identified a subgroup of SGS3-LIKE proteins including FDM1-5 and IDN2 as key components of RNA-directed DNA methylation pathway (RdDM). Although FDM1 and IDN2 bind RNAs with 5' overhangs, their functions in the RdDM pathway remain to be examined. Here we show that FDM1 interacts with itself and IDN2. Gel filtration suggests that FDM1 may exist as a homodimer in a heterotetramer complex *in vivo*. The XH domain of FDM1 mediates the FDM1-FDM1 and FDM1-IDN2 interactions. Deletion of the XH domain disrupts FDM1 complex formation and results in loss-of-function of FDM1. These results demonstrate that XH domain-mediated complex formation of FDM1 is required for its function in RdDM. In addition, FDM1 binds unmethylated but not methylated DNAs through its coiled-coil domain. RNAs with 5' overhangs does not compete with DNA for binding by FDM1, indicating that FDM1 may bind DNA and RNA simultaneously. These results provide novel insight on how FDM1 functions in RdDM.

Introduction

In plants and animals, DNA methylation often associates with transcriptional silencing (TGS) and is thought to play key roles in maintaining genome stability (Feng *et al.,* 2010, Moazed 2009, Zhang and Zhu 2011). In Arabidopsis, a class of ~ 24 nt repeat associated small RNAs (ra-siRNAs) directs *de novo* DNA methylation at their homologous loci through an RNA-directed DNA methylation pathway (RdDM) (Feng *et al.,* 2010, Moazed 2009, Zhang and Zhu 2011). The framework of RdDM has been established through identification and characterization of genes involved in this process (Feng *et al.,* 2010, Moazed 2009, Zhang and Zhu 2011). The RNAse III enzyme DICER-LIKE 3

(DCL3) produces ra-siRNAs from dsRNAs converted by RNA-dependent RNA polymerase 2 from single-stranded RNAs (Xie *et al.,* 2004), which may be produced by plant specific DNA-directed RNA polymerase IV (Pol IV) from RdDM target loci (Herr *et al.,* 2005, Kanno *et al.,* 2005, Onodera *et al.,* 2005, Pontier *et al.,* 2005). ARGONAUTE 4 (AGO4) binds ra-siRNA to form an AGO4-ra-siRNA complex (Havecker *et al.,* 2010, Zheng *et al.,* 2007, Zilberman *et al.,* 2003), which is recruited to chromatin by the interaction of AGO4 and plant specific DNA-directed RNA polymerase V (Pol V) (El-Shami *et al.,* 2007) and/or base-pairing between siRNA and Pol V-dependent transcripts (Wierzbicki *et al.,* 2008, Wierzbicki *et al.,* 2009). The recruitment of AGO4 to some low-copy-number loci also requires DNA-directed RNA polymerase II (Pol II) (Zheng *et al.,* 2009). After loaded into chromatin, AGO4 is thought to recruit the Domains Rearranged Methyltransferase 2 (DRM2) that catalyzes *de novo* cytosine DNA methylation at symmetric CG, CHG (H is adenine, thymine or cytosine) and asymmetric CHH context (Cao and Jacobsen 2002, El-Shami *et al.,* 2007, Wierzbicki *et al.,* 2009). The KOW-CONTAINING TRANSCRIPTION FACTOR 1/SPT5-LIKE protein (KTF1/SPT5L) is required for RdDM and interacts with AGO4 to help the recruitment of DRM2 (Bies-Etheve *et al.,* 2009, He *et al.,* 2009). Recruitment of SPT5L to Pol V-dependent transcript and chromatin is AGO4 independent (Rowley *et al.,* 2011). CLASSY 1 (CLSY1), a chromatin-remodeling protein, and SAWADEE HOMEODOMAIN HOMOLOG 1 (SHH1)/DNA-BINDING TRANSCRIPTION FACTOR 1 (DTF1) are essential for ra-siRNA accumulation and DNA methylation (Law *et al.,* 2011, Liu *et al.,* 2011, Smith *et al.,* 2007). These three proteins are co-purified with Pol IV, indicating that they form a complex to function (Law *et al.,* 2011). DEFECTIVE

IN RNA-DIRECTED DNA METHYLATION 1 (DRD1; a chromatin-remodeling protein), DEFECTIVE IN MERISTEM SILENCING 3 (DMS3; a protein containing a hinge domain of structural maintenance of chromosome proteins), and RNA-DIRECTED DNA METHYLATION 1 (RDM1; a methylated DNA binding protein) are required for the generation of Pol V-dependent transcripts and for RdDM (Gao *et al.,* 2010, Kanno *et al.,* 2008, Kanno *et al.,* 2004, Law *et al.,* 2010). It has been shown that DRD1, DMS3 and RDM1 function as a complex in RdDM (Law *et al.,* 2010). RDM1 also interacts with AGO4 and DRM2 and may help recruit the silencing complex to chromatin (Cao *et al.,* 2010).

Recent studies reveal that six homolog proteins including FACTOR of DNA METHYLATION 1 (FDM1), 2, 3, 4, 5 and INVOLVED IN DE NOVO 2 (IDN2, also called RDM12) act redundantly in the RdDM pathway in *Arabidopsis* (Ausin *et al.,* 2009, Xie *et al.,* 2012, Zheng *et al.,* 2010). These proteins belong to the plant specific SGS3-LIKE protein family, whose founder members are *Arabidopsis* SGS3 and its rice homolog X1(Bateman 2002, Mourrain *et al.,* 2000). SGS3 is an essential component in post-transcriptional gene silencing (Mourrain *et al.,* 2000). It may stabilize RNA intermediates generated during trans-acting siRNA biogenesis by its RNA binding ability (Peragine *et al.,* 2004). SGS3 contains an XS domain and a coiled-coil domain from N- to C-terminus (Bateman 2002). In contrast, FDMs and IDN2 possess two additional domains, an N-terminal zinc finger domain and an XH domain that is named as X-homolog domain with unknown function (Ausin *et al.,* 2009, Xie *et al.,* 2012, Zheng *et al.,* 2010). Like SGS3, IDN2 and FDM1 bind dsRNAs with 5’ overhangs (Ausin *et al.,*

2009, Xie *et al.* 2012, Zheng *et al.* 2010). However, the *in vivo* substrates of FDM1 and IDN2 remain to be identified although they are proposed to stabilize the duplex generated by base pairing between ra-siRNA and Pol V-dependent transcript (Xie *et al.,* 2012, Austin *et al.,* 2009).

In this study, we report that FDM1 acts as a complex in RdDM. FDM1 interacts with both itself and IDN2. Gel filtration analysis suggests that FDM1 exists as a homodimer in a heterotetramer complex that may contain IDN2 *in vivo*. The FDM1 complex formation depends on its XH domain. The mutant FDM1 protein lacking its XH domain fails to form a complex and is unable to complement the DNA methylation defects of *fdm1-1 fdm2-1*, demonstrating that XH-domain mediated complex formation of FDM1 is required for its function in RdDM. FDM1 binds DNA *in vitro* through its coiled-coil domain. RNAs with 5' overhangs do not abolish the DNA binding ability of FDM1, indicating that FDM1 may bind both DNA and RNA simultaneously. Through functional analyses of FDM1 protein domains, this study extends our understanding on the RdDM pathway.

Results:

FDM1 interacts with itself and IDN2

FDM1 and FDM2 share high identity (~93% identity and 96% similarity). However, *fdm1-1* (null mutation) but not *fdm2-1* (null mutation) alone reduces DNA methylation, indicating that FDM1 may have a major role in RdDM. In addition, expression of FDM1

but not FDM2 in *fdm1-1 fdm2-1* is sufficient to recover the defect of DNA methylation to wild-type levels. This provides an advantage to study FDM1 function without the effect of FDM2 *in vivo* by expression of FDM1 mutants. Thus, we focused on FDM1 in this study. In order to gain insight into how FDM1 acts in RdDM, we tested the interaction of FDM1 with known RdDM components including DRM2, DMS3, RDR2, SPT5L, FDM1 and IDN2 using the pGBKT7/pGADT7 two-hybrid system. In this system, a protein of interest is fused with a DNA binding domain in the pGBKT7 plasmid, while the potential interactor is fused with a transcriptional activation domain in the pGADT7 vector. If two proteins interact, the DNA binding domain associates with the transcriptional activation domain after co-transformed into yeast cells. This activates the expression of a report gene that produces Adenine (Ade) and thus enables the growth of the yeast strain under the absence of Ade. Co-transformation of pGADT7- FDM1/pGBKT7-FDM1 and pGADT7-FDM1/pGBKT7-IDN2 pairs enabled the growth of yeast cell under the absence of Ade (Figure 3-1a). In contrast, yeast cells failed to grow on –Ade medium after co-transformation of pGADT7/pGBKT7-FDM1, pGADT7/pGBKT7-IDN2 and pGBKT7/pGADT7-FDM1 pairs (Figure 3-1a). These results indicated that FDM1 might interact with itself and IDN2. This assay did not detect the interaction of FDM1 with DRM2, DMS3, RDR2 and SPTL5 (Figure 3-1b). The FDM1-RDR2 interaction result from this assay is consistent with the *in vitro* pull down results (Xie *et al.,* 2012).

The XH domain of FDM1 is necessary for FDM1-FDM1 and FDM1-IDN2 interactions

To identify protein domains of FDM1 responsible for the interaction, we generated a series of truncation mutants of FDM1 in pGADT7 (Figure 3-2a): lacking the XH domain (FDM1-T1), XH domain alone (FDM1-T2), lacking the XH and coiled-coil domains (FDM1-T3) and lacking the zinc-finger and XS domain (FDM1-T4). We tested the interaction of these truncated FDM1 mutants with full length FDM1 and IDN2 using the yeast two-hybrid assay described above. FDM1-T2 and FDM1-T4 were able to interact with FDM1 and IDN2, respectively, because co-transformation of these pairs enabled yeast cell to grow under the absence of Ade (Figure 3-2b and 3-2c). In contrast, both FDM1-T1 and FDM-T3 did not interact with FDM1 and IDN2, respectively. These results indicated that the XH domain of FDM1 is necessary for FDM1-FDM1 and FDM1-IDN2 interactions. However, the yeast cells containing FDM1/FDM1-T2 (XH domain alone) and IDN2/FDM1-T2 grew slower than those containing FDM1/FDM1 and IDN2/FDM1, respectively (Figure 3-1 and 3-2). This result indicated that the full-strength FDM1-FDM1 and FDM1-IDN2 interactions might require additional protein domains.

To validate the function of XH domain in protein-protein interactions, we replaced Tryptophan 605 (W605) and Glutamic acids 617 (E617) with Alanine (A) in the XH domain of FDM1, respectively (FDM1-T5; Figure 3-2a). These two amino acids are conserved in XH domains and hence play important roles in mediating protein-protein interactions. As shown in Figure 2d, FDM1-T5 did not interact with FDM1 and IDN2. This result confirmed that the XH-domain of FDM1 is necessary for FDM1-FDM1 and FDM1-IDN2 interactions.

FDM1 pulls down FDM1 and IDN2 *in vitro*

To further confirm FDM1-IDN2 interaction, we conducted an *in vitro* pull down assay. We expressed the recombinant IDN2 protein fused with a maltose-binding protein epitope at its N-terminus (MBP-IDN2), FDM1 fused with an N-terminal GST tag (GST-FDM1) and controls MBP, GST and GST-FDM1ΔXH (FDM1 lacking XH domain) in *E.coli*, respectively. After expression, protein extracts containing MBP-IDN2 were mixed with extracts containing GST-FDM1 and reciprocal pull down was then performed with amylose resin or glutathione beads. To avoid the DNA or RNA-mediated protein interactions, we treated the samples with Micrococcal nuclease that digests both DNA and RNA. The enrichment of MBP-IDN2 in GST-FDM1 complex and GST-FDM1 in MBP-IDN2 complex was detected using antibodies against MBP or GST, respectively (Figure 3-3a and 3-3b). In contrast, GST and GST-FDM1ΔXH failed to pull down MBP-IDN2 and MBP did not pull down GST-FDM1 (Figure 3-3a and 3-3b). To validate FDM1-FDM1 interaction, we mixed protein extracts containing YFP-FDM1 or YFP -FDM1ΔXH with extracts containing GST-FDM1 or GST, respectively, and performed reciprocal pull down assay. GST-FDM1 and YFP-FDM1 could reciprocally pull down each other (Figure 3-3c and 3-3d), while YFP -FDM1ΔXH and GST did not interact with GST-FDM1 and YFP-FDM1, respectively (Figure 3-3c and 3-3d).

FDM1 forms a tetramer *in vitro*

The yeast two-hybrid and pull down analyses suggest that FDM1 interacts with itself through its XH-domain. Thus, we examined whether FDM1 forms a dimer or an

oligomer complex. We first expressed recombinant FDM1 fused with a C-terminal 6XHis tag (FDM1-His) and a truncated FDM1-His lacking XH (FDM1ΔXH-His) in *E.coli* and purified the resulting proteins. The FDM1-His or FDM1ΔXH-His was then analyzed by size-exclusion HPLC. The timed elution fractions were then separated in SDS-PAGE and probed with antibodies recognizing His tag. The column was calibrated with Bio-Rad protein standards. We obtained information on the relative size of FDM1 complex by comparing fractions of FDM1 with peak elution times of standard proteins. FDM1-His had a peak elution of 114 to 118 minutes (Figure 3-4a), suggesting that FDM1-His may exist as a ~300 KDa tetramer complex. In contrast, FDM1ΔXH-His eluted from 144 to 148 minutes corresponding to the size of FDM1ΔXH monomer (~60 KDa; Figure 3-4a). These analyses revealed that FDM1 forms a tetramer complex that requires the XH domain for its formation.

Because FDM1 also interacts with IDN2, we next tested whether incubation of IDN2 and FDM1 generates a larger complex or tetramer with expectation to get insight into the nature of FDM1-IDN2 complex. We purified MBP-IDN2 and removed the MBP tag. However, incubation of IDN2 with FDM1-His still produced a tetramer (Figure 3-4b). This result indicated that FDM1 and IDN2 might form a tetramer *in vitro*. However, the copy numbers of FDM1 and IDN2 in the complex remain to be determined.

FDM1 protein exists as a dimer in a tetramer complex *in vivo*

To get into the FDM1 complex *in vivo*, we analyzed *Arabidopsis* protein extracts containing YFP-FDM1 by size-exclusion HPLC. The YFP-FDM1 complemented the

DNA methylation defects in *fdm1-1*. The anti-YFP antibody detected the presence of YFP-FDM1 in a ~350 KDa complex as calculated by standard curve, which was produced using protein standard elution time (Figure 3-4c). The calculated molecular mass for YFP-FDM1 is ~ 100 KDa and for untagged FDM proteins and IDN2 are ~ 75 KDa. Thus, the ~ 350 KDa equals to the molecular mass of two copies of YFP-FDM1 and two copies other untagged FDM proteins or IDN2. This result suggested that FDM1 might exist as a homodimer in a heterotetramer complex.

The XH domain is required for the function of FDM1 in RdDM.

Next, we examined if the XH domain was required for FDM1 function in RdDM. We generated transgenic *fdm1-1 fdm2-1* containing either *35S::YFP-FDM1* or *35S:YFP-FDM1ΔXH* lacking the XH domain. In previous studies, we showed that expression of *FDM1* under the direction of its native promoter is sufficient to complement the DNA methylation defects of *fdm1-1 fdm2-1* (Xie *et al.,* 2012). Thus, using *fdm1-1 fdm2-1* enabled us to test the function of XH domain of FDM1 without effects of FDM2, which has a 96% similarity with FDM1. The transcript levels of transgenes and their products were similar in all four transgenic lines (Figure 3-5a and 3-5b). We examined the methylation levels of *ATSN1* and *ING5* in two transgenic *fdm1-1 fdm2*-1 lines harboring *35S::YFP-FDM1* and two transgenic *fdm1-1 fdm2-1* harboring *35S:YFP-FDM1ΔXH* using methylation sensitive HaeIII restriction enzyme digestion followed by PCR. Less DNA methylation at *ATSN1* and *IGN5* results in less PCR product after HaeIII digestion because it cuts unmethylated but not methylated DNA. As shown in Figure 3-5c, the *35S::YFP-FDM1* transgene recovered DNA methylation content of *ATSN1* and *IGN5* in

fdm1-1 fdm2-1 to WT levels. In contrast, the DNA methylation levels of *fdm1-1 fdm2-1* harboring *35S:YFP-FDM1ΔXH* were comparable with those in *fdm1-1 fdm2-1*. Consistent with this, the silencing of *AtSN1* transcription was not restored in *fdm1-1 fdm2-1* harboring *35S:YFP-FDM1ΔXH* (Figure 3-5d). These results demonstrated that the XH domain is essential for the function of FDM1 in RdDM.

FDM1 binds unmethylated but not methylated DNA

Protein sequence analyses showed that the coiled-coil domain of FDM1 has ~ 50% similarities to a portion of SMC (structural maintenance of chromosomes) protein from *Methanocaldococcus sp. FS406-22* (Figure 3-6a). As SMC proteins bind DNAs, this finding prompts us to test whether FDM1 binds DNA using a GST-pull down assay. This method reduces the background signal because it eliminates the unbound probes. Others and we have used this method to study protein-nucleic acid interaction (Yu *et al.,* 2008, Jiao *et al.,* 2002). We incubated purified GST-FDM1 with a 50 bp P^{32} labeled DNA fragment and a 50 nt P^{32} labeled single-stranded DNA (ssDNA) that corresponds to a fragment of *AtSN1* DNA (Figure 3-6b). After washing, the DNAs were extracted from beads and separated on a native PAGE gel. The GST-FDM1 but not GST alone retained the 50 bp DNA fragment (Figure 3-6b). However, FDM1 was unable to bind the ssDNA (Figure 3-6b). Addition of unlabelled DNA with same sequences eliminated the radioactive signals. These results indicated that FDM1 binds DNA (Figure 3-6b). FDM1 also bound a DNA fragment containing a poly(A) strand and a poly(T) strand (Figure 3-6d). This result suggested that DNA binding of FDM1 is not sequence specific. However, FDM1 did not bind methylated DNA (Figure 3-6d). To identify protein domains required

for DNA binding ability of FDM1, we expressed and purified a series of truncated FDM1 proteins fused with a N-terminal GST tag (Figure 3-6e and 3-6f). The truncated FDM1 lacking a portion of coiled-coil domain but not other domains failed to bind DNA (Figure 3-6f). In addition, the coiled-coil domain itself was able to bind DNA (Figure 3-6f). Based on these results, we proposed that the coiled-coil domain is necessary and sufficient for DNA binding of FDM1.

We have shown that FDM1 binds the RNA with 5' overhangs, which depends on the XS domain. This raised a question of whether FDM1 can bind DNA and RNA simultaneously. To address this question, we examined if addition of unlabelled RNAs with 5' overhangs affects the DNA binding ability of FDM1. If FDM1 binds DNA and RNA at the same time, addition of RNAs shall not eliminate DNA binding of FDM1. As shown in Figure 6C, addition of RNAs with 5' overhang did not affect DNA retention of FDM1.

Discussion

Studies on FDM1 and IDN2 have suggested that they may act in the downstream RdDM, presumably by stabilizing the duplex of ra-siRNA-Pol V-dependent transcripts (Ausin *et al.,* 2009, Xie *et al.,* 2012, Zheng *et al.,* 2010). In this study, we demonstrate that FDM1 exists in a complex for its proper function in RdDM and is an RNA and DNA binding protein.

Yeast two-hybrid and *in vitro* protein pull down experiments show that FDM1 interacts with IDN2. Given its high similarity with FDM1, FDM2 most likely interacts with IDN2 as well. While this manuscript was in preparation, two other groups found that IDN2 complex contain IDN2 PARALOG 1 (IDP1)/IDN2-LIKE1 (IDNL1) and IDP2/IDNL2 (Austin *et al.,* 2012, Zhang *et al.,* 2012). IDP1/IDNL1 and IDP2/IDNL2 are synonymous to FDM1 and FDM2, respectively. These results demonstrate that FDM1/IDP1/IDNL1 and FDM2/IDP2/IDNL2 form a complex with IDN2. We but not Zhang et al., (2012) detect the FDM1/IDP1-FDM1/IDP1 interaction in a yeast two-hybrid assay. This discrepancy may be due to the fact that different yeast strains were used (PJ694A vs AH109). Protein pull-down (Figure 3-3) and gel filtration (Figure 3-4) experiments further confirmed the FDM1-FDM1 interaction. FDM1 forms a homotetramer *in vitro* but may exist as a homodimer in a tetramer complex *in vivo* (Figure 3-4). Multidimensional protein identification technology (MudPIT) analysis shows that IDN2 may be the only partner of IDNL1/FDM1 (Austin *et al.,* 2012). Crystal structure and yeast two-hybrid analyses reveal that IDN2 lacking XH domain forms a homodimer *in vitro* (Austin *et al.,* 2012, Zhang *et al.,* 2012). These results suggest that FDM1 and IDN2 form a heterotetramer containing an FDM1 dimer and an IDN2 dimer. FDM2 is in the IDN2 complex and is highly similar to FDM1, indicating the presence of an IDN2-IDN2-FDM2-FDM2 tetramer. The presence of these two complexes is consistent with the functional redundancy of FDM1 and FDM2 (Xie *et al.,* 2012). However, it is possible that IDN2-IDN2-FDM1-FDM2 exists at low amount so that MudPIT cannot detect IDNL2/FDM2 in the IDNL1/FDM1 complex.

Whole genome bisulfite sequence analysis reveals that DNA methylation patterns are similar in *idn2-1, idnl1-1 idnl2-1* (*fdm1 fdm2*) and *idn2-1 idnl1-1 idnl2-1*, indicating that IDN2, FDM1/IDNL1 and FDM2/IDNL2 mostly likely function together (Austin *et al.,* 2012). IDN2, FDM1/IDNL1 and FDM2/IDNL2 affect most DRM2 targets and few non-DRM2 targets, indicating that they mainly act in RdDM pathway (Austin *et al.,* 2012). The DNA methylation defect in *idn2-1*, *idnl1-1 idnl2-1* and *idn2-1 idnl1-1 idnl2-1* is weaker than that in *drm2*.

This may be due to the redundant functions of homologs of IDN2, FDM1 and FDM2. Indeed, three FDM1 homologs, FDM3, FDM4 and FDM5 act redundantly with FDM1 (Xie *et al.,* 2012). Among of them, FDM3 and FDM4 are in the subfamily of IDN2, whereas FDM5 is grouped with FDM1 and FDM2. The IDN2 complex does not contain FDM3, FDM4 and FDM5 (Ausin *et al.,* 2012, Zhang *et al.,* 2012), raising the possibility that other FDM complexes may exist. Perhaps, the IDN2/ FDM1 (FDM2) complex plays a major role in RdDM, while others have minor functions, because loss-of-function either of IDN2 or FDM1 alone causes DNA methylation defection while lacking other FDM proteins alone does not (Xie *et al.,* 2012). Alternatively, they may have different roles with IDN2/FDM1 complex.

The function of XH domain was previously unknown. We found that FDM1 protein lacking its XH domain or harboring mutations in its XH domain failed to interact with itself or with IDN2 (Figure 3-2, 3 and 4). The XH domain of FDM1 by itself interacts with FDM1 and IDN2, demonstrating that the XH domain of FDM1 functions in mediating protein-protein interaction. In addition, IDN2 without functional XH domain

fails to interact with IDP1/FDM1 (Zhang *et al.,* 2012). Both FDM1 and IDN2 lacking the XH domain failed to rescue DNA methylation defects in their mutants (Figure 3-5; Zhang *et al.,* 2012), respectively, demonstrating that the XH-domain mediated complex formation is essential for their function in RdDM.

FDM1/IDNL1 or FDM2/IDNL2 cannot replace IDN2 in their complexes because a strong *idn2-1* allele has a similar DNA methylation defects as *idnl-1 idnl-2* and *idn2-1 idnl1-1 idnl2-1* (Austin et al., 2012). However, it is reasonable that a weak *idn2-3* mutation will further reduce DNA methylation in *fdm1-1* and *fdm2-1*, respectively (Xie *et al.,* 2012), because the function of IDN2/FDM1/FDM2 complex will be further impaired in the double mutants. What causes the difference between FDM1 and IDN2? For both FDM1 and IDN2, the XH domain mediates protein-protein interaction and the XS domain binds dsRNAs with 5' overhangs, indicating that they may not causes the differences between FDM1 and IDN2. We find that the coiled-coil domain of FDM1 binds DNA and is not required for FDM1-FDM1 and FDM1-IDN2 interactions. In contrast, coiled-coil domain of IDN2 is shown to mediate IDN2-IDN2 interaction (Zhang *et al.,* 2012, Austin *et al.,* 2012). Thus, the coiled-coil domain of FDM1 is biochemically different from that of IDN2 and may be the factor to distinguish FDM1 from IDN2. Given the high similarity between FDM1 and FDM2, this most likely is the cause for the difference between FDM2 and IDN2 as well.

FDM1 and IDN2 bind RNAs with 5' overhang through its XS domain (Ausin *et al.,* 2009, Xie *et al.,* 2012; Zhang *et al.,* 2012). In addition, FDM1 binds DNA in a non-

sequence specific manner through the coiled-coil domain and DNA binding of FDM1 cannot be competed by the RNA, indicating that FDM1 may bind DNA and RNA simultaneously. These results have advanced the model for IDN2/FDM1 function (Austin *et al.,* 2009, Zheng *et al.,* 2010, Xie *et al.,* 2012, Austin *et al.,* 2012). The XS domain of FDM1 (FDM2) and IDN2 may bind the duplex of AGO4-bound ra-siRNA and Pol V-dependent transcript (Austin *et al.,* 2009, Zheng *et al.,* 2010, Xie *et al.,* 2012, Austin et al., 2012). This binding will recruit FDM1 (FDM2)-IDN2 complex to RdDM loci. Subsequently, the coiled-coil domain of FDM1 (FDM1) binds the DNA. Binding of FDM1-IDN2 complex to the RNA duplex and RdDM target loci may have two roles that are not mutually exclusive. One is to prevent the potential cleavage of Pol V-dependent transcript by the AGO4-rasiRNA complex, which may disrupt the AGO4-chromatin interaction. However, the levels of Pol V transcripts are not affected by *fdm1 fdm2* and *idn2* mutations, (Xie *et al.,* 2012; Austin *et al.,* 2012), arguing against this possibility. The other is that the FDM1 complex may provide a marker for DRM2 to recognize. However, FDM1 does not bind methylated DNA, indicating that FDM1 complex may be required for the initiation but not reinforcement of DNA methylation. The yeast two-hybrid assay does not identify the FDM1-DRM2 interaction, suggesting that other factors may be involved. Clearly this model needs to be examined using FDM1 mutant deficient in DNA and/or RNA binding.

Materials and methods

Plant materials and growth conditions

Plants were grown at 22°C under long day condition (16 hour light/8 hour night). *fdm1-1* (SALK_075813), *fdm2-1* (SAIL_291_F01) and *fdm1-1 fdm2-1* are in Columbia genetic background (Xie *et al.,* 2012).

Plasmid Construction

YFP cDNA was cloned into binary vector pMDC32 to generate pMDC32-YFP(Curtis and Grossniklaus 2003). Then the *FDM1* and *FDM1ΔXH* (lacking XH domain) cDNAs were PCR amplified and cloned into pMDC32-YFP to generate *p35S::FDM1-YFP* and *p35S::FDM1ΔXH-YFP* constructs, respectively. The *FDM1*, truncated *FDM1*, *IDN2*, *RDR2*, *DRM2*, *SPTL5* and *DMS3* cDNAs were PCR amplified and cloned into pGADT7 and/or pGBKT7 vector to constructions used for yeast two-hybrid assay. The full-length FDM1 and truncated FDM1 cDNAs were PCR amplified cloned into pGEX-2TK or pET28 (a) vectors to generate GST or 6XHIS fusion constructions. The IDN2 cDNA was PCR amplified and cloned into pMAL-c5X vector to generate *MBP-IDN2* fusion construct.

Plant Transformation

p35S::FDM1-YFP and *p35S::FDM1ΔXH-YFP* were transformed into *fdm1-1* or *fdm1-1 fdm2-1*, respectively. The T1 transgenic plants were selected with hygromycin resistance.

Yeast two-hybrid assay

Various plasmid pairs were co-transformed into yeast strain AH109. SD –Leu –Trp medium was used to select yeast containing the plasmid pairs. The resulting clones were

diluted in 50 μl water and 5 μl was used for spot assay on SD –Leu –Trp –Ade plates. The interactions of FDM1-FDM1 and FDM1-IDN2 activate the expression of Ade, which enables the growth of AH109 cells in Ade minus plates. 1

Protein expression

GST, MBP or HIS tagged proteins were expressed in *E. coli.* BL21 and extracted as described (Xie *et al.* 2012). *YFP-FDM1* and *YFP-FDM1ΔXH* were transiently expressed in tobacco *N. benthamiana* and extracted according to Yu *et al.,* (2008)

Protein Pull down Assay

Protein extracts containing GST, GST-FDM1 or GST-FDM1ΔXH were mixed with equal volume of protein extracts containing MBP-IDN2, YFP and YFP-FDM1, respectively. The mixed lysate was incubated with anti-GFP (and GFP variants) antibodies coupled to protein A agarose beads (Clontech), amylose resin (NEB) or glutathione Sepharose 4B beads for 4 hours.

The precipitates were washed with extraction buffer for 5 times and separated on SDS-PAGE gel and blotted with antibodies recognizing MBP, GST or YFP tag.

Gel filtration

FDM1-6XHIS and FDM1ΔXH-6XHIS were purified using Ni-resin according to manufacture's instruction. After elusion from Ni-resin, 100 μl protein solution was passed through a 0.22 μm filter and loaded onto column. The gel filtration was carried out on an HPLC system and the HiPrep 16/60 Sephacryl S-300 HR column (GE Health)

at a rate of 0.5 ml/min and 0.5 ml solution of fractions were collected every minute. For gel filtration of *Arabidopsis* protein extracts, collected fractions were precipitated with acetone at −20°C overnight and resuspended in SDS loading buffer. Fractions were solved in 8% SDS-PAGE gel and analyzed by western blotting using antibodies recognizing HIS or YFP. Protein standards (Bio-rad) used to calibrate the column contain five size standards and the elution time for each peak is: 670kDa at 94 min, 158kDa at129 min, 44kDa at 150 min, 17kDa at 173 min, 1.35kDa at 233 min.

DNA methylation and RT-PCR analysis

The DNA methylation assay was performed as described (Xie *et al.,* 2012). Genomic DNAs extracted from flowers were digested with HaeIII. 5% of digested DNA was used for PCR amplification of *AtSN1* and *IGN5*. Simultaneously, undigested genomic DNA was amplified as the quantity control. After DNase I treatment, 5 μg of total RNAs from inflorescences were used to synthesize cDNA with SuperScript III (Invitrogen) using oligo-dT. The diluted cDNA was used to amplify *AtSN1* by PCR. The amplification of *UBQ5* was used as a loading control.

DNA binding assay

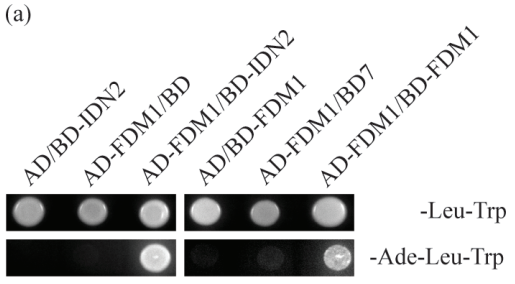
GST-FDM1 and GST-tagged FDM1 mutants were purified according to (Xie et al., 2012). 5’ overhanging dsRNA probe was generated as described (Xie *et al.,* 2012). A synthesized 50-nt single strand DNA fragment corresponding to a portion of AtSN1 DNA was labeled in its 5’ end using T4 polynucleotide kinase (NEB) in the presence of [α^{32}P]ATP. Annealing this ssDNA with its complementary strand produced double-

stranded DNA. The DNA and RNA binding assays were performed as previously described (Jiao *et al.,* 2002). Methylated DNA and its unmethylated control were synthesized at IDT as described (Ito et al., 2003).

Figures

Figure 3-1. Determining the interaction of FDM1 with other components in RdDM

(a) Interactions of FDM1 with FDM1 and IDN2. The growth of yeast cell (AH109) on adenine-deficient medium (-Ade–Leu–Trp) shows the interaction of FDM1 with FDM1 and IDN2. pGADT7 (AD) and pGBKT7 (BD) plasmids contain the activation and DNA binding domains of GAL4, respectively. Paired AD and BD fusion constructs were co-transformed into yeast AH109 cells. The transformants were selected with synthetic dropout medium (–Leu–Trp) and spotted on adenine-deficient medium (-Ade–Leu–Trp). (b) Summary of Yeast two-hybrid analyses. "+" Indicates interactions; "-" indicates non-interactions. FDM1 did not interact with DRM2, DMS3, SPT5L and RDR2.



(b)

Yeast-two hybrid screening

	FDM1	IDN2	DRM2	DMS3	SPT5L	RDR2
FDM1	+	+	−	−	−	−

Figure 3-2. The XH-domain mediates FDM1-FDM1 and FDM1-IDN2 interactions. (a) Schematic structure of the full length and truncated FDM1 proteins used for yeast-two hybrid assay. FDM1-T1: truncated FDM1 protein lacking XH domain; FDM1-T2: XH domain alone; FDM1-T3: truncated FDM1 protein containing ZF and XS domain; FDM1-T4: truncated FDM1 protein containing Coiled-coil domain and XH domain. FDM1-T5: Tryptophan 605 (W605) and Glutamic acids 617 (E617) were replaced with Alanine (A), respectively. (b) Interaction analyses of truncated FDM1 proteins with IDN2 in yeast. (c) Interaction analyses of truncated FDM1 with FDM1 in yeast AH109 cells. (d) Interactions of FDM1 containing point mutations with FDM1 and IDN2. Mutated FDM1 was cloned into pGADT7 (AD). IDN2 and FDM1 were in pGBKT7 (BD), respectively. The paired AD and BD fusion constructs were co-transformed yeast. The positive clones selected in –Leu–Trp were spotted on -Ade–Leu–Trp medium.

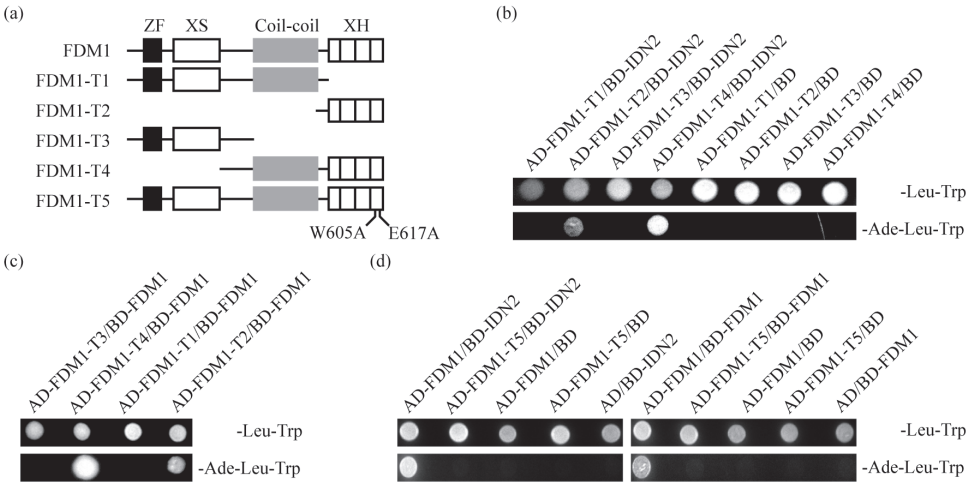


Figure 3-3. *In vitro* FDM1-FDM1 and FDM1-IDN2 interactions

(a) GST-FDM1 but not GST/GST-FDM1ΔXH pulls down MBP-IDN2 protein. (b) MBP-IDN2 pulls down GST-FDM1 but not GST/GST-FDM1ΔXH. GST, GST-FDM1 or GST-FDM1ΔXH extracts were separately mixed with MBP or MBP-IDN2 extracts, respectively, to generate GST/MBP, GST/MBP-IDN2, GST-FDM1/MBP, GST-FDM1/MBP-IDN2, GST-FDM1ΔXH/MBP or GST-FDM1ΔXH/MBP-IDN2 mixtures. Protein mixtures were incubated with glutathione sepharose 4B beads or amylose resin to capture GST fusion proteins or MBP fusion proteins, respectively. MBP fusion proteins and GST fusion proteins were detected by Western blot using MBP antibody and GST-antibody respectively. Bait: proteins were captured by glutathione beads (a) or amylose resin (b). Prey: proteins associated with the bait. (c) GST-FDM1 pulls down YFP-FDM1 but not YFP/YFP-FMD1ΔXH. (d) YFP-FDM1 but not YFP-FMD1ΔXH pulls down GST-FDM1. YFP, YFP-FDM1 or YFP-FDM1ΔXH extracts were separately mixed with GST-FDM1 or GST, respectively, to generate YPP/GST-FDM1, YFP/GST, YFP-FDM1/GST-FDM1, YFP-FDM1/GST, YFP-FDM1ΔXH/GST-FDM1 and YFP-FDM1ΔXH/GST mixtures. Protein mixtures were incubated with glutathione beads or Anti-GFP antibody conjugated to agarose-A beads to capture GST fusion proteins or YFP-fusion proteins, respectively. YFP fusion proteins and GST fusion proteins were detected by Western blot. Bait: proteins were captured by glutathione beads (c) or GFP-antibody (d). Prey: proteins associated with the bait.

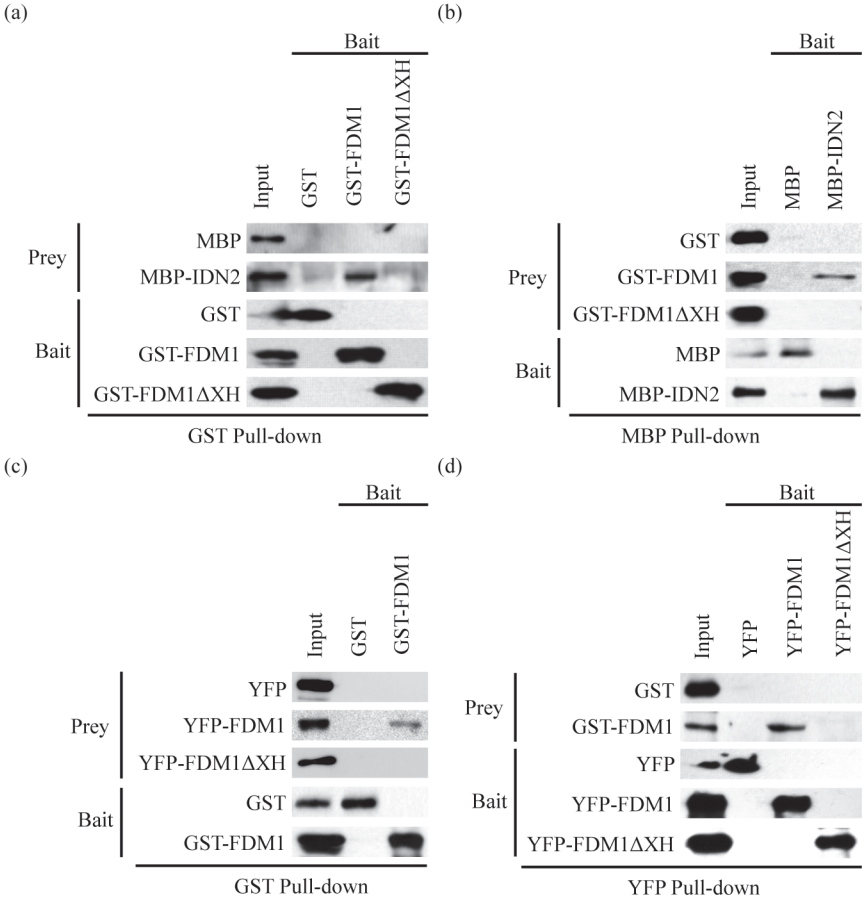

(a)
Bait
Input
GST
GST-FDM1
GST-FDM1ΔXH
Prey
MBP
MBP-IDN2
Bait
GST
GST-FDM1
GST-FDM1ΔXH
GST Pull-down
(b)
Bait
Input
MBP
MBP-IDN2
Prey
GST
GST-FDM1
GST-FDM1ΔXH
Bait
MBP
MBP-IDN2
MBP Pull-down
(c)
Bait
Input
GST
GST-FDM1
Prey
YFP
YFP-FDM1
YFP-FDM1ΔXH
Bait
GST
GST-FDM1
GST Pull-down
(d)
Bait
Input
YFP
YFP-FDM1
YFP-FDM1ΔXH
Prey
GST
GST-FDM1
Bait
YFP
YFP-FDM1
YFP-FDM1ΔXH
YFP Pull-down

Figure 3-4. Gel filtration analysis of FDM1 complex

(a) XH-domain dependent tetramer formation of FDM1 *in vitro*. (b) FDM1 exists in a tetramer complex *in vivo* (c) The effect of IDN2 on FDM1 complex formation. Purified FDM1-6HIS, FDM1-6HIS/IDN2, FDM1ΔXH-6HIS or *Arabidopsis* extracts containing YFP-FDM1 were separated by HPLC. Eluted fractions were separated by SDS-PAGE gel and detected by western blot using Anti-HIS antibody or Anti-YFP antibodies. Elution times of protein standards are shown on the top.

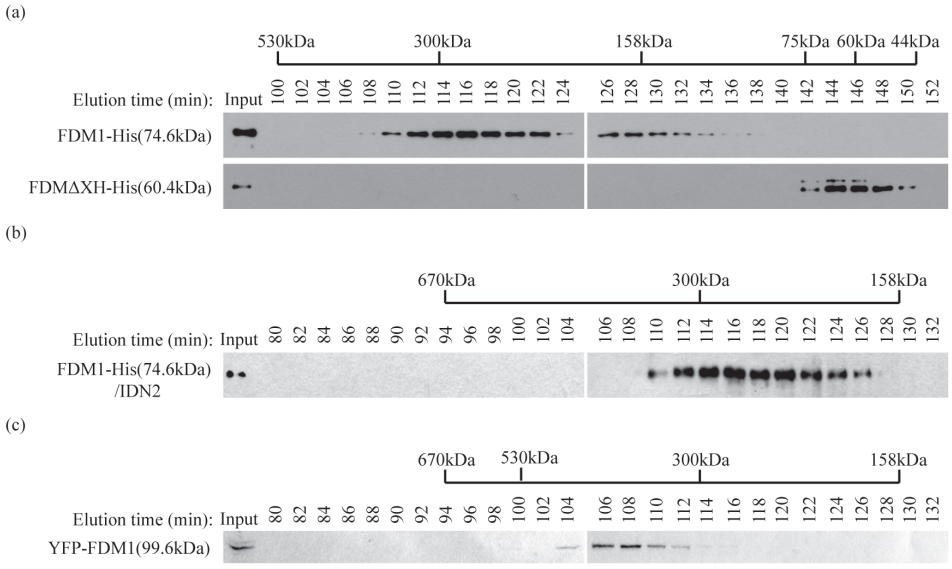


Figure 3-5. The XH domain is required for the function of FDM1 in RdDM. (a) and (b) Deletion of XH domain has not effect on the expression of *FDM1*. The transcript levels of YFP-FDM1 and FDM1ΔXH were determined by RT-PCR. Amplification of *UBIQUITIN5* (At3g26650; *UBQ5*) with or without reverse transcription (-RT) is shown as a control. The protein levels of YFP-FDM1 and FDM1ΔXH were determined by western blot. Heat shock protein 70 (HSP70) was blotted as a loading control. (c) Expression of *YFP-FDM1ΔXH* does not rescue the DNA methylation defects at AtSN1 and IGN5 loci in *fdm1-1 fdm2-1*. HaeIII digested genomic DNAs from various genotypes were used for PCR amplification of *AtSN1* and *IGN5*, whereas undigested genomic DNAs were used as loading controls. (d) Expression of *YFP-FDM1ΔXH* does not silence the expression of AtSN1 in *fdm1-1 fdm2-1*. *AtSN1* Transcripts were detected by RT-PCR. Amplification of *UBQ5* with or without reverse transcription (-RT) is shown as a control.

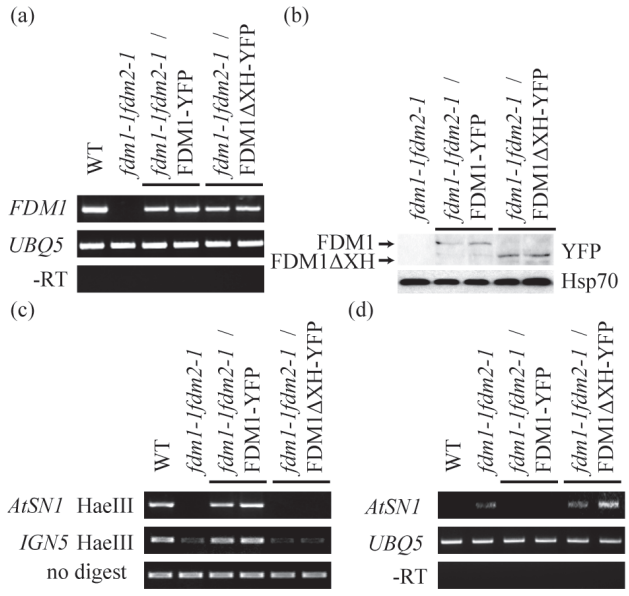
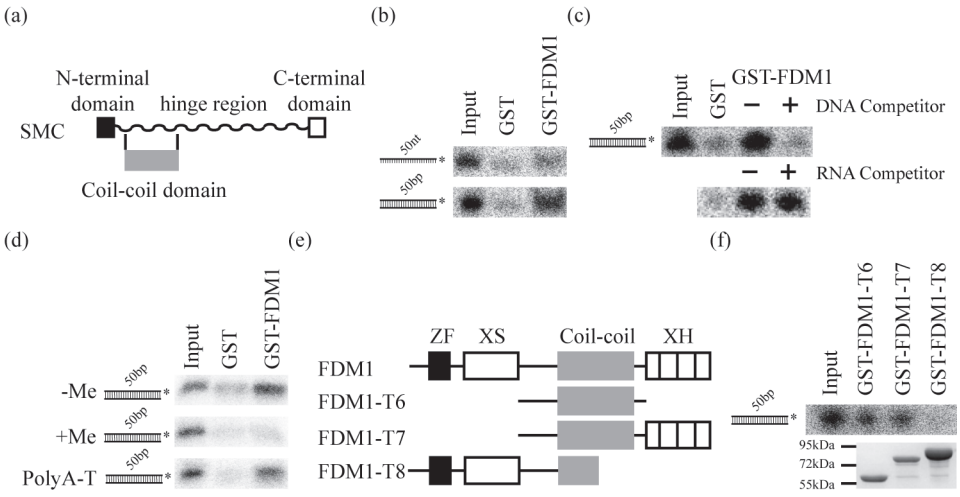


Figure 3-6. FDM1 binds DNA through its coil-coil domain.

(a) Diagrams show that coil-coil domain shares similarities with a portion of SMC (structural maintenance of chromosomes) protein. (b) and (c) FDM1 binds double-stranded DNA but not single-stranded DNA. The structure of various probes is shown on the right. (d) DNA binding specificity of FDM1. –me: unmethylated DNA control; +me: cytosine methylated DNA. Poly A-T: A DNA fragment contains a Poly(A) strand and a Poly(T) strand. (e) Diagrams of truncated GST-FDM1 used for DNA binding assay. FDM1-T6: The coiled-coil domain of FDM1 alone; FDM1-T7: truncated FDM1 containing only the coil-coil and XH domains; FDM1-T8: truncated FDM1 lacking the XH and a portion of coil-coil domains. (f) The coiled-coil domain is necessary and sufficient for DNA binding of FDM1. Purified proteins used in the binding assay were resolved in SDS-PAGE gel and stained with Coomassie Blue and are shown below the DNA binding gel. The protein molecular weights are indicated on the left *: Radioactive labeled DNA strand. Approximately 50 μg protein was used for the binding assay. 150X unlabeled DNAs of the same sequence or RNAs with 5' overhang were used for the competition assay.



References

1. Ausin, I., Mockler, T. C., Chory, J., & Jacobsen, S. E. (2009). IDN1 and IDN2 are required for de novo DNA methylation in *Arabidopsis* thaliana. Nat Struct Mol Biol, 16(12), 1325-1327.

2. Ausin, I., Greenberg, M. V., Simanshu, D. K., Hale, C. J., Vashisht, A. A., Simon, S. A., Jacobsen, S. E. (2012). INVOLVED IN DE NOVO 2-containing complex involved in RNA-directed DNA methylation in Arabidopsis. Proc Natl Acad Sci U S A. doi: 1206638109

3. Bateman, A. (2002). The SGS3 protein involved in PTGS finds a family. BMC Bioinformatics, 3, 21.

4. Bies-Etheve, N., Pontier, D., Lahmy, S., Picart, C., Vega, D., Cooke, R., & Lagrange, T. (2009). RNA-directed DNA methylation requires an AGO4-interacting member of the SPT5 elongation factor family. EMBO Rep, 10(6), 649-654.

5. Cao, X., & Jacobsen, S. E. (2002). Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. Proc Natl Acad Sci U S A, 99 Suppl 4, 16491-16498.

6. Cao, X., Aufsatz, W., Zilberman, D., Mette, M. F., Huang, M. S., Matzke, M., & Jacobsen, S. E. (2003). Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation. Curr Biol, 13(24), 2212-2217.

7. Curtis, M. D., & Grossniklaus, U. (2003). A gateway cloning vector set for high-throughput functional analysis of genes in planta. Plant Physiol, 133(2), 462-469.

8. El-Shami, M., Pontier, D., Lahmy, S., Braun, L., Picart, C., Vega, D., Lagrange, T. (2007). Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. Genes Dev, 21(20), 2539-2544.

9. Feng, S., Jacobsen, S. E., & Reik, W. (2010). Epigenetic reprogramming in plant and animal development. Science, 330(6004), 622-627.

10. Fukunaga, R., & Doudna, J. A. (2009). dsRNA with 5' overhangs contributes to endogenous and antiviral RNA silencing pathways in plants. EMBO J, 28(5), 545-555.

11. Gao, Z., Liu, H. L., Daxinger, L., Pontes, O., He, X., Qian, W., Zhu, J. K. (2010). An RNA polymerase II- and AGO4-associated protein acts in RNA-directed DNA methylation. Nature, 465(7294), 106-109.

12. Havecker, E. R., Wallbridge, L. M., Hardcastle, T. J., Bush, M. S., Kelly, K. A., Dunn, R. M., Baulcombe, D. C. (2010). The Arabidopsis RNA-directed DNA methylation argonautes functionally diverge based on their expression and interaction with target loci. Plant Cell, 22(2), 321-334.

13. He, X. J., Hsu, Y. F., Zhu, S., Wierzbicki, A. T., Pontes, O., Pikaard, C. S., Zhu, J. K. (2009). An effector of RNA-directed DNA methylation in arabidopsis is an ARGONAUTE 4- and RNA-binding protein. Cell, 137(3), 498-508.

14. Herr, A. J., Jensen, M. B., Dalmay, T., & Baulcombe, D. C. (2005). RNA polymerase IV directs silencing of endogenous DNA. Science, 308(5718), 118-120.

15. Ito, M., Koike, A., Koizumi, N., & Sano, H. (2003). Methylated DNA-binding proteins from Arabidopsis. Plant Physiol, 133(4), 1747-1754.

16. Jiao, X., Trifillis, P., & Kiledjian, M. (2002). Identification of target messenger RNA substrates for the murine deleted in azoospermia-like RNA-binding protein. Biol Reprod, 66(2), 475-485.

17. Kanno, T., Mette, M. F., Kreil, D. P., Aufsatz, W., Matzke, M., & Matzke, A. J. (2004). Involvement of putative SNF2 chromatin remodeling protein DRD1 in RNA-directed DNA methylation. Curr Biol, 14(9), 801-805.

18. Kanno, T., Huettel, B., Mette, M. F., Aufsatz, W., Jaligot, E., Daxinger, L., Matzke, A. J. (2005). Atypical RNA polymerase subunits required for RNA-directed DNA methylation. Nat Genet, 37(7), 761-765.

19. Kanno, T., Bucher, E., Daxinger, L., Huettel, B., Bohmdorfer, G., Gregor, W., Matzke, A. J. (2008). A structural-maintenance-of-chromosomes hinge domain-containing protein is required for RNA-directed DNA methylation. Nat Genet, 40(5), 670-675.

20. Law, J. A., Ausin, I., Johnson, L. M., Vashisht, A. A., Zhu, J. K., Wohlschlegel, J. A., & Jacobsen, S. E. (2010). A protein complex required for polymerase V transcripts and RNA- directed DNA methylation in Arabidopsis. Curr Biol, 20(10), 951-956.

21. Law, J. A., Vashisht, A. A., Wohlschlegel, J. A., & Jacobsen, S. E. (2011). SHH1, a homeodomain protein required for DNA methylation, as well as RDR2, RDM4, and chromatin remodeling factors, associate with RNA polymerase IV. PLoS Genet, 7(7), e1002195.

22. Liu, J., Bai, G., Zhang, C., Chen, W., Zhou, J., Zhang, S., Zhu, J. K. (2011). An atypical component of RNA-directed DNA methylation machinery has both DNA methylation-dependent and -independent roles in locus-specific transcriptional gene silencing. Cell Res, 21(12), 1691-1700.

23. Moazed, D. (2009). Small RNAs in transcriptional gene silencing and genome defence. Nature, 457(7228), 413-420.

24. Mourrain, P., Beclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J. B., Vaucheret, H. (2000). Arabidopsis SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance. Cell, 101(5), 533 542.

25. Onodera, Y., Haag, J. R., Ream, T., Costa Nunes, P., Pontes, O., & Pikaard, C. S. (2005). Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. Cell, 120(5), 613-622.

26. Pontier, D., Yahubyan, G., Vega, D., Bulski, A., Saez-Vasquez, J., Hakimi, M. A., Lagrange, T. (2005). Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in Arabidopsis. Genes Dev, 19(17), 2030-2040.

27. Rowley, M. J., Avrutsky, M. I., Sifuentes, C. J., Pereira, L., & Wierzbicki, A. T. (2011). Independent chromatin binding of ARGONAUTE4 and SPT5L/KTF1 mediates transcriptional gene silencing. PLoS Genet, 7(6), e1002120.

28. Smith, L. M., Pontes, O., Searle, I., Yelina, N., Yousafzai, F. K., Herr, A. J., Baulcombe, D. C. (2007). An SNF2 protein associated with nuclear RNA

silencing and the spread of a silencing signal between cells in Arabidopsis. Plant Cell, 19(5), 1507-1521.

29. Wierzbicki, A. T., Haag, J. R., & Pikaard, C. S. (2008). Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. Cell, 135(4), 635-648.

30. Wierzbicki, A. T., Ream, T. S., Haag, J. R., & Pikaard, C. S. (2009). RNA polymerase V transcription guides ARGONAUTE4 to chromatin. Nat Genet, 41(5), 630-634.

31. Xie, Z., Johansen, L. K., Gustafson, A. M., Kasschau, K. D., Lellis, A. D., Zilberman, D., Carrington, J. C. (2004). Genetic and functional diversification of small RNA pathways in plants. PLoS Biol, 2(5), E104.

32. Xie, M., Ren, G., Costa-Nunes, P., Pontes, O., & Yu, B. (2012). A subgroup of SGS3-like proteins act redundantly in RNA-directed DNA methylation. Nucleic Acids Res.

33. Yu, B., Bi, L., Zheng, B., Ji, L., Chevalier, D., Agarwal, M., Chen, X. (2008). The FHA domain proteins DAWDLE in Arabidopsis and SNIP1 in humans act in small RNA biogenesis. Proc Natl Acad Sci U S A, 105(29), 10073-10078.

34. Zhang, H., & Zhu, J. K. (2011). RNA-directed DNA methylation. Curr Opin Plant Biol, 14(2), 142-147.

35. Zhang, C. J., Ning, Y. Q., Zhang, S. W., Chen, Q., Shao, C. R., Guo, Y. W., He, X. J. (2012). IDN2 and Its Paralogs Form a Complex Required for RNA-Directed DNA Methylation. PLoS Genet, 8(5), e1002693.

36. Zheng, X., Zhu, J., Kapoor, A., & Zhu, J. K. (2007). Role of Arabidopsis AGO6 in siRNA accumulation, DNA methylation and transcriptional gene silencing. EMBO J, 26(6), 1691-1701.

37. Zheng, B., Wang, Z., Li, S., Yu, B., Liu, J. Y., & Chen, X. (2009). Intergenic transcription by RNA polymerase II coordinates Pol IV and Pol V in siRNA-directed transcriptional gene silencing in Arabidopsis. Genes Dev, 23(24), 2850-2860.

38. Zheng, Z., Xing, Y., He, X. J., Li, W., Hu, Y., Yadav, S. K., . . . Zhu, J. K. (2010). An SGS3-like protein functions in RNA-directed DNA methylation and transcriptional gene silencing in Arabidopsis. Plant J, 62(1), 92-99.

39. Zilberman, D., Cao, X., & Jacobsen, S. E. (2003). ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. Science, 299(5607), 716-719.

CHAPTER 4

Regulation of miRNA abundance by RNA binding protein TOUGH in *Arabidopsis*

PNAS. **(2012) Vol. 109(31):12817-1282**1

Guodong Ren, Meng Xie, Yongchao Dou, Shuxin Zhang, Chi Zhang

and Bin Yu

Abstract

miRNAs are regulators of gene expression in plants and animals. Their biogenesis is precisely controlled to secure normal development of organisms. Here we report that TOUGH (TGH) is a novel component of DCL1-HYL1-SE complex that processes of primary transcripts of miRNAs (pri-miRNAs) into miRNAs in *Arabidopsis*. Lack of TGH impairs multiple DCL activities *in vitro* and reduces the accumulation of miRNAs and siRNAs *in vivo*. TGH is an RNA binding protein, binds pri-miRNAs and pre-miRNAs *in vivo* and contributes to pri-miRNA-HYL1interaction. These results indicate that TGH might regulate abundance of miRNAs through promoting the DCL1 cleavage efficiency and/or recruitment of pri-miRNAs.

Introduction

Small RNAs, including microRNAs (miRNAs) and small interfering RNAs (siRNAs), are sequence-specific regulators of gene expression in plants and animals (1). MiRNAs are derived from imperfect stem-loop transcripts, called primary-miRNAs (pri-miRNAs), which are predominately produced by DNA-dependent RNA polymerase II, whereas siRNAs are processed from perfect or near perfect long double-stranded RNAs (dsRNAs) (2). After generation, miRNA and siRNA are loaded into an RNA-induced silencing complex (RISC) containing the Argonaute protein to guide posttranscriptional or transcriptional gene silencing (1).

In animals, pri-miRNAs are first processed to pre-miRNAs in the nucleus by the microprocessor containing Drosha and a dsRNA-binding protein DGCR8 (1). The resulting pre-miRNAs are then processed by Dicer in the cytoplasm to produce mature

miRNAs (1). It has emerged that the activities of Drosha and Dicer are controlled to regulate miRNA expression in response to developmental and environmental signals (3). In *Arabidopsis*, DCL1, a dsRNA-binding protein HYL1 and a zinc finger protein SERRATE (SE) form a complex to process pri-miRNAs in the nucleus to pre-miRNAs and then to mature miRNAs (4-6). The accumulation of miRNAs in *Arabidopsis* also requires DDL, which was proposed to stabilize pri-miRNAs and to facilitate their processing (7). Recently, two cap-binding proteins, CBP80/ABH1 and CBP20, were found to be required for both pre-mRNA splicing and pri-miRNA processing (8, 9). Plants also encode several classes of endogenous siRNAs including the natural anti-sense transcript derived siRNA (nat-siRNA), siRNA derived from repetitive DNA sequences (rasiRNA), and trans-acting siRNA (ta-siRNA) (10). In *Arabidopsis*, the generation of these siRNAs from long dsRNAs involves DCL1 homologs DCL2, DCL3 and DCL4, which produce 22nt, 24nt and 21nt siRNAs, respectively (11-13).

In this report, we show that TOUGH (TGH) is an important factor for miRNA and siRNA biogenesis. Loss-of-function TGH in *tgh-1* reduces the activity of multiple DCLs *in vitro* and the accumulation of miRNA and siRNAs *in vivo*. In miRNA pathway, TGH associates with the DCL1 complex, binds pri-miRNAs and pre-miRNAs. TGH is required for the efficient *in vivo* interaction between pri-miRNA and HYL1. These data suggest that TGH assists DCLs to efficiently process and/or recruit the precursors of miRNAs and siRNAs.

Results

TGH is required for the accumulation of miRNAs and siRNAs in *Arabidopsis*

Three facts prompted us to test whether TGH acts in miRNA pathway. First, TGH is an evolutionarily conserved protein across plant and animal kingdoms (14), agreeing with the fact that many components involved in miRNA biogenesis are conserved in eukaryotes (1). Second, TGH contains a G-patch and a Suppressor-of-White-APricot (SWAP) domain that often exist within RNA metabolism related proteins (14) (figure 4-6A). Finally, like *dcl1, ddl, hyl1* and *abh1* that are deficient in miRNA pathway, the *tgh* mutants exhibit pleiotropic developmental defects such as smaller plant size, altered leaf shape, short stature, increased branches, disordered node distribution and reduced fertility (14-20) (figure 4-6B).

To determine whether TGH functions in miRNA biogenesis, we examined the accumulation of various DCL1-dependent miRNAs in inflorescences of *tgh-1* (SALK_053445), which contains a T-DNA insertion in the 11th intron and is a potential null allele (14) (Fig. S1A). The levels of all tested miRNAs were reduced in *tgh-1* by 50%-70% relative to the wild-type control (Columbia-0; Wt; figure 4-1A). The expression of miR172* was also reduced in *tgh-1* (figure 4-1A). Expressing a genomic copy of *TGH* driven by its native promoter fused with a *HA* tag at its C-terminal (*TGH::TGH-HA*) fully restored the levels of these miRNAs and miRNA172* (figure 4-1A), demonstrating that lack of *TGH* in *tgh-1* was responsible for the defects in miRNA accumulation. We also checked the levels of several miRNAs in mature leaves. All of them were less accumulated in *tgh-1* than in Wt (figure 4-61C).

Next, we asked whether TGH plays a role in the accumulation of rasiRNAs and ta-siRNAs. We found that both DCL4-dependent ta-siRNAs, TAS1-siR255 and TAS3-5'D8(+) , DCL2-dependent IR71 and DCL3-dependent rasiRNAs were reduced in abundance in *tgh-1* compared to those in Wt, and the reduction was rescued by the *TGH* transgene (figure 4-1B). In addition, the levels of DCL4-dependent miR822 (21) were also lower in *tgh-1* than in Wt and the defect was restored by the *TGH* transgene (figure 4-1A).

We further compared the transcript levels of several miRNA targets, *CUC1*, *PHV*, *SAMT*, *PPR* and a ta-siRNA target *ARF3* between Wt and *tgh-1*, which should inform whether *tgh-1* impaired miRNA and ta-siRNA function. The transcript levels of these miRNA targets were slightly increased in *tgh-1* relative to Wt (figure 4-6D).

TGH does not affect miRNA precision

Although Northern blot showed that TGH affects the accumulation of miRNAs, it could not tell whether miRNA precision requires TGH. To address this question, we performed Illumina deep sequencing analysis of small RNA libraries constructed from inflorescences of WT and *tgh-1*. The data set was deposited into NCBI (GSE38600). We focused our analysis on miRNAs. The abundance of most miRNAs was reduced in *tgh-1* relative to Wt in two biological replicates (figure 4-1C). This analysis further confirmed that TGH is required for the accumulation of miRNAs. We next evaluated whether TGH affected processing precision. According to Liu et al (22), imprecise miRNAs were defined as those that did not fall within ±2 bases of the annotated mature miRNA(s) or

miRNA*(s) positions. Because evaluation on miRNA precision depends on sequencing depth (22), we only analyzed the highly expressed miRNAs. Like Wt, *tgh-1* contained very low ratio of imprecise miRNAs, indicating that TGH may be not required for the accurate cleavage of pri-miRNAs.

Multiple DCL activities are impaired in *tgh-1*

To determine at which step TGH may act in miRNA biogenesis, we examined the levels of pri-miRNAs in Wt and *tgh-1*. Quantitative RT-PCR (qRT-PCR) analyses showed that the levels of pri-miRNAs at 6 loci were increased by 1.5 to 2.5-fold in *tgh-1* relative to that in Wt (figure 4-2A). This result suggested a potential defect of DCL1 activity in *tgh-1*. We also compared the levels of pri-miRNA from each member of miR159, miR167 and miR171 between Wt and *tgh-1*, with the expectation to inform whether TGH equally affects the processing of each member of miRNA families. Although *tgh-1* increased the levels of these pri-miRNAs, its effects on individual pri-miRNA were varied (figure 4-7B).

It has been established that DCL1 and DCL3 are responsible for the production of 21 and 24 nt small RNAs in an *in vitro* dsRNA processing assay using *Arabidopsis* protein extracts, respectively (23). We adapted this assay to test whether DCL1 and DCL3 activities are impaired in *tgh-1*. A radioactive labeled dsRNA (460 bp) was incubated with protein extracts from young flower buds of *tgh-1* or Wt. The reactions were stopped at 40, 80 and 120 minutes, and the RNAs from each reaction were extracted and resolved on a polyacrylamide gel. The production of small RNAs by *tgh-1* protein extract was

lower than that by Wt (figure 4-2B). Quantitative analysis revealed that the overall DCL processing activity in *tgh-1* was about 40% of that in Wt (figure 4-2D). The RNAs extracted from 120 minute-reaction were further resolved on a long PAGE gel to separate the 24nt and 21 nt small RNAs. The production of both 24 and 21 nt small RNAs was lower in *tgh-1* extracts than in Wt (figure 4-2C). These observations indicated that both DCL1 and DCL3 activities are impaired in *tgh-1*. To test the effects of *tgh-1* on DCL1-mediated miRNA maturation, we compared processing of a short form of pri-miR162b (predicted stem loop with 6 nt arms at each end; figure 4-7A) between *tgh-1* and Wt protein extracts. As a control of pri-miRNA processing, we included *dcl1-9*, which is a weak allele of *dcl1* and has reduced miRNA production, as a control. Like *dcl1-9*, *tgh-1* reduced pri-miR162b processing efficiency relative to Wt (figure 4-2E).

TGH associates with the DCL1 complex

There are several possible ways for TGH to affect DCL1 activities. We first analyzed the expression level of several key genes in miRNA biogenesis by qRT-PCR. The abundance of *DDL*, *CBP20* and *CBP80* were comparable between Wt and *tgh-1* (figure 4-3A). The expression levels of *DCL1*, *SE* and *HEN1* were slightly increased in *tgh-1* compared with Wt, whereas the levels of HYL1 were slightly decreased (figure 4-3B). However, *tgh-1* had no effect on the protein level of HYL1 and DCL1 (figure 4-3B).

Next, we tested the association of TGH with DCL1 using co-IP/pull down assay. We expressed the recombinant TGH protein fused with a maltose-binding protein epitope at its N-terminus (MBP-TGH) in *E.coli* and the DCL1 protein fused with a yellow fluorescent protein (YFP) in *Nicotiana benthamiana* (figure 4-3C) (7). We mixed the

MBP-TGH and DCL1-YFP protein extracts and performed reciprocal pull down assays with amylose resin and a GFP antibody conjugated to protein A-agarose beads, respectively. Antibodies against GFP and MBP epitope detected the enrichments of DCL1-YFP in MBP-TGH precipitates and MBP-TGH in DCL1-YFP complexes, respectively (figure 4-3A and 4-3B), indicating the TGH-DCL1 interaction. TGH is a putative RNA-binding protein raising the possibility that the TGH-DCL1 association might be RNA-mediated. RNase A treatment abolished the RNA-mediated FDM1-AGO4 interaction (figure 4-8D) (24) but not TGH-DCL1 interaction (figure 4-3A and 4-3B). As controls, we performed reciprocal pull downs to test the YFP/MBP, YFP/MBP-TGH, and MBP/DCL1-YFP interactions. We did not detect any interactions among these proteins (figure 4-3A and 4-3B). We further tested the HYL1-TGH and SE-TGH associations using pull down assay. MBP-TGH but not MBP pulled down HYL1 and SE from *Arabidopsis* protein extracts (figure 4-3C). The control protein HSP70 was not detected in the MBP-TGH precipitates. Because TGH affects 24 nt siRNA production, we tested co-immunoprecipitation between TGH and DCL3. We detected the presence of MBP-TGH but not MBP in the DCL3 immunoprecipitates (figure 4-8E).

To ascertain the association between TGH and the DCL1 complex, we performed a bimolecular fluorescence complementation (BiFC) assay. In this assay, we fused protein partners to the N-terminal fragment of Venus (nVenus) or C-terminal fragment of cyan fluorescent protein (cCFP), respectively, and introduced paired proteins into tobacco cells by infiltration. The interaction of the two protein partners will generate a functional YFP leading to fluorescence (25). Similar methods have been previously used to investigate

the interactions among DCL1, HYL1 and SE (4, 5). BiFC signals produced from the TGH-SE, TGH-DCL1, TGH-HYL1 and SE-DCL1 (positive control) interactions were observed in distinct nuclear speckles (figure 4-3D). In contrast, only weak fluorescence signals were observed from the control AGO1-TGH pair (figure 4-3D). These results indicated that TGH is a component of the pri-miRNA processing complex.

TGH binds both pri-miRNAs and pre-miRNAs

The presence of putative RNA binding domains in TGH suggested that TGH might be an RNA binding protein. We performed a pull-down assay to examine the interaction between TGH and pri-miR162b interaction, which was used for in *vitro* processing assay (figure 4-7A). MBP and TGH-MBP expressed in *E.coli* were purified with amylose resin (figure 4A). TGH-MBP but not MBP was able to retain pri-miR162b and addition of unlabelled pri-miR162b was able to wash off the radioactive signal (figure 4-4B). We also generated a radioactive-labeled pre-miR162b, which has a 2 nt 3' overhang (figure 4-7A), by *in vitro* transcription and examined its interaction with TGH. TGH interacted with the pre-miR162b. However, TGH-MBP couldn't bind a ~460 bp double-stranded RNA (dsRNA) (figure 4-4B), indicating that TGH may be an ssRNA binding protein. In fact, TGH bound a ~100 nt of *UBIQUITIN 5* mRNA from 5' end CDS (*UBQ5*) *in vitro* (figure 4-4B).

Next, we tested TGH- pri-miRNA and TGH-miRNA associations *in vivo*. Seedlings of *tgh-1* complementation plants harboring the *TGH::TGH-HA* transgene were subjected to RNA immunoprecipitation (RIP) (26). RT-PCT detected all the tested pri-miRNAs were

present in the TGH-HA complex but not in the immunoprecipitates from non-transgenic plants and "no antibody" controls (figure 4-4C). We did not find the interaction between TGH and RNA controls *AtSN1B* RNA, which is transcribed from the flanking region of AtSN1 locus (26), *npc72* (27) and *UBQ5* mRNA (figure 4-4C). This result indicated that TGH might specifically interact with some RNAs *in vivo*. However, we did not detect *AtSN1A* RNA (figure 4-4C), which likely is a ra-siRNA generating RNA, in the TGH-HA complex. An explanation is that TGH might transiently interact with the DCL3 complex. Alternatively, it may due to that the substrates of DCL3 are dsRNAs. To examine the association of TGH with pre-miRNA *in vivo,* TGH-bound RNAs were ligated to a 3' adaptor and then reverse transcription and nested PCR were performed to detect the pre-miRNA (figure 4-4D). This assay allowed us to detect pre-miR172a and pre-miR166a in the TGH complex (figure 4-4D).

tgh-1 impairs the HYL1-pri-miRNA interaction

Based on the association of TGH with pri-miRNA and its processing complex, we tested whether TGH contributes to HYL1-pri-miRNA interaction. HYL1-pri-miRNA interaction is essential for pri-miRNA processing (28). We examined HYL1-pri-miRNA interaction in Wt and *tgh-1* by RIP using antibody against HYL1. A similar amount of HYL1 was obtained from the protein extracts of *tgh-1* and Wt (figure 4-5A). RT-PCR and qRT-PCR analysis revealed that the amount of HYL1-bound pri-miR167a and pri-miR171a was reduced in *tgh-1* relative to Wt (figure 4-5B and 4-5C). We included *hyl1-2* as a negative control in this experiment. No HYL1 and its associated RNAs were immunoprecipitated by HYL1 antibody from *hyl1-2* (figure 4-9).

Discussion

In conclusion, TGH is an important component of miRNA and siRNA biogenesis. Several lines of evidences demonstrate that TGH has a role in promoting miRNA maturation. The facts that lack of TGH in *tgh-1* reduces the accumulation of miRNAs and increases the levels of pri-miRNAs and the association of TGH with the DCL1 complex, pri- miRNAs and pre-miRNAs demonstrate that TGH has a role in promoting miRNA maturation. However, TGH shall have additional important functions in plants because *tgh-1* has severe morphological phenotypes whereas its effects on the levels of miRNAs appear to be less than *dcl1-9*.

In the miRNA pathway, TGH may have two non-mutually exclusive activities. First, TGH may contribute to the interaction between pri-miRNA and DCL1 complex, which is supported by the reduced amount of pri-miRNA in the HYL1 complex from *tgh-1*. Second, TGH may have a role in modulating DCL1 activity, as DCL1-dependent *in vitro* pri-miRNA and dsRNA processing is impaired in the TGH-depleted extracts. However, TGH may not affect miRNA precision as *tgh-1* contains very low ratio of imprecise miRNAs. TGH affects the accumulation of DCL4-dependent miR822. The reduction of ta-siRNA and ra-siRNA levels indicates that TGH may have a role in siRNA biogenesis. However, the direct role of TGH in ta-siRNA processing needs further investigation, because DCL1-dependent miRNAs is also required for ta-siRNA biogenesis (29, 30). The reduction of DCL3-dependent 24 nt small RNA production in *tgh-1* protein extracts indicates that TGH may act as a co-factor of DCL3 to facilitate dsRNA processing (figure

4-2). However, TGH may not contribute to the DCL3-dsRNA association as it does not bind dsRNAs *in vitro*. Clearly, this needs to be further examined.

TGH is an evolutionarily conserved protein in plant and animals. Given the similarity of small RNA pathways among different organisms, it will not be a surprise that the TGH homologs from other organisms have a role in RNA silencing. The reduced expression of *TGH* homolog from *C. elegans* has been shown to cause either embryonic lethality or developmental defects in genome-wide RNAi screens (31), consistent with the role of miRNA in regulating developmental processes of plants and animals.

Materials and Methods

Plant materials

A ~5.5 kb TGH genomic fragment containing the TGH coding and promoter regions was amplified by PCR with primers TGHg-GW F/R and cloned into Gateway vector pEG301 to produce a pTGH:TGH-HA plasmid. The resulting plasmid was transformed into *tgh-1*. Basta resistance was used to select the transgenic plants.

Plasmid construction

TGH cDNA was amplified by RT-PCR and cloned into pMAL-c5x (NEB) to generate an MBP-TGH plasmid construct. MBP was amplified by PCR using the pMAL-c5x plasmid DNA as template and cloned into pET43a+ (Novagen) to generate a pM6H construct. *TGH* cDNA was then amplified by RT-PCR and cloned into to the pM6H vector to

generate a TGH-MBP-6xHIS construct. cDNAs of *DCL*1, *HYL1*, *SE*, *DCL1-9*, *AGO1* and *AtCoilin* were cloned into the pSAT1-nVenus-C vector, respectively. The resulting plasmids were cut with AscI restriction enzyme to release desired DNA fragments, which were subsequently cloned into the binary vector pPZP- ocs-bar-RCS2-2 to generate the nVenus tagged DCL1, HYL1, SE, DCL1-9, AGO1 and AtCoilin constructs. cDNAs of TGH and SE were cloned into or pSAT4-cCFP-C, respectively. The DNA fragments containing TGH or SE from the resulting plasmids were released by I-SceI restriction enzyme treatment and subsequently cloned into the pPZP- ocs-bar-RCS2-2 plasmid to generate cCFP tagged TGH and SE constructs. DCL3 cDNA were amplified with primer DCL3GW F/R and cloned into pEG101.

Small RNA sequencing

Total RNA was extracted from inflorescence tissue. Small RNAs with 15-30nt in size were purified from 200ug total RNA by denatured Polyacrylamide gel according to the reference. Small RNA libraries were prepared and sequenced using Illumina Genome Analyzer IIx following the standard protocol. The small RNA reads were trimmed for adaptor sequence using Perl scripts and mapped to either the *Arabidopsis* genome (TAIR 9.0, for miRNA abundance) or miRNA hairpin sequences (from miRBase v1.8, for miRNA imprecision) using Bowtie program. The sequences of miRNA and miRNA* sequences were obtained from miRBase. Comparison of miRNA abundance was calculated by using EdgeR with trimmed mean of M values (TMM) normalization method. The total numbers of perfectly aligned reads, except reads aligned to t/r/sn/snoRNA, were used for normalization.

RNA analysis

5′-End-labeled ^{32}P antisense LNA oligonucleotides were used to probe miRNAs and siRNAs. For quantitative RT-PCR (qRT-PCR) analysis of pri-miRNAs and miRNA target transcripts, RNA was reverse transcribed by the Superscript III reverse transcriptase (Invitrogen) and a oligo-T18 primer to generate cDNA. qRT-PCR was performed in triplicate using SYBR Green kit (Bio-Rad) on an iCycler (Bio-Rad) apparatus.

RNA immunoprecipitation

RNA immunoprecipitation(RIP) were performed as described. Briefly, 2g *Arabidopsis* infloresence was crosslinked with 1% formaldehyde by vacuum infiltration for 40 minutes and quenched by adding glycine to 0.125M for 10 minutes. The nuclei were then extracted and suspended in 400μl Nuclei Lysis Buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS) and sonicated 5 times. Debris was removed by centrifugation at 16,000 g for 10 min. Protein concentration was determined by Bradford assay (BioRad) and equal amount of protein was used for RIP analysis. 60μl aliquot of supernatant (10ul was saved for input) was diluted with 540μl RIP Dilution Buffer (1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl pH 8.0, 167 mM NaCl). After preclear with Protein A agarose beads, 20μl protein A agarose conjugated -anti-HA beads or Protein A agarose beads (for no Antibody controls) were added and incubated overnight . Immunoprecipitates were washed five times with RIP Washing buffer (150 mM NaCl, 20 mM Tris-HCl pH 8.0, 2 mM EDTA, 1% Triton X-100, 0.1% SDS). Immune complexes were subsequently eluted

with 500μl Elution Buffer (100mM NaHCO3, 1% SDS) with occasionally shaking for 30 min at 65 ºC. Crosslinking was reversed at 65 ºC for 2h in the presence of 20 μg Proteinase K (Invitrogen) and 200mM NaCl. RNAs were then extracted and used for RT-PCR analysis. For quantitative analysis of HYL1-bound pri-miRNAs, the pri-miRNA amount in HYL1 precipitates were normalized to that in 10% input as described (12). The percentage of input were calcaulted as 100*2 (Ct of input-Ct of IP)*0.1 (0.1 is the dilution factor; 10% of input was used for quantitative analysis). T4 RNA ligase (BioLab) -mediated 3´ adapter primer ligation was performed. RT was performed using primer P1 recognizing the 3' adaptor. Nested PCR was performed first with primers P1 and P2, and then with P3 and P4.

Dicer activity assay

Dicer activity assay was performed according to (23). DNA template for dsRNA and pri-miR162b was amplified using T7 promoter anchored primers. The DNA templates for dsRNAs contain the T7 promoter at both ends. Resulting DNAs were used for *in vitro* transcription under the presence of α-^{32}P UTP. RNAs were resolved on 6% native PAGE gel and eluted with buffer containing 300 mM NaCl and 15 mM EDTA. After passing Spin-X filter, purified RNAs were precipitated with ethanol. For Dicer activity assay, RNAs were incubated with 30 μg protein in 20 μl reaction buffer containing 100mM NaCl, 1mM ATP, 0.2mM GTP, 1.2mM MgCl2, 25mM creatine phosphate, 30 μg/ml creatine kinase, and 4 U Rnase Inhibitor at room temperature. RNAs were extracted, precipitated and resolved on PAGE gel. Radioactive signals were detected with a phosphor imager and quantified by ImageQuant V5.2.

BiFC assay

Paired constructs were co-expressed in *N. benthamiana* leaves for 40 hrs and subjected to confocal microscopy (Olympus Fluoview 500 workstation; Olympus America Inc) for imaging. BiFC were excited at 488 nm and detected with a narrow barrier filter (BA505–525 nm). Nuclei were visualized by DAPI staining.

Figures

Figure 4-1. *tgh-1* reduces the accumulation of miRNAs and siRNAs. (A) The accumulation of miRNAs and miR172* in three genotypes. (B) The accumulation of siRNAs in three genotypes. For miR159/319: Upper band, miR159; lower band miR319. The numbers indicate the relative abundance of small RNAs among the three genotypes and represent the mean of three repeats ($p<0.05$). U6 blot was used as a loading control. Col-0, the wild-type control (Wt) for *tgh-1*; *tgh-1+TGH*, *tgh-1* harboring TGH genomic DNA. (C) Deep sequencing analysis of miRNAs in *tgh-1* and Wt. The miRNA abundance was calculated as reads per million (RPM) and a $\log_2$-transformed ratio of *tgh-1*/Wt were plotted. Each circle represents one miRNA. Thick lines indicate median values.

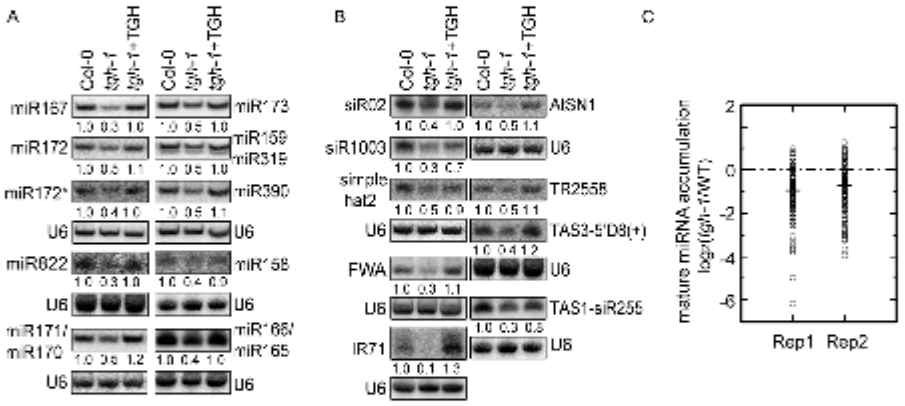


Figure 4-2. *tgh-1* impairs multiple DCL activities. (A) Increased pri-miRNA levels in inflorescences of *tgh-1*. The levels of pri-miRNAs in *tgh-1* were normalized to those of *UBIQUITIN 5* and compared with Col. Error bars indicate standard deviation of three technical replications. *:p<0.05; **: p<0.01 (B) and (C) Reduced production of siRNAs from dsRNAs in the *tgh-1* protein extracts. Numbers below indicated the siRNA production in *tgh-1* relative to the control. (D) Quantification of overall siRNA production in *tgh-1* extracts relative to the control extracts. Data are presented as mean and standard deviation (n=7). ***:p<0.001. (E) The pri-miR162b processing. Numbers indicate overall miRNA production in *tgh-1* and *dcl1-9* extracts relative to their respective control extracts and represent the mean of three experiments (P<0.05).

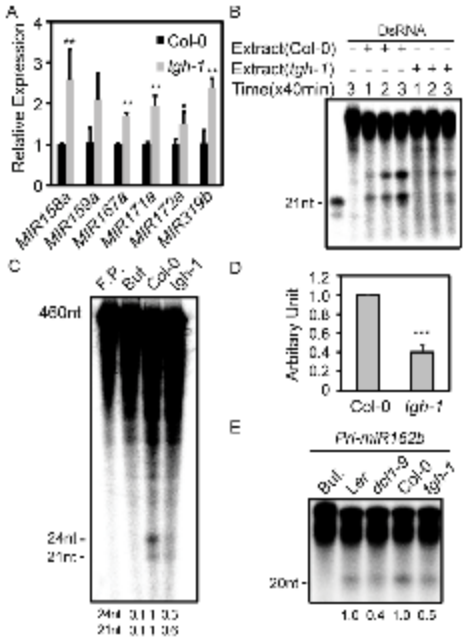


Figure 4-3. TGH associates with the DCL1 complex. (A) MBP-TGH pulls down DCL1-YFP. (B) DCL1-YFP pulls down MBP-TGH. (C) MBP-TGH pulls down HYL1 and SE. Protein precipitates were analyzed by Western blot using anti-MBP, anti-GFP and anti-HYL1 antibodies, respectively. 1/100 input was used for MBP-TGH and MBP. 1/50 input was used for YFP, DCL1-YFP and HYL1. (D) BiFC analysis between TGH and the components of DCL1 complex. TGH, and SE were fused with cCFP, respectively, whereas DCL1, HYL1, SE, AGO1 were fused with nVenus, respectively. Respective pair of cCFP and nVenus fusion proteins was co-infiltrated into leaves and fluorescence signals were examined ~40 hours after co-infiltration. The interaction of paired proteins will result in yellow fluorescence (green color in the picture). More than 30 nuclei were examined for each pair and a graph was shown. DNA was stained with DAPI to visualize the nuclei (blue color).

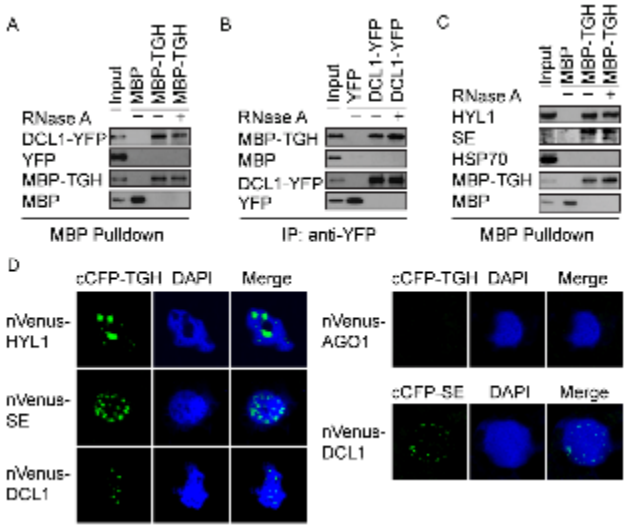


Figure 4-4. TGH is an RNA binding protein. (A) The TGH-MBP and MBP proteins used in the *in vitro* RNA binding assay. The proteins were detected by comassie brillant blue staining. (B) TGH binds pri-miR162b and pre-miR162b *in vitro*. (C) TGH binds pri-miRNA in *vivo*. C:Col-0. T: *tgh-1* harboring a *TGH::TGH-HA* transgene. No Ab: no antibody. 1/8 of immunoprecipitates were analyzed by western blot. No RT was performed with the pri-miR167a primers. Input RNA=5%. (D) TGH binds pre-miRNAs *in vivo*. RT was performed with primer P2. The first round PCR was done with primer P1 and P2. The Second round PCR was performed with primer P3 and P4, which recognize the junction between pre-miRNA and the adaptor. Open box: adaptor; light/dark grey box: miRNA/miRNA*; Black box: region between miRNA and miRNA*.

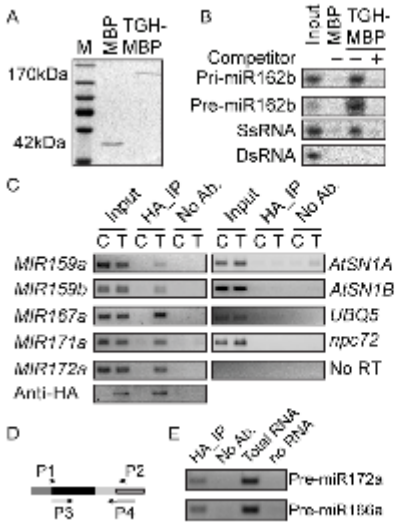


Figure 4-5. TGH contributes to *in vivo* HYL1-pri-miRNA interaction. A) Detection of HYL1 protein after immunoprecipitation. Immunoprecipitation was performed with the anti-HYL1 antibody. (B) and (C) The association between HYL1 and pri-miR171a and pri-miR167a was impaired in *tgh-1*. C:Col-0; t: *tgh-1*. No Ab: no antibody. 1/8 of immunoprecipitates were analyzed by western blot. Input=2% of total input. The amount of pri-miR167a and pri-miR171a was determined by qRT-PCR and normalized to the input. *AtSN1B* was used as a negative control. *:p<0.05; **:p<0.01.

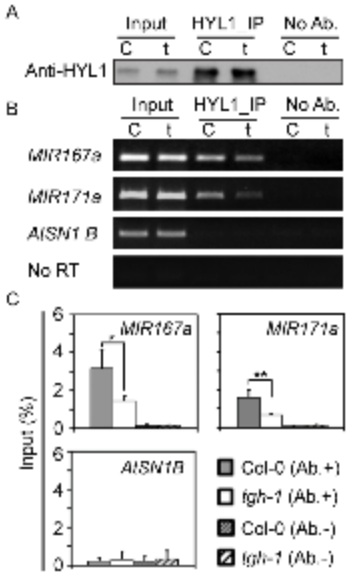


Figure 4-6. Phenotypes of *tgh-1*. (A) Schematic structure of TGH protein. DUF1604: domain of unknown function 1604; SWAP: Suppressor-of-White-Apricot. (B) Inflorescence stem structure of Col-0, *tgh-1* and *se-1*. Inflorescence stem structure of Col-0, *tgh-1* and *se-1*. Red arrowheads indicate two siliques emanating from the same node, while black arrows indicate fertile fruits in *tgh-1*. (C) The accumulation of miRNAs in *tgh-1* was reduced leaf tissues. miRNAs were detected by northern blot. U6 RNA served as a loading control. (D) The transcript levels of miRNA and ta-siRNA targets in *tgh-1* and Wt. The levels of target transcripts in *tgh-1* were normalized with *UBQUITIN5* (*UBQ5*) and compared with those in Wt. The Wt value is 1. Error bars indicate standard deviation of three technical replications. The experiment was repeated once with similar results.

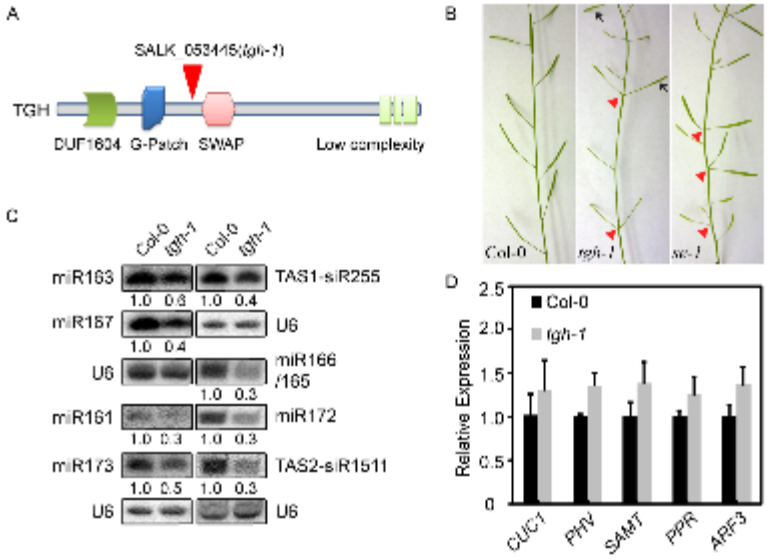


Figure 4-7. The effects of TGH on miRNA pathway. (A) Schematic diagram of the pri-miR162b and pre-miR162b. (B) Levels of pri-miRNAs in *tgh-1* compared to Wt. n.d.: Not detected. *UBQ5* was used as a reference control. The Wt value is 1. Error bars indicate standard deviation of three technical replications. The experiment was repeated once with similar results.

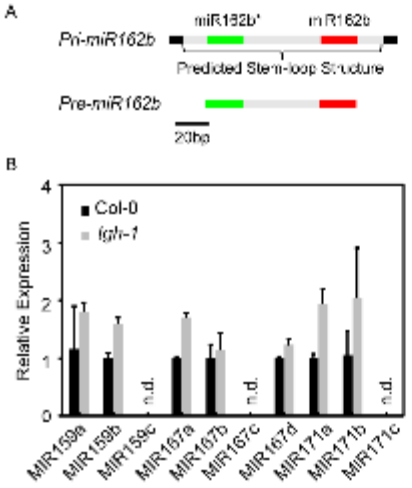


Figure 4-8. The role of TGH in miRNA pathway. (A) Expression levels of miRNA biogenesis pathway related genes in *tgh-1*. *UBQ5* was used as a reference control. Error bars indicate standard deviation of three technical replications. The experiment was repeated once with similar results. (B) DCL1 and HYL1 protein levels in *tgh-1*. DCL1-9 and HSC70 were included as an internal control. (C) The protein extracts containing DCL1-YFPor YFP were resolved on SDS-PAGE gel. DCL1-YFP and YFP were detected by Western blot using anti-YFP antibody. (D) Positive control of RNase treatment, *Arabidopsis* extracts containing myc-AGO4 were mixed with protein extracts containing GST or GST-FDM1 and captured with glutathione beads. GST and GST-FDM1 were visualized by Coomassie Blue staining. Protein precipitates were resolved on an SDS-polyacrylamide gel and detected by Western blotting with anti-Myc antibody. (E) DCL3 co-immunoprecipitates with TGH. Protein extracts containing DCL3-YFP or YFP were mixed with MBP or MBP-TGH and captured anti-GFP antibody conjugated to agarose beads with or without RNaseA. Proteins were detected with anti-GFP and anti-MBP antibodies, respectively.

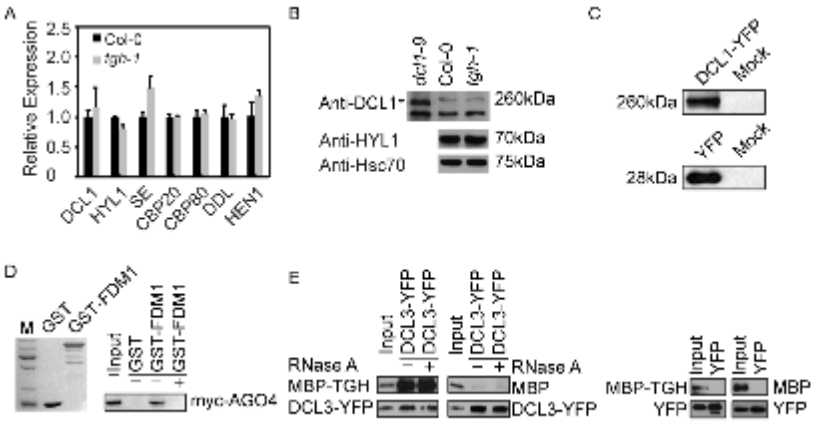
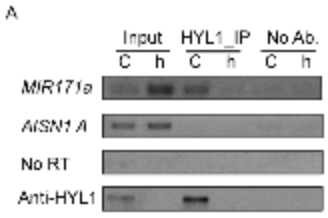


Figure 4-9. Association of HYL1 with pri-miRNA in Col-0 and *hyl1-2*. C:Col-0 control. h: *hyl1-2*. No Ab: no antibody control. 1/8 of immunoprecipitates were further analyzed by western blot. Input protein=2% of total input proteins. The pri-miR171a was detected by RT-PCR. *AtSN1B* was used as a negative control.



Reference:

1. Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116(2):281-297.

2. Chen X (2005) MicroRNA biogenesis and function in plants. *FEBS letters* 579(26):5923-5931.

3. Krol J, Loedige I, & Filipowicz W (2010) The widespread regulation of microRNA biogenesis, function and decay. *Nature reviews* 11(9):597-610.

4. Fang Y & Spector DL (2007) Identification of nuclear dicing bodies containing proteins for microRNA biogenesis in living Arabidopsis plants. *Curr Biol* 17(9):818-823.

5. Fujioka Y, Utsumi M, Ohba Y, & Watanabe Y (2007) Location of a possible miRNA processing site in SmD3/SmB nuclear bodies in Arabidopsis. *Plant Cell Physiol* 48(9):1243-1253.

6. Song L, Han MH, Lesicka J, & Fedoroff N (2007) Arabidopsis primary microRNA processing proteins HYL1 and DCL1 define a nuclear body distinct from the Cajal body. *Proc Natl Acad Sci U S A* 104(13):5437-5442.

7. Yu B*, et al.* (2008) The FHA domain proteins DAWDLE in Arabidopsis and SNIP1 in humans act in small RNA biogenesis. *Proc Natl Acad Sci U S A* 105(29):10073-10078.

8. Laubinger S*, et al.* (2008) Dual roles of the nuclear cap-binding complex and SERRATE in pre-mRNA splicing and microRNA processing in Arabidopsis thaliana. *Proc Natl Acad Sci U S A* 105(25):8795-8800.

9. Gregory BD*, et al.* (2008) A link between RNA metabolism and silencing affecting Arabidopsis development. *Dev Cell* 14(6):854-866.

10. Vazquez F (2006) Arabidopsis endogenous small RNAs: highways and byways. *Trends Plant Sci* 11(9):460-468.

11. Bouche N, Lauressergues D, Gasciolli V, & Vaucheret H (2006) An antagonistic function for Arabidopsis DCL2 in development and a new function for DCL4 in generating viral siRNAs. *EMBO J* 25(14):3347-3356.

12. Xie Z, Allen E, Wilken A, & Carrington JC (2005) DICER-LIKE 4 functions in trans-acting small interfering RNA biogenesis and vegetative phase change in Arabidopsis thaliana. *Proc Natl Acad Sci U S A* 102(36):12984-12989.

13. Henderson IR*, et al.* (2006) Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nat Genet* 38(6):721-725.

14. Calderon-Villalobos LI*, et al.* (2005) The evolutionarily conserved TOUGH protein is required for proper development of Arabidopsis thaliana. *Plant Cell* 17(9):2473-2485.

15. Hugouvieux V, Kwak JM, & Schroeder JI (2001) An mRNA cap binding protein, ABH1, modulates early abscisic acid signal transduction in Arabidopsis. *Cell* 106(4):477-487.

16. Jacobsen SE, Running MP, & Meyerowitz EM (1999) Disruption of an RNA helicase/RNAse III gene in Arabidopsis causes unregulated cell division in floral meristems. *Development* 126(23):5231-5243.

17. Chen X, Liu J, Cheng Y, & Jia D (2002) HEN1 functions pleiotropically in Arabidopsis development and acts in C function in the flower. *Development* 129(5):1085-1094.

18. Clarke JH, Tack D, Findlay K, Van Montagu M, & Van Lijsebettens M (1999) The SERRATE locus controls the formation of the early juvenile leaves and phase length in Arabidopsis. *Plant J* 20(4):493-501.

19. Lu C & Fedoroff N (2000) A mutation in the Arabidopsis HYL1 gene encoding a dsRNA binding protein affects responses to abscisic acid, auxin, and cytokinin. *Plant Cell* 12(12):2351-2366.

20. Morris ER, Chevalier D, & Walker JC (2006) DAWDLE, a forkhead-associated domain gene, regulates multiple aspects of plant development. *Plant Physiol* 141(3):932-941.

21. Rajagopalan R, Vaucheret H, Trejo J, & Bartel DP (2006) A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. *Genes Dev* 20(24):3407-3425.

22. Liu C, Axtell MJ, & Fedoroff NV (2012) The helicase and RNaseIIIa domains of Arabidopsis DCL1 modulate catalytic parameters during microRNA biogenesis. *Plant Physiol*.

23. Qi Y, Denli AM, & Hannon GJ (2005) Biochemical specialization within Arabidopsis RNA silencing pathways. *Mol Cell* 19(3):421-428.

24. Xie M, Ren G, Costa-Nunes P, Pontes O, & Yu B (2012) A subgroup of SGS3-like proteins act redundantly in RNA-directed DNA methylation. *Nucleic Acids Res* 40(10):4422-4431.

25. Ghosh I, Hamilton AD, & Regan L (2000) Antiparallel leucine zipper-directed protein reassembly: Application to the green fluorescent protein. *Journal of the American Chemical Society* 122(23):5658-5659.

26. Wierzbicki AT, Haag JR, & Pikaard CS (2008) Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell* 135(4):635-648.

27. Ben Amor B*, et al.* (2009) Novel long non-protein coding RNAs involved in Arabidopsis differentiation and stress responses. *Genome Res* 19(1):57-69.

28. Yang SW*, et al.* (2010) Structure of Arabidopsis HYPONASTIC LEAVES1 and its molecular implications for miRNA processing. *Structure* 18(5):594-605.

29. Allen E, Xie Z, Gustafson AM, & Carrington JC (2005) microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell* 121(2):207-221.

30. Yoshikawa M, Peragine A, Park MY, & Poethig RS (2005) A pathway for the biogenesis of trans-acting siRNAs in Arabidopsis.*Genes Dev* 19(18):2164-2175.

31. Kamath RS*, et al.* (2003) Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. *Nature* 421(6920):231-237.

CHAPER 5

CDC5, a DNA binding protein, positively regulates posttranscriptional processing and/or transcription of primary microRNA transcritpts

PNAS. (2013) Vol. 110(43):17588-17593

Shuxin Zhang, Meng Xie, Guodong Ren and Bin Yu

Abstract

CDC5 is a MYB-related protein that exists in plants, animals and fungi. In *Arabidopsis*, CDC5 regulates both growth and immunity through unknown mechanisms. Here, we show that CDC5 from *Arabidopsis* positively regulates the accumulation of miRNAs that control many biological processes including development and adaptations to environments in plants. CDC5 interacts with both the promoters of genes encoding miRNAs (*MIR*) and the DNA-dependent RNA polymerase II (Pol II) and positively regulates *MIR* transcription and the occupancy of Pol II at *MIR* promoters. In addition, CDC5 interacts with DCL1, which generates miRNAs from their primary transcripts (pri-miRNAs), and is required for efficient pri-miRNA processing. These results demonstrate dual roles of CDC5 in miRNA biogenesis: functioning as a positive transcription factor of *MIR* and/or acting as a component of the DCL1 complex to enhance pri-miRNA processing.

Introduction

microRNAs (miRNAs) and small interfering RNAs (siRNAs) are ~ 22-nucleotide (nt) non-coding RNAs that regulate various biological processes including development, metabolism and immunity in plants and animals (1-3). miRNAs and siRNAs are generated from primary miRNA transcripts (pri-miRNAs) containing stem-loop structures and long perfect double-stranded RNAs (dsRNAs), respectively (1-3). They are associated with members of the Argonaute protein family to repress gene expression at posttranscriptional and/or transcriptional levels (1-3). Beyond miRNAs, plants encode

two major classes of siRNAs, siRNAs derived from repeated DNAs (ra-siRNAs) and trans-acting siRNAs (ta-siRNAs) (4-6).

Studies in *Arabidopsis* have established the framework of miRNA biogenesis in plants (1-3).

In *Arabidopsis*, pri-miRNAs are majorly transcribed by DNA-dependent RNA polymerase II (Pol II) with assistances of the mediator complex and the transcription factor Negative on TATA less2 (NOT2; 7, 8). After transcription, pri-miRNAs are processed by an RNAase III enzyme called DICER-LIKE1 (DCL1) to miRNA precursors (pre-miRNAs) and then to mature miRNAs (9, 10). The efficient processing of pri-miRNA by DCL1 requires SERRATE (SE; a zinc finger protein), TOUGH (TGH; an RNA binding protein) and a dephosphorylated HYPONASTIC LEAVES1 (HYL1; a double-stranded RNA binding protein) that form a complex with DCL1 (11-18). SE and HYL1 also promote the processing accuracy of pri-miRNAs (19). Four other proteins, DAWDLE (DDL; an RNA binding protein), Cap-Binding Protein 20 (CBP20), CBP80 and NOT2, which are associated with the DCL1 complex (8, 20-22), also function in miRNA biogenesis. Recent studies also suggest that the correct localization of DCL1 requires NOT2 and MODIFIER OF SNC1, 2 (MOS2; an RNA binding protein) (8, 23). In addition, the accumulation of a subset of miRNAs requires a proline rich protein named SICKLE (SIC) (24).

The Cell Division Cycle 5 (CDC5) protein is a conserved protein that exists in animal, plants and fungi (25). It was first isolated from *Schizosaccharomyces pombe* as a cell

cycle regulator. Because CDC5 contains homolog sequences to MYB transcription factor and binds DNA *in vitro* (26-28), it is thought to function as a putative transcription factor. In human and yeast, CDC5 has been shown to act as a component of spliceosome to participate in mRNA splicing (29, 30). In *Arabidopsis*, CDC5 binds DNA and is required for normal plant development and plant immunity to bacteria infection (31, 32). However, how CDC5 functions in *Arabidopsis* is unclear.

Here, we show that CDC5 plays important roles in the biogenesis of miRNAs and siRNAs in *Arabidopsis*. CDC5 interacts with both Pol II and the promoters of genes encoding miRNAs (*MIR*). Consequently, impairment of CDC5 reduces the *MIR* promoter activity and the occupancy of Pol II in the *MIR* promoter. In addition, CDC5 is associated with the DCL1 complex and is required for efficient pri-miRNA processing. Based on these results, we conclude that CDC5 positively regulates processing and/or transcription of pri-miRNAs.

Results

CDC5 is required for the accumulation of miRNAs and siRNAs

In *cdc5-1*, a T-DNA insertion disrupts the expression of *CDC5*, resulting in multiple developmental defects such as smaller plant size, altered leaf shape, later flowering and sterility (31, 32). We reasoned that *cdc5-1* might impair miRNA accumulation since the alteration in miRNA levels often causes pleiotropic developmental defects (33, 34). We thus performed northern blot analysis to examine miRNA abundance in inflorescences of

cdc5-1 and Columbia-0 (Col; wild-type control). The levels of all 9 examined miRNAs (miR166/165, miR167, miR159/319, miR390, miR171, miR172, miR173, miR156 and miR163) were reduced in *cdc5-1* when compared to those in Col (Figure 5-1A and Figure 5-7A). A *CDC5-YFP* transgene driven by the *CDC5* promoter (*pCDC5::CDC5-YFP*) fully restored miRNA levels (Figure 5-8A), demonstrating that *cdc5-1* is responsible for the reduction of miRNA abundance. In addition, *cdc5-1* exhibited a similar effect on levels of several examined miRNAs in leaves as in inflorescences (Figure 5-1B and Figure 5-7B). We also tested the effect of *cdc5-1* on the accumulation of endogenous siRNAs. The levels of all examined siRNAs including two trans-acting siRNAs (ta-siRNAs), TAS1-siR255 and TAS2-siR1511 and siRNAs derived from repetitive DNAs (rasiRNAs), siR02, siR1003, cluster 4, IR71 and TR2588 were lower in *cdc5-1* than in Col (Figure 5-1C and Figure 5-7C).

We next examined the effects of *cdc5-1* on miRNA and ta-siRNA function by analyzing the expression levels of miRNA targets using quantitative RT-PCR (qRT-PCR). The transcript levels of several targets of miRNAs or ta-siRNAs (*ARF8*, *CUC1*, *MYB65*, *PPR*, *SPL6*, *SPL10* and *ARF3*) were increased in *cdc5-1* relative to Col (figure 8B). However, it is possible that *cdc5-1* has more impacts on some other targets.

CDC5 regulates the transcription of genes encoding miRNAs (*MIR*)

We next performed qRT-PCR to examine the levels of seven pri-miRNAs (pri-miR158a, pri-miR159a, pri-miR167a, pri-miR171a, pri-miR172a, pri-miR172b and pri-miR173) in Col and *cdc5-1*. The levels of examined pri-miRNAs were decreased in *cdc5-1* relative

to Col (figure 5-2A). The reduced levels of pri-miRNAs and miRNAs in *cdc5-1* can result from impaired transcription and/or posttranscriptional processing of pri-miRNAs. Alternatively, CDC5 may act after miRNA maturation. We first determined whether CDC5 regulates *MIR* transcription by examining the effect of *cdc5-1* on the expression of a *GUS* reporter gene driven by *MIR172b* promoter (*pMIR172b::GUS*) (20). We have used this system to determine the function of DDL in regulating *MIR* transcription (20). If CDC5 is indeed a positive transcription regulator of *MIR*, *cdc5-1* will negatively affect the expression of *GUS*. We crossed *cdc5-1* with a Col transgenic line, which contains the *pMIR172b::GUS* transgene (20). In F2 generation, we obtained $CDC5^{+}$ (*CDC5/CDC5* or *CDC5/cdc5*) and *cdc5-1* genotypes containing *pMIR172b::GUS*. GUS staining on these plants revealed that the GUS activity was lower in *cdc5-1* than in $CDC5^{+}$ (figure 5-2B). qRT-PCR analysis confirmed that *GUS* mRNA levels in *cdc5-1* were reduced relative to those in $CDC5^{+}$(figure 5-2C).

CDC5 is required for Pol II occupancy at the promoter of *MIR*

To confirm that CDC5 is a positive transcription factor of *MIR*, we monitored the occupancy of RNA polymerase II (Pol II) at promoters of *MIR166a*, *MIR167a, MIR171a* and *MIR172b* in *cdc5-1* and Col by chromatin immunoprecipitation (ChIP) using an antibody against the second largest subunit of Pol II (RPB2) as described by Kim et al (7). We included a "no-antibody" ChIP as a negative control. After ChIP, the *MIR166a*, *MIR167a*, *MIR171a* and *MIR172b* promoter fragments were examined by qPCR. Like previously reported (7), the promoter regions of these four *MIR*s but not Pol II C1 (a genomic fragment between At2g17470 and At2g17460; 7) were enriched in RPB2

immunoprecipitates relative to " no antibody" control in Col. *cdc5-1* reduced the occupancy of Pol II at these regions relative to Col (Figure 5-3A and 3B). We also examined whether *cdc5-1* affected the occupancy of Pol II at DCL1 promoter. The result showed that the association of Pol II with DCL1 promoter was not significantly changed (figure 5-9A). These data further supported that CDC5 positively regulates *MIR* transcription in *Arabidopsis*.

CDC5 interacts with *MIR* promoters

To understand how CDC5 regulates the transcription of *MIR*, we examined whether CDC5 binds the promoter of *MIRs* since CDC5 is a putative *MYB* domain-containing transcription factor and has a DNA binding activity (27). We performed ChIP using an antibody against YFP on *cdc5-1* complementation line containing *pCDC5::CDC5-YFP* (figure 5-7A) and Col. qPCR analysis showed that *MIR166a*, *MIR167a*, *MIR171a* and *MIR172b* promoter fragments were enriched in CDC5-YFP immunoprecipitates but not in Col and "no-antibody" controls (Figure 5-3C and 5-3D). In addition, CDC5 did not bind the promoter of DCL1 (Figure 5-9B). These results suggested that CDC5 is associated with *MIR* promoters.

CDC5 interacts with Pol II

The association of CDC5 with *MIR* promoters and the reduced Pol II occupancy in *MIR* promoters in *cdc5-1* suggest that CDC5 may positively regulates *MIR* transcription by promoting the recruitment of Pol II to their promoters, which predicts a potential CDC5-Pol II interaction. Thus, we tested the association of CDC5 with Pol II through reciprocal

co-immunoprecipitation (co-IP). We extracted proteins from *cdc5-1* complementation line expressing *pCDC5::CDC5-YFP* and Col control expressing a *YFP* transgene. IP was performed with either anti-YFP antibody or anti-RPB2 antibody. Western blots detected RPB2 in the CDC5-YFP immunoprecipitates and CDC5-YFP in the RPB2 immunoprecipitates, respectively (Figure 5-4 A and 5-4B). In contrast, the interaction between YFP and RPB2 were not detected. In addition, protein G beads without antibody failed to pull down either CDC5-YFP or RPB2. These results suggested a CDC5-Pol II association. Both CDC5 and Pol II bind DNAs, suggesting that the CDC5-Pol II interaction may depend on DNA. However, DNAse I treatment during IP had no obvious effect on CDC5-Pol II interaction (Figure 5-4C). The CDC5-Pol II interaction suggested that CDC5 might pull down Pol II-associated promoters. However, we did not observe the occupancy of CDC5 at DCL1 promoter, indicating that the Pol II amount in CDC5 immunoprecipitates may be tiny such that the DCL1 promoter is not detectable in the CDC5 immunoprecipitates.

CDC5 is required for efficient pri-miRNA processing

We next asked whether CDC5 has a role in pri-miRNA processing by examining the effect of *cdc5-1* on the processing of pri-miR162b using an *in vitro* assay (13, 35). A radioactive labeled pri-miR162b probe (*MIR162b*; predicted stem-loop of miR162b with 6-nt arms at each end; figure 5-5A; 13) was first generated by *in vitro* transcription under the presence of [α-^{32}P] UTP. Radioactive labeled *MIR162b* was then incubated with protein extracts from young flower buds of *cdc5-1* and Col, respectively. After reactions were stopped at 50, 100 and 150 min, RNAs were extracted and resolved on a denaturing

polyacrylamide gel. The protein extracts of *cdc5-1* generated less miR162b than that of Col (Figure 5-5B). Quantitative analysis at 100 min time point showed that the DCL1 activity in *cdc5-1* was ~ 50% of that in Col (Figure 5-5C). These results suggested that CDC5 positively contributes to the DCL1 activity.

CDC5 is associated with the DCL1 complex

There are at least two possible ways by which CDC5 contributes to the DCL1 activity. It may positively regulate the transcription of other genes involved in miRNA biogenesis or act as a component of the DCL1 complex. To clarify these possibilities, we first examined the transcript levels of several known genes involved in miRNA biogenesis including *CBP80*, *CBP20*, *DDL*, *HYL1*, *DCL1*, *HEN1* and *SE* by qRT-PCR. The expression levels of these genes were slightly increased in *cdc5-1* relative to Col (Figure 5-10A). Western blot analysis showed that the protein levels of DCL1 and HYL1 were comparable in *cdc5-1* with those in Col (Figure 5-10B and 10C).

Next we tested the interaction of CDC5 with the DCL1 complex through a bimolecular fluorescence complementation (BiFC) assay. We have used this assay to determine the association of TGH with the DCL1 complex (13). The protein partners were fused to the N-terminal fragment of Venus (nVenus) or C-terminal fragment of cyan fluorescent protein (cCFP) under the control a Cauliflower mosaic virus 35S promoter and co-introduced into *Nicotiana benthamiana* (*N. benthamiana)*. In this assay, generation of a functional yellow fluorescent protein (YFP) indicates the potential interaction between proteins (36). The CDC5-DCL1, CDC5-SE and SE-DCL1 (positive control) but not

AGO1-CDC5 (negative control) interactions were observed (Figure 5-6A). In addition, weak YFP signals were produced from the CDC5-HYL1 pair, indicating a weak or no interaction between CDC5 and HYL1 (Figure 5-6A).

We performed co-immunoprecipitation (co-IP) assay to confirm the BiFC results. The DCL1-YFP fusion protein and YFP were expressed in *N. benthamiana*, respectively, whereas recombinant CDC5 fused with a maltose-binding protein epitope tag at its N-terminus (MBP-CDC5) and MBP were expressed in *E.coli* BL21 (13). Then anti-YFP antibody conjugated with protein G agarose beads was incubated with the protein mixture containing MBP-CDC5 / DCL-YFP, MBP-CDC5/YFP or MBP/DCL1-YFP to capture the DCL1-YFP or YFP complex. We were able to detect MBP-CDC5 but not MBP in the DCL1-YFP complex (Figure 5-6C). In contrast, YFP did not pull-down either MBP or MBP-CDC5 (Figure 5-6C). In addition, RNase A treatment did not impair the CDC5-DCL1 interaction although it abolished an RNA-mediated AGO4-FDM1 interaction (Figure 5-6C and Figure 5-10D). These results indicated that the CDC5-DCL1 interaction maybe not RNA-mediated.

We further determined the protein domains of DCL1 that mediate the DCL1-CDC5 interaction. Five different DCL1 fragments named F1 (aa1-468 covering amino terminus to helicase domain 1), F2 (aa465-840; helicase domain 2), F3 (aa835-1330; domain of unknown Function and PAZ domain:), F4 (aa1328-1700; RNaseIIIa+IIIb domains), and F5 (aa1729-1909; dsRNA binding domains I+II) were expressed in *Nicotiana benthamiana*, respectively, as described (37; Figure 5-6B). CDC5-YFP was able to pull

down F2 (Helicase domain 2) and F5 (dsRNA binding domains I+II) but not other fragments (Figure 5-6D).

We next examined the interactions of CDC5 with SE and HYL1. CDC5 and SE but not CDC5 and HYL1 were able to pull down each other, which was not affected by RNAse A treatment (Figure 5-6E and 5-6F). In addition, the interactions among controls were not detected (Figure 5-6E and 5-6F). The interaction of CDC5 with SE and DCL1 suggested that CDC5 is a component of DCL1 complex. However, we did not detect the HYL1-CDC5 interaction (Figure 5-6F). This was not unexpected as CDC5 may be weakly associated with the DCL1 complex or its association with HYL1 may need bridge proteins. In fact, NOT2 has been shown to interact with DCL1 and SE but not HYL1 (8).

Discussion

In conclusion, we show that CDC5, a MYB-related and evolutionarily conserved protein, is an important player in miRNA biogenesis. This is evidenced by reduced transcript levels and processing efficiency of pri-miRNAs and less accumulation of miRNAs in *cdc5-1*. Impairment of CDC5 function causes both immunity and pleiotropic development defects, which agrees with the crucial roles of miRNAs in regulating multiple biological processes (31, 32). However, it is possible that the regulation of genes other than small RNAs by CDC5 also contributes to the observed phenotypes of *cdc5-1*.

Based on studies of CDC5 homologs in other organisms, the roles of plant CDC5 in transcription have been speculated (31, 32). This study provides direct evidences to

support that CDC5 is a positive transcription factor. The facts that CDC5 does not bind the DCL1 promoter and that *cdc5-1* does not significantly affect the occupancy of Pol II at the DCL1 promoter suggest that CDC5 maybe not a general transcription factor. Rather, it may affect the expression of a subset of genes. CDC5 interacts with Pol II, suggesting the occupancy of CDC5 at promoters may depend on Pol II. However, CDC5 is a DNA binding protein (27) and does not interact with DCL1 promoter, supporting that CDC5 may directly bind *MIR* promoters. *cdc5-1* reduces *MIR* promoter activity and the occupancy of Pol II at *MIR* promoters, suggesting that CDC5 may have a direct role in promoting the transcription of *MIR* by recruiting Pol II to their promoters. It is possible that CDC5 also contributes to Pol II activity through its interaction with Pol II. However, *cdc5-1* does not significantly affect DCL1 transcript levels as well as the occupancy of Pol II at its promoter, suggesting that the CDC5-Pol II interaction by itself maybe not sufficient to regulate the Pol II activity. Whether the CDC5-Pol II interaction is required for the regulation of *MIR* transcription needs to be further investigated.

CDC5 also has a role in promoting miRNA maturation. This is unlikely to be caused by the reduced transcription of key genes involved in miRNA biogenesis since their transcript levels are slightly increased in *cdc5-1*. Rather, CDC5 may act as a component of the DCL1 complex to enhance pri-miRNA processing efficiency based on the association of CDC5 with the DCL1 complex and the fact that *cdc5-1* reduces the processing efficiency of pri-miR162b *in vitro*. CDC5 interacts with the helicase and dsRNA binding domains of DCL1, which regulate the DCL1 activity (10, 38). Structure studies have revealed that the interaction of human dicer with other proteins can cause

dicer conformational change and therefore improve its activity (39). Thus, it is possible that CDC5 may regulate DCL1 activity through its interaction with DCL1.

In summary, our study reveals that CDC5 can positively regulate processing and transcription of pri-miRNAs. CDC5 unlikely regulates the transcription of all *MIRs* since it maybe not a general transcription factor. Thus, CDC5 may only regulate some pri-miRNAs at both transcriptional and posttranscriptional levels. However, CDC5 may have a general role in regulating pri-miRNA processing since it acts as co-factor of DCL1. In addition, CDC5 is majorly expressed in the proliferating cells (32), suggesting that CDC5 may have cell specific activities on miRNA accumulation. CDC5 is also required for the accumulation of ra-siRNAs and ta-siRNAs. It is unclear whether CDC5 has a direct role in ta-siRNA biogenesis as the generation of ta-siRNAs requires miRNAs. Based on the function of CDC5 in the miRNA pathway, CDC5 may have two contributions, which are not mutually exclusive, to the production of ra-siRNAs. First, it may affect Pol IV activity that is thought to produce the precursor RNAs of ra-siRNAs. Second, it may regulate the DCL3 activity that generates 24 nt ra-siRNAs from long dsRNAs. Clearly, these two possibilities need to be examined in the near future.

Materials and Methods

Plant Materials

The *cdc5-1* (SAIL_207_F03) that is in Columbia genetic background was obtained from *Arabidopsis* Biological Resources Center (ABRC) (31,32). Transgenic line harboring

pMIR172b::GUS (20) was crossed to *cdc5-1.* In F2 generation, *CDC5*$^{+}$ (*CDC5/CDC5* and *CDC5/cdc5-1*) and *cdc5-1* containing *pMIR172b::GUS* were identified by genotyping of *cdc5-1* and *GUS.*

RNA Analysis

Northern Blot analysis of small RNAs and qRT- PCR analysis of pri-miRNA and miRNA targets transcription levels were performed as described (13).

Plasmid Construction

A ~ 5.2 Kb genomic DNA covering *CDC5* coding region and promoter from Col genome was amplified by PCR and cloned to pMDC204 to generate the *pCDC5::CDC5-YFP* construct. A full-length *CDC5* cDNA was amplified by RT-PCR and ligated to pMAL-c5x (NEB) to produce the *MBP-CDC5* plasmid. *CDC5* cDNA was cloned into pSAT4-C-CFP. The *CDC5-C-CFP* fragment was then released by I-SceI restriction enzyme digestion and subsequently cloned into the pPZP-ocs-bar-RCS2-2 vector. SE cDNA was amplified by RT-PCR and cloned into pEarleyGate203 vector to generate the SE-MYC construct. The truncated DCL1 (F1 to F5)-MYC plasmids were obtained from the laboratory of Dr. Y. Adam Yuan at National University of Singapore (12).

Plant complementation

The *pCDC5::CDC5-YFP* plasmid was transformed into *CDC5/cdc5-1*. The transgenic plants were selected using Hygromycin resistance. In T2 generation, *cdc5-1* harboring *pCDC5::CDC5-YFP* was identified by genotyping of YFP and *cdc5-1*.

Chromatin immunoprecipitation (ChIP) assay

ChIP was performed as described by Kim et al (7). Three biological replicates were performed. Anti-RPB2 and anti-GFP and GFP variants antibodies (Clontech) were used for immunoprecipitation. Quantitative PCR (qPCR) were performed on DNAs co-purified with Pol II or CDC5.

Co-Immunoprecipitation (Co-IP) assay

For Pol II-CDC5 co-IP, protein extracts from plants expressing *pCDC5::CDC5-YFP* or *YFP* were incubated with anti-GFP (and GFP variants; Clontech) antibodies or anti-RBP2 coupled to protein G-agarose beads for 4 hours at 4 °C. After five-time washing, the proteins in the immunoprecipitates were subjected to western blot analysis using anti-GFP antibody and anti-RBP2 antibody, respectively. For the interactions of CDC5 with components of DCL1 complex, MBP-CDC5 and MBP were expressed in BL21 and extracted followed the manufacturer's protocol (New England Biolabs; NEB) while DCL1-YFP, truncated DCL1-MYC (F1 to F5), SE-MYC and YFP alone were expressed in *N. benthamiana* (20). HYL1 and CDC5-YFP were obtained from inflorescences of Col and plants expressing *pCDC5::CDC5-YFP*, respectively. Anti-GFP (and GFP variants) and anti-MYC antibodies were used to capture and detect corresponsive YFP and MYC tagged proteins, respectively. Anti-HYL1 and anti-MBP antibodies (NEB) were used to detect HYL1 and MBP-tagged proteins, respectively, in western blot.

Dicer Activity Assay

Pri-miR162b was prepared by *in vitro* transcription under the presence of [α-^{32}P] UTP. *In vitro* dicer activity assay was performed according to Qi et al and Ren et al (13, 35). Radioactive signals were quantified with ImageQuant version 5.2.

BiFC Assay

Paired cCFP and nVenus constructs were co-infiltrated into *N. benthamiana* leaves. After 48 hours, yellow fluorescence signals and Chlorophyll auto fluorescence signals were exited at 488 nm and detected by confocal microscopy (Fluoview 500 workstation; Olympus) with a narrow barrier filter (BA505–525 nm).

Figures

Figure 5-1. *cdc5-1* reduces the accumulation of miRNAs and siRNAs. (A) miRNA abundance in inflorescences of *cdc5-1* and Columbia (Col) . **(B)** miRNA abundance in leaves of *cdc5-1* and Col. **(C)** siRNA abundance in inflorescences *cdc5-1* and Col. Col: wild-type control of *cdc5-1*. Small RNAs were detected by Northern Blot. After Northern blot, the radioactive signals were detected with phosphor imager and quantified with ImageQuant (V5.2). To determine relative abundance of small RNAs in *cdc5-1*, the amount of a miRNA or siRNA in *cdc5-1* was normalized to U6 RNA and compared with that in Col. The value of miRNAs or siRNAs in Col was set as 1. The number below *cdc5-1* indicated the relative abundance of miRNAs or siRNAs, which is the average value of three repeats ($P<0.05$; except for siR255 in Figure 5-1C; t-test). For miR159/319: upper band, miR159; lower band, miR319.

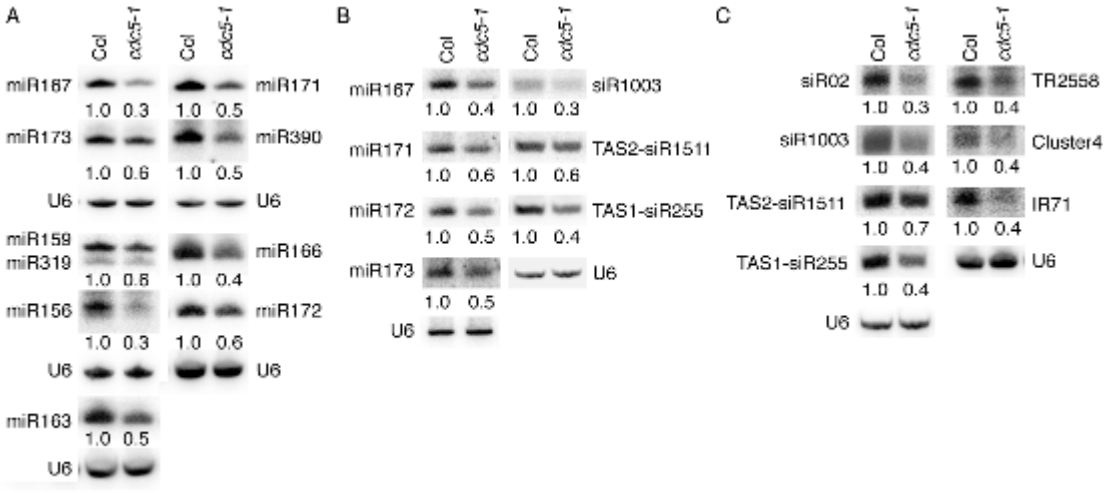


Figure 5-2. *cdc5-1* reduces the promoter activity of genes encoding miRNAs (*MIR*). **(A)** The transcript levels of various pri-miRNAs in inflorescences of *cdc5-1* and Col determined by quantitative RT-PCR (qRT-PCR). The abundance of pri-miRNAs in *cdc5-1* was normalized to that of *UBQUITIN5* (*UBQ5*), and compared with that in Col. Value of Col was set to 1. Standard deviation of three technical replications was shown as error bars. (B) The levels of GUS in *CDC5*$^{+}$ and *cdc5-1* harboring *MIR172b*::GUS . *CDC5*$^{+}$*: CDC5/CDC5 or CDC5/cdc5-1.* Twenty plants containing *GUS* were analyzed for each of *CDC5*$^{+}$ and *cdc5-1* genotypes. An image for each genotype is shown. (C) The transcript levels of *GUS* driven by *MIR172b* promoter in *CDC5*$^{+}$ and *cdc5-1*. *GUS* transcript levels were determined by qRT-PCR. The *GUS* mRNA levels in *cdc5-1* were normalized to *UBQ5* and compared with those in *CDC5*$^{+}$. *:P<0.05; **:P<0.01; ***:P<0.001 (t-test).

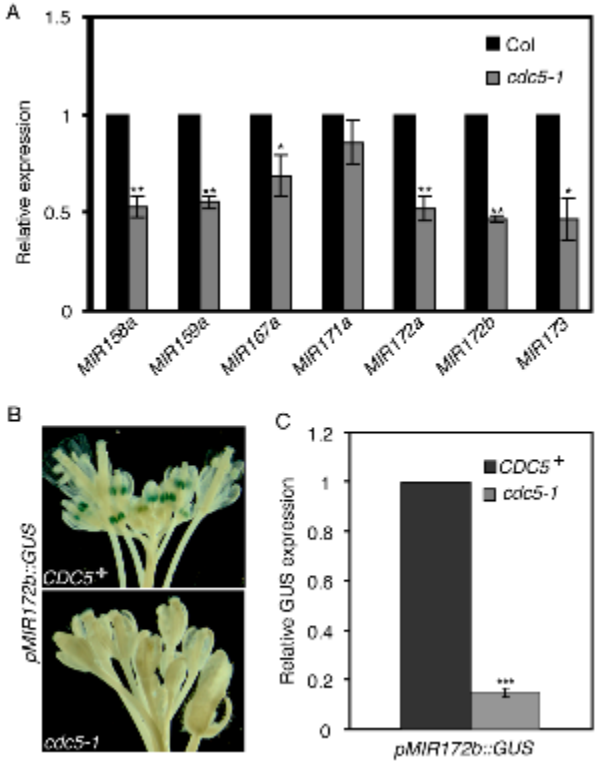


Figure 5-3. CDC5 is required for the recruitment of Pol II to *MIR* promoters. (A) and (B) The occupancy of Pol II at various *MIR* promoters detected by ChIP using anti-RBP2 antibody in *cdc5-1* and Col. **(C)** and **(D)** The association of CDC5 with various *MIR* promoter detected by ChIP using anti-YFP antibody in plants containing *pCDC5::CDC5-YFP*. DNAs co-purified with CDC5 or Pol II were analyzed with qRT-PCR. The intergenic region between At2g17470 and At2g17460 (Pol II C1) that is not occupied by Pol II was used as a negative control. ChIP with no antibodies was performed as another control. Means and standard derivations of three technical repeats are presented and three biological replicates gave similar results. Please note that the results of Pol II C1 in RBP2 ChIP (A, and B) and in CDC5 ChIP (C and D) were showed twice, respectively, for control purpose. *:P<0.05 (t-test).

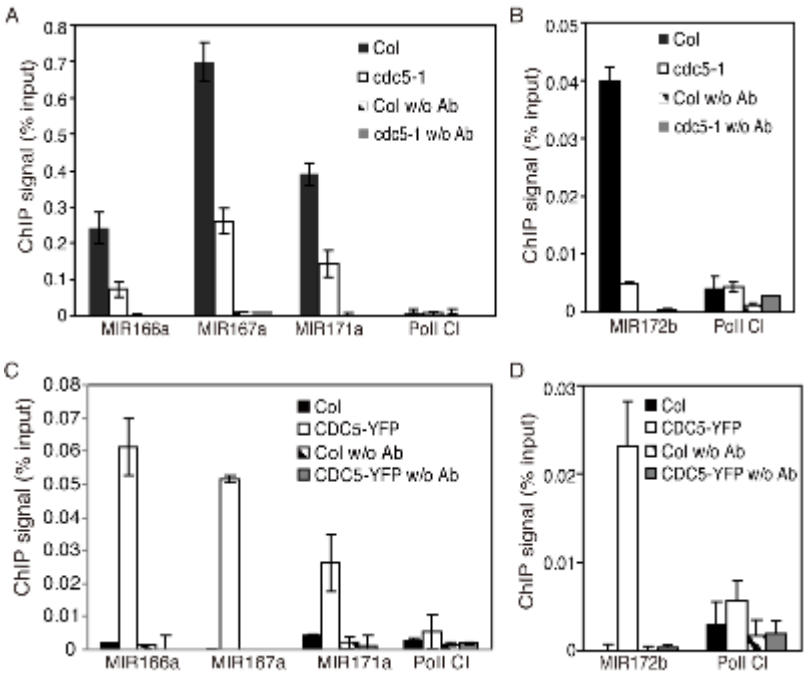


Figure 5-4. CDC5 interacts with Pol II. (A) and **(B)** Co-immunoprecipitation (Co-IP) between CDC5-YFP and Pol II. **(C)** Co-IP between CDC5-YFP and Pol II is DNA independent. Proteins extracts isolated from inflorescences of plants containing CDC5-YFP or YFP were used to perform IP using Anti-YFP or Anti-RBP2. The proteins in the extracts were indicated on top of the picture. YFP, CDC5-YFP and RBP2 were detected by western blot using anti-YFP antibody and anti-RPB2, respectively, and labeled on the left side of the picture. Two percent of input proteins were used for RPB2 while twenty percent input proteins were used for YFP and DCL1-YFP, respectively.

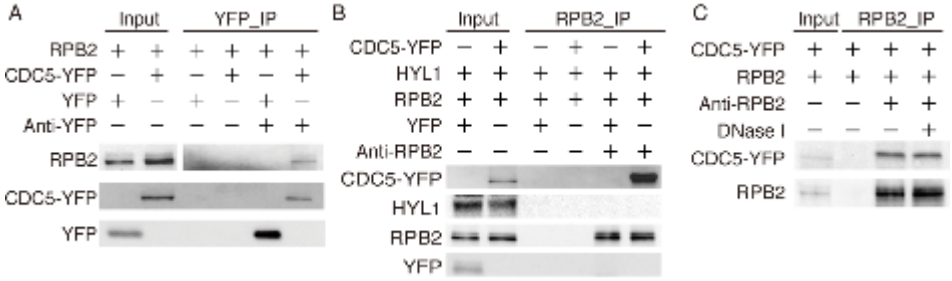


Figure 5-5. *cdc5-1* reduces the DCL1 activity. (A) Schematic diagram of the pri-miR162b used *in vitro* processing assay. **(B)** Pri-miR162b processing by protein extracts from *cdc5-1* and Col. After reaction, RNAs were extracted, resolved on PAGE gel and detected with a phosphor imager. **(C)** Quantification of miR162 production in *cdc5-1* relative to Col. The Quantitative analysis was performed for the reaction stopped at 100 min as shown in (B). The radioactive signal of miR162 was quantified with an ImageQuant software (V5.2) and then normalized to input to determine the amount of miR162 produced by *cdc5-1* or Col protein extracts ($miR162_{cdc5\text{-}1}$ or $miR162_{Col}$). The relative level of miR162 produced by *cdc5-1* was calculated as $miR162_{cdc5\text{-}1}$ divided by $miR162_{Col}$. The value of $miR162_{Col}$ was set as 1. The value represents mean of three repeats (*** $P < 0.001$; t-test).

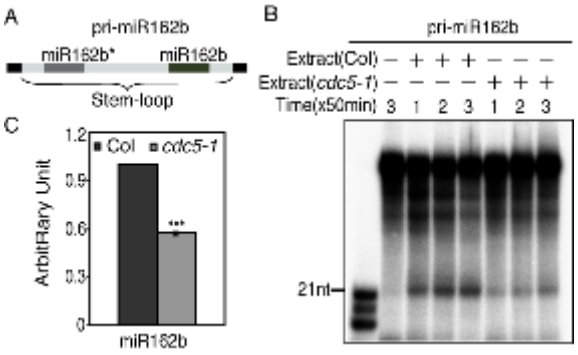


Figure 5-6. CDC5 interacts with the DCL1 complex. (A) BiFC analysis of CDC5 with DCL1, SE, HYL1 and AGO1. Respective pairs of cCFP (cCFP-CDC5, cCFP-SE) and nVenus (nVenus-DCL1, nVenus-HYL1, nVenus-SE and nVenus-AGO1) fused proteins were co-infiltrated into *N. benthamiana* leaves. Yellow fluorescence (green in image) signals were examined at 48h after infiltration by confocal microscopy. Arrow indicates the BiFC signal. The red spot was inflorescence from chlorophyll. 30 nuclei were examined for each pair and an image is shown. **(B)** Schematic diagram of DCL1 domains and truncated DCL1 fragments used for protein interaction assay. **(C)** Co-immunoprecipitation between CDC5 and DCL1. The protein pairs in the protein extracts were indicated by the labels on the left side of and on top of the picture. DCL1-YFP/YFP and MBP-CDC5/MBP were detected by western blot using anti-YFP and anti-MBP, respectively, and labeled on the left side of the picture. One percent input proteins were used for MBP-CDC5 and MBP. Twenty percent input proteins were used for DCL1-YFP and YFP, respectively. **(D)** Co-immunoprecipitation between CDC5 with the helicase and dsRNA binding domains of DCL1. Truncated DCL1 proteins fused with a myc tag at their N-terminus were expressed in *N. benthamiana* leaves. The protein pairs in the protein extracts were indicated by the labels on the left side of and on top of the picture. Anti-myc antibody was used to detect myc fusion proteins in western blots. Labels on left side of picture indicate proteins detected by western blot. Five percent input proteins were used for MYC tagged proteins while twenty percent inputs were used for DCL1-YFP and YFP, respectively. Please note only an IP picture was shown for CDC5-YFP and YFP, respectively. **(E)** and **(F)** Co-immunoprecipitation between CDC5 and SERRATE (SE). The protein pairs in the protein extracts were indicated by the labels on

the left side of and on tope of the picture. Proteins detected by western blot were indicated on the left side of the picture. Two percent of input proteins were used for SE-MYC. Twenty percent inputs proteins were used for MBP and YFP tagged proteins, respectively. Please note only an IP picture was shown for CDC5-YFP and YFP, respectively.

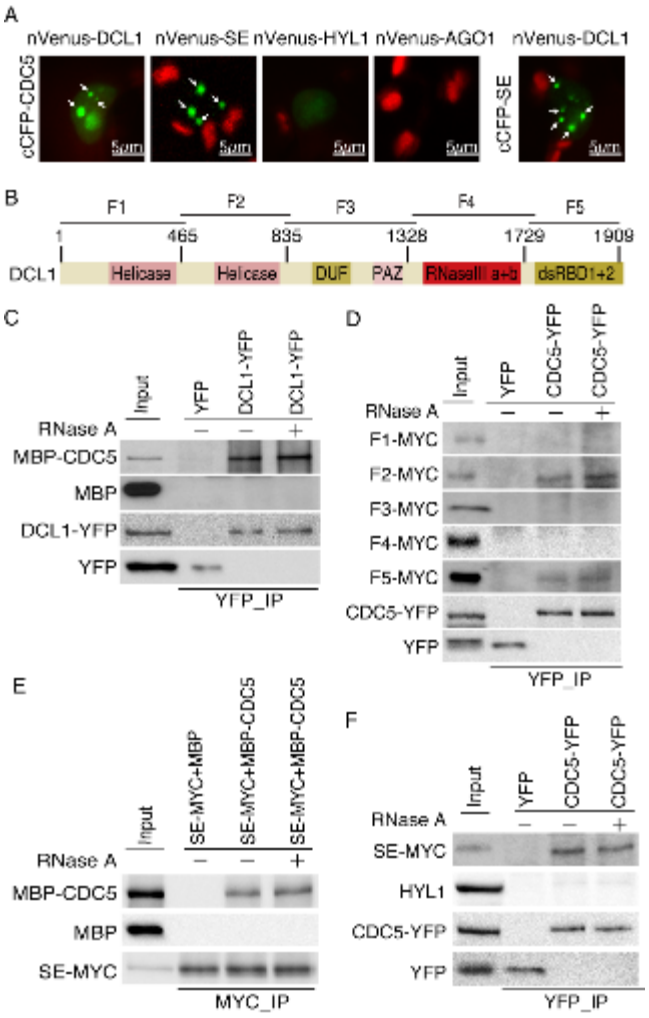


Figure 5-7. Quantification of miRNA and siRNA abundance. (A) miRNA abundance in inflorescences of *cdc5-1* and Col. **(B)** miRNA abundance in leaves of *cdc5-1* and Col. **(C)** siRNA abundance in inflorescences *cdc5-1* and Col. The amount of miRNAs or siRNAs in *cdc5-1* was quantified with ImageQuant (V5.2) was normalized to U6 RNA and compared with that in Col (normalized to U6 as well). The value represents mean of three repeats. t-test was used for comparison. *:P<0.05. **:P<0.01.

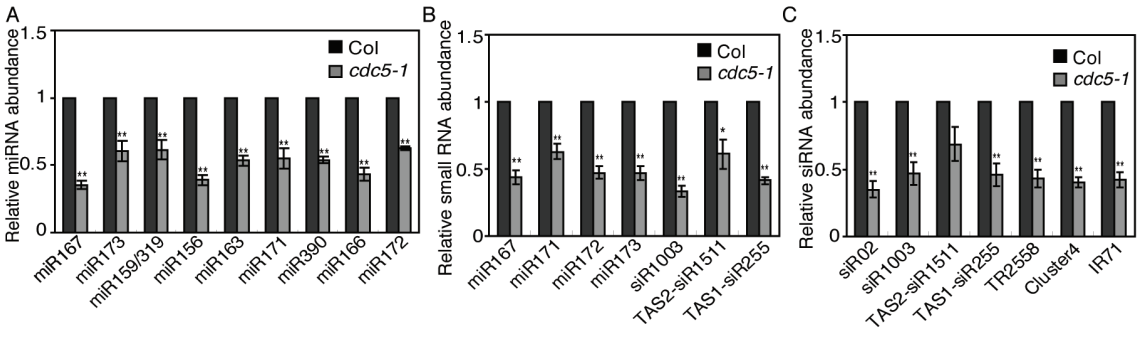


Figure 5-8. The effects of *cdc5-1* on the accumulation of miRNAs and target transcripts. **(A)** CDC5 recovers the miRNA abundance in *cdc5-1*. U6 RNA was probed for loading control. Number represents the relative abundance of miRNAs in Col (wild-type control), *cdc5-1* and two complementation lines (*cdc5-1+CDC5*). **(B)** *cdc5-1* increases the transcript levels of miRNA and ta-siRNA targets . The levels of target transcripts in *cdc5-1* were normalized with *UBQUITIN5* (*UBQ5*) and compared with those in Col. Value of Col is 1. Standard deviations of three technical replications are shown as error bars. A similar result was produced with an additional biological replicate. *:P<0.05; **:P<0.01.

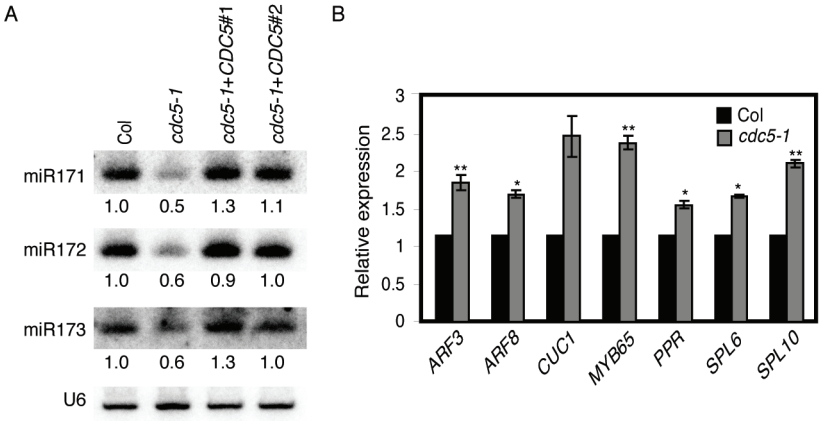


Figure 5-9. The occupancy of Pol II and CDC5 at *DCL1* promoter. (A) The occupancy of Pol II at *DCL1* promoter detected by ChIP using anti-RBP2 antibody in *cdc5-1* and Col. **(B)** The occupancy of CDC5 at *DCL1* promoter detected by ChIP using anti-YFP antibody in plants containing *pCDC5::CDC5-YFP*. DNAs co-purified with CDC5 or Pol II were analyzed with qPCR. Means and standard derivations of three technical repeats are presented. t-test was used for statistic analysis.

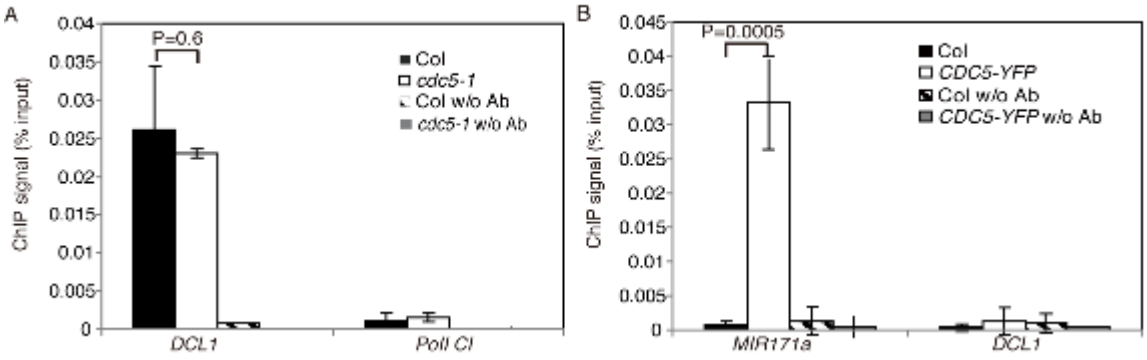
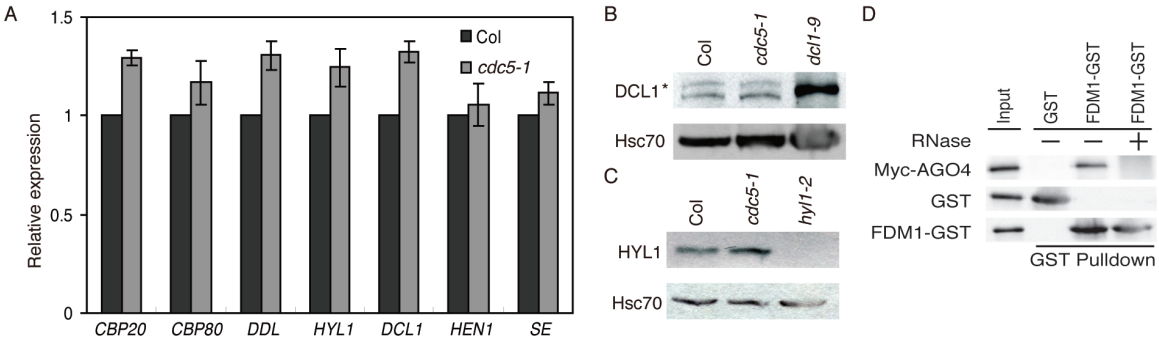


Figure 5-10. The effects of *cdc5-1* on the expression of several genes involved in miRNA biogenesis. (A) Transcript levels of several genes involved in miRNA biogenesis determined by qRT-PCR in *cdc5-1* and Col. *UBQ5* was used as a reference control. Error bars represent standard deviation of three technical replications. The experiment was repeated once with similar results. **(B)** DCL1 and **(C)** HYL1 protein levels detected by western blot in *cdc5-1* and Col. *dcl1-9* containing a truncated DCL1 protein and *hyl1-2* lacking of HYL1 were used as controls. **(D)** RNase A treatment abolished the AGO4-FDM1 interaction. Proteins extracts containing myc-AGO4/GST or myc-AGO4/GST-FDM1 incubated with glutathione beads to capture GST or GST-FDM1 complex. After pull down, proteins were detected by western blot. The proteins detected by western blot were labeled left side of the picture.



References:

1. Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116(2):281-297.
2. Chen X (2005) MicroRNA biogenesis and function in plants. *FEBS letters* 579(26):5923-5931.
3. Voinnet O (2009) Origin, biogenesis, and activity of plant microRNAs. *Cell* 136(4):669-687.
4. Vazquez F (2006) Arabidopsis endogenous small RNAs: highways and byways. *Trends Plant Sci* 11(9):460-468.
5. Brodersen P & Voinnet O (2006) The diversity of RNA silencing pathways in plants. *Trends Genet* 22(5):268-280.
6. Herr AJ & Baulcombe DC (2004) RNA silencing pathways in plants. *Cold Spring Harb Symp Quant Biol* 69:363-370.
7. Kim YJ*, et al.* (2011) The role of Mediator in small and long noncoding RNA production in Arabidopsis thaliana. *EMBO J* 30(5):814-822.
8. Wang L*, et al.* (2013) NOT2 proteins promote polymerase II-dependent transcription and interact with multiple MicroRNA biogenesis factors in Arabidopsis. *Plant Cell* 25(2):715-727.
9. Park W, Li J, Song R, Messing J, & Chen X (2002) CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in Arabidopsis thaliana. *Curr Biol* 12(17):1484-1495.
10. Kurihara Y & Watanabe Y (2004) Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. *Proc Natl Acad Sci U S A* 101(34):12753-12758.

11. Song L, Han MH, Lesicka J, & Fedoroff N (2007) Arabidopsis primary microRNA processing proteins HYL1 and DCL1 define a nuclear body distinct from the Cajal body. *Proc Natl Acad Sci U S A* 104(13):5437-5442.

12. Manavella PA*, et al.* (2012) Fast-forward genetics identifies plant CPL phosphatases as regulators of miRNA processing factor HYL1. *Cell* 151(4):859-870.

13. Ren G*, et al.* (2012) Regulation of miRNA abundance by RNA binding protein TOUGH in Arabidopsis. *Proc Natl Acad Sci U S A* 109(31):12817-12821.

14. Fujioka Y, Utsumi M, Ohba Y, & Watanabe Y (2007) Location of a possible miRNA processing site in SmD3/SmB nuclear bodies in Arabidopsis. *Plant Cell Physiol* 48(9):1243-1253.

15. Vazquez F, Gasciolli V, Crete P, & Vaucheret H (2004) The nuclear dsRNA binding protein HYL1 is required for microRNA accumulation and plant development, but not posttranscriptional transgene silencing. *Curr Biol* 14(4):346-351.

16. Han MH, Goud S, Song L, & Fedoroff N (2004) The Arabidopsis double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. *Proc Natl Acad Sci U S A* 101(4):1093-1098.

17. Yang L, Liu Z, Lu F, Dong A, & Huang H (2006) SERRATE is a novel nuclear regulator in primary microRNA processing in Arabidopsis. *Plant J* 47(6):841-850.

18. Lobbes D, Rallapalli G, Schmidt DD, Martin C, & Clarke J (2006) SERRATE: a new player on the plant microRNA scene. *EMBO reports* 7(10):1052-1058.

19. Dong Z, Han MH, & Fedoroff N (2008) The RNA-binding proteins HYL1 and SE promote accurate in vitro processing of pri-miRNA by DCL1. *Proc Natl Acad Sci U S A* 105(29):9970-9975.

20. Yu B*, et al.* (2008) The FHA domain proteins DAWDLE in Arabidopsis and SNIP1 in humans act in small RNA biogenesis. *Proc Natl Acad Sci U S A* 105(29):10073-10078.

21. Gregory BD*, et al.* (2008) A link between RNA metabolism and silencing affecting Arabidopsis development. *Dev Cell* 14(6):854-866.

22. Laubinger S*, et al.* (2008) Dual roles of the nuclear cap-binding complex and SERRATE in pre-mRNA splicing and microRNA processing in Arabidopsis thaliana. *Proc Natl Acad Sci U S A* 105(25):8795-8800.

23. Wu X*, et al.* (2013) A role for the RNA-binding protein MOS2 in microRNA maturation in Arabidopsis. *Cell Res* 23(5):645-657.

24. Zhan X*, et al.* (2012) Arabidopsis proline-rich protein important for development and abiotic stress tolerance is involved in microRNA biogenesis. *Proc Natl Acad Sci U S A* 109(44):18198-18203.

25. Ohi R*, et al.* (1998) Myb-related Schizosaccharomyces pombe cdc5p is structurally and functionally conserved in eukaryotes. *Mol Cell Biol* 18(7):4097-4108.

26. Ohi R*, et al.* (1994) The Schizosaccharomyces pombe cdc5+ gene encodes an essential protein with homology to c-Myb. *EMBO J* 13(2):471-483.

27. Hirayama T & Shinozaki K (1996) A cdc5+ homolog of a higher plant, Arabidopsis thaliana. *Proc Natl Acad Sci U S A* 93(23):13371-13376.

28. Bernstein HS & Coughlin SR (1998) A mammalian homolog of fission yeast Cdc5 regulates G2 progression and mitotic entry. *J Biol Chem* 273(8):4666-4671.

29. Burns CG, Ohi R, Krainer AR, & Gould KL (1999) Evidence that Myb-related CDC5 proteins are required for pre-mRNA splicing. *Proc Natl Acad Sci U S A* 96(24):13789-13794.

30. McDonald WH, Ohi R, Smelkova N, Frendewey D, & Gould KL (1999) Myb-related fission yeast cdc5p is a component of a 40S snRNP-containing complex and is essential for pre-mRNA splicing. *Mol Cell Biol* 19(8):5352-5362.

31. Palma K*, et al.* (2007) Regulation of plant innate immunity by three proteins in a complex conserved across the plant and animal kingdoms. *Genes Dev* 21(12):1484-1493 .

32. Lin Z*, et al.* (2007) AtCDC5 regulates the G2 to M transition of the cell cycle and is critical for the function of Arabidopsis shoot apical meristem. *Cell Res* 17(9):815-828.

33. Jacobsen SE, Running MP, & Meyerowitz EM (1999) Disruption of an RNA helicase/RNAse III gene in Arabidopsis causes unregulated cell division in floral meristems. *Development* 126(23):5231-5243.

34. Lu C & Fedoroff N (2000) A mutation in the Arabidopsis HYL1 gene encoding a dsRNA binding protein affects responses to abscisic acid, auxin, and cytokinin. *Plant Cell* 12(12):2351-2366.

35. Qi Y, Denli AM, & Hannon GJ (2005) Biochemical specialization within Arabidopsis RNA silencing pathways. *Mol Cell* 19(3):421-428.

36. Ghosh I, Hamilton AD, & Regan L (2000) Antiparallel leucine zipper-directed protein reassembly: Application to the green fluorescent protein. *Journal of the American Chemical Society* 122(23):5658-5659.

37. Machida S & Yuan YA (2013) Crystal Structure of Arabidopsis thaliana Dawdle Forkhead-Associated Domain Reveals a Conserved Phospho-Threonine Recognition Cleft for Dicer-Like 1 Binding. *Mol Plant.*

38. Liu C, Axtell MJ, & Fedoroff NV (2012) The helicase and RNaseIIIa domains of Arabidopsis DCL1 modulate catalytic parameters during microRNA biogenesis. *Plant Physiol*.

39. Lau PW*, et al.* (2012) The molecular architecture of human Dicer. *Nat Struct Mol Biol* 19(4):436-440.

CHAPTER 6

Methylation protects miRNAs from AGO1-associated activity that uridylates 5' RNA fragments generated by AGO1 cleavage

PNAS. **(Pending acception in Editorial Board).**

Guodong Ren*, Meng Xie*, Shuxin Zhang, Carissa Vinovskis, Xuemei Chen and Bin Yu

Abstract

In plants, methylation catalyzed by HEN1 (small RNA methyl transferase) prevents microRNAs (miRNAs) from degradation triggered by uridylation. How methylation antagonizes uridylation of miRNAs *in vivo* is not well understood. In addition, 5' RNA fragments (5' fragments) produced by miRNA-mediated RNA cleavage can be uridylated in plants and animals. However, the biological significance of this modification is unknown and enzymes uridylating 5' fragments remain to be identified. Here, we report that in *Arabidopsis*, HEN1 SUPRESSOR1 (HESO1, a miRNA nucleotidyl transferase) uridylates 5' fragments to trigger their degradation. We also show that AGO1, the effector protein of miRNAs, interacts with HESO1 through its PAZ and PIWI domains, which bind the 3' end of miRNA and cleave the target mRNAs, respectively. Furthermore, HESO1 is able to uridylate AGO1-bound miRNAs *in vitro* and miRNA uridylation *in vivo* requires a functional AGO1 in *hen1,* in which miRNA methylation is impaired, demonstrating that HESO1 can recognize its substrates in the AGO1 complex. Based on these results, we propose that methylation is required to protect miRNAs from AGO1-associated HESO1 activity that normally uridylates 5' fragments.

Introduction

microRNAs (miRNA) and small interfering RNAs (siRNAs), ~ 20-25 nucleotides (nt) in size, are important regulators of gene expression. miRNAs and siRNAs are derived from imperfect hairpin transcripts and perfect long double-stranded RNAs, respectively (1, 2) . miRNAs and siRNAs are then associated with Argonaute (AGO) proteins to

repress gene expression through target cleavage and/or translational inhibition (3). The cleavage of target mRNAs usually occurs at a position opposite to the 10th and 11th nucleotides of miRNAs, resulting in a 5' RNA fragment (5' fragment) and a 3' fragment (4). In *Arabidopsis*, the major effector protein for miRNA-mediated gene silencing is AGO1, which possesses the endonuclease activity required for target cleavage (5-7). In Drosophila, the exosome removes the 5' fragments through its 3'-to-5' exoribonuclease activity (8). How 5' fragments are degraded in higher plants remains unknown. It has been shown that the 5' fragments are subject to untemplated uridine addition at their 3' termini (uridylation) in both animals and plants (9). However, the biological significance of this modification remains unknown due to lack of knowledge of the enzymes targeting 5' fragments for uridylation.

Uridylation plays important roles in regulating miRNA biogenesis. In animals, TUT4, a terminal uridyl transferase is recruited by Lin-28 (an RNA binding protein) to the let-7 precursor (pre-let-7), resulting in uridylation of pre-let-7 (10, 11). This modification impairs the stability of pre-let-7, resulting in reduced levels of let-7. In addition, mono-uridylation has been shown to be required for the processing of some miRNA precursors (12). Deep sequencing analysis reveals that precursor uridylation is a widespread phenomenon occurring in many miRNA families in animals (13). Uridylation also regulates the function and stability of mature miRNAs and siRNAs in both animals and plants (14-16). Uridylation of miR26 in animals reduces its activity without affecting its stability (17). In contrast, uridylation of some siRNA in *C. elegans* restricts them to CSR-1 (an AGO protein) and reduces their abundance, which is required for proper

chromosome segregation (18). In the green algae *Chlamydomonas reinhardtii* (*C. reinhardtii*) and the flowering plant *Arabidopsis*, uridylation causes the degradation of miRNAs and siRNAs (19-21). Enzymes that uridylate miRNAs and siRNAs have been identified in both animals and plants. In humans and *C. elegans*, terminal uridyl transferases ZCCHC6, ZCCHC11, TUT1 and other enzymes have been shown to uridylate miRNAs in a miRNA sequence-specific manner (22) while HESO1 acts on most of miRNAs and siRNAs in *Arabidopsis* (20, 21). Nevertheless, it is unclear how these terminal uridyl transferases recognize their targets.

Here we show that HESO1 catalyzes the uridylation of 5' fragments that are produced by AGO1-mediated cleavage of miRNA target RNAs. Uridylation of the 5' fragment of *MYB33* (a target of miR159; *MYB33-5'*) is impaired in *heso1-2,* resulting in increased abundance of *MYB33-5'*. In addition, the proportion of *MYB33-5'* with 3' truncation is increased in *heso1-2* when compared with those in wild-type plants. These results demonstrate that HESO1-mediated uridylation triggers 5' fragment degradation through a mechanism that may be different from 3'-to-5' trimming activity. Furthermore, we show that HESO1 interacts with AGO1 and is able to uridylate AGO1-bound miRNAs *in vitro*. Based on these observations, we propose that HESO1 can uridylate AGO1-associated 5' fragments and miRNAs, resulting in their degradation.

Results:

HESO1 uridylates 5' RNA fragments generated by miRNA-mediated cleavage

HESO1 possesses terminal uridyl transferase activity on 21 nt small RNAs *in vitro* (20, 21). However, whether HESO1 acts on other RNAs is not known. To address this question, we generated a [^{32}P] labeled single-stranded RNA (ssRNA; ~100 nt), which corresponds to a portion of *UBQ5* mRNA through *in vitro* transcription. HESO1 lengthened this ssRNA in the presence of UTP (Figure 6-1A). This result suggested that HESO1 might have substrates other than small RNAs, and therefore, prompted us to test whether 5' fragments are also substrates of HESO1. We compared 5' fragment uridylation in the null *heso1-2* mutant (20) with that in Landsberg *erecta* (L*er*; wild type control of *heso1-2*) using a 3' al-RACE (adaptor-ligation mediated rapid amplification of cDNA ends) approach. Total RNAs from L*er* or *heso1-2* were isolated, ligated to a 3' adapter and reverse transcribed with a primer recognizing the 3' adapter. Semi-nested PCR was subsequently performed to amplify 5' fragments generated by AGO1 slicing of *MYB DOMAIN PROTEIN 33* (*MYB33-5'*), *AUXIN RESPONSE FACTOR 10* (*ARF10-5'*), and *LOST MERISTEMS 1* (*LOM1-5'*), which are targets of miR159, miR160 and miR171, respectively (23-26). PCR products of the expected sizes were gel-purified, cloned and sequenced (Figure 6-7). 75%, 59.1% and 26.5% of *MYB33-5'*, *ARF10-5'* and *LOM1-5'* were uridylated in L*er*, respectively (Figure 6-1B and 1C). In contrast, the proportions of uridylated *MYB33-5'*, *ARF10-5'* and *LOM1-5'* were reduced to 5.9%, 23.8% and 12.9% in *heso1-2*, respectively (Figure 6-1B and 1C). Furthermore, the 3' tail length of 5' fragments was reduced in *heso1-2* compared with that in L*er* (1-3nt vs 1-15 nt; Figure 6-1C). These results together with the *in vitro* activity analysis (Figure 6-1A and 6-1C) demonstrated that HESO1 catalyzes uridylation of 5' fragments generated by miRNA-mediated cleavage. However, the presence of uridylated 5' fragments in the null

heso1-2 mutant (Figure 6-1C) indicated that additional HESO1 homolog(s) might also act on 5' fragments.

HESO1-mediated uridylation triggers the degradation of the 5' fragment of *MYB33* generated by AGO1 cleavage

Next, we examined whether uridylation induced the degradation of 5' fragments using *MYB33* as a reporter RNA. *MYB33* was selected because the majority of its 5' fragments (*MYB33-5'*) are uridylated (Figure 6-1C) (9). We compared the accumulation of *MYB33-5'* in *heso1-2* with that in L*er* by Northern blotting with probes recognizing *MYB33-5'* (Figure 6-2A). To determine the specificity of probe for *MYB33-5'*, we included a *myb33* mutant, in which a T-DNA insertion abolished the transcription of *MYB33* (26). We were able to detect *MYB33-5'* in L*er* and *heso1-2* but not in *myb33*. The levels of *MYB33-5'* increased in *heso1-2* relative to those in L*er* (Figure 6-2B). This could be a result of the enhanced cleavage of *MYB33* by AGO1 or decreased degradation of *MYB33-5'*. If increased levels of *MYB33-5'* were caused by enhanced target cleavage, the abundance of *MYB33-3'* would increase as well. Our data showed that the levels of *MYB33-3'* were similar in *heso1-2* to those in L*er* (Figure 6- 2B), indicating that miRNA-mediated *MYB33* cleavage did not increase in *heso1-2*. Consistent with this observation, the levels of miR159 were not altered and the abundance of *MYB33* was only slightly elevated in *heso1-2* (Figure 6- 2B, 2C and S2A). Thus, we concluded that HESO1-mediated uridylation promotes 5' fragment degradation.

heso1-2* increases the proportion of 3' truncated *MYB33-5'

Next we asked whether uridylation could trigger 3'-to-5' degradation of *MYB33-5'* as 5' fragments can be degraded from the 3' end by the exosome in Drosophila and in the green algae *Chlamydomonas reinhardtii* (*C. reinhardtii*) (8, 27). The 3' ends of both capped and uncapped *MYB33-5'* in L*er* and *heso1-2* were examined separately since they both contain U-tails (9). We used a cRACE (circularized rapid amplification of cDNA ends, Figure 6-3A-3C) approach to analyze the 3' ends. Two ligation experiments were performed. In the first set of experiments, RNAs were self-ligated to analyze uncapped *MYB33-5'*, whose 5' mono-phosphate allows self-ligation (Figure 6-3A). In contrast, the self-ligation of capped *MYB33-5'*was blocked by the cap structure (Figure 6-3A). In the second set of experiments, total RNAs were treated with CIP (Alkaline Phosphatase, Calf Intestinal), which removes the 5' mono-phosphate and thus inhibits self-ligation of uncapped 5' fragments (Figure 6-3B). The resulting RNAs were further treated with TAP (tobacco acid pyrophosphatase) to remove the cap structure of capped RNAs, resulting in RNAs with a 5' mono-phosphate. After this step, RNAs were ligated, which enabled us to analyze the capped 5' fragments (Figure 6-3B). Nested RT-PCR was then performed using the ligation products generated from these two sets of experiments as templates (Figure 6-3C and Figure 6-8B). RT-PCR products were directly cloned and sequenced. Both capped and uncapped *MYB33-5'* contained U-tails in L*er* (Figure 6-3D and 3E). However, the relative levels of uridylated *MYB33-5'* in the capped population was lower than those in the uncapped population in L*er* (Figure 6-3D and 3E). The relative levels of uridylated *MYB33-5'* in both capped and uncapped populations were reduced in *heso1-2* when compared with L*er* (Figure 6-3D and 3E), consistent with our alRACE results (Figure 6-1C). We compared the levels of 3' truncated *MYB33-5'* in *heso1-2* and L*er*. If

uridylation triggered 3'-to-5' degradation, lack of uridylation in *heso1-2* should reduce the proportion of 3' truncated *MYB33-5'*. However, the proportion of both capped and uncapped 5' fragments with 3' truncation increased in *heso1-2* relative to L*er* (59.1% *vs* 47.1% for capped ones; 48% *vs* 19.1% for uncapped ones; Figure 6-3F), suggesting that 3' trimming of 5' fragments may compete with uridylation. We also examined whether *heso1-2* had any effect on the 5'-to-3' truncation of uncapped *MYB33-5'*. However, no obvious changes for the positions of 5' truncation were observed in *heso1-2* relative to L*er* (Figure 6- 3D).

Exoribonuclease 4 (XRN4) can degrade 5' fragments

Studies have shown that exoribonucleases are involved in the degradation of RNA products generated by miRNA-mediated cleavage in Drosophila and *C. reinhardtii* (8, 27). We therefore asked whether exoribonucleases have roles in degrading 5' fragments in *Arabidopsis*. We examined whether XRN4, which is a major cytoplasmic 5'-to-3' exoribonuclease in *Arabidopsis* (28, 29), could degrade *MYB33-5'*. The levels of *MYB33-5'* in *xrn4-5*, in which a T-DNA insertion completely abolished XRN4 function (29), were higher than those in Col (wild-type control) by Northern blotting. In contrast, the full-length *MYB33* transcript was not obviously affected by *xrn4-5* (figure 6-9), suggesting that the 5' fragments are subjected to 5'-to-3' degradation in *Arabidopsis*. We also tested the function of the exosome components CSL4 and RRP6L in *MYB33-5'* degradation. Northern blotting showed that the levels of *MYB33-5'* in *csl4-1* and *rrp6l1-1 rrp6l2-1 rrp6l3-1* were comparable with those in Col (Figure 6-9), suggesting that CSL4 and RRP6L may not be involved in 5' fragment degradation.

HESO1 interacts with AGO1

Next we asked how HESO1 recognizes miRNAs and 5' fragments. Since both miRNAs and 5' fragments are associated with AGO1 during the cleavage process, we hypothesized that HESO1 might interact with AGO1 to recognize its substrates. Consistent with this hypothesis, AGO1 is associated with uridylated miRNAs (15, 30). We first examined whether HESO1 co-localized with AGO1. We co-expressed HESO1 fused with a red fluorescence protein (HESO1-RFP) and AGO1 fused with a yellow fluorescence protein (AGO1-YFP-HA) in *Nicotiana benthamiana* (*N. benthamiana*). The yellow fluorescence signal produced from AGO1-YFP overlapped with the red fluorescence signal generated by HESO1-RFP (Figure 6-4A), indicating that HESO1 and AGO1 might be associated with each other.

To confirm the AGO1-HESO1 interaction, we performed reciprocal co-immunoprecipitation (Co-IP) assays. We transiently expressed HESO1-YFP (20) in leaves of *N. benthamiana,* mixed the HESO1-YFP containing protein extracts with the AGO1 containing protein extracts from *Arabidopsis* inflorescence and performed IP with either anti-AGO1 antibody (Figure 6-4B and Figure 6-S4A) or anti-YFP antibody (Figure 6-4C). We were able to detect HESO1-YFP (~95 KDa) in the AGO1 immunoprecipitates and AGO1 (~120 KDa) in the HESO1-YFP immunoprecipitates (Figure 6-4B and 6-4C). In contrast, YFP (~26 KDa) and AGO1 did not co-IP with each other (Figure 6-4B and 4C). In addition, Protein A beads without antibody failed to pull down either AGO1 or HESO1-YFP (Figure 6-4B and 6-4C). As both AGO1 and HESO1 recognize RNAs, it is

possible that the AGO1-HESO1 interaction might be RNA-mediated. To test this, we treated the protein extracts with RNase A during the immunoprecipitation. We used this assay previously to show the RNA-dependent FDM1-AGO4 interaction (32). This treatment did not abolish the AGO1-HESO1 interaction, suggesting that HESO1 may interact with AGO1 in an RNA-independent manner (Figure 6-10B).

We next asked which domains of AGO1 interact with HESO1. We expressed five N-terminal 10XMYC-fused AGO1 fragments named FL (Full-length; ~150 KDa), A1 (AA 1-390; the N-terminal domain; ~ 80Kda), A2 (aa 381-530; the PAZ domain; ~40 KDa), A3 (aa 521-700; the L2-MID domain; ~45 KDa) and A4 (aa 671-1050; the PIWI domain; ~75 KDa) (Figure 6-4D) individually in *N. benthamiana*, and performed co-IP with HESO1-YFP. The PAZ and PIWI domains (A2 and A4) but not the N-terminal and L2-MID domains interacted with HESO1 (Figure 6-4E). We also identified the protein domains of HESO1 that mediate the AGO1-HESO1 interaction. Two fragments of HESO1 (Figure 6-4F), an N-terminal fragment, which covers the poly A polymerase domain (PAP/25A) and the PAP-associated domain (aa 1-320;T1; ~63 KDa), and a C-terminal fragment that contains the PAP-associated domain and the glutamine rich region (aa 200-511;T2; ~ 62 KDa), were fused with YFP at their C-terminus, expressed in *N. benthamiana* and analyzed for interactions with AGO1. The results showed that T1 but not T2 interacted with AGO1 (Figure 6-4G).

HESO1 acts on AGO1-bound miRNAs

The AGO1-HESO1 interaction suggested that HESO1 might act on miRNA in the AGO1 complex. If so, uridylation of miRNAs may require a functional AGO1. To test this, we crossed *ago1-27* carrying a point mutation in the PIWI domain of AGO1 into the null *hen1-1* mutant and examined the status of 3' tailing of miRNAs in *ago1-27 hen1-1*. Northern blotting revealed that the tailing of miR159/319 and miR171/170 was dramatically impaired in *ago1-27 hen1-1* when compared with *hen1-1* (Figure 6-5A). Consistent with this result, the *ago1-11* mutation also reduces the tailing of many miRNAs in *hen1-2* (33). These results supported that HESO1 may uridylate miRNAs after AGO1 loading. We therefore examined whether HESO1 could act on AGO1-bound miRNA *in vitro*. We transiently expressed AGO1-YFP in *N. benthamiana* and immunoprecipitated the AGO1 complex using anti-AGO1 antibodies conjugated to protein A-agarose beads (Figure 6-11A). The resulting AGO1 complex was incubated with 5' [^{32}P] labeled miR166a (unmethylated), to assemble the AGO1-miR166a complex, and unbound miR166a was removed through washing. AGO1-miR166a (Figure 6-11B) was subsequently incubated with MBP-HESO1 or MBP in the presence of UTP. After washing, miR166a was extracted from the AGO1 complex and separated in a denaturing PAGE gel. miR166a was lengthened by MBP-HESO1 but not MBP, indicating that HESO1 is able to target AGO1-bound miRNA in *vitro* (Figure 6-5B). It should be noted that endogenous *N. benthamiana* HESO1 might be co-immunoprecipitated with AGO1 as well. However, its amount might be too low to contribute to the lengthening of AGO1-bound miR166a in our assay since no obvious activity was detected in the control reaction (Figure 6-5B).

Discussion

In this study, we show that HESO1, a miRNA nucleotidyl transferase, uridylates 5' fragments produced by miRNA-mediated target cleavage. We also reveal that HESO1 associates with AGO1 and acts on AGO1-bound miRNAs *in vitro*. Since both miRNAs and 5' fragments are associated with AGO1 during the cleavage process, we propose that HESO1 can uridylate its substrates in the AGO1 complex (Figure 6-6). However, the 3' end of a miRNA may be protected by the PAZ domain of AGO1, which may reduce its exposure to HESO1. It is tempting to speculate that the uridylation of unmethylated miRNAs by HESO1 may depend on base-pairing between miRNAs and their targets *in vivo* since base-pairing with targets is predicted to release the 3' end of miRNAs from the PAZ domain (34). Consistent with this notion, miRNA uridylation is blocked when AGO1 function is impaired in *hen1* (Figure 6-5A) (33) and extensive complementarity between targets and miRNAs triggers miRNA tailing in animals (35). However, the majority of miRNAs are normally methylated in plants, which prevents HESO1 function and, therefore, maintains the recycling of miRNA-AGO1 complex (15, 20, 21, 36). Lack of HESO1 cannot completely eliminate uridylated 5' fragments and miRNAs (20, 21), indicating one or more HESO1 homologs may function redundantly with HESO1 in the miRNA pathway.

The abundance of 5' fragments is increased in *heso1-2* relative to L*er*, demonstrating that uridylation induces the degradation of 5' fragments (Figure 6-2B and Figure 6-6B). How does uridylation trigger 5' fragment degradation? In Drosophila and *C. reinhardtii*, it has been observed that 5' fragments can be degraded through 3'-to-5' exonulcease activities

(8, 27). However, the relative levels of 5’ fragments with 3’ truncation in both capped and uncapped 5’ fragment populations in *heso1-2* are increased when compared with those in L*er*, suggesting that uridylation may trigger activities other than 3’-to-5’ exonucleases in *Arabidopsis* (Figure 6-3 and Figure 6-6). In fact, oligouridylation could prevent RNA from 3’ to 5’ degradation in vitro (37). However, we cannot rule out the possibility that 3’-to-5’ degradation activities triggered by uridylation are highly progressive such that no or few 3’ truncation intermediates are accumulated in *vivo*. 5’ fragments with 5’ truncation exist in both *heso1* and L*er*, suggesting that 5’-to-3 degradation of 5’ fragments may occur. Indeed, XRN4 can degrade the 5’ fragments. However, it is possible that the 5’-to-3’ truncation of 5’ fragment occurs independently of uridylation since lack of uridylation has no obvious effects on 5’-to-3’ truncation of 5’ fragments. The presence of capped and uncapped *MYB33-5’* with 3’ truncation indicates that they both can be degraded through 3’-to-5’ degradation activities (Figure 6-3), which may be a slow process and compete with HESO1 for substrates in *Arabidopsis* (Figure 6-3). The enzymes degrading 5’ fragments from 3’-to-5’ remain to be identified as the abundance of MYB33-5’ is not altered in exosome mutants *rrp6l1 rrp6l2 rrp613* and *csl4* (Figure 6-S3). In humans and yeast, uridylation has been shown to induce decapping of some RNAs followed by degradation (37-39). The ratio of uridylated *MYB33-5’* in uncapped population is higher than that in capped population in L*er* suggesting that uridylation may also have a role in stimulating decapping. Clearly, this possibility needs to be examined in the near future.

Materials and methods

Materials

The *myb33* (CS851168), *xrn4-5* (CS829864), *csl4-1* (SALK_004562), *rrp6l1-1* (Salk_004432), *rrp6l2-2* (Salk_113786) and *rrp6l3-1* (SALK_018102) mutants were all in the Col-0 background and were obtained from the *Arabidopsis* Biological Resources Center. The *heso1-2* mutant is in the L*er* background (20).

Plasmid

HESO1 and AGO1 CDS were amplified by RT-PCR and cloned into Gateway vector pB7WGR2,0 (40) and pEarleyGate 101 (41) to generate HESO1-RFP and AGO1-YFP-HA, respectively. To express truncated AGO1 and HESO1, different AGO1 fragments (A1-A4) and HESO1 fragments (T1 and T2) were PCR amplified and cloned into the Gateway vectors pGWB521 (42) and pEarleyGate101 to generate YFP (YFP fused at C terminus)- and 10xMYC (10xMYC fused at N terminus)-tagged proteins, respectively.

Protein expression, confocal microscopy, protein size-fractionation and co-immunoprecipitation

Protein expression in *N. benthamiana* and the *E. coli* strain BL21, confocal microscopy and co-immunoprecipitation were performed as described (43). The affinity purified anti-AGO1 antibodies recognizing the N-terminal peptide of AGO1 (N-MVR KRRTDAPSC-C; 6) were produced by GenScript (Piscataway, NJ). Anti-GFP (Clontech) and anti-AGO1 were pre-coupled to protein A agarose beads (Santa Cruz) and used for IP analyses. Anti-GFP, Anti-MYC, and Anti-AGO1 antibodies were used for western blot detection of the respective proteins.

AGO1-miR166a assembly and terminal uridyl transferase assay

The AGO1-miR166a complex was prepared according to (5) and used for an in vitro terminal uridyl transferase assay (20) .

Al-RACE and cRACE

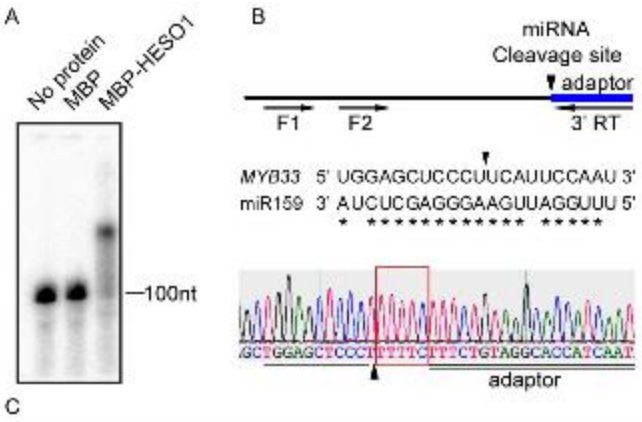
Al-RACE and cRACE were performed according to (9) with some modifications. In the al-RACE experiment, 5µg total RNA was first ligated to 100 pmol RNA adaptor by T_4 RNA ligase. In the cRACE experiment, 5µg treated (CIP followed by TAP) or non-treated RNAs were subjected to self-ligation. First strand cDNA was synthesized using the 3' RT primer (for al-RACE) or the R1 primer (for cRACE). First round PCR was performed using 3´RT/F1 (for al-RACE) or R1/F1 (for cRACE). Then 1µl PCR product was diluted for 50 times and used for the second round of PCR using 3´RT/F2 (For al-race) or R2/F2 (for cRACE) and F2. The PCR products were cloned into pGEM-T Easy Vector (Promega) and sequenced.

Northern blot

Small RNA Northern blot was conducted as described (44). To detect *MYB33-5'*or *MYB33- 3'* by Northern blot, 30µg total RNAs were resolved by electrophoresis on a 1.2% denaturing-formaldehyde agarose gel and transferred onto Zeta-probe membranes (Bio-Rad). Membranes were UV cross-linked and hybridized with probes recognizing *MYB33-5'* or *MYB33-3'*. Radioactive signals were detected using a Typhoon 9500 phosphorimager.

Figures

Figure 6-1. HESO1 uridylates 5' fragments. (A) HESO1 uridylates a long single-stranded RNA (ssRNA) *in vitro*. A 5′-end [^{32}P] labeled ssRNA was incubated with buffer, MBP or MBP-HESO1 in the presence of UTP for 120 minutes, and products were resolved on a denaturing polyacrylamide gel. **(B)** Uridine addition (red rectangle) at the 3' end of the cleavage site of *MYB33-5'*(▲or▼). =: the adaptor. **(C)** Uridylation of 5' fragments in L*er* and *heso1-2*. Uridines in lowercase indicate that they can alternatively be considered as templated addition. The numbers of clones for each modification were shown in (). Clones: numbers of sequenced clones. Ratio: frequency of clones with 3' end modifications among sequenced clones.



C

Target/miRNA	Genotype	untemplated 3' end	Total Clones/Ratio
MYB33/miR159	L*er*	u(3),uu,uuU(4), uuUUA,uuUUC(3)	16(75%)
	heso1-2	u	17(5.9%)
ARF10/miR160	L*er*	U(5),UU(3),UUA,UUUC, UUUUA,UUUUU AAUUUUUUUU	22(59.1%)
	heso1-2	U(3),UU,UUU	21(23.8%)
LOM1/miR171	L*er*	UUU(2), UUC, AAAAU(2),AAAGU, UUUCUUUUUUUU(2), UUUCUUUUUUUUUUU	34(26.5%)
	heso1-2	U(3), UU	31(12.9%)

Figure 6-2. HESO1-mediated uridylation triggers the degradation of *MYB33-5'*. (A) A schematic diagram of the *MYB33* cDNA showing the positions of probes used for northern blotting analyses. The filled circle represents the stop codon. ▲: cleavage site. (**B**) The abundance of *MYB33-5'* was higher in *heso1-2* than in L*er*. *MYB33* RNAs were detected by Northern blotting using probes (shown in (A)) recognizing *MYB33-5'* or *MYB33-3'* generated by AGO1-mediated cleavage. FL: Full-length *MYB33* transcripts; *myb33*: a mutant allele of *MYB33*, in which a T-DNA insertion disrupts the transcription of *MYB33* (26). The levels of cleavage products in *heso1-2* were normalized to full-length transcripts and compared with those in L*er*. **(C)** Northern Blot analysis of miR159 in L*er* and *heso1-2*. U6 RNA was probed as a loading control. Note that the miR159 probe also recognizes miR319.

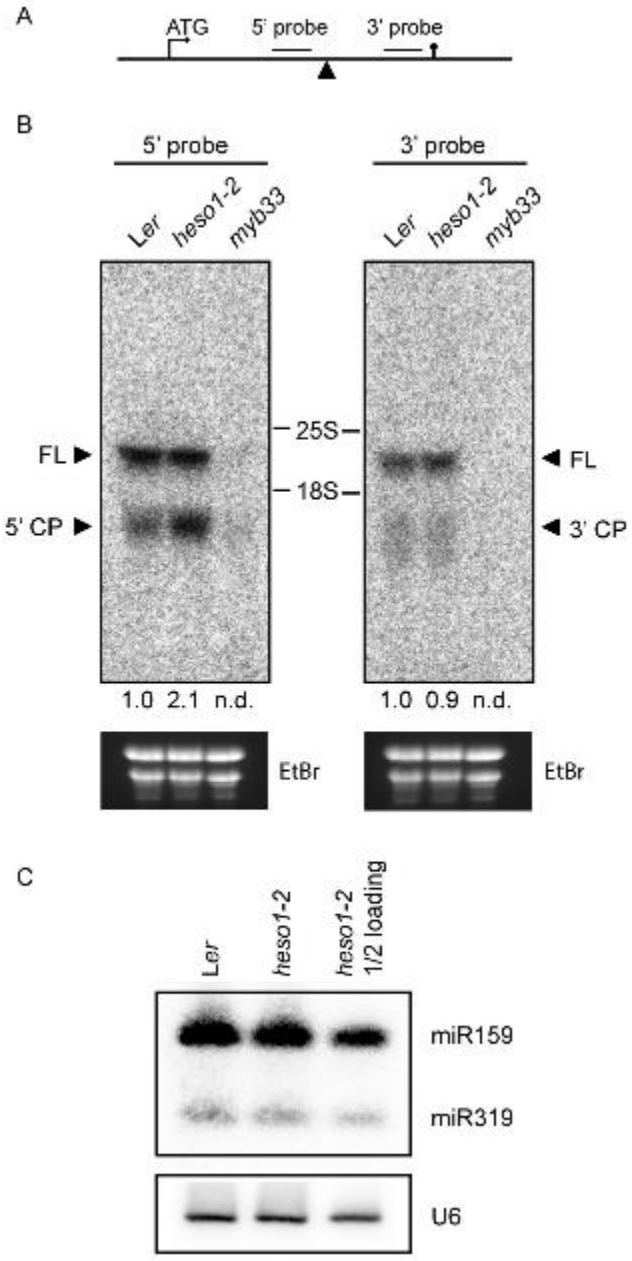

A
ATG
5' probe
3' probe
B
5' probe
3' probe
Ler
hesO1-2
myb33
Ler
hesO1-2
myb33
FL
5' CP
25S
18S
FL
3' CP
1.0 2.1 n.d.
1.0 0.9 n.d.
EtBr
EtBr
C
Ler
hesO1-2
hesO1-2 1/2 loading
miR159
miR319
U6

Figure 6-3. cRACE analysis of *MYB33-5'*. (A) and **(B)** Schematic diagrams of cRACE followed by nested RT-PCR (cRT-PCR) used to analyze capped (black) or uncapped (gray) *MYB33-5'*. CIP: Alkaline Phosphatase (Calf Intestinal). TAP: Tobacco acid pyrophosphatase. **(C)** A schematic diagram of the *MYB33* cDNA showing the positions of primers for nested RT-PCR. ▼: Cleavage site. **(D)** Analyses of 5' and 3' ends of *MYB33*-5'. The 3' end signature (Y-axis) of individual *MYB33-5'* clones was plotted against its 5' end position (X-axis). The values on the X-axis indicate the 5' positions of individual *MYB33-5'* clones relative to the translation start site that is set as +1. The positive values on the Y-axis indicate the lengths (nt) of 3' tailing while the negative values on the Y-axis represent the degree of 3' truncation that is calculated as $-Log_2(-N+1)$ (N represents the distance between the 3' end position of *MYB33-5'* with 3' truncation to the miRNA cleavage site, which is set as 0). Note: The reason to use $Log_2(-N+1)$ instead of $log_2 N$ is to include clones with one nucleotide truncation on the plot. Different colors were used to distinguish clones with the same 5' end signature (1^{st}, Black; 2^{nd}, Red; 3^{rd}, Blue; 4^{th}, Cyan; 5^{th}, Pink). 5' UTR: 5' untranslated region. CDS: Coding sequence. **(E)** The frequency of 3' end uridylation in L*er* and *heso1-2*. (**F**) The proportions of 3' truncated *MYB33-5'* in *heso1-2* and L*er*. The proportion indicates the frequency of 3' truncated clones among all sequenced clones of cRT-PCR products. n: numbers of sequenced clones.

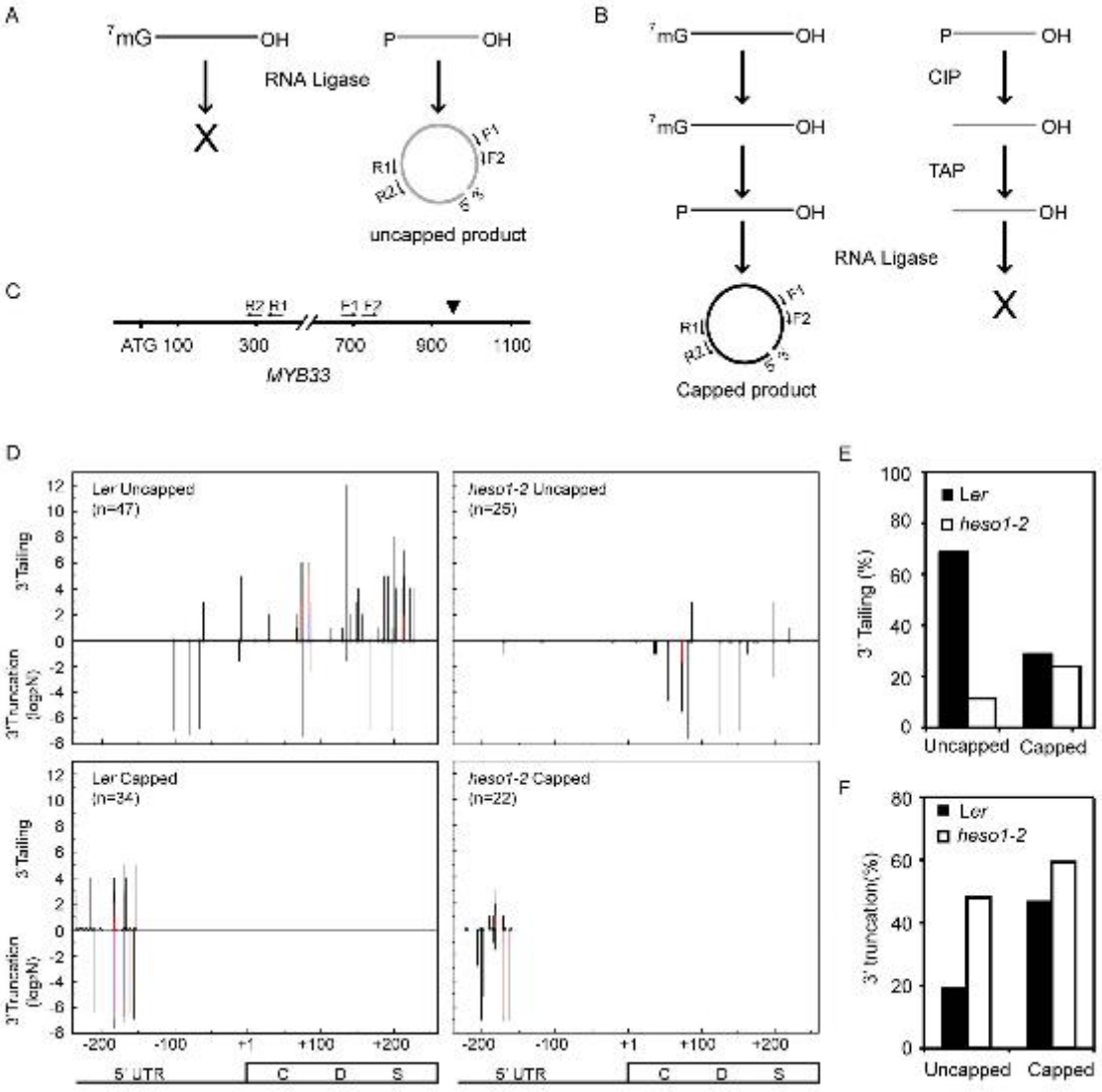

A
7mG OH
P OH
RNA Ligase
X
F1
F2
R1
R2
uncapped product
B
7mG OH
7mG OH
P OH
P OH
CIP
OH
TAP
OH
RNA Ligase
X
F1
F2
R1
R2
Capped product
C
R2 R1
F1 F2
ATG 100
300
700
900
1100
MYB33
D
Ler Uncapped
(n=47)
hesо1-2 Uncapped
(n=25)
Ler Capped
(n=34)
hesо1-2 Capped
(n=22)
3' Tailing
3' Truncation (log2N)
-200
-100
+1
+100
+200
5' UTR
C
D
S
E
Ler
hesо1-2
3' Tailing (%)
Uncapped
Capped
F
Ler
hesо1-2
3' truncation(%)
Uncapped
Capped

Figure 6-4. HESO1 interacts with AGO1. (**A**) Co-localization of HESO1-RFP and AGO1-YFP. HESO1-RFP and AGO1-YFP fusion proteins were co-infiltrated into *N. benthamiana* leaves and RFP and YFP fluorescence signals were monitored 48h after infiltration by confocal microscopy. (**B**) HESO1-YFP co-immunoprecipitates (Co-IPs) with AGO1. (**C**) AGO1 co-IPs with HESO1-YFP. The protein mixtures containing AGO1/HESO1-YFP or AGO1/YFP were incubated with anti-AGO1-protein A-agarose beads and anti-YFP-protein A-agarose beads to capture AGO1, HESO1-YFP and YFP, respectively. (**D**) A schematic diagram of AGO1 domains and truncated AGO1 fragments used for co-IP assays. (**E**) A diagram of truncated HESO1 fragments used for co-IP assays. (**F**) HESO1 co-IPs with the PAZ and PIWI domains of AGO1. Anti-YFP-protein A agarose beads were incubated with the protein extracts containing HESO1-YFP and full-length AGO1 or a truncated AGO1 fragment (indicated on the left or right side of the picture) to capture the HESO1-YFP complex. Full-length AGO1 and truncated AGO1 fragments were fused with 10xMYC at their N-termini. Please note only one IP picture was shown for HESO1-YFP. (**G**) The N-terminal region of HESO1 interacts with AGO1. Both IP and co-IP signals were detected by western blot analyses, ~10% input (for detecting IP signals) and ~1% input (for detecting co-IP signals) were analyzed in parallel.

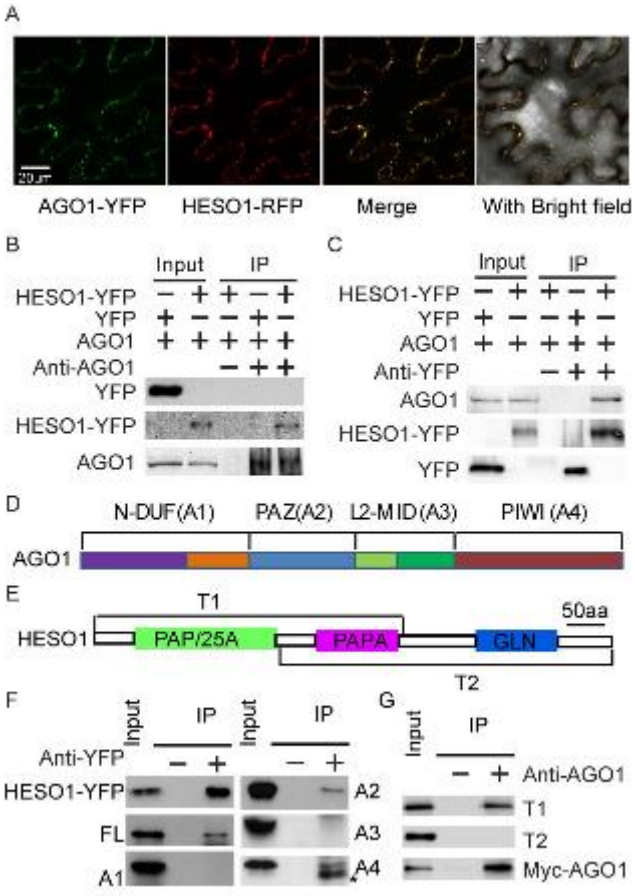

A
AGO1-YFP
HESO1-RFP
Merge
With Bright field
B
Input
IP
HESO1-YFP
YFP
AGO1
Anti-AGO1
YFP
HESO1-YFP
AGO1
C
Input
IP
HESO1-YFP
YFP
AGO1
Anti-YFP
AGO1
HESO1-YFP
YFP
D
N-DUF(A1)
PAZ(A2)
L2-MID(A3)
PIWI (A4)
AGO1
E
T1
50aa
HESO1
PAP/25A
PAPA
GLN
T2
F
Input
IP
Input
IP
Anti-YFP
HESO1-YFP
FL
A1
A2
A3
A4
G
Input
IP
Anti-AGO1
T1
T2
Myc-AGO1

Figure 6-5. HESO1 is able to uridylate an AGO1 bound miRNA *in vitro*. (**A**) The uridylation of miR159/319 and miR171/170 was reduced in *ago1-27 hen1-1*. (**B**) HESO1 lengthens AGO1-bound miR166a. The AGO1-miR166a complex or miR166a alone was incubated with HESO1-MBP or MBP in a reaction buffer containing UTP for 30 minutes. After the reactions, miR166a was extracted and separated by denaturing PAGE. MiR166a was [^{32}P] labeled at the 5' end using T4 Polynucleotide Kinase. Fp: Free probe.

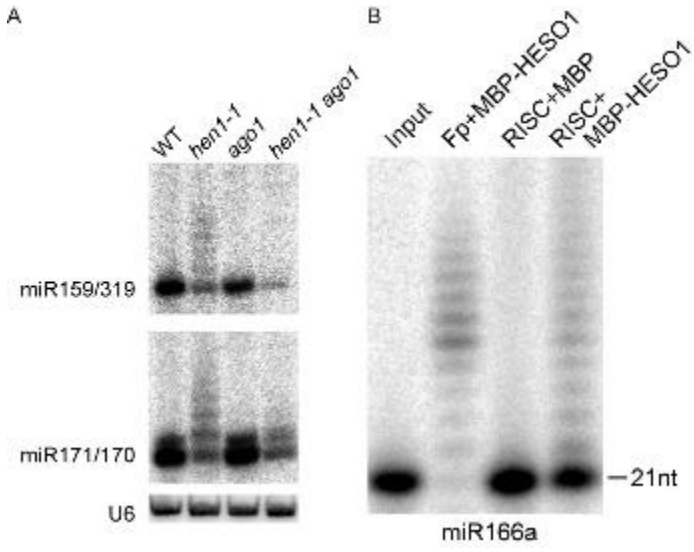


Figure 6-6. A proposed model for HESO1 function in *Arabidopsis*. (**A**) HESO1 uridylates unmethylated miRNAs to lead to its degradation. (**B**) HESO1 uridylates the 5' fragment to promote its degradation. Both 3'-to-5' trimming activities and HESO1 target 5' fragments and unmethylated miRNAs. HESO1-mediated uridylation triggers the degradation of 5' fragments through a mechanism that is likely different from 3'-to-5' trimming activities. Me: 3' methyl group; H: HESO1; Blue oval: AGO1.

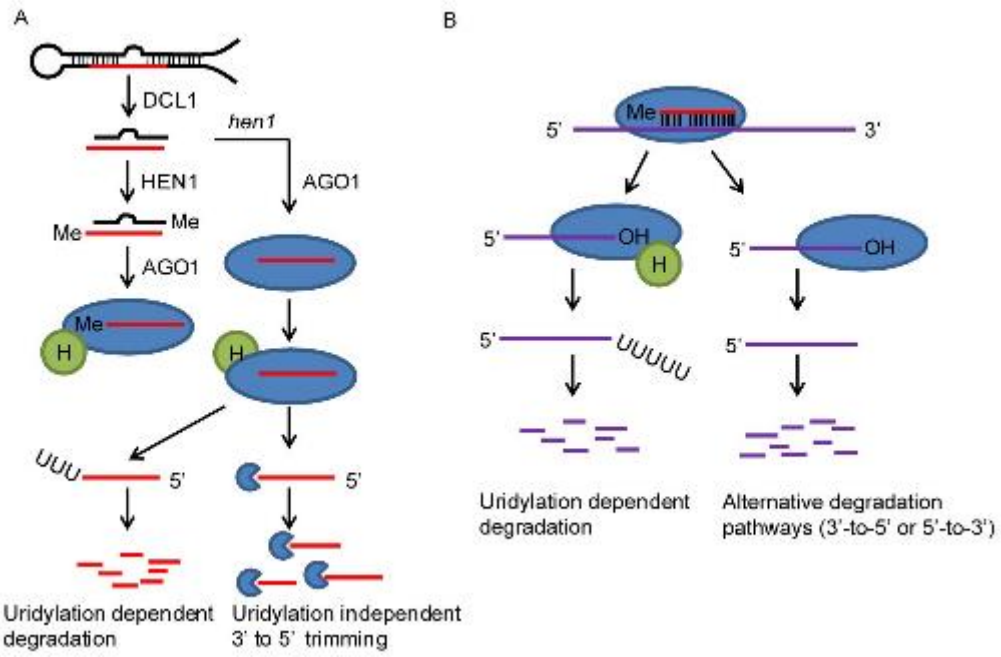


Figure 6-7. Al-RACE cloning of 5' fragments. Total RNAs from L*er* or *heso1-2* were ligated to a 3' RNA adaptor and subjected to 3' al-RACE, which was followed by RT-PCR. The nested-PCR products were resolved in a 1.5% agarose gel. DNAs of the expected size were gel purified before cloning (white box).

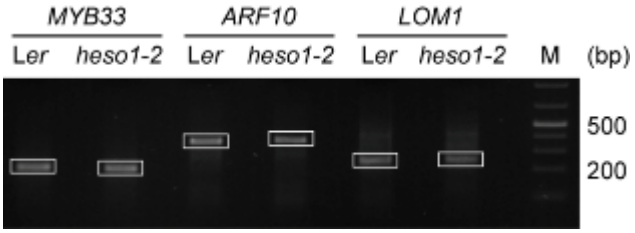


Figure 6-8. cRACE cloning of capped and uncapped *MYB33-5'*. (**A**) Quantitative RT-PCR analysis of *MYB33* transcripts using primers that span the miRNA cleavage site. (**B**) RT-PCR analysis of cRACE products of uncapped and capped *MYB33-5'* in L*er* and *heso1-2*. Total RNAs with or without the sequential treatment by CIP and TAP were subjected to self-ligation (See Fig. 2A and 2B). The nested-PCR products were resolved in a 1.5% agarose gel.

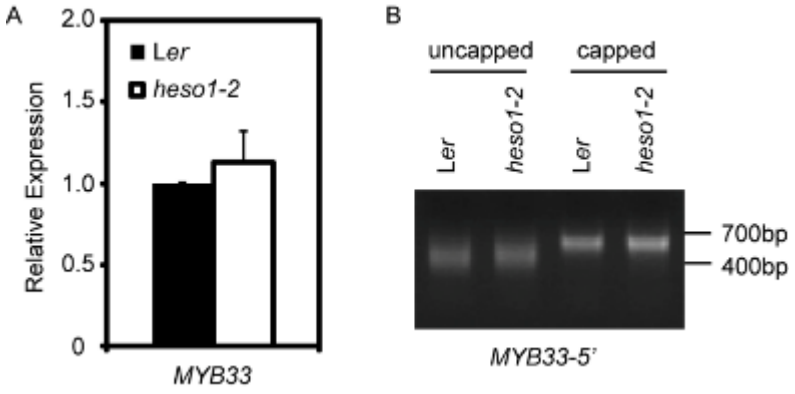


Figure 6-9. The accumulation of *MYB33-5'* is increased in *xrn4-5*. *MYB33* RNAs in Col, *xrn4-5, rrp6l1 rrp6l2 rrp6l3* (*rrp6l* triple) and *csl4-1* were detected by Northern blotting using the 5' probe shown in Fig. 2A. FL: full-length *MYB33* transcripts. 5' CP: 5' Cleavage product. *: non-specific signal.

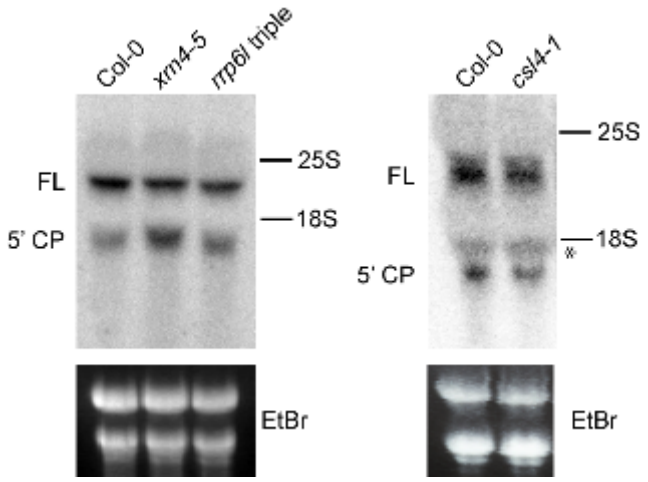


Figure 6-10. HESO1 interacts with AGO1 in an RNA-independent manner. (**A**) Examination of anti-AGO1 antibodies by western blot. The *ago1-36* mutant, a null allele of *ago1*, was used as a negative control. 1:2000 dilution of anti-AGO1 was used for the western blot. RbcL was visualized by staining with Coomassie Brilliant Blue (CBB). (**B**) The HESO1-AGO1 interaction is resistant to the RNase A treatment.

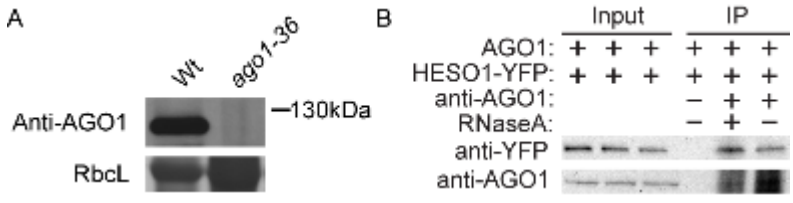
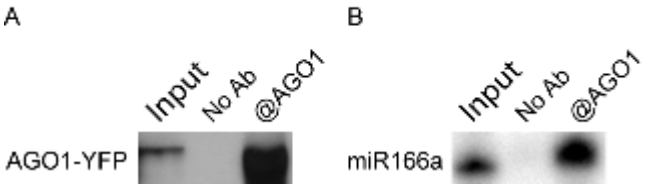


Figure 6-11. Assembling of the AGO1-miR166a complex *in vitro*. (**A**) Immunoprecipitation of AGO1-YFP by anti-AGO1 coupled to protein A beads. Proteins were resolved on an SDS-polyacrylamide gel and detected by western blot with an anti-YFP antibody (Covance). (**B**) Detection of [^{32}P] labeled miR166a in the AGO1 complex.



References

1. Chen XM (2009) Small RNAs and Their Roles in Plant Development. *Annu Rev Cell Dev Bi* 25**,** 21-44.
2. Kim VN, Han J, Siomi MC (2009) Biogenesis of small RNAs in animals. *Nat Rev Mol Cell Bio* 10**,** 126-139.
3. Hutvagner G, Simard MJ (2008) Argonaute proteins: key players in RNA silencing. *Nat Rev Mol Cell Bio* 9**,** 22-32.
4. Schwab R*, et al.* (2005) Specific effects of MicroRNAs on the plant transcriptome. *Dev Cell* 8**,** 517-527.
5. Baumberger N, Baulcombe DC (2005) Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits rnicroRNAs and short interfering RNAs. *P Natl Acad Sci USA* 102**,** 11928-11933.
6. Qi Y, Denli AM, Hannon GJ (2005) Biochemical specialization within Arabidopsis RNA silencing pathways. *Mol Cell* 19**,** 421-428.
7. Vaucheret H, Vazquez F, Crete P, Bartel DP (2004) The action of ARGONAUTE1 in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development. *Gene Dev* 18**,** 1187-1197.
8. Orban TI, Izaurralde E (2005) Decay of mRNAs targeted by RISC requires XRN1, the Ski complex, and the exosome. *Rna* 11**,** 459-469.
9. Shen BZ, Goodman HM (2004) Uridine addition after microRNA-directed cleavage. *Science* 306**,** 997-997.

10. Hagan JP, Piskounova E, Gregory RI (2009) Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells. *Nat Struct Mol Biol* 16**,** 1021-1033.
11. Heo I*, et al.* (2009) TUT4 in Concert with Lin28 Suppresses MicroRNA Biogenesis through Pre-MicroRNA Uridylation. *Cell* 138**,** 696-708.
12. Heo I*, et al.* (2012) Mono-Uridylation of Pre-MicroRNA as a Key Step in the Biogenesis of Group II let-7 MicroRNAs. *Cell* 151**,** 521-532.
13. Newman MA, Mani V, Hammond SM (2011) Deep sequencing of microRNA precursors reveals extensive 3 ' end modification. *Rna* 17**,** 1795-1803.
14. Ameres SL, Hung JH, Xu J, Weng ZP, Zamore PD (2011) Target RNA-directed tailing and trimming purifies the sorting of endo-siRNAs between the two Drosophila Argonaute proteins. *Rna* 17**,** 54-63.
15. Li JJ, Yang ZY, Yu B, Liu J, Chen XM (2005) Methylation protects miRNAs and siRNAs from a 3 '-end uridylation activity in Arabildopsis. *Curr Biol* 15**,** 1501-1507.
16. Kamminga LM*, et al.* (2010) Hen1 is required for oocyte development and piRNA stability in zebrafish. *Embo J* 29**,** 3688-3700.
17. Jones MR*, et al.* (2009) Zcchc11-dependent uridylation of microRNA directs cytokine expression. *Nat Cell Biol* 11**,** 1157-U1258.
18. van Wolfswinkel JC*, et al.* (2009) CDE-1 Affects Chromosome Segregation through Uridylation of CSR-1-Bound siRNAs. *Cell* 139**,** 135-148.
19. Ibrahim F*, et al.* (2010) Uridylation of mature miRNAs and siRNAs by the MUT68 nucleotidyltransferase promotes their degradation in Chlamydomonas. *P Natl Acad Sci USA* 107**,** 3906-3911.

20. Ren GD, Chen XM, Yu B (2012) Uridylation of miRNAs by HEN1 SUPPRESSOR1 in Arabidopsis. *Curr Biol* 22**,** 695-700.
21. Zhao YY*, et al.* (2012) The Arabidopsis Nucleotidyl Transferase HESO1 Uridylates Unmethylated Small RNAs to Trigger Their Degradation. *Curr Biol* 22**,** 689-694.
22. Wyman SK*, et al.* (2011) Post-transcriptional generation of miRNA variants by multiple nucleotidyl transferases contributes to miRNA transcriptome complexity. *Genome Res* 21**,** 1450-1461.
23. Rhoades MW*, et al.* (2002) Prediction of plant microRNA targets. *Cell* 110**,** 513-520.
24. Mallory AC, Bartel DP, Bartel B (2005) MicroRNA-directed regulation of Arabidopsis AUXIN RESPONSE FACTOR17 is essential for proper development and modulates expression of early auxin response genes. *Plant Cell* 17**,** 1360-1375.
25. Llave C, Xie ZX, Kasschau KD, Carrington JC (2002) Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. *Science* 297**,** 2053-2056.
26. Millar AA, Gubler F (2005) The Arabidopsis GAMYB-like genes, MYB33 and MYB65, are MicroRNA-regulated genes that redundantly facilitate anther development. *Plant Cell* 17**,** 705-721.
27. Ibrahim F, Rohr J, Jeong WJ, Hesson J, Cerutti H (2006) Untemplated oligoadenylation promotes degradation of RISC-cleaved transcripts. *Science* 314**,** 1893-1893.
28. Kastenmayer JP, Green PJ (2000) Novel features of the XRN-family in Arabidopsis: evidence that AtXRN4, one of several orthologs of nuclear Xrn2p/Rat1p, functions in the cytoplasm. *P Natl Acad Sci USA* 97**,** 13985-13990.

29. Souret FF, Kastenmayer JP, Green PJ (2004) AtXRN4 degrades mRNA in Arabidopsis and its substrates include selected miRNA targets. *Mol Cell* 15**,** 173-183.

30. Zhao YY, Mo BX, Chen XM (2012) Mechanisms that impact microRNA stability in plants. *Rna Biol* 9**,** 1218-1223.

31. Curtis MD, Grossniklaus U (2003) A gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant physiology* 133**,** 462-469.

32. Xie M, Ren GD, Costa-Nunes P, Pontes O, Yu B (2012) A subgroup of SGS3-like proteins act redundantly in RNA-directed DNA methylation. *Nucleic Acids Res* 40**,** 4422-4431.

33. Zhai JX*, et al.* (2013) Plant MicroRNAs Display Differential 3'- Truncation and Tailing, Modifications Which Are ARGONAUTE1-Dependent and Conserved Across Species. *Plant Cell* 25**,** 2417-2428.

34. Yuan YR*, et al.* (2005) Crystal structure of A-aeolicus Argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. *Mol Cell* 19**,** 405-419.

35. Ameres SL*, et al.* (2010) Target RNA-Directed Trimming and Tailing of Small Silencing RNAs. *Science* 328**,** 1534-1539.

36. Yu B*, et al.* (2005) Methylation as a crucial step in plant microRNA biogenesis. *Science* 307**,** 932-935.

37. Song MG, Kiledjian M (2007) 3 ' Terminal oligo U-tract-mediated stimulation of decapping. *Rna* 13**,** 2356-2365.

38. Mullen TE, Marzluff WF (2008) Degradation of histone mRNA requires oligouridylation followed by decapping and simultaneous degradation of the mRNA both 5 ' to 3 ' and 3 ' to 5 '. *Gene Dev* 22**,** 50-65.

39. Rissland OS, Norbury CJ (2009) Decapping is preceded by 3 ' uridylation in a novel pathway of bulk mRNA turnover. *Nat Struct Mol Biol* 16**,** 616-U656.

40. Karimi M, Inze D, Depicker A (2002) GATEWAY((TM)) vectors for Agrobacterium-mediated plant transformation. *Trends Plant Sci* 7**,** 193-195.

41. Earley KW*, et al.* (2006) Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J* 45**,** 616-629.

42. Tanaka Y, Nakamura S, Kawamukai M, Koizumi N, Nakagawa T (2011) Development of a series of gateway binary vectors possessing a tunicamycin resistance gene as a marker for the transformation of Arabidopsis thaliana. *Bioscience, biotechnology, and biochemistry* 75**,** 804-807.

43. Ren GD*, et al.* (2012) Regulation of miRNA abundance by RNA binding protein TOUGH in Arabidopsis. *P Natl Acad Sci USA* 109**,** 12817-12821.

44. Park W, Li JJ, Song RT, Messing J, Chen XM (2002) CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in Arabidopsis thaliana. *Curr Biol* 12**,** 1484-1495.

CHAPTER 7

CONCLUSIONS

7.1 FDM1 and FDM2 are involved in RdDM

FDM1 and *FDM2* display a highly correlated expression pattern with known components of RdDM, such as AGO4, NRPE1, and RDR2. FDM1 and FDM2 act redundantly in DNA methylation, accumulation of Pol V-dependent rasiRNAs and silencing of RdDM loci. FDM1 and FDM2 mutants display reduced DNA methylation and siRNA levels. The results that FDM1 and FDM2 are not required for the accumulation of POL V- and POL II-dependent scaffold transcripts suggest that FDM1 and FDM2 may be involved in DNA methylation downstream of POL V transcription.

FDM1 and FDM2 are potential RNA-binding proteins with four domains: zinc-finger, XH, Coil-coil, and XS domain. To study the detailed function of FDM1 and FDM2 in RdDM, we studied the biochemical features of FDM1 and functions of each domain of FDM1. We found that FDM1 acts as a complex in RdDM. FDM1 interacts with both itself and IDN2. Gel filtration analysis suggests that FDM1 exists as a homodimer in a heterotetramer complex that may contain IDN2 *in vivo*. XH domain is necessary for the formation of FDM1 complex. The mutant FDM1 protein lacking its XH domain fails to form a complex and is unable to complement the DNA methylation defects of *fdm1-1 fdm2-1*, demonstrating that XH-domain mediated complex formation of FDM1 is required for its function in RdDM. FDM1 binds DNA *in vitro* through its coiled-coil domain. RNAs with 5' overhangs do not abolish the DNA binding ability of FDM1, indicating that FDM1 may bind both DNA and RNA simultaneously. Through functional analyses of FDM1 protein domains, this study extends our understanding on the RdDM pathway.

In order to furtherly study roles of FDM1 and FDM2 in RdDM, RNA immunoprecipitation and DNA immunoprecipitation following deep sequencing are necessary to determine the RNA and DNA targets of FDM1 and FDM2 *in vivo*. Additionally, this whole genome study is possible to unveil new RdDM targets. On the other hand, the crystal structure of FDM1 and FDM2 will be able to reveal the process of how FDM1 complex recognize and bind to dsRNA substrates from AGO4-siRNA-POL V transcript complex.

7.2 Functions of TOUGH and CDC5 are partially overlapped

We studied the function of TOUGH in miRNA biogenesis. We show that TOUGH (TGH) is an important factor for miRNA and siRNA biogenesis. Loss-of-function TOUGH in *tgh-1* reduces the activity of multiple DCLs *in vitro* and the accumulation of miRNA and siRNAs *in vivo*. The results that TOUGH associates with the DCL1 complex, binds pri-miRNAs and pre-miRNAs, and is required for the efficient *in vivo* interaction between pri-miRNA and HYL1 suggest that TGH assists DCLs to efficiently process and/or recruit the precursors of miRNAs and siRNAs.

We found that CDC5 is also involved in miRNA biogenesis but has a different role from TOUGH. Similar to TOUGH mutants, CDC5 mutants showed reduced miRNA levels. Our results suggest that CDC5 may have dual roles in miRNA biogenesis. The fact that CDC5 interacts with both the promoters of genes encoding *MIR* and POL II and positively regulates *MIR* transcription and the occupancy of Pol II at *MIR* promoters demonstrate that CDC5 is a transcription factor that regulates POL II transcription. On

the other hand, similar to TOUGH, CDC5 interacts with DCL1 and is required for efficient pri-miRNA processing, demonstrating that CDC5 acts as a component of the DCL1 complex to enhance pri-miRNA processing.

Besides miRNA, our results demonstrate that TOUGH and CDC5 are required for the accumulation of some siRNAs. However it is unclear whether TOUGH and CDC5 have a direct role in siRNA biogenesis or not. It is also possible that there is a cross talk between miRNA biogenesis and siRNA biogenesis pathway. In this way, TOUGH and CDC5 may indirectly regulate siRNA production. These two possibilities need to be examined in the near future.

7.3 AGO1 is required for HESO1-triggered miRNA uridylation and degradation.

Previously, our lab reported that in *Arabidopsis*, HESO1 uridylates 5' fragments to trigger their degradation. In this work, we show that AGO1 interacts with HESO1 through its PAZ and PIWI domains, which bind the 3' end of miRNA and cleave the target mRNAs, respectively. Furthermore, HESO1 is able to uridylate AGO1-bound miRNAs *in vitro* and miRNA uridylation *in vivo* requires a functional AGO1 in *hen1,* in which miRNA methylation is impaired, demonstrating that HESO1 can recognize its substrates in the AGO1 complex. Based on these results, we propose that methylation is required to protect miRNAs from AGO1-associated HESO1 activity that normally uridylates 5' fragments.

In our future work, homologs of HESO1 in *Arabidopsis* will be studied. We will test whether they have overlapping functions in miRNA uridylation and degradation or not. On the other hand, we will determine and compare the substrate preferences of these nucleotidyl transferases.