universität wien

DISSERTATION

Titel der Dissertation

RNA-directed DNA Methylation in Arabidopsis:
Do secondary siRNAs act in trans?
Atypical DNA methylation in genes encoding cysteine-rich peptides

Verfasser

Wanhui You, MSc.

angestrebter akademischer Grad

Doctor of Philosophy (PhD)

Wien, 2013

Studienkennzahl lt. Studienblatt:	A 094 490
Dissertationsgebiet lt. Studienblatt:	Molekulare Biologie
Betreuerin / Betreuer:	Dr. Marjori Matzke

Acknowledgements

First of all, I would like to thank my Ph.D supervisor Marjori Matzke for her giving me the opportunity to work in the lab, thanks for her constant encouragement, scientifically support throughout my Ph.D study, and special thanks for her patience during the revision of my thesis.

I would like to thank my Ph.D committee members, Werner Aufsatz, Hisashi Tamaru and Andrea Barta, for their constant support, useful discussions, and also for giving valuable feedback to the VBC Ph.D programme.

I would like to thank all the people in the Matzke Lab, both former and current members, for their nice help and providing a friendly and perfect scientific working environment. Thanks to Antonius Matzke for his technical assistance and discussions in the plasmids construction and establishments of transgenic plants. Also special thanks to him for helping me with the German abstract of my Ph.D thesis. Thanks to Tatsuo Kanno for his useful information and nice discussions when I arrived in the lab. Thanks to Agata Tyczewska and Matthew Spencer for their support and discussions in the CRP project. Thanks to Changho Eun, Taku Sasaki, and Zdravko Lorkovic for their daily scientific discussions and technical help, additionally to Zdravko for his valuable comments on my thesis. Thanks to Johannes van der Winden, Ulf Naumann, and also other colleagues from the institute for their German support during the life in Vienna.

I would also like to thank GMI/OEAW and FWF for financially supporting my Ph.D study in Gregor Mendel Institute of Plant Molecular Biology.

Finally, I would like to express special thanks to my parents, my wife for their love, encouragement, and constant support.

Zusammenfassung

RNA-dirigierte DNS Methylierung (RdDM) ist ein epigenetischer Prozess, der zuerst in Pflanzen aufgeklärt wurde, wobei doppelstängige RNA (dsRNA) in 24 nt siRNAs verarbeitet werden, um de novo Methylierung von sowohl symmetrischen als auch asymmetrischen Cytosinen, umgeben von jedweder Sequenz, an der homologen DNA Region zu bewirken.

Kapitel I: Im RNA-dirigierenden DNA Methylierungs (RdDM) Pfad führen kleine RNA Moleküle, die sogenannten siRNAs, Cytosin Methylierung an die passenden Stellen im Genom. An einigen Stellen kann durch die Produktion der sogenannten sekundären siRNAs, die für Ihre Biogenese auf Pol IV and RDR2 angewiesen sind, die Ausbreitung der Methylierung über den ursprünglich modifizierten Bereich hinaus erfolgen. Hier zeigen wir, dass sekundäre siRNAs nicht nur in cis an der Stelle wo sie hergestellt werden Methylierung bewirken können, sondern auch in trans an ungekoppelten homologen Zielstellen, sofern diese Sequenzen durch Pol II transkribiert werden um eine "nascent" RNA zu erzeugen. Pol II Transkription scheint die Amplifikation von den siRNAs an der Zielstelle zu unterstützen, vermutlich durch Erhöhung der siRNAs auf Niveaus, die ausreichend sind um RdDM zu induzieren. Nicht-transkribierte Zielstellen, die in der Anwesenheit von in trans wirkenden sekundären siRNAs unmethyliert verbleiben, können trotzdem von reichlichen Mengen von Haarnadel-abgeleiteten siRNAs, die unabhängig von Pol IV und RDR2 sind, methyliert werden. So scheint es, dass die Transkription einer Zielstelle durch Pol II für die siRNA abhängige DNA Methylierung nicht erforderlich ist, sondern eher für die Vervielfachung der siRNAs, höchst wahrscheinlich durch ihre bekannte Fähigkeit Pol IV zu rekrutieren. Die Wichtigkeit von Pol II Transkription für die siRNA Vervielfachung ist vielleicht besonders offensichtlich, wenn der ursprüngliche Auslöser, der in unseren Experimenten aus in trans wirkenden sekundären siRNAs bestand, in geringem Überschuss vorhanden ist.

Kapitel II: DNA Methylierung ist eine epigenetische Modifikation, die in der Stilllegung von Transposons beteiligt ist, um die Integrität des Genoms zu erhalten. In

Pflanzen werden Transposons durch CG, CHG, CHH Methylierung modifiziert, verursacht durch kleine interferierende RNA-dirigierter DNA Methylierung (RdDM). Im Gegensatz dazu haben ein Drittel der Protein-kodierenden Gene CG Methylierung, die siRNA unabhängig ist, in ihrer für Aminosäuren kodierenden Sequenz (Gen-Körper). Die Funktion dieser Gen-Körper Methylierung bleibt unbekannt. Hier berichten wir über atypische, für Transposon ähnliche DNA Methylierung im Gen Körper von einigen Genen die reich an Cystein sind (CRPs-Cystein reiche Peptide). Wenigstens ein Teil dieser Methylierung wird von RdDM verursacht. Die Transposon-ähnliche Gen Körper Methylierung ist wahrscheinlich mit der Stilllegung der CRP Gene in sporophytischen Geweben verbunden, da die Methylierung in Synergid Zellen, in denen die *CRP* Gene spezifisch exprimiert werden, verloren geht. Das Transposon-ähnliche Methylierungs-Muster einiger *CRP* Gene könnte auf einen Retrogen Ursprung hinweisen.

Summary

RNA-directed DNA Methylation (RdDM) is an epigenetic process first elucidated in plants, whereby double-strands RNAs (dsRNAs) are processed into 24 nt siRNAs to direct *de novo* DNA methylation of both symmetric and asymmetric cytosines in all sequence contexts at the homologous DNA region.

Chapter I: In the RNA-directed DNA methylation (RdDM) pathway, small RNAs (siRNAs) guide cytosine methylation at the matching region of the genome. At some loci, spreading of methylation beyond the originally modified region can occur through the production of so-called secondary siRNAs, which rely on Pol IV and RDR2 for their biogenesis. Here we show that Pol IV/RDR2-dependent secondary siRNAs can not only induce methylation in cis at the site where they are generated but also in trans at unlinked homologous target sites, provided these sequences are transcribed by Pol II to produce a 'nascent' RNA. Pol II transcription appears to support amplification of siRNAs at the target site, which presumably increases siRNAs to levels that are sufficient to induce RdDM. Non-transcribed target sites that remain unmethylated in the presence of trans-acting secondary siRNAs can nevertheless become methylated by abundant hairpin-derived siRNAs that are independent of Pol IV and RDR2. Thus, transcription of a target locus by Pol II does not seem to be required for siRNA-guided DNA methylation but rather for amplification of siRNAs, most likely through the known ability of Pol II to recruit Pol IV. The importance of Pol II transcription for siRNA amplification may be particularly apparent when the initial trigger, which in our experiments consisted of trans-acting secondary siRNAs, is present at low abundance.

Chapter II: DNA methylation is an epigenetic modification involved in transposon silencing to maintain genome integrity. In plants, transposons are modified by CG, CHG, CHH methylation which results from small interfering RNAs directed DNA methylation (RdDM). By contrast, about one third of protein-coding genes contain CG methylation, which is siRNA-independent, in their gene body. The function of this gene body methylation remains unknown. Here we demonstrate atypical, transposon like DNA methylation in the gene body region of several genes encoding cysteine rich

peptides (CRPs). At least part of this DNA methylation is due to RdDM. The transposon-like gene body methylation is likely associated with silencing of *CRP* genes in sporophytic tissues, because methylation is lost in synergid cells where *CRP* genes are specifically expressed. The transposon-like methylation pattern of some *CRP* genes may reflect a retrogene origin.

Table of Contents

1. Introduction

Epigenetics, a term that was first coined by Conrad Waddington (1905-1975) in 1942, was defined as "the branch of biology which studies the causal interactions between genes and their products, which bring phenotype into being" (Waddington, 1942). However, recently this definition is narrowed into the "the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence" by molecular biologists (Riggs *et al*. 1996). Nowadays, the majority of today's epigenetic research is focused on studying the epigenetic regulation of gene expression, which involves gene silencing, genome imprinting, reprogramming, and so on. These diverse epigenetic phenomena are primarily caused by DNA methylation, histone modification, and non-coding RNAs in different organisms.

In plants, the key components of epigenetic mechanisms in gene regulation are also including DNA methylation, histone modifications, and non-coding RNAs, which will be briefly introduced below.

1.1 DNA methylation and demethylation

DNA methylation is an epigenetic modification that involves the addition of a methyl group to the fifth carbon of the cytosine pyrimidine ring or the number 6 nitrogen of the adenine purine ring, which has been only to be found in bacteria, so far. In *Arabidopsis thaliana* (*Arabidopsis*) DNA methylation can be found in all sequence contexts CG, CHG, CHH, where H represents any nucleotide but guanine. These different sorts of DNA methylation are carried out or maintained by different DNA methyltransferases, including METHYLTRANSFERASE 1 (MET1), which is responsible for maintenance of CG methylation (Kankel *et al*., 2003), CHROMOMETHYLASE 3 (CMT3), a plant specific DNA methyltransferase whose function depends upon H3K9 (the lysine residue at position 9 of histone H3) methyltransferase KRYPTONITE (KYP), contributes to the maintenance of CHG

methylation (Jackson *et al*., 2002) and DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) is required for *de novo* DNA methylation, which is guided by the components of RNA directed DNA methylation (RdDM) pathway, also acting redundantly with CMT3 to maintain non-CG DNA methylation (Cao *et al*., 2002; Matzke and Birchler 2005). Additionally, two SWI2/SNF2 chromatin remodelling factors, DECREASE IN DNA METHYLATION 1 (DDM1) and DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1 (DRD1) are also found to play a role in regulating DNA methylation. DDM1，a ATP-dependent DNA helicase is required for maintenance of genomic DNA methylation (Jeddeloh *et al*., 1999), and DRD1 was shown to be functional not only in RdDM mediated *de novo* DNA methylation pathway, but also in the maintenance of non-CG DNA methylation (Kanno *et al*., 2004; Kanno *et al*., 2005).

Table 1. Genes that affect DNA methylation or demethylation in *Arabidopsis*.

Gene	AGI	Function	References
DRM2	At5g14620	*De novo* DNA methyltransferase, also maintenance of CHH methylation	Cao X *et al*, 2002
MET1	At5g49160	Maintenance of CG methylation	Kankel MW *et al*, 2003
CMT3	At1g69770	Maintenance of CHG methylation	Lindroth AM *et al*, 2001
DDM1	At5g66750	Maintenance of CG, CHG and CHH methylation	Jeddeloh *et al*., 1999
DRD1	At2g16390	*De novo* or maintenance of non-CG methylation	Kanno *et al*., 2004, 2005
ROS1	At2g36490	DNA demethyltrasferase	Gong Z *et al*, 2002
DME	At5g04560	DNA demethyltransferase	Choi Y *et al*, 2002
DML2	At3g10010	DNA demethyltransferase	Penterman *et al*., 2007
DML3	At4g34060	DNA demethyltransferase	Penterman *et al*., 2007

The table gives a brief summary of the major genes that are involved in *de novo* or maintenance DNA methylation and also DNA demethylation in *Arabidopsis*.

In *Arabidopsis*, DNA methylation is primarily enriched in transposable elements and other repetitive DNA sequences in the pericentromeric heterochromatin region to maintain genome integrity. However, about one-third protein-coding genes contain CG methylation in their gene body compared with only 5% CG methylation in the promoter region as uncovered by high throughput sequencing technology (Zhang *et al*., 2006; Cokus *et al*., 2008; Lister *et al*., 2008).

DNA methylation in *Arabidopsis* is also reversible, which includes both active and passive demethylation pathways. In *Arabidopsis*, active DNA demethylation depends on four DNA glycosylases: REPRESSOR OF SILENCING 1 (ROS1), DEMETER (DME), DME-LIKE 2 (DML2), and DME-LIKE 3 (DML3). *ROS1* was identified from a genetic screen and it functions as increasing gene silencing of a transgene construct and also some endogenous target genes (Gong *et al*., 2002). A histone acetyltransferase named INCREASED DNA METHYLATION (IDM1) has recently been shown to play a role in the *ROS1* mediated DNA demethylation pathway (Qian *et al*., 2012). *DME* is expressed in the central cell of the female gametophyte and is required for gene imprinting in central cells and also genome-wide demethylation in endosperm after fertilization (Choi *et al*., 2002). Recently DME was found to be functional not only in central cells, but also has been shown to play a role in active demethylation in male gametophyte, which contributes to the regulation of transposable elements (Schoft *et al*., 2011; Ibarra *et al*., 2012). DML2 and DML3 are two DNA glycosylases that can remove methylated cytosines from DNA by cleavage of the phosphodiester backbone (Penterman *et al*., 2007).

In addition to active DNA demethylation mediated by DME, passive loss of DNA methylation is also observed during plant gametophyte development. This takes place in the absence of DNA methyltransferases MET1 during DNA replication, resulting in a reduction of DNA methylation (Saze *et al*., 2003). Additionally, MET1 is observed to be down regulated, resulting in a loss of CG methylation in central cells during female gametophyte development (Jullien *et al*., 2008). Passive demethylation probably plays an important function in the establishment of gene imprinting. Genes that affect DNA methylation or demethylation in *Arabidopsis* are listed in Table 1.

1.2 Histone modifications

Chromatin is made up of nucleosomes, which consist of 147 base pairs of DNA wrapped around a histone octamer containing two copies of H2A, H2B, H3, and H4 (Luger *et al*., 1997). Chromatin based gene regulation includes histone modification, histone variants and histone remodelling. Histone modifications, including acetylation, methylation, phosphorylation etc. are important epigenetic markers for regulation of chromatin structure and gene expression. These post-translational modifications occur primarily on H3 and H4 tails at lysine (K) or arginine (R) residues, which are associated with both transcriptional activation and repression (Liu *et al*., 2010).

Histone methylation has been well studied in all kinds of organisms. It is catalyzed by histone methyltransferases (HMTs) and can take place on different lysines of H3 or H4, such as K4, K9, K27, K36, and K79 of histone H3, and K20 of histone H4 by adding one, two or three methyl groups (me1, me2 or me3), whereas arginine methylation has only mono- and di- states. All known HMTs have an evolutionarily conserved SET catalytic domain named after initially identified *Drosophila* HMTs, SUPPRESSOR OF VARIEGATION (Su (var) 3-9), ENHANCER OF ZESTE (E (z)) and TRITHORAX (TRX) (Jenuwein *et al*., 1998), except for DISRUPTOR OF TELOMERIC SILENCING (DOT1), which is a specific enzyme for H3K79 methylation first discovered in *Saccharomyces cerevisiae* (Feng *et al*., 2002). In *Arabidopsis*, histone H3K4 methylation mediated by two Trithorax group (TrxG) protein homologs *ARABIDOPSIS* TRITHORAX 1 (ATX1) and *ARABIDOPSIS* TRITHORAX 2 (ATX2) is shown to generally associate with transcriptional activation. ATX1 is required for H3K4 tri-methylation (Alvarez-Venegas *et al*., 2003), while ATX2 is responsible for H3K4 di-methylation (Saleh *et al*., 2008). By contrast, histone H3K9 and H3K27 methylation represent a transcriptional repressive state. H3K9 methylation is catalyzed by suppressor of variegation (Su (var) 3-9) homologs in *Arabidopsis*: KYP (also known as SUVH4), SUVH5 and SUVH6 (Jackson *et al*., 2002; Ebbs *et al*., 2005; Ebbs and Bender. 2006). It is usually linked with DNA methylation together to play a role in transcriptional gene silencing and maintenance of genome stability. For example, loss of DNA methylation in *met1* mutant leads to decreased

H3K9 di-methylation (Johnson *et al*., 2002; Soppe *et al*., 2002; Tariq *et al*., 2003). In the *kyp* mutant, loss of CHG methylation, which is catalysed by CMT3, is shown to be caused by the reduction of H3K9 di-methylation (Jackson *et al*., 2002; Malagnac *et al*., 2002).

By contrast, mono- and di-methylation of H3K27 are independent of DNA methylation; only H3K27 tri-methylation is dependent on CG methylation (Mathieu *et al*., 2005). H3K27 mono-methylation is catalysed by *ARABIDOPSIS* TRITHORAX RELATED PROTEIN 5 (ATXR5) and *ARABIDOPSIS* TRITHORAX RELATED PROTEIN 6 (ATRX6) (Jacob *et al*., 2009). H3K27 tri-methylation is thought to be catalysed by three Enhancer of Zeste (E (z)) homologs: CURLY LEAF (CLF), MEDEA (MEA), and SWINGER (SWN) in *Arabidopsis*, which are important factors in polycomb group protein complex 2 (PRC2) (Schubert *et al*., 2006; Grossniklaus *et al*., 1998; Katz *et al*., 2004). Although they are widely thought as H3K27 tri-methylation methyltransferases, direct biochemical evidence is still missing. In *Arabidopsis*, H3K27 tri-methylation mediated by distinct PRC2 complex containing MEA, CLF, and SWN, together with three Suppressor of Zeste (Su (z)) 12 homologs: FERTILIZATION INDEPENDENT SEED 2 (FIS2), EMBRYONIC FLOWER 2 (EMF2), and VERNALIZATION 2 (VRN2), which have been implicated in playing an important role during plant development (Kohler *et al*., 2003; Jullien *et al*., 2006; Goodrich *et al*., 1997; Henderson and Dean, 2004).

Histone demethylation has also been observed in *Arabidopsis*. INCREASE IN BONSAI METHYLATION 1 (IBM1), a Jumonji C (JmjC) domain–containing protein is shown to affect the H3K9 di-methylation and CHG methylation at the target gene site (Saze *et al*., 2008; Miura *et al*., 2009). Another JmjC domain containing protein RELATIVE OF EARLY FLOWERING 6 (REF6) is found to specifically demethylate H3K27 di- and tri-methylation, which has a function in regulating flowing time, and also some other developmental process (Noh *et al*., 2004; Lu *et al*., 2011).

In addition to histone methylation, histone acetylation (Aufsatz *et al*., 2002; Earley *et al*., 2006), and ubiquitination have also been shown to play an important role in epigenetic regulation in *Arabidopsis* (Sridhar *et al*., 2007). Genes that are involved in

histone modification in *Arabidopsis thaliana* are briefly summarized in Table 2.

Table 2. Genes involved in histone modification in *Arabidopsis*.

Gene	AGI	Function	References
ATX1	AT2G31650	H3K4 tri-methylation methyltransferase	Alvarez-Venegas *et al*., 2003
ATX2	AT1G05830	H3K4 di-methylation methyltransferase	Saleh *et al*., 2008
SUVH4	AT5G13960	H3K9 di-methylation Methyltransferase	Jackson *et al*., 2002
SUVH5	AT2G35160	H3K9 di-methylation Methyltransferase	Ebbs and Bender. 2006
SUVH6	AT2G22740	H3K9 di-methylation methyltransferase	Ebbs *et al*., 2005
IBM1	AT3G07610	H3K9 demethylase	Saze *et al*., 2008
ATXR5	AT5G09790	H3K27 mono-methylation methyltransferases	Jacob *et al*., 2009
ATXR6	AT5G24330	H3K27 mono-methylation methyltransferases	Jacob *et al*., 2009
CLF	AT2G23380	H3K27 tri-methylation methyltransferases	Schubert *et al*., 2006
MEA	AT1G02580	H3K27 tri-methylation methyltransferases	Grossniklaus *et al*., 1998
SWN	AT4G02020	H3K27 tri-methylation methyltransferases	Katz *et al*., 2004
REF6	AT3G48430	H3K27 demethylase	Lu *et al*., 2011
HDA6	AT5G63110	histone deacetylase	Aufsatz *et al*., 2002
SUP32/UBP26	AT3G49600	H2B deubiquitination enzyme	Sridhar *et al*., 2007

The table lists the genes responsible for the histone modification in *Arabidopsis*.

1.3 Non-coding RNAs

Non-coding RNAs are functional RNA molecules that cannot be translated into protein and they include both long and small non-coding RNAs. In *Arabidopsis*, small non-coding RNAs, which play an important role in regulation of many biological

processes, are mainly divided into two different classes: microRNAs (miRNAs) and small interfering RNAs (siRNAs).

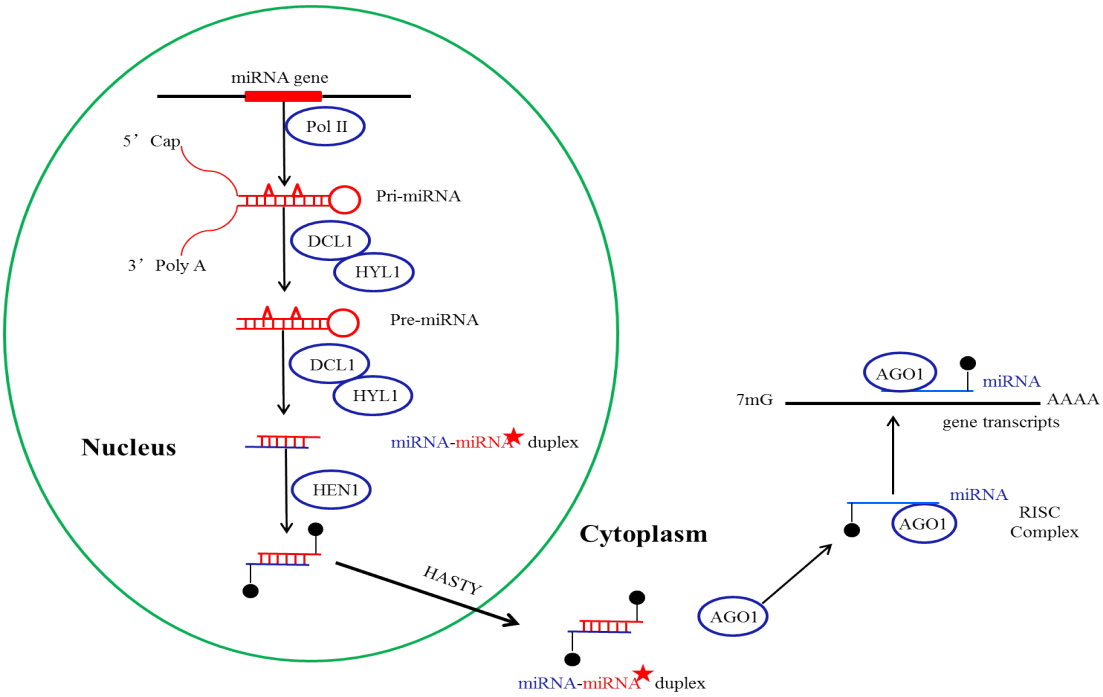

Figure 1. Biogenesis of miRNAs in *Arabidopsis*. The microRNA gene is transcribed by Pol II into pri-miRNA, which is a single strand that can be formed into an imperfectly matched hairpin structure with 5' capped and poly-adenylated in the 3' end. Pri-miRNA is processed into pre-miRNA, which is further processed into a miRNA duplex by DCL1 and HYL1. The duplex is methylated by HEN1, and then transferred to the cytoplasm in a HASTY dependent manner, where miRNA associates with AGO1 to form an RNA induced silencing complex (RISC) that directs either mRNA cleavage or translational inhibition.

miRNAs, usually are 21 nt RNAs, arising from non- protein coding miRNA genes, which generate single strand RNA (ssRNA) that base pairs to form an imperfectly matched hairpin structure called pri-miRNA, then cleaved by *DCL1* to form pre-miRNAs and miRNAs duplex (Park *et al*., 2002; Finnegan *et al*., 2003; Xie *et al*., 2003; Ramachandran and Chen. 2008). After processing into 21 nt siRNAs, a RNA methylase HUA-ENHANCER (HEN1) adds a methyl group to the 2′ hydroxyl groups at the 3′ ends of these miRNAs, which stabilize 21 nt miRNAs (Yu *et al*., 2005). The guide strand in the miRNAs duplex is loaded onto a protein complex named RNA-

induced silencing complex (RISC) which includes an Argonaute protein AGO1 to repress target genes by guiding mRNA degradation or translational inhibition through base-pairing (Vaucheret *et al*., 2004; Baumberger and Baulcombe. 2005; Chen. 2005) (Figure 1). The major functions of miRNAs have been shown in hormone biosynthesis, cell signalling, and morphogenesis during plant development (Chen. 2005; Chen. 2009).

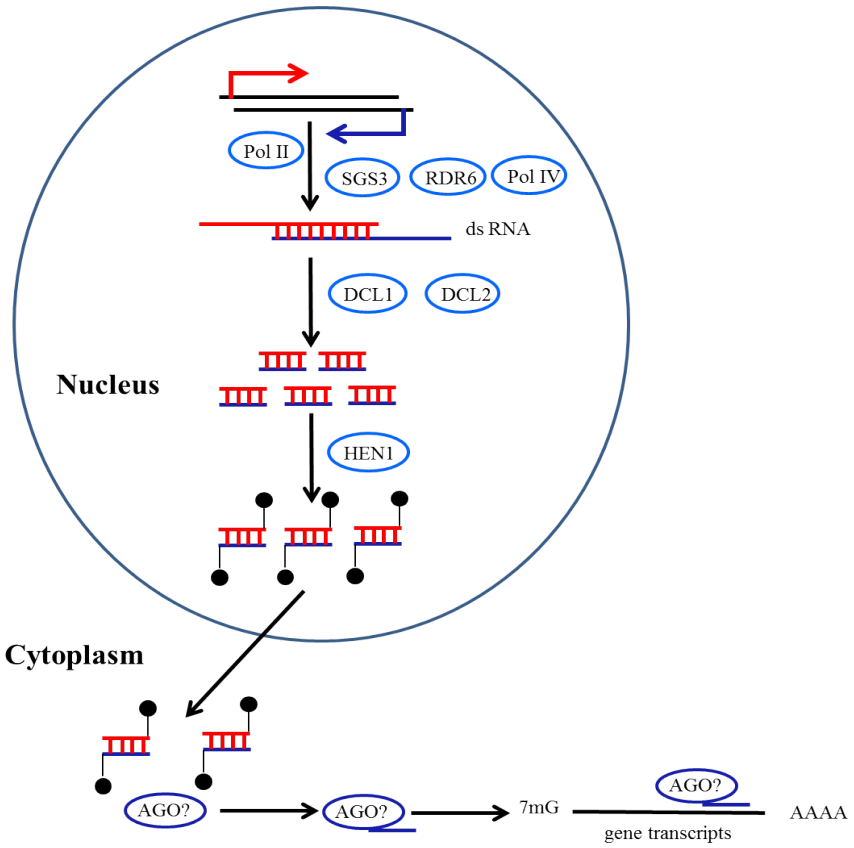

Figure 2. Biogenesis of nat-siRNAs in *Arabidopsis*. Natural antisense transcript-derived siRNAs (nat-siRNAs) derived from Pol II transcribed bidirectional transcripts, also requires SGS3, RDR6, Pol IV, DCL1 and DCL2 in *Arabidopsis*.

Nat-siRNAs arise from the loci that produce *cis*-natural antisense transcripts in the *Arabidopsis* genome. Sense and antisense transcripts transcribed by Pol II together with RDR6, SUPPRESSOR OF GENE SILENCING (SGS3), and also a plant specific

polymerase Pol IV can form dsRNAs, which are processed by DICER-LIKE 1 (DCL1) or DICER-LIKE 2 (DCL2) into nat-siRNAs, which play a role in abiotic or biotic stress response (Borsani *et al*., 2005; Katiyar-Agarwal *et al*., 2006) (Figure 2).

Trans-acting siRNAs (ta-siRNAs) comprise a special class of small RNA in *Arabidopsis*, which are generated from TAS (*Trans*-acting siRNA) transcripts cleaved by miRNAs. In brief, Pol II transcribed TAS transcripts are targeted by miRNAs, cleaved by a RISC complex including AGO1, and these short products are stabilized by SGS3. The cleaved fragments from TAS genes can be copied by RDR6 to form dsRNAs, which are processed by DCL4 to generate 21 or 22nt ta-siRNAs acting *in trans*, which regulate the target genes rather than their originating loci (Peragine *et al*., 2004; Vazquez *et al*., 2004; Allen *et al*., 2005; Yoshikawa *et al*., 2005) (Figure 3).

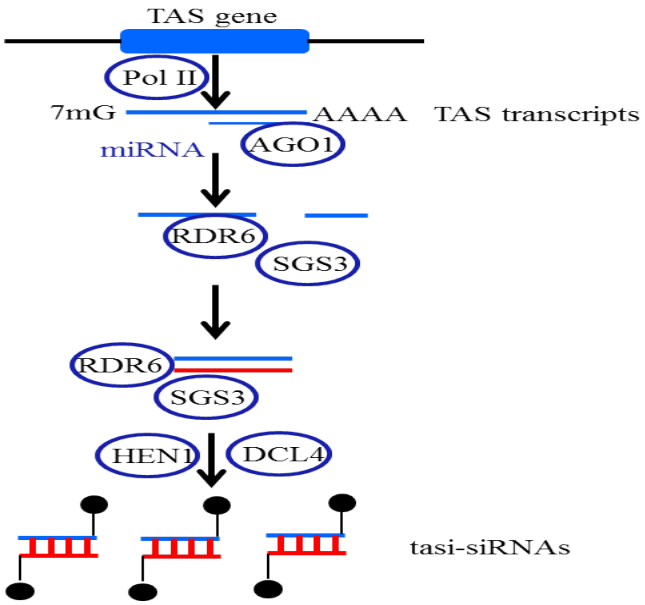

Figure 3. Biogenesis of trans-acting siRNAs in *Arabidopsis*. The transcripts transcribed by Pol II from the TAS loci are cleaved by miRNA, which can be used as a template by RDR6 to generated dsRNA, also requires SGS3. The dsRNA is processed into tasi-RNAs by DCL4.

Heterochromatin-associated 24 nt siRNAs are generated from dsRNAs mainly transcribed from transposable elements or other repetitive sequences. The 24 nt siRNAs

play an important role in the RdDM pathway to guide *de novo* DNA methylation in all sequence contexts at the homologous region of the genome. In the current hypothetical model, RdDM is divided into two parts that are controlled by two different plant specific polymerases, Pol IV and Pol V. The first part is biogenesis and amplification of siRNAs, which is controlled by the plant specific polymerase Pol IV. The second part is *de novo* DNA methylation, which requires another plant specific polymerase Pol V (Law and Jacobsen 2010; Haag and Pikaard 2011). The biogenesis and amplification of 24nt siRNAs requires Pol IV (produce single strand RNA), *RDR2* (generate double strand RNA), *DCL3* (process dsRNA into 24 nt siRNAs) (Herr *et al*., 2005; Kanno *et al*., 2005; Onodera *et al*., 2005; Pontier *et al*., 2005; Xie *et al*., 2004; Henderson *et al*., 2006; Daxinger *et al*., 2009). After being processed into 24 nt siRNAs, the RNA methylase HEN1 adds a methyl group to the 2′ hydroxyl groups at the 3′ ends of these siRNAs, stabilizing 24 nt siRNAs (Yu *et al*., 2005). Afterwards, one strand from 24nt siRNAs binds to AGO4, which interacts with another plant specific polymerase Pol V through a conserved WG/GW domain (Zilberman *et al*., 2003; Qi *et al*., 2006; Li *et al*., 2006; El-Shami *et al*., 2007).

In the process of *de novo* DNA methylation, Pol V transcribes intergenic non-coding regions, assisted by the DDR complex (Law *et al*., 2010), including *DRD1* (Kanno *et al*., 2004; Kanno *et al*., 2005), DEFECTIVE IN MERISTEM SILENCING 3 (DMS3) (Kanno *et al*., 2008), RNA-DIRECTED DNA METHYLATION 1 (RDM1) (Gao *et al*., 2010). DEFECTIVE IN MERISTEM SILENCING 4 (DMS4, also known as RDM4 (RNA-DIRECTED DNA METHYLATION 4)) might also be involved in generating non-coding transcripts that base-pair with siRNAs associating with AGO4 (Kanno *et al*., 2010; He *et al*., 2009). This interaction is stabilized by IDN2 (Ausin *et al*., 2009) and KOW DOMAIN-CONTAINING TRANSCRIPTION FACTOR 1 (KTF1 or SPT5L (SUPPRESSOR OF TY INSERTION 5-LIKE)), which interacts with AGO4 through WG/GW domain (Bies-Etheve *et al*., 2009; He *et al*., 2009). In addition to the Pol V dependent non-coding transcripts, recent findings show non-coding transcripts from some loci that can be transcribed by RNA polymerase Pol II, which act as a scaffold to recruit RdDM components, such as AGO4, DMS4, RDM1. Thus, Pol II also plays a role in RdDM pathway at some loci (Zheng *et al*., 2009). Furthermore, the

RDM1 in the DDR complex also interacts with both AGO4 and DRM2, which indicates this association might assist *DRM2* to find the target loci (Gao *et al*., 2010) (Figure 4).

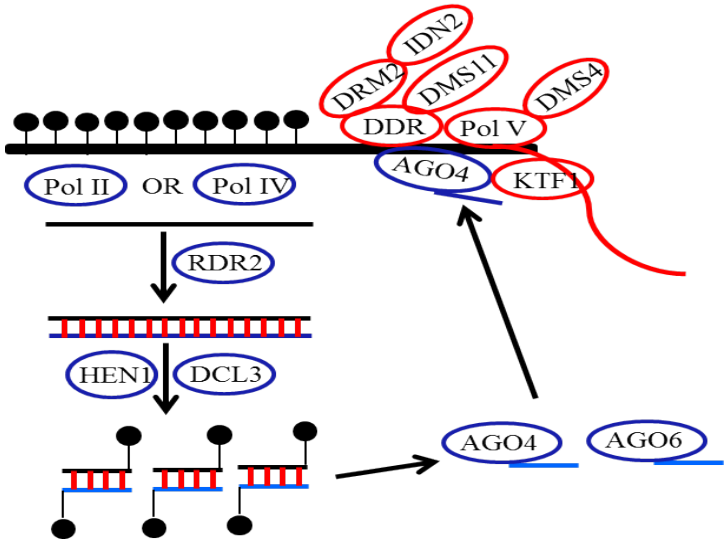

Figure 4. Heterochromatin associated 24 nt siRNAs induced DNA methylation. Methylated region (indicated by black solid ball) in transposons or repeat elements recruit Pol IV or Pol II to generate single strand RNA, which can be transcribed into dsRNA by RDR2. The dsRNA is subsequently cleaved by DCL3 into siRNAs, associating with AGO4 or AGO6, interact with RdDM components to direct DRM2 to methylate the target sequences.

1.4 Aim of thesis

1.4.1 Chapter 1: Do secondary siRNAs act in *trans*?

RNA-directed DNA Methylation (RdDM) is an epigenetic process first elucidated in plants (Wassenegger *et al*., 1994; Mette *et al*., 1999; Mette *et al*., 2000), whereby double-stranded RNAs (dsRNAs) are processed into 24 nt siRNAs to direct *de novo* DNA methylation of both symmetric and asymmetric cytosines in all sequence contexts at the homologous DNA region.

Our lab has established a two component transgene system comprising a target transformant line (T line) and a double transformant line containing both target and

silencer constructs (T+S line) in *Arabidopsis*. To identify factors involved in the RdDM pathway, we performed a forward genetic screen using the T+S line (Kanno *et al*., 2008; Daxinger *et al*., 2009; Naumann *et al*., 2010; Eun *et al*, 2011; Lorković *et al* 2012; Eun *et al*., 2012). In addition to the mutants we identified in our forward genetic screen (reviewed by Eun *et al*., 2012), a previous study carried out in our lab using the T+S system also uncovered an interesting 'methylation spreading' phenomenon involving the stepwise production of Pol IV/RDR2-dependent secondary siRNAs (Daxinger *et al.*, 2009). These secondary siRNAs arose from the sequence directly downstream of the target enhancer region that becomes methylated in *trans* by Pol II-transcribed hairpin-derived siRNAs (primary siRNAs) originating from the silencer locus. By contrast, the Pol IV/RDR2-dependent secondary siRNAs are made in *cis* and they act in *cis* to induce DNA methylation of the downstream region (Daxinger *et al.*, 2009). An interesting question not answered by Daxinger and coworkers is whether the secondary siRNAs, similarly to the RNA polymerase II (Pol II) transcribed hairpin-derived siRNAs, can also act in *trans* to trigger DNA methylation of homologous DNA sequences. Additionally, it remains another important question whether RdDM is based on an RNA-RNA or RNA-DNA interaction (Wassenegger *et al*., 1994; Jones *et al*., 1999; Wierzbicki *et al*., 2008).

Therefore, the first part in my thesis is trying to answer these questions mentioned above in RdDM pathway.

1.4.2 Chapter 2: Atypical DNA methylation in genes encoding cysteine-rich peptides

To identify targets of RdDM in *Arabidopsis*, methylation-sensitive amplification polymorphism (MSAP) has been performed by former Ph.D student Lucia Daxinger in our lab (Daxinger dissertation 2008, You *et al*., 2012). One fragment identified in this assay was found to be a satellite-like sequence in Repbase named ATSAT5, which is defined by a 2,196 bp long consensus sequence (Jurka. 2006). Based on a further BLAST analysis, four euchromatic copies of ATSAT5 fragments were identified

throughout the *Arabidopsis* genome. Although SAT5 was originally thought to be a non-protein coding repeat, our further analysis showed that genes encoding small cysteine-rich peptides (CRPs) and also *CRP* pseudogenes are embedded in the SAT5 repeat monomer in the annotation of the *Arabidopsis* genome sequence. Preliminary data from bisulfite sequencing showed that the euchromatic *CRP* gene embedded in SAT5 fragment on chromosome 4 was targeted by RdDM components. Further small RNA Northern blot and deep sequencing of small RNAs confirmed that the *CRP* gene on chromosome 4 is the target of RdDM components. However, DNA methylation pattern in the *CRP* gene is found to be different from the typical gene body DNA methylation, which carries all sequence contexts CG, CHG, CHH methylation instead of only CG methylation (Zhang *et al*., 2006, Cokus *et al*., 2008, Lister *et al*., 2008). This atypical DNA methylation in the gene body region of the *CRP* gene is more like the pattern seen in transposable elements.

Therefore, my aim in this part of the study is to investigate why the *CRP* gene has a special DNA methylation pattern in the gene body region, what is the putative function of this unusual DNA methylation, and also the function of the *CRP* gene in *Arabidopsis*.

Chapter I------ Do secondary siRNAs act in-trans?

2.1 Results of Chapter I

DNA methylation induced by secondary siRNAs in *trans*

Our lab has established a two component transgene system comprising a target transformant line (T line) and a double transformant line containing both target and silencer constructs (T+S line) (Figure 1). To identify factors involved in the RdDM pathway, we performed a forward genetic screen using the T+S line (Kanno *et al*., 2008; Daxinger *et al*., 2009; Naumann *et al*., 2010; Eun *et al*, 2011; Lorković *et al* 2012; Eun *et al*., 2012).

In addition to the thirteen mutants we identified in our forward genetic screen (reviewed by Eun *et al*., 2012), a previous study carried out in our lab using the T+S system also uncovered an interesting 'methylation spreading' phenomenon involving the stepwise production of Pol IV/RDR2-dependent secondary siRNAs (Daxinger *et al.*, 2009). These secondary siRNAs arose from the sequence directly downstream of the target enhancer region that becomes methylated in *trans* by Pol II-transcribed hairpin-derived siRNAs (primary siRNAs) originating from the silencer locus. By contrast, the Pol IV/RDR2-dependent secondary siRNAs are made in *cis* and they act in *cis* to induce DNA methylation of the downstream region (Daxinger *et al.*, 2009) (Figure 1). An interesting question not answered by Daxinger and coworkers is whether the secondary siRNAs, similarly to the RNA polymerase II (Pol II) transcribed hairpin-derived siRNAs, can also act in *trans* to trigger DNA methylation of homologous DNA sequences.

To investigate this question, two new transgene constructs were assembled. These constructs were designed to include 88 bp of the downstream region that is targeted for methylation by *cis*-acting secondary siRNAs. Following introduction of these constructs into the original T+S line, thus producing so-called 'triple transformants', this 88 bp

segment could be used as a new target sequence to test whether the Pol IV/RDR2-dependent secondary siRNAs could also act in *trans* to induce DNA methylation. A feature of the two constructs is that they were designed to test whether transcription of the 88 bp target sequence was required for siRNA-induced DNA methylation. This was accomplished by either including or not including the 88 bp sequence in a Pol II

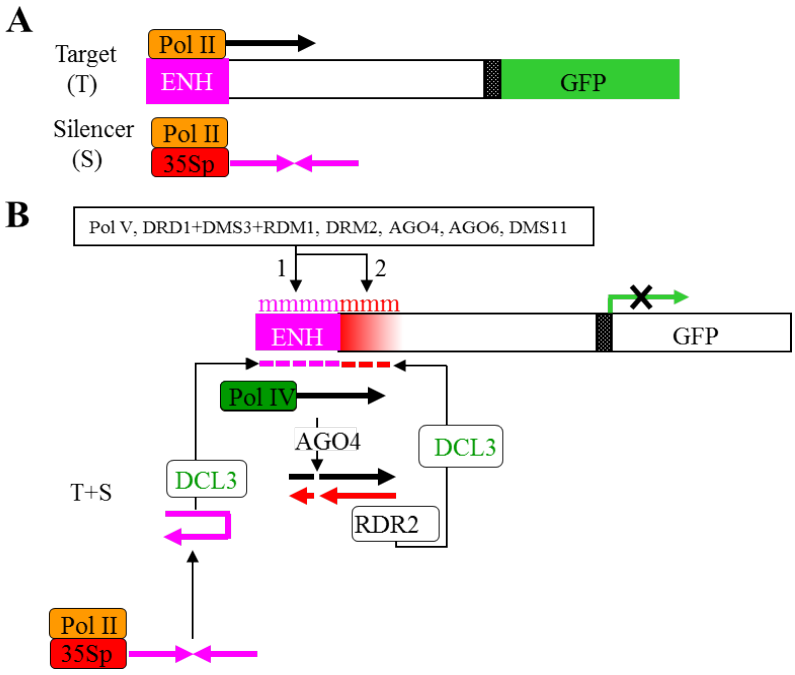

Figure 1 Two-component (T+S) transgene silencing system and model for stepwise production of Pol IV/RDR2-dependent secondary siRNAs. (A) The target (T) construct contains an enhancer (distal region that is targeted for methylation is shown in pink) upstream of a minimal promoter (hatched) and a gene encoding green fluorescent protein (GFP). The silencer (S) construct contains an inverted DNA repeat (IR) of the target enhancer region downstream of the 35S promoter of cauliflower mosaic virus. Transcription of the IR by Pol II produces a hairpin RNA that is cleaved into hairpin-derived primary small interfering RNAs (siRNAs), which induce *de novo* DNA methylation at the homologous target region in *trans*. **(B)** T+S: The hairpin RNA encoded at the silencer locus is processed by DCL3 into primary siRNAs (pink dashes), which associate with the Pol V complex to methylate the target region (primary methylation: pink 'm') and induce transcriptional silencing of the *GFP* reporter gene. In the proposed stepwise pathway of production of secondary siRNAs, the methylated region is hypothesized to recruit Pol IV to transcribe a single stranded 'nascent' RNA, which is cleaved by the slicer activity of AGO4 into a template that can be used by RDR2 to generate dsRNA. The resulting Pol IV/RDR2-dependent dsRNA is processed by DCL3 into secondary siRNAs (red dashes), which associate with RdDM complex to induce secondary DNA methylation (red 'm') at the homologous region in *cis* (Daxinger *et al*., 2009).

transcription unit. In the first construct, the 88 bp fragment was placed upstream of the maize ubiquitin promoter (Ubi pro) followed by a gene encoding dsRED fluorescent protein (88bp-Ubi pro-dsRed) (Figure 2A, bottom). In this construct, the 88bp fragment is not included in the transcription unit that is under the control of the ubiquitin promoter. In the second construct, the 88 bp fragment is located between ubiquitin promoter and *dsRED* gene, in which case it can be transcribed together with *dsRED* by Pol II from the ubiquitin promoter (Ubi pro-88bp-dsRed) (Figure 2A, top). The sequence of the 88bp fragment is shown in Figure 2B.

The reason that I assembled these two different constructs is to potentially test whether RdDM involves a siRNA-DNA or siRNA-RNA interaction (Wierzbicki *et al.*, 2008). Therefore, the former construct was named as 'SD' (siRNA-DNA interaction), the latter construct was named as 'SR' (siRNA-RNA interaction) (Figure 2A). Each construct was introduced individually into the original T line and the T+S line to produce new double transformants and triple transformant (TT) lines, respectively (Figure 2C).

Transformants with a single copy of the respective construct containing the 88 bp target fragment (as determined by a 3:1 segregation of the linked phosphinothricin-resistance marker) were used for further analysis. To analyze DNA methylation of the 88 bp fragment in the 'SR' or 'SD' constructs, six individual transformants from the T1 generation, which are hemizygous for insertion of the 'SR' or 'SD' construct, were used for bisulfite sequencing analysis. In the lines containing either the 'SR' or 'SD'construct and only the original T locus, no DNA methylation of the 88 bp target sequence was detected (data not shown). This result was expected because the T locus alone does not produce secondary siRNAs in the absence of the original S locus, which is required to produce hairpin-derived primary siRNAs that trigger DNA methylation at the enhancer region, presumably leading to recruitment of Pol IV pathway components that generate secondary siRNAs (Daxinger et al., 2009) (Figure 1).

For the new TT lines, which contain both the original T and S loci – thus allowing synthesis of secondary siRNAs (Figure 1) - as well as either the 'SD' or 'SR' construct, differing results were obtained. In TT lines containing the 'SR' construct, no

methylation in the 88 bp sequence could be detected from any of the six individual transformants tested (data not shown). By contrast, in TT lines bearing the 'SD'

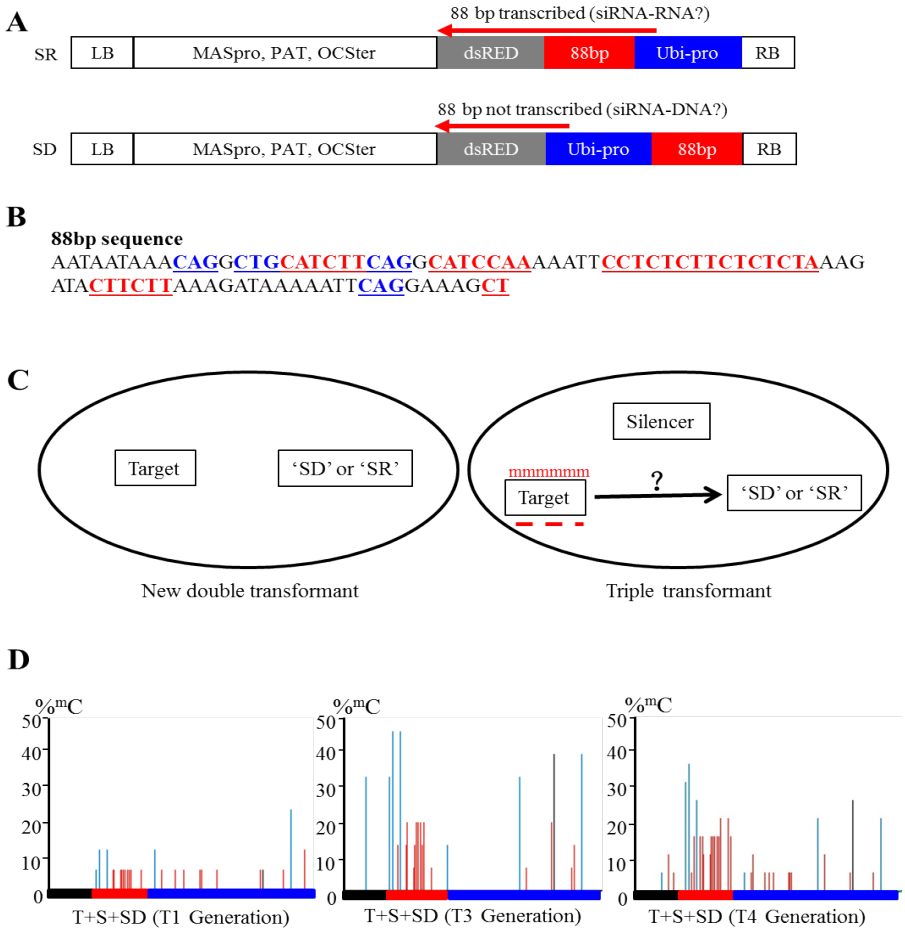

Figure 2 Schemes of 'SR' and 'SD' constructs and DNA methylation analysis of 'SD' in triple transformants. (A) The 'SR' construct carries a ubiquitin promoter driving transcription of the 88bp fragment (targeted by secondary siRNAs) together with a gene encoding dsRED. In the 'SD' construct, the 88bp target sequence is located upstream of the ubiquitin promoter followed by the gene encoding dsRed. Both constructs were inserted into the MPO binary vector. (B) Sequence of the 88 bp fragment and sequence context of cytosine (C) residues. CHG sites are shown in blue bold characters, and CHH sites are shown in red bold characters. **(C)** Schemes describing the new double transformant containing the original T (Target) locus and either the 'SD' or 'SR' construct (left), and the triple transformant (TT) that contains T and S (silencer) as well as either the 'SD' or 'SR' loci (right). **(D)** The graphs show the percentage of DNA methylation at individual Cs in the TT line 4 (TT-4) at the T1, T3 and T4 generations. The black box represents the right border of the MPO binary vector, the red box represents the 88 bp target region and the blue box represents the ubiquitin promoter region. The black lines indicate CG methylation, blue lines indicate CHG methylation and red lines indicate CHH methylation. The results from at least 15 cloned sequences are shown.

construct, low levels of methylation in the 88 bp sequence were detected in three of the six lines analyzed. This methylation generally increased in T3 and T4 generations (which are homozygous for the 'SD' insertion) and although it was concentrated in the 88 bp target region, methylation tended to spread to some extent downstream of the 88 bp target sequence (Figure 2D). The different results obtained with the 'SR' and 'SD' constructs suggest that the 88 bp sequence is resistant to methylation when it is within the protein-coding transcript initiating from the ubiquitin promoter ('SR' construct) but is susceptible to methylation when it is not within this transcription unit ('SD' construct).

Pol II transcribed 'nascent' RNAs are detected from methylated 'SD' plants

Bisulfite sequencing results indicated that DNA methylation spreads from the 88 bp target sequence into the downstream ubiquitin promoter region in the TT lines carrying the 'SD' construct. This finding was reminiscent of the methylation-spreading phenomenon involving Pol IV/RDR2-dependent secondary siRNAs production that was observed in our previous T+S double transformant system (Daxinger *et al.*, 2009). Therefore, I designed primers to check for transcripts originating from the 88 bp target sequence and the flanking region by RT- PCR (Figure 3A). Transcripts were indeed detected in the two lines that displayed methylation in the 88 bp target sequence (TT-1 and TT-4), but not in the other two lines that remained unmethylated (TT-2 and TT-3) (Figure 3B).

Thus, methylation of the 88 bp target region in the 'SD' containing TT lines was correlated with the presence of a non-coding 'nascent' RNA extending through the 88 bp target region. The 'nascent' RNA is presumably transcribed by Pol II because it was still detectable in mutants defective in either the largest subunit of Pol IV or of Pol V (*nrpd1* and *nrpe1*, respectively) (Figure 3C).

Although 'nascent' RNA can be detected in *nrpd1* and *nrpe1* mutants, CHG and

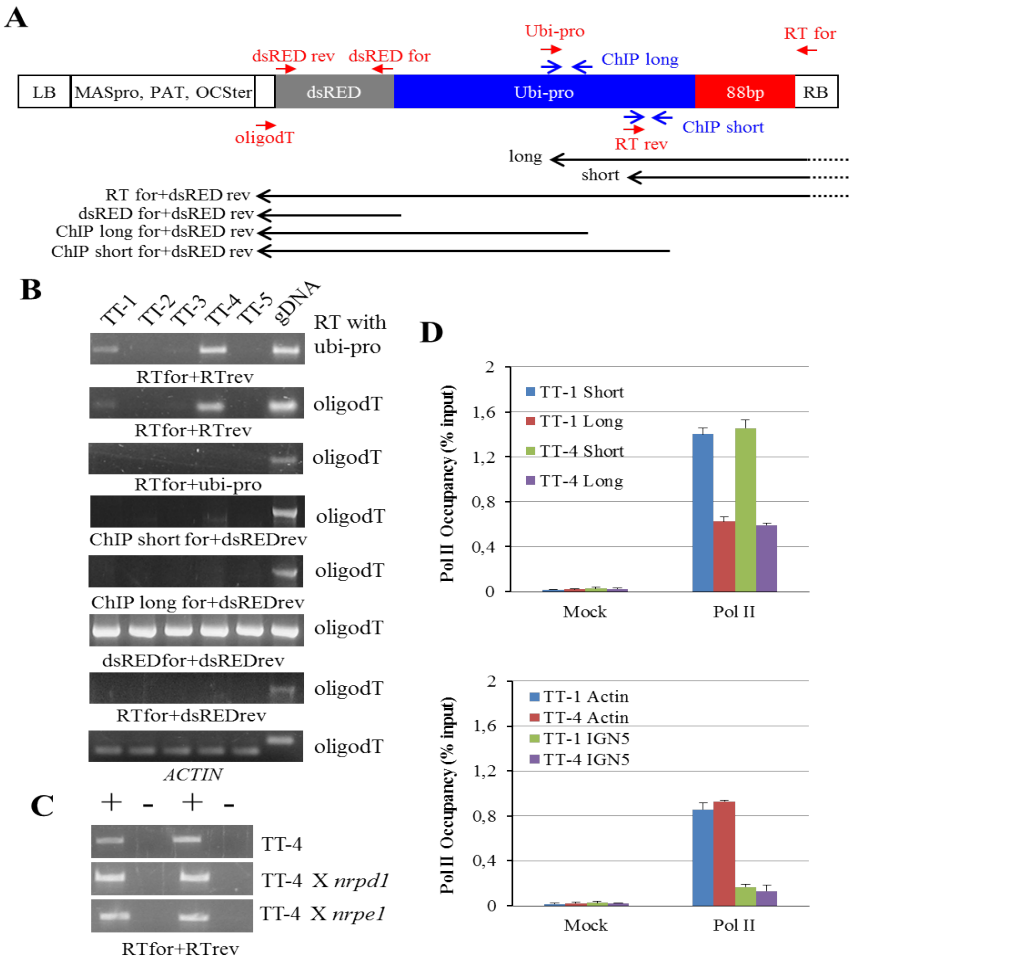

Figure 3 Analysis of 'nascent' RNA in triple transformant lines carrying the original T and S loci as well as the 'SD' construct. **(A)** Schematic presentation of 'SD' construct with indicated positions of primer sets used for 'nascent' RNA detection by RT-PCR (as shown in black solid lines) and ChIP analysis of Pol II occupancy (Blue primer sets). **(B)** 'Nascent' RNA detection in TT line containing the 'SD' construct with different primer sets shown in (A). The short product is included in the methylated region induced by *trans*-acting secondary siRNAs, which is amplified with RT forward plus RT reverse. The long product is partially overlapping with the short product, which is amplified with RT forward plus Ubi-pro oligo. dsRED transcripts were detected with dsREDfor and dsRED rev. However, no read-through transcriptional products were detected with the primer sets RT for, ChIP short for, and ChIP long for plus dsRED rev, genomic DNA was used as a positive control. *Actin* (At3g18780) is shown as a constitutively expressed control. **(C)** 'Nascent' RNA detection with Ubi-pro oligo in TT line 4 after introgressing in *nrpd1* and *nrpe1* mutations. '-' lanes indicate minus RT control. **(D)** Pol II occupancy at the region transcribed to produce nascent RNA. 'Short' indicates the region in the short product amplified from nascent RNA RT-PCR (ChIP primer used is bottom blue arrow). 'Long' indicates the region in the long product that cannot be amplified from nascent RNA RT-PCR (ChIP primer used is top blue arrow). ChIP was performed by using anti-Pol II antibody and Pol II co-purified DNA was quantified by real-time PCR. Two biological replicates were performed and Standard deviations were calculated from three technical repeats.

CHH methylation at the 88 bp target region, which is presumably induced by *trans*-acting secondary siRNAs synthesized at the original T locus in the presence of the S locus, was dramatically reduced in the *nrpd1* and *nrpe1* mutants (Figure 4). Because the secondary siRNAs are below detection levels in these two mutants (Kanno et al., 2008; Daxinger et al., 2009), these results support a role for the secondary siRNAs inducing methylation in *trans* of the 88 bp target region. They also substantiate the notion that the 'nascent' RNA itself is not sufficient to induce or maintain the DNA methylation.

The continued presence of the 'nascent RNA' in *nrpd1* and *nrpe1* mutants indicates that this RNA is transcribed by a polymerase other than Pol IV or Pol V, the best candidate being Pol II. To test this, I performed chromatin immunoprecipitation (ChIP) to determine whether Pol II can be detected at the region that produces 'nascent'

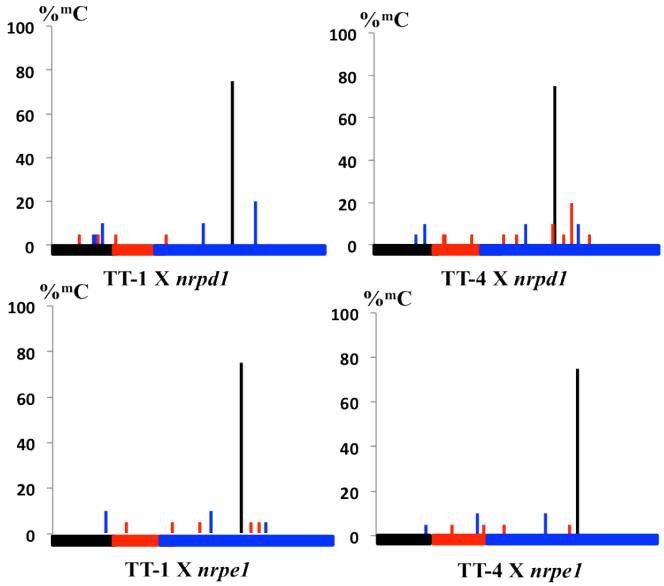

Figure 4 Loss of non-CG methylation induced by *trans*-acting secondary siRNAs in *nrpd1* and *nrpe1*mutants. The graphs show the percentage of DNA methylation at individual cytosines in TT lines containing the 'SD' constructs after introgressing in *nrpd1* and *nrpe1* mutations. The black box represents the right border of MPO binary vector, the red box represents the 88bp region and the blue box represents the ubiquitin promoter region. The black lines indicate CG methylation, blue lines indicate CHG methylation and red lines indicate CHH methylation. The results from at least 15 cloned sequences are shown.

RNAs. This analysis showed that Pol II is bound to the shorter region more than to the longer region (Figure 3D). Taken together, these findings suggest that a Pol II-generated non-coding 'nascent' transcript across the 88 bp target region is required for methylation of this region by *trans*-acting Pol IV/RDR2-dependent secondary siRNAs.

Hairpin-derived siRNAs can induce DNA methylation at targets that cannot be methylated by secondary siRNAs

It was interesting that not all of the TT lines containing the 'SD' construct could be methylated in the presence of *trans*-acting secondary siRNAs. As described above, methylation of the 88 bp target sequence was only observed in two out of four of these lines: TT-1 and TT-4, but not in TT-2 and TT-3. To test whether the 88 bp target sequence in the 'SD' construct of lines TT-2 and TT-3 can in principal become methylated in the presence of siRNAs or whether this sequence was only resistant to methylation by the *trans*-acting Pol- IV/RDR2 dependent secondary siRNAs, I backcrossed the TT-2 line with non-transgenic plants to segregate away the original T and S loci. I then introduced into a plant containing only the 'SD' construct a new transgene construct containing an inverted DNA repeat of the 88 bp sequence under the control of the cauliflower mosaic virus 35S promoter. The resulting Pol II-generated hairpin RNA can be processed into primary siRNAs that trigger RdDM (Mette *et al*., 1999; Mette et al., 2000).

Strikingly, a subsequent DNA methylation analysis showed that the 88 bp target sequence in the TT-2 line can indeed be methylated in the presence of the hairpin-derived primary siRNAs. This DNA methylation was more concentrated in the 88 bp sequence region compared with DNA methylation triggered by Pol IV/RDR2 dependent secondary siRNAs in the TT-1 and TT-4 lines (Figure 5, compare with Figure 2B). Additionally, no 'nascent' RNA was detected in the methylated TT-2 line (Figure 6A), indicating that methylation of the 88 bp target region by hairpin-derived primary siRNAs does not require a corresponding a Pol II-transcribed nascent RNA. By contrast, as shown above, such a nascent RNA was detected in the two TT lines in which the 88

bp target sequence could be methylated by *trans*-acting secondary siRNAs (TT-1 and TT-4, Figure 3B). These results demonstrate that the 88 bp target region in the TT-2 line is not in principal recalcitrant to methylation by *trans*-acting siRNAs. However, even though this methylation does not require detectable amounts of a Pol II-generated nascent RNA, it may require some special feature associated with hairpin-derived primary siRNAs that is not found with secondary siRNAs, for example, high abundance or presence of additional size classes (21-nt and/or 22-nt) in addition to the DCL3-generated 24-nt siRNAs (discussed further below).

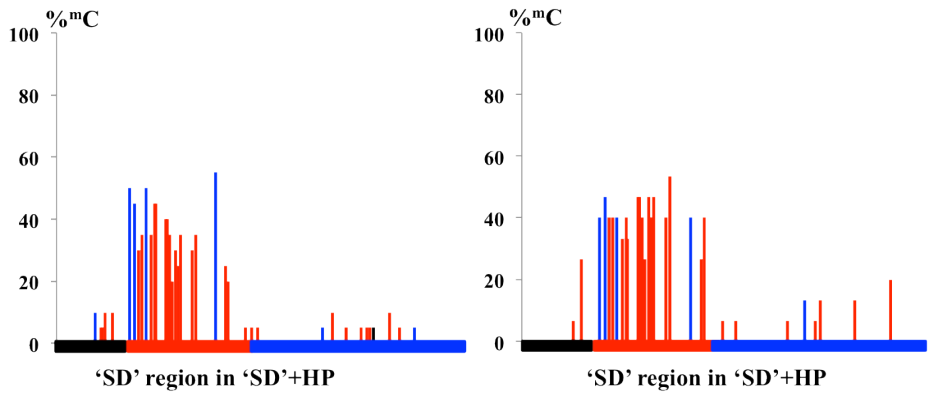

Figure 5 DNA methylation of the 88 bp target region in the 'SD' construct in line TT-2 as induced by hairpin-derived siRNAs. The graphs show the percentage of DNA methylation, incuded by *trans*-acting, hairpin-derived siRNAs, at individual cytosine residues in the 'SD' construct. Left, DNA methylation in the T1 generation; right, DNA methylation in the T3 generation. The black box represents the right border of MPO binary vector, the red box represents the 88bp target region and the blue box represents the ubiquitin promoter region. The black lines indicate CG methylation, blue lines indicate CHG methylation and red lines indicate CHH methylation. The results from at least 15 cloned sequences are shown.

Presence of nascent RNA at the 88 bp target sequence correlates not only with methylation by *trans*-acting secondary siRNAs but also with amplification of secondary siRNAs

To examine siRNAs in the TT lines, I performed a Northern blot analysis using a probe specific for the 88 bp region. Interestingly, 24-nt siRNAs were present at higher levels in the two 'SD' lines (TT-1 and TT-4) in which the 88 bp target region became

methylated in the presence of *trans*-acting secondary siRNAs and which produced the 'nascent' RNA comprising the 88 bp sequence (Figure 6B, 6C) than in the other two unmethylated 'SD' lines (TT-2 and TT-3). In TT-2 and TT-3, the intensity of the 24-nt siRNA signal was approximately the same as that observed for secondary siRNAs in the original T+S line (Figure 6B, 6C). These observations suggest amplification of the siRNAs homologous to the 88 bp target sequence in the methylated TT-1 and TT-4 lines. If so, one hypothesis is that the 'nascent' RNA contributes to the Pol IV/RDR2-

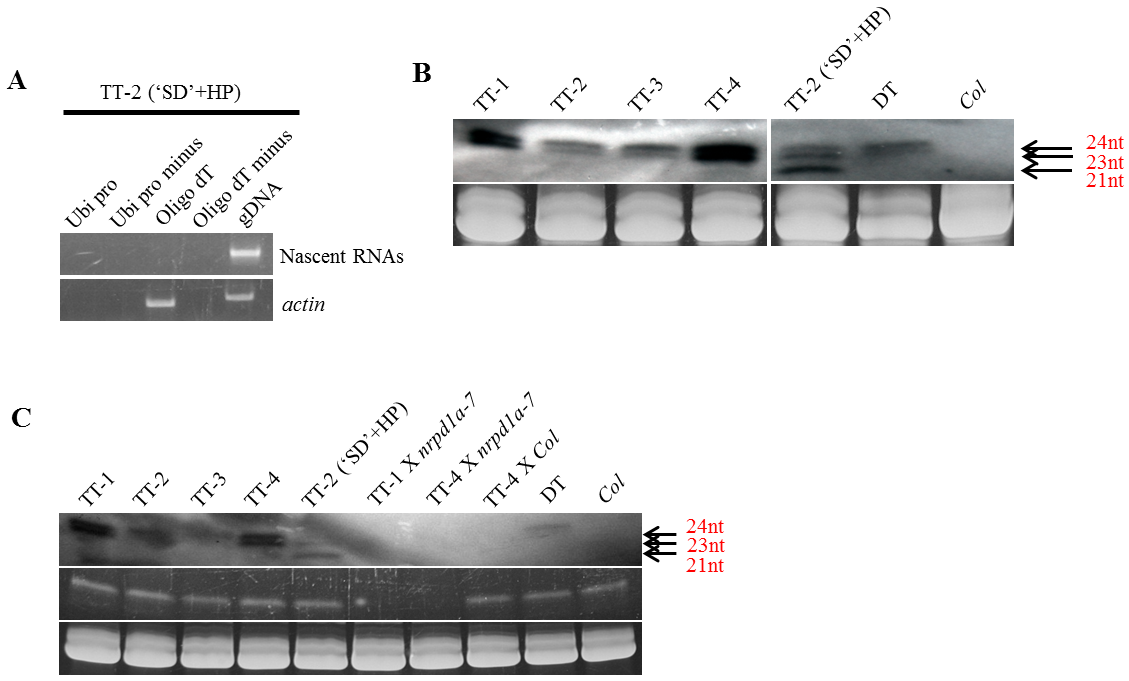

Figure 6 RNA profiles in 88 bp hairpin transgenic line and TT lines. (A) No 'nascent' RNA can be detected by RT-PCR in the TT-2 line (introgressed by 88 bp hairpin construct) that contains heavy DNA methylation in the 88bp target region. RT-PCR has been performed with both Ubi-pro oligo and oligo dT. 'Nascent' RNA (top) and actin control (bottom). **(B)** Northern blots of secondary siRNAs in TT line (23, 24-nt siRNAs indicated by black arrow in the right) and hairpin-derived siRNAs (21-, 23-, 24-nt siRNAs indicated by black arrow in the right) in the new hairpin transgene line (top). Original T+S transgenic line was used as positive control. *Col* ecotype was used as a negative control. Ethidium bromide staining of the major RNA on the gels is shown as a loading control (bottom). **(C)** No secondary siRNAs can be detected in the crossed plants between methylated TT-1, TT-4 and *nrpd1*, and also TT-4 and *Col* ecotype without T and S loci by Northern blots (top). No endogenous 24-nt siRNAs can be detected in the *nrpd1* mutant crossed lines (middle). Ethidium bromide staining of the major RNA on the gels is shown as a loading control (bottom).

dependent amplification of siRNAs, which probably can trigger DNA methylation at the 88 bp sequence region in *cis* to form a methylation loop. Interestingly, in an *nrpd1* mutant background, 'nascent' RNA could be detected, but no secondary siRNAs were observed on Northern blots (Figure 6C). Previous DNA methylation data showed that no CHG and CHH methylation was found in the *nrpd1* mutant line, which is consistent with this Northern blot result.

If a 'methylation loop' is indeed established at the 88 bp region such that siRNAs are amplified in *cis*, then the amplified siRNAs and DNA methylation may still be detected when the 88 bp target is segregated away from the original T and S loci. To obtain plants containing only the 'SD' construct and lacking the original T and S loci, I backcrossed the TT-4 plant with wild-type *Col-0* and screened for progeny containing only the 'SD' construct. Although the 'nascent' RNA in these plants is still detectable (Figure 7A), no secondary siRNAs were detected in the backcrossed line (Figure 6C). There was also less DNA methylation detected in the 88bp target region in the 'SD' construct in the same plant (Figure 7B, compared with Figure 2D T4 generations).

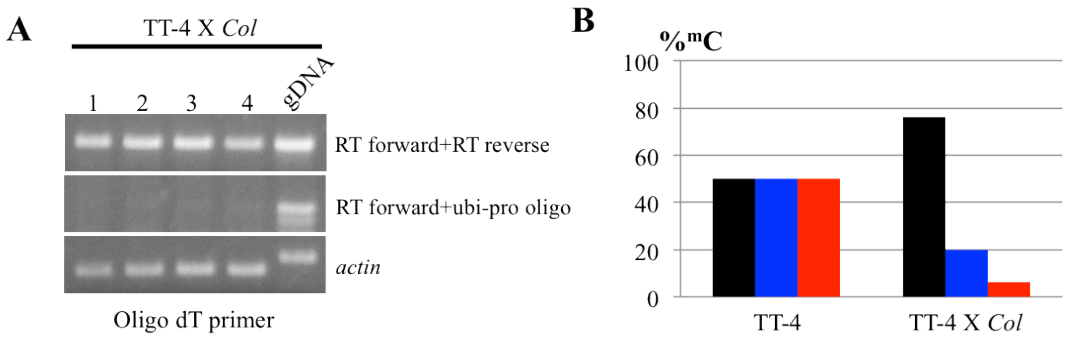

Figure 7 'Nascent' RNA and DNA methylation analysis in plants obtained by crossing TT-4 and *Col.* (**A**) 'Nascent' RNA is still detectable in the backcrossed plant between TT-4 and *Col*. *Actin* (At3g18780) is shown as a constitutively expressed control. (**B**) The graphs show the comparison of overall levels of methylation in CG (Black), CHG (blue) and CHH (red) nucleotide between TT-4 and the backcrossed plant. The level of DNA methylation in all sequence contexts from TT-4 line was set as a standard value 50% in this experiment. The level of DNA methylation in TT-4 X Col was calculated based on the standard value.

Taken together, our results suggest that the use of the 'nascent' RNA to amplify siRNAs is probably triggered by DNA methylation in the 88 bp sequence region, indicating 'nascent' RNA acts downstream of the secondary siRNAs induced DNA methylation process. In this hypothetical model, once the 88 bp region by *trans*-acting secondary siRNAs is methylated, the 'nascent' RNA acts together with Pol IV pathway components to generate in *cis* more siRNAs that maintain methylation in the 88 bp region (Figure 8).

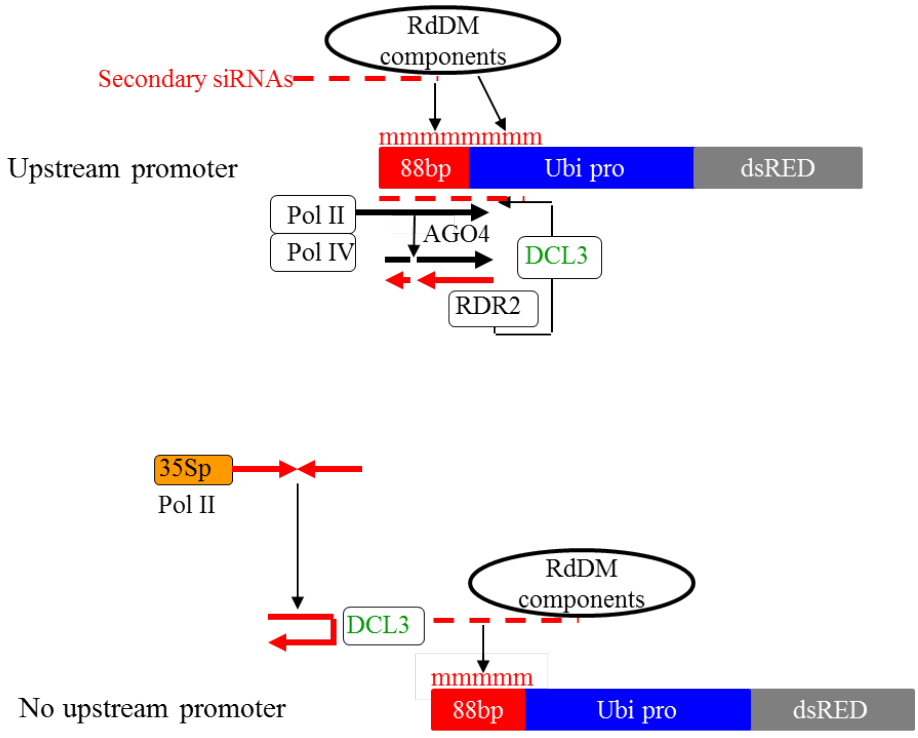

Figure 8 Model of chromosomal localization hypothesis in *trans*-acting secondary siRNAs induced DNA methylation. In the methylated TT lines (TT-1 and TT-4), the 'SD' construct is hypothetically inserted into the downstream of an unknown promoter in flanking plant DNA (top). 'Nascent' RNA is transcribed by Pol II under the control of the unknown upstream promoter. Once the 88 bp target region becomes methylated by *trans*-acting secondary siRNAs (red dashes above) generated from the original T locus, the 'nascent' RNA acts together with Pol IV pathway components including AGO4, RDR2, DCL3 to form a amplification loop of siRNAs (red dashes below), which associate with RdDM components, acting in *cis* to trigger DNA methylation at the homologous region. By contrast, in the unmethylated TT-2 line, because of no promoter located in plant DNA upstream of the 'SD' construct, 88 bp sequence can only be target by hairpin-derived siRNAs (bottom).

The hairpin-derived siRNAs detected in my study were 21-, 23- and 24 nt in length, which is somewhat different from 21-, 22-, and 24 nt reported previously (Daxinger *et al*., 2009). Strikingly, the amount of 23- and 24 nt siRNAs were as abundant as the Pol IV/RDR2-dependent secondary siRNAs in the plants that cannot be methylated (Figure 6B, 6C). These results suggest that differences in abundance cannot explain the different abilities of *trans*-acting secondary siRNAs and hairpin-derived siRNAs to induce methylation of the 88 bp target sequence in the TT-2 line. As an alternate hypothesis, it is possible that 21-nt siRNAs might also have a function within a mixed-size population of hairpin-derived siRNAs to induce DNA methylation (Wu *et al*., 2012, Pontier *et al*., 2012).

Can DNA methylation at the 88 bp target region of 'SR' be induced by Pol II transcribed hairpin-derived siRNAs?

The results of the DNA methylation analysis described above demonstrated that hairpin-derived primary siRNAs can efficiently induce DNA methylation at the 88 bp region in the TT-2 containing the 'SD' construct. By contrast, these regions could not be methylated by Pol IV/ RDR2-dependent *trans*-acting secondary siRNAs, perhaps because they lacked a 'nascent' RNA comprising the 88 bp sequence (see above). Therefore, it was interesting to know whether the 88 bp region in the 'SR' construct, which was resistant to methylation in the presence of *trans*-acting secondary siRNAs in all TT lines tested, can also be methylated by *trans*-acting siRNAs produced from Pol II-transcribed hairpin RNAs.

To investigate this question, I introduced the 88 bp hairpin construct, in which an inverted repeat of the 88 bp sequence is under the control of the 35S promoter, into a new double transformant carrying both the original T locus (but not the S locus) and the 'SR' construct, thus producing a new triply transformed line. Bisulfite sequencing results demonstrated that the 88 bp region at the original T locus, which becomes methylated by *cis*- acting secondary siRNAs generated by Pol IV and RDR2 in the presence of the S locus (Daxinger *et al*., 2009) (Figure 1), can also acquire at least some

methylation induced by the 88 bp hairpin-derived primary siRNAs in the absence of the S locus. The T locus thus served as an internal control to demonstrate that the hairpin-derived siRNAs could indeed induce methylation of a homologous target sequence. By contrast, in the same plant, the 88 bp region in the 'SR' construct acquired substantially less methylation (Figure 9). Thus, hairpin-derived siRNAs, which have been shown to be capable of inducing methylation of a homologous target sequence are less able to trigger methylation of the 88 bp target sequence when it is part of a Pol II transcript encoding a protein (dsRED) in the 'SR' construct. These results may be relevant for the question of how Pol II-transcribed protein-coding genes are protected from RdDM.

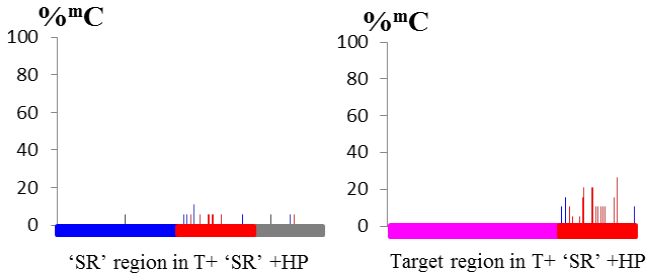

Figure 9 DNA methylation of 88 bp target sequence in the 'SR' and T locus in a triply transformed line containing the original T locus, 'SR' construct and construct encoding 88bp hairpin RNA. The graphs show the percentage of DNA methylation at individual C residues in the 88 bp target sequence of the 'SR' construct and the T locus induced by *trans*-acting hairpin-derived primary siRNAs. In the left graph depicting the 'SR' construct, the blue box represents the ubiquitin promoter region, the red box represents the 88bp target region and the grey box represents the gene encoding dsRED. In the right graph representing the T locus, the pink box represents the target enhancer region and the red box represents the 88bp downstream region. The black lines indicate CG methylation, blue lines indicate CHG methylation and red lines indicate CHH methylation. The results from at least 15 cloned sequences are shown.

3.1 Discussion of chapter I

Function of Pol II transcribed 'nascent' RNA in DNA methylation induced by *trans*-acting secondary siRNAs

Our studies show that Pol IV/RDR2-depenedent secondary siRNAs, which are able to induce methylation in *cis* at the site where they are generated, are also able to act in *trans* to induce methylation of an unlinked homologous target sequence, provided this sequence is transcribed by Pol II to produce a non-coding, 'nascent' RNA. Furthermore, we show that the transcription of 'nascent' RNA is not dependent on the DNA methylation in the 88 bp target region of the 'SD' construct. More interestingly, 'nascent' RNA is still detected in either an *nrpd1* or *nrpe1* mutant, or in the backcrossed plants (TT-4x*Col*), suggesting that Pol II, probably recruited by an upstream promoter in flanking plant DNA, is required for the 'nascent' RNA transcription. Additionally, more siRNAs are generated in the methylated plants triggered by *trans*-acting secondary siRNAs (TT-1 and TT-4), in which 'nascent' RNA is detected, compared with the unmethylated TT lines (TT-2 and TT-3) or original double transgenic line (T+S). These data suggest that 'nascent' RNA contributes to the production/amplification of siRNAs.

A chromosomal localization hypothesis has been formulated to describe how the 'nascent' RNA is transcribed and the function of 'nascent' RNA in the process of DNA methylation triggered by *trans*-acting secondary siRNAs (Figure 8 in the Results). In the hypothetical model, 'nascent' RNA transcribed by Pol II, which is recruited by an unknown upstream promoter in flanking plant DNA at the T-DNA insertion site, is required for the amplification of siRNAs (Vaucheret 2005). DNA methylation at the 88bp region is probably required for the processing of Pol II transcribed 'nascent' RNA into siRNAs, which act downstream to initiate an siRNA amplification loop. These amplified siRNAs act in *cis* to induce DNA methylation of the 88 bp target region, thus forming a positive feedback loop that contributes to the *de novo* methylation or maintenance methylation at the 88bp target region in the 'SD' construct. By contrast, in the TT-2 lines, due to their lacking amplified siRNAs from 'nascent' RNA, no DNA

methylation at the 88bp target region in the 'SD' construct can be induced by secondary siRNAs in TT-2 or TT-3. However, the same target that cannot be methylated by secondary siRNAs is shown to be methylated by hairpin-derived siRNAs, indicating that hairpin-derived siRNAs are more efficient than the Pol IV/RDR2 dependent siRNAs in inducing RdDM. Furthermore, no 'nascent' RNA is detected in the methylated plants, which supports our hypothesis that described above.

However, it remains another possibility that the 'SD' construct is inserted into a heterochromatic region of the plant genome where it is easier to recruit RdDM components (or they are more enriched) that associate with *trans*-acting secondary siRNAs to trigger DNA methylation at the homologous region (Figure 1). In TT-1 and TT-4, DNA methylation at 88 bp target region is induced by *trans*-acting secondary siRNAs probably based on a chromosomal localization effect (Simmer *et al*., 2010). The methylated region might recruit Pol IV, acting together with RDR2 and DCL3 to initiate an amplification loop of siRNAs, which is independent of Pol II transcription (Herr *et al*. 2005; Kanno *et al*. 2005; Onodera *et al*. 2005; Pontier *et al*. 2005; Vaucheret 2005), which can enforce the DNA methylation at the 88 bp target region in the 'SD' construct. Concerning the detection of 'nascent' RNA but no DNA methylation in the backcrossed plants (TT-4 x Col) and methylated TT lines introgressed into either an *nrpd1* or *nrpe1* mutant background, these data indicate that the polymerase in charge of the transcription of 'nascent' RNA might be switched into Pol II from Pol IV or Pol V.

In *Schizosaccharomyces pombe*, Pol II is required for siRNAs processing and heterochromatin assembly at the pericentromeric region (Kato *et al*., 2005). Additionally, a recent study using Pol V chromatin immunoprecipitation with massively parallel sequencing showed that majority of 24 nt siRNA biogenesis and CHH methylation is coupled with Pol V binding regions in *Arabidopsis* (Wierzbicki *et al*., 2012). However, Pol V binding sites that are not associated with DNA methylation or siRNAs biogenesis are still detected at some loci (Wierzbicki *et al*., 2012), which provides evidence that Pol II might be functional instead of Pol V in these regions (Zheng *et al*., 2009). Our work shows that 'nascent' RNA transcribed from a non-coding region by Pol II plays a role in the biogenesis of siRNAs, which are crucial for

de novo or maintenance DNA methylation at the target region.

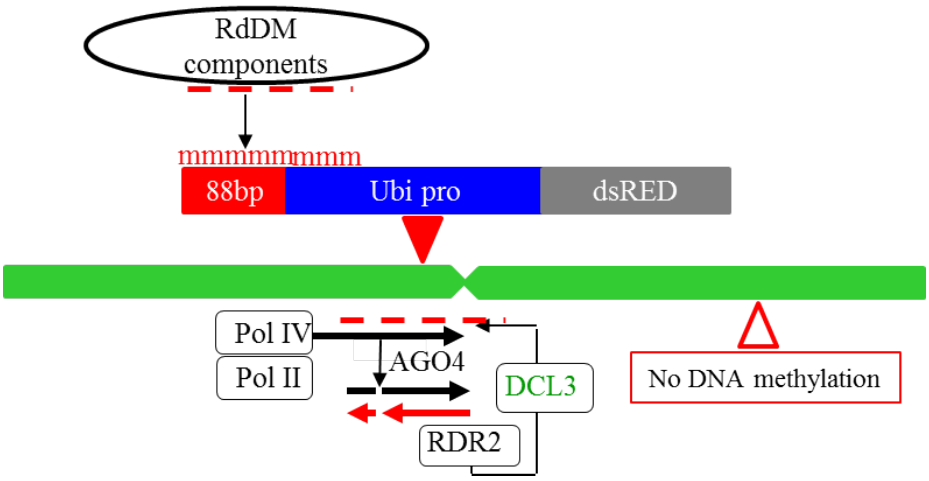

Figure 1 Model of chromosomal localization hypothesis. The scheme describes that *trans*-silencing depends on the insertion site of the 'SD' construct in *Arabidopsis* genome. *Trans*-silencing is only efficient when the target gene (shown as red solid triangle) is located near heterochromatic region, if the target gene locates in a euchromatin region (shown as red hollow triangle), no DNA methylation can be induced by *trans*-acting secondary siRNAs. Secondary siRNAs generated from target locus (red dashes) associate with Pol V complex to induce DNA methylation at the 88bp target region of the 'SD' construct in *trans*, methylated sequences recruit Pol IV to initiate a transcription of nascent RNA, which generates more siRNAs together with AGO4, RDR2, and DCL3 to form a positive feedback loop that maintains the DNA methylation.

RNA-DNA or RNA-RNA interaction based RNA directed DNA methylation pathway

The phenomenon of RdDM is first discovered in the transgenic tobacco plants carrying potato spindle tuber *viroid* (PSTVd) cDNAs (Wassenegger *et al*., 1994). The replicating viroid RNA is shown to trigger DNA methylation on its own cDNAs probably based on recognition of RNA-DNA hybrid, suggesting an RNA-DNA interaction based RdDM process (Wassenegger *et al*., 1994; Wassenegger, 2000).

Furthermore, a transgene promoter was shown to be targeted by dsRNA, becoming methylated and resulting in transcriptional gene silencing, which provides experimental evidence that RdDM pathway require siRNAs generated from dsRNA (Mette *et al*., 1999; Mette *et al*., 2000). These studies also indicate that RdDM potentially is functional through a siRNAs-DNA interaction.

Recent studies in *Arabidopsis* identified two plant-specific RNA polymerases Pol IV and Pol V that play key roles in RdDM (Herr *et al*., 2005, Kanno *et al*., 2005, Onodera *et al*., 2005, Pontier *et al*., 2005). Therefore, an alternative hypothesis suggests that RdDM occurs through a RNA-RNA or RNA-DNA interaction based on the function of Pol V in RdDM pathway. In the currently proposed model, 24 nt siRNAs associating with AGO4 either interact with Pol V transcripts, which serve as a scaffold to recruit RdDM components that guide DNA methylation at the homologous DNA region (RNA-RNA interaction) or directly base pair with the homologous DNA (RNA-DNA interaction), recruiting RdDM components to induce DNA methylation (Wierzbicki *et al*., 2008). However, further or more direct evidence in support of one of the two models needs to be investigated in the future.

In our transgenic lines, both Pol IV/RDR2-dependent secondary siRNAs and hairpin-derived siRNAs are shown to induce DNA methylation at the transgenic target sequence, which is not included in Pol II transcription unit rather than within transcription unit. These findings provide indirect evidence that recognition of a target sequence, in which DNA methylation can be triggered by Pol IV/RDR2 dependent siRNAs, is based on an RNA-DNA interaction, which is consistent with the previous finding that DNA methylation in the transgene was found to be associated with an RNA-DNA interaction (Jones *et al*., 1999).

Role of 21 nt siRNAs in RdDM pathway

24 nt siRNAs are usually considered the most important siRNAs involved in RdDM, which induces CG, CHG and CHH methylation. Interestingly, several recent, independent studies have shown that 21 nt siRNAs might be also correlated with

activation of transposable elements or DNA methylation at some loci (Slotkin *et al*., 2009; Dowen *et al*., 2012; Wu *et al*., 2012; Pontier *et al*., 2012). These studies revealed that the activation of *Athila* family LTR retrotransposon is associated with 21 nt siRNAs rather than 24 nt siRNAs in pollen (Slotkin *et al.*, 2009). Additionally, a striking increase of 21 nt siRNAs, but not 24 nt siRNAs, is observed during the transcriptional activation of some transposons under the stress of salicylic acid (Dowen *et al*., 2012). However, detailed mechanisms about the function of 21 nt siRNAs in the regulation of transposable elements remain unclear. Here, we provide evidence suggesting that hairpin-derived 21-, 23-, 24 nt siRNAs can methylate the 88 bp target region that cannot be methylated by Pol IV/RDR2 dependent *trans*-acting 24 nt siRNAs. This finding indicates that Pol II-transcribed hairpin-derived siRNAs are more efficient than Pol IV/RDR2 dependent secondary siRNAs in inducing RdDM. Considering that the abundance of 23-24 nt siRNAs is almost the same in the populations of hairpin-derived siRNAs and Pol IV/RDR2 dependent siRNAs, our findings suggest that 21nt siRNAs might also play a role in triggering DNA methylation at 88bp target region in the 'SD' construct, which is consistent with the recent studies in TAS loci and NERD target loci (Wu *et al*., 2012, Pontier *et al*., 2012). Because of *de novo* DNA methylation is primarily triggered by 24 nt siRNAs, these findings suggest a new role of 21 nt siRNAs that probably required for guiding DNA methylation at some genome loci. However, the detailed mechanism by which 21 nt siRNAs are involved in DNA methylation remains need to be addressed.

Chapter II------Atypical DNA methylation in genes encoding cysteine-rich peptides

2.2 Results of chapter II

CRP genes embedded in ATSAT5 repeats as RdDM targets

To identify targets of RNA-directed DNA methylation (RdDM) in *Arabidopsis thaliana* (*Arabidopsis*), methylation sensitive amplification polymorphism (MSAP) has been performed by former Ph.D student Lucia Daxinger in our lab. Two different restriction enzymes were used. One is methylation sensitive (*Eco*T22I: recognition sequence ATGCAT) and the other is methylation-insensitive (*Mse*I: recognition sequence AATT). This allowed us to identify genomic regions that are differentially methylated at asymmetric CHH sites in wild-type plants compared to a mutant defective in the largest subunit of Pol V (*nrpe1*) (Daxinger 2008; You *et al*., 2012). One fragment identified in this assay was found to be a satellite-like sequence in Repbase named

Table.1 Single dispersed copies of SAT5 repeats

Chromosome	Sequence context	Gene coordinate	Length(bp)	CRP gene embedded
2	intergenic	7845025-7844080	946	At2g18042 pseudogene
3	transposon-rich	14681750-14683927	2,178	At3g42565 gene
4	transposon and gene	6038150-6037139	1,012	At4g09545 gene (*CRP4*)
5	intergenic	24543172-24542491	682	At5g60978 pseudogenes

Four single dispersed copies of the SAT5 repeat were detected in a BLAST search using the 2196 bp SAT5 consensus sequence as a query sequence (Jurka, 2006). One nearly full-length copy is in a transposon-rich region close to the centromere of chromosome 3. Two truncated copies are present in relatively gene-rich regions on chromosomes 2 and 5, respectively. A third truncated copy is present in a region containing both genes and transposons on chromosome 4. The AGI numbers of the embedded *CRP* genes or pseudogenes are shown.

ATSAT5 (referred to hereafter as SAT5), which is defined by a 2,196 bp long consensus sequence (Jurka 2006).

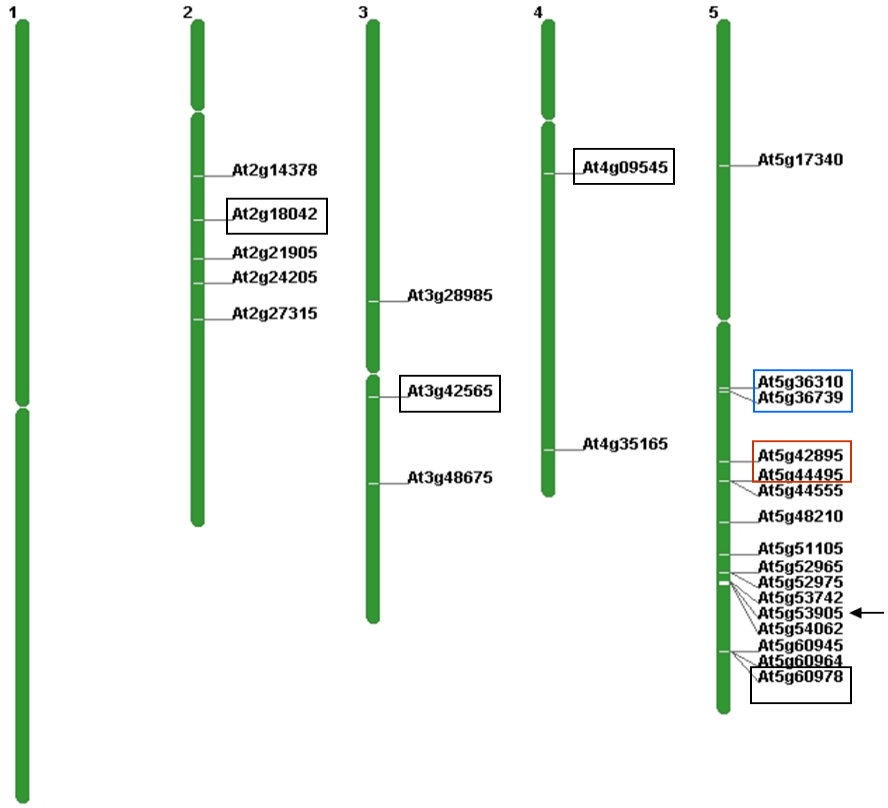

Figure 1 Chromosomal location of CRP3600 family members. The CRP3600 subgroup has 57 members, 32 of which are embedded in SAT5 repeat monomers arranged in a repeat array close to the centromere of chromosome 5; the AGI numbers of only the first and last copies of the array are shown (blue box). The four single dispersed copies of the SAT5 monomer containing different *CRP* genes or pseudogenes on chromosomes 2, 3, 4 and 5 are boxed in black. The pseudogenes on chromosome 2 (At2g19042) and chromosome 5 (At5g60978) were identified in the original MSAP analysis. The chromosome 4 gene (*CRP4*: At4g09545) has been analyzed in our study, which also considers two other genes in the CRP3600 subgroup that are not embedded in a SAT5 repeat (*CRP5a*: At5g42895 and *CRP5b*: At5g44495) (boxed in red). The single member of the CRP3600 subgroup containing introns (Atg53905) is indicated by the arrow. The figure was generated using the Chromosome Map Tool on the TAIR website (http://www.arabidopsis.org/jsp/ChromosomeMap/tool.jsp).

Based on a further BLAST analysis, four euchromatic copies of SAT5 fragments that are distributed throughout the *Arabidopsis thaliana* genome were identified (Table 1) (You et al., 2012). Although SAT5 was originally thought to be a non-protein coding

repeat, our further analysis showed that genes encoding small cysteine-rich peptides (CRPs) and also *CRP* pseudogenes are embedded in the SAT5 repeat monomer in the annotation of the *Arabidopsis* genome sequence. The *Arabidopsis* genome contains 825 *CRP* genes and pseudogenes that have been classified into different homology subgroups by hidden Markov model analysis (Silverstein et al., 2007).

The particular subgroup associated with the SAT5 repeat has been termed CRP3600, which has 57 members in Arabidopsis (Silverstein *et al*., 2007) (Figure 1). The CRP3600 subgroup has also been referred to as DUF1278 (domain of unknown function) family and ECA1 gametogenesis-related proteins (Jones-Rhoades *et al*., 2007). In addition to 32 copies embedded in the block of SAT5 repeat monomers close to the centromere on chromosome 5 (Figure 1, blue box), there are in addition 25 single copies of the CRP3600 family (not all of which are embedded in a SAT5 repeat monomer) distributed on four out of the five *Arabidopsis* chromosomes (Figure 1). Eighteen of these are predicted to encode proteins, including a gene we have termed *CRP4* (At4g09545), which is embedded in a truncated, intergenic copy of SAT5 on chromosome 4. The remaining seven CRP3600 family members are likely pseudogenes, including two embedded in truncated SAT5 copies in chromosomes 2 (At2g18042, identified from MSAP) and 5 (At5g60978), respectively (Figure 1, black box).

RdDM dependent CHH methylation of SAT5 and embedded *CRP4* gene on chromosome 4

Because the SAT5 copy on chromosome 2 was identified as a Pol V target in the MSAP assay (Daxinger 2008), we wondered whether other copies of SAT5 can be also targeted by RdDM components. To investigate that question, I performed a bisulfite sequence analysis on the truncated SAT5 monomer (1,012 bp) on chromosome 4, which includes 369 bp of *CRP4* coding region as well as 430 bp upstream and 213 bp downstream sequences. The bisulfite sequencing results showed a non-uniform distribution of methylation that appeared to be more concentrated within the *CRP4* coding region (Figure 2). Similar results were also observed in other RdDM mutants

that were examined (Figure 3B). There are two possible explanations for the non-uniform distribution of methylation. One is the presence of different numbers of cytosines available for methylation in CG and CHG sequence contexts along the SAT5 segment. The numbers were relatively low in the regions upstream (3 CGs, 6 CHGs) and downstream (2 CGs, 1 CHG) but considerably higher in the *CRP4* coding region (11 CGs, 20 CHGs) (Table 2). The other possibility is the tendency for CHH trinucleotides to be more efficiently methylated within the *CRP4* coding region than in flanking sequences.

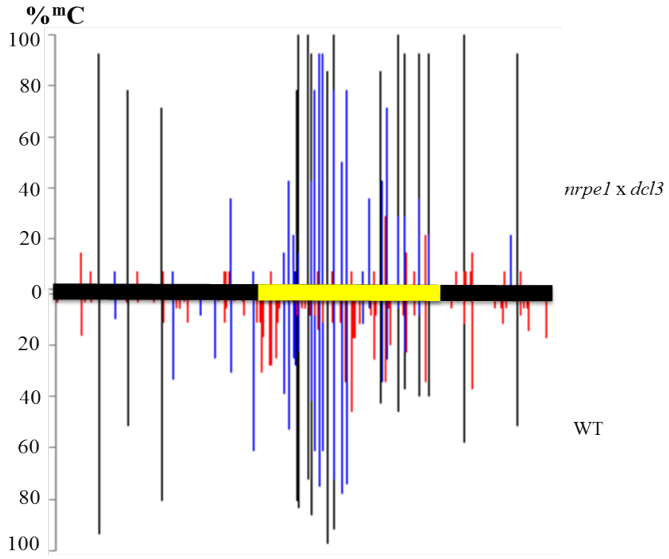

Figure 2 DNA methylation of SAT5 repeat monomer containing the *CRP4* gene. Bisulfite sequencing analysis of DNA methylation of the SAT5 repeat monomer containing the *CRP4* gene (At4g09545) in wild-type plants (bottom) and an *nrpe1 dcl3* double mutant (top). Methylation occurs in all three sequence contexts (CG, black; CHG, blue; CHH, red) and is concentrated in the *CRP4* coding region (yellow bar). Primarily CHH methylation is reduced in the *nrpe1 dcl3* double mutant. In a few cases, blue and black bars appear to be overlapping. These represent adjacent CG and CHG nucleotides that are not resolved in the figure format. CG methylation is always higher than CHG methylation. The results from at least 15 cloned sequences are shown.

Although the number of CHH sites is approximately the same in the upstream region and *CRP4* coding region (51 and 52, respectively), methylation was only

detected at around 22 CHHs (~ 43%) in the upstream region compared to 44 (~ 85%) in the *CRP4* coding region (Table 2). In the downstream region, methylation was detected at 19 out of 28 CHHs (~ 68%) (Table 2). In addition, the overall level of CHH methylation in the *CRP4* coding region was 12% compared to ~ 3% and ~ 6% in the upstream and downstream regions, respectively (Figure 3A).

Table 2. Percent methylation of individual cytosines in SAT5 repeat monomer containing the *CRP4* gene

Col-0	methylated CG	unmethylated CG	methylated CHG	unmethylated CHG	methylated CHH	unmethylated CHH
upstream	3/3 （100%）	0/3	6/6 （100%）	0/6	22/51 **(43%)**	29/51 (57%)
CRP4 cds	11/11 （100%）	0/11	20/20 (100%)	0/20	44/52 **(85%)**	8/52 (15%)
downstream	2/2 （100%）	0/2	0/1	1/1 (100%)	19/28 **(68%)**	9/28 (32%)

The total number of cytosines and those that are methylated or unmethylated in each sequence context (CH, CHG, CHH) are shown for each region of the SAT5 repeat monomer containing the *CRP4* gene [upstream, *CRP4* coding sequence (cds), downstream]. The percentage of methylated or unmethylated cytosines is shown in parentheses except where the value is 0 %. Percentages for methylated cytosines in a CHH context (indicative of RNA-directed DNA methylation) are shown in bold. The difference for the upstream region vs. *CRP4* cds is significant ($p = 0.000013$; two-tailed Fisher's exact test http://www.graphpad.com/quickcalcs/index.cfm). For the downstream region, the difference is not significant ($p = 0.093$) but this may reflect the lower number of cytosines in the downstream region.

Because CHH methylation is a hallmark of RdDM, these findings suggested that the *CRP4* coding region is targeted by this epigenetic pathway. Supporting this idea, the overall level of CHH methylation in the *CRP4* coding region was reduced ~ 65-80% in various mutants defective in RdDM (Figure 3B). A previous Northern blot analysis with a SAT5 probe demonstrated that the 24-nt siRNAs originating from SAT5 fragments were Pol IV and RDR2 dependent (You *et al*., 2012). Subsequent deep sequencing of small RNAs also revealed that production of unique 24-nt siRNAs from the coding region of *CRP4*, *CRP5a* and *CRP5b* were dependent on Pol IV and RDR2 (You *et al*., 2012). Taken together, the results on CHH methylation and unique siRNA accumulation

indicated that *CRP* genes are the targets of Pol IV- and Pol V-dependent RdDM.

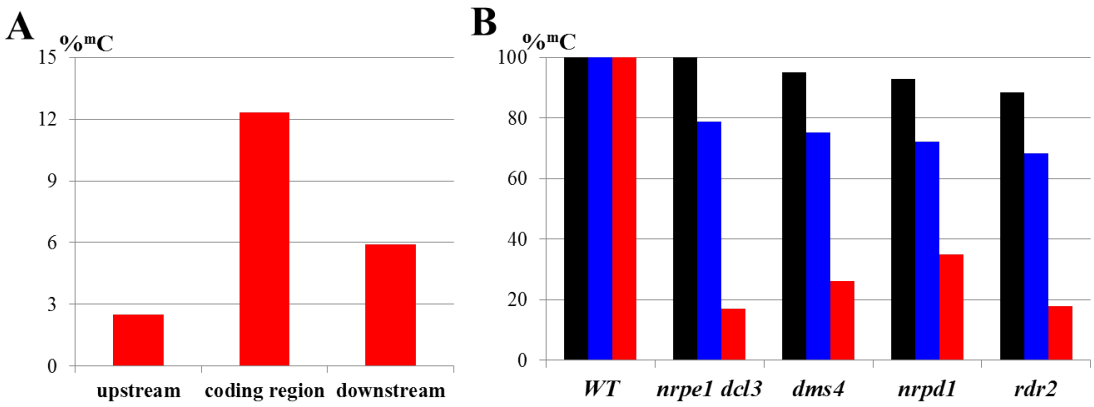

Figure 3 CHH methylation in *CRP4* coding region. (A) Average percent methylation of cytosines in a CHH context in different parts of the SAT5 repeat monomer containing the *CRP4* gene. CHH methylation is highest in the *CRP4* coding region. **(B)** Overall levels of methylation in CG (Black), CHG (blue) and CHH (red) nucleotide groups in the *CRP4* coding region in wild type plants (WT) and mutants defective in RNA-directed DNA methylation (*nrpe1 dcl3* double mutant and *dms4*, *nrpd1*, and *rdr2* single mutants).

Expression of the *CRP4* gene in the female gametophyte

A previous genome-wide expression profiling analysis demonstrated that many *CRP* genes from the DUF1278 subgroup, including *CRP5a* and *CRP5b*, are expressed in the synergid cell region of the female gametophyte (Jones-Rhoades *et al*., 2007). However, the expression patterns of the four CRP3600 subgroup members embedded in single copy SAT5 repeat monomers (two genes and two pseudogenes) (Table 1) has not yet been reported. Therefore, I first performed RT-PCR analysis of *CRP* genes with mRNA isolated from both leaves and flowers in wild-type (WT) and several RdDM mutant plants. The RT-PCR results showed that the four CRP3600 subgroup members are not appreciably expressed in leaves of either WT or RdDM mutant plants, but are expressed in flower tissues of both WT and RdDM mutant plants (Figure 4A).

To obtain a more detailed view of the cell type-specific expression pattern, I generated constructs encoding a GUS reporter protein and a CRP4-GFP fusion protein, both under the control of an endogenous *CRP4* promoter fragment, which comprises 921 bp upstream of the ATG start codon of the *CRP4* gene (Figure 4B, top). After

introducing these constructs into wild type *Arabidopsis* plants, *GUS* and *CRP4-GFP* gene expression was observed specifically in the synergid cell region of the female gametophyte, indicating that the *CRP4* promoter is active exclusively in these cells (Figure 4B, bottom). Thus, the RT-PCR experiments, GUS assay and GFP localization data all indicate that the *CRP4* gene is expressed in the synergid cells of female gametophyte.

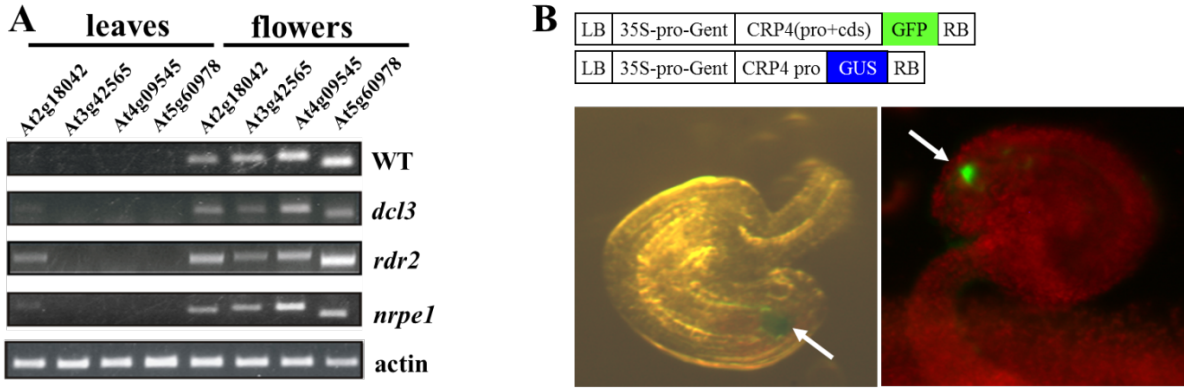

Figure 4 Flower and synergid cell-specific expression of *CRP* genes and the *CRP4* promoter. (A) Semi-quantitative RT-PCR demonstrates expression of endogenous *CRP* genes (At3g49565, At4g09545) and *CRP* pseudogenes (At2g18042, At5g60978) in flowers but not in leaves. With the exception of the pseudogene At2g18042, which shows some reactivation in leaves in mutants defective in RNA-directed DNA methylation (*dcl3*, At3g43920; *rdr2*, At4g11130; *nrpe1*, At2g40030), none of the genes or pseudogenes are appreciably expressed in these mutants. Actin (At3g18780) is shown as a constitutively expressed control. **(B)** Scheme of CRP4-GFP fusion and GUS fusion constructs, GUS assay and GFP fusion protein localization. The *CRP4 -GFP* fusion gene and the *GUS* reporter gene are controlled by the endogenous *CRP4* promoter (top). Photographs of ovules from transgenic Arabidopsis plants transformed with constructs containing *CRP4-GFP* (bottom right) and *GUS* (bottom left) reporter genes under the control of the *CRP4* promoter. Expression occurs exclusively in the synergid cell region of the female gametophyte (white arrows). Identical results were obtained in multiple independent transgenic lines obtained for each construct.

DNA methylation in the *CRP4* gene body region of a transgenic copy from sporophyte tissues

Previous results showed that the endogenous *CRP4* gene body region can be methylated in leaves in the presence of 24-nt siRNAs originating from the *CRP4* coding region (You *et al*., 2012). Therefore, it was interesting to see whether the transgenic

copies encoding the two CRP4-GFP fusion proteins, under the control of 35S promoter and *CRP4* promoter, respectively, can also become methylated in the *CRP4* coding region. Bisulfite sequencing analysis with DNA from leaves showed that the *CRP4* coding region in the transgenic copy under the control of the 35S promoter is unmethylated (Figure 5A). By contrast, dense methylation, which is similar to that detected in the coding region of the endogenous *CRP4* gene, was seen in the *CRP4*

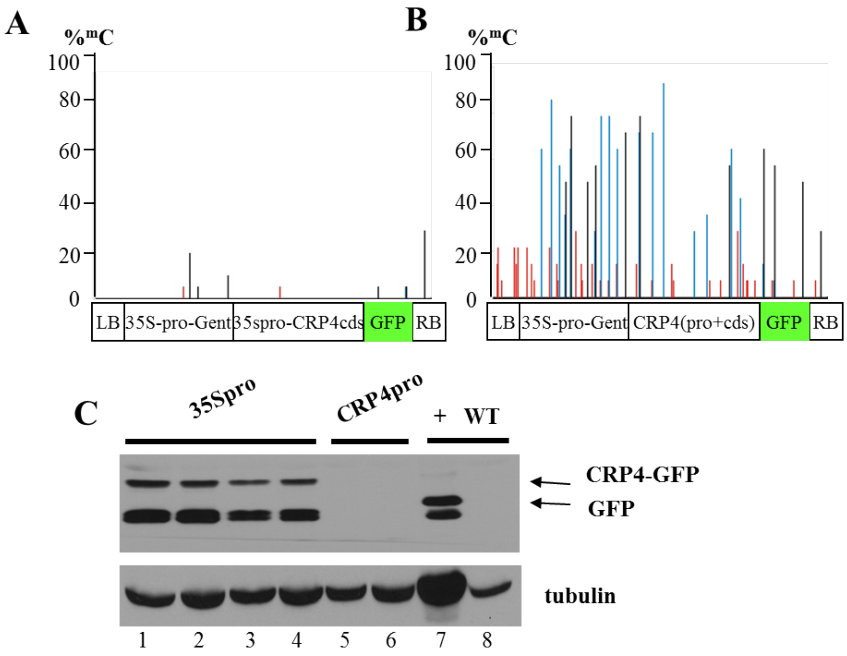

Figure 5 *CRP4* coding sequence is unmethylated in leaves when transcribed from 35S promoter to produce translatable mRNA. Bisulfite sequencing analysis showed methylation in leaves of the *CRP4* coding region in a *CRP4-GFP* fusion gene under the control of the constitutive 35S promoter **(A)** and the synergid cell-specific *CRP4* promoter **(B)**. Scheme of constructs in the transgenic lines were shown under the graph, respectively. The sequenced region shown corresponds to *CRP4* coding region that is indicated by the yellow bar in Figure 2. The bisulfite sequencing results from at least 15 cloned sequences are shown. **(C)** Western blot detection of the CRP-GFP fusion protein in seedlings from four independent transgenic lines transformed with a 35S promoter-CRP4-GFP fusion construct (lanes 1–4, top band, top arrow). The '+' lane shows a transgenic line that express GFP as a positive control (top band, bottom arrow). The lower band in these lanes probably corresponds to a degradation product that reacts with the GFP antibody. A CRP4-GFP fusion protein is not detected in transgenic plants containing a *CRP4-GFP* fusion gene under the control of the *CRP4* promoter (lanes 5 and 6; results from two independent lines shown). Tubulin is shown as a constitutively expressed loading control.

coding region portion of the *CRP4* promoter-driven *CRP4-GFP* fusion gene in the transgenic lines (Figure 5B).

I also performed bisulfite sequencing of the *GFP* coding sequence in the *CRP4* promoter-directed *CRP4-GFP* fusion construct, however, no methylation was observed (data not shown). This result suggested that the DNA methylation in the *CRP4* gene body region is specifically induced by siRNAs originating from *CRP4* coding region.

A CRP4-GFP fusion protein was detectable on Western blots using a GFP antibody in unmethylated transgenic plants containing a *CRP4-GFP* fusion gene under the control of the 35S promoter, indicating productive Pol II transcription of a translatable *CRP4-GFP* mRNA from the 35S promoter (Figure 5C). By contrast, no CRP4-GFP fusion protein was detected in plants containing a *CRP4-GFP* fusion gene under the control of the *CRP4* promoter (Figure 5C), even though *CRP4-GFP* transcripts can be detected in these plants (Figure 6A)

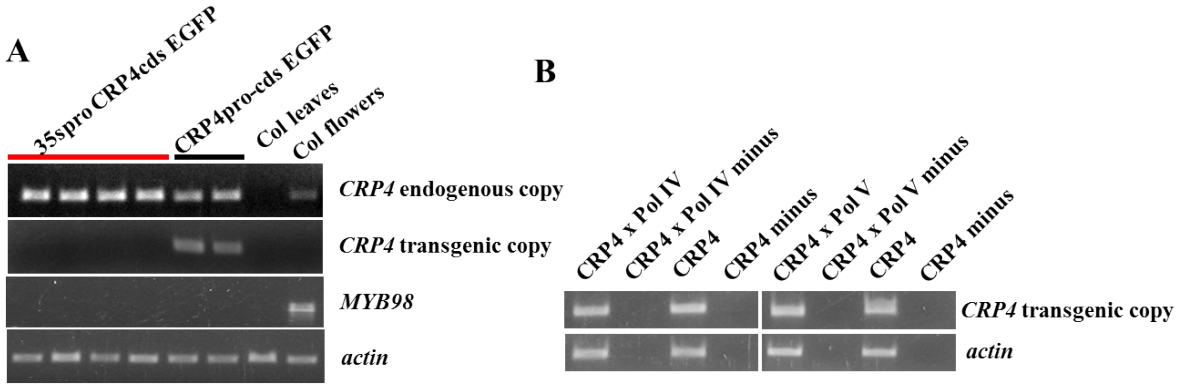

Figure 6 Pol II transcribed *CPR4* transgenic copy in sporophyte tissues. (A) Semi-quantitative RT-PCR demonstrates that transgenic *CRP4* copy can be expressed in the seedlings, but not the *MYB98,* which is a synergid cell specific transcription factor. The product of *CRP4* endogenous copy is only part of *CRP4* coding region, and *CRP4* transgenic copy includes the entire coding region. Thus, the PCR product amplified from *CRP4* endogenous copy is included in the transgenic copy, which was also confirmed by sequencing. **(B)** Semi-quantitative RT-PCR demonstrates that transgenic *CRP4* copy can also be expressed in the plants after crossing with Pol IV or Pol V mutant. Actin expression is shown as a constitutively expressed control. Minus means RT reaction without reverse transcriptase.

A possible explanation for this result is that these 'unproductive' (not translated)

transcripts are transcribed by an RNA polymerase other than Pol II, such as Pol IV or Pol V. To verify which polymerase is indeed responsible for this transcription, crosses were made between *CRP4pro CRP4- GFP* fusion lines and mutants in the largest subunits of Pol IV and Pol V, *nrpd1a* and *nrpe1,* respectively. The *CRP4-GFP* transcript was still detected after introducing these mutations, thus confirming that the 'unproductive' transcripts were transcribed by Pol II rather than Pol IV or Pol V (Figure 6B).

Semi-quantitative RT-PCR results showed that *MYB98*, which is exclusively expressed in synergid cells and is required for pollen tube guidance and synergid cell differentiation (Kasahara *et al*., 2005), was not expressed when *CRP4* transgenic copy can be transcribed by Pol II in sporophyte tissues (Figure 6A). This result indicates that *MYB98* is not required for the transcription of *CRP4* transgenic copy, which suggests there are some other unknown transcription factor(s) involved in this transcription.

DNA methylation results obtained with DNA from synergid cells that was isolated through laser micro-dissection (performed by Matthew Spencer, a postdoc in the lab) showed reduced methylation in the gene body region of the endogenous *CRP4* gene and *CRP4-GFP* fusion gene, correlating with *CRP4* expression in synergid cells (You *et al*., 2012). Former DNA methylation data showed that the Pol II transcribed *CRP4* transgenic copy cannot be translated, perhaps due to dense DNA methylation in *CRP4* gene body region (Figure 5). Therefore, these results suggest that DNA methylation, particular CG and CHG methylation, might be involved in the regulation of *CRP4* expression in both sporophyte and reproductive tissues.

Hairpin structure induced silencing of *CRP4* in synergid cells

Previous studies showed that the probable functions of *CRP* genes include playing a role in plant-microbe defense mechanisms (Wheeler et al., 2009) or during plant reproductive development (Marshall et al., 2011), when they may act as as pollen tube attractants or intercellular signaling molecules (Okuda et al., 2009, Grossniklaus 2011). In our study, the CRP4 GUS and GFP reporter assays showed a synergid cell-specific

expression pattern, indicating a putative function of CRP4 during female gametophyte development in *Arabidopsis*. To investigate the probable function of CRP4 in synergid cells, I assembled hairpin structure constructs designed to knock out *CRP4* gene expression in synergid cells (Figure 7A). A GFP hairpin construct that is controlled by 35S promoter was used as a positive control. After introducing the construct containing CRP4pro-cds HP into former *CRP4-GFP* fusion gene controlled by *CRP4* own promoter transgenic lines (see Figure 4B top), no GFP signal was observed in the embryo sac from the homozygote transformant (Figure 7B right), which indicated that *CRP4* gene was silenced most probably through a post transcriptional gene silencing pathway (PTGS) by the hairpin structure of *CRP4* coding region.

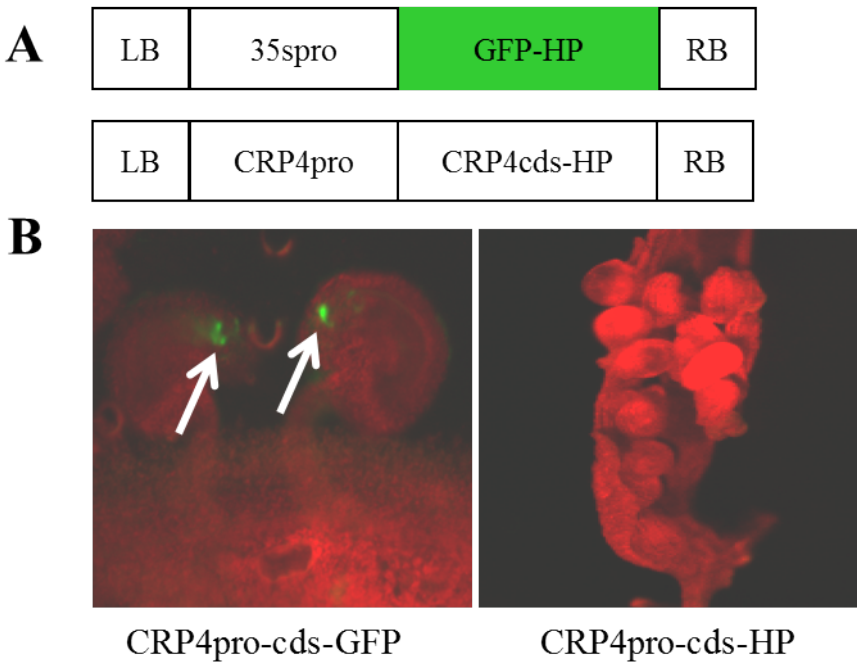

Figure 7 *CRP4* hairpin structure leads to the silencing of *CRP4-GFP* fusion gene in synergid cells. (A) Schemes of *CRP4* and GFP hairpin structure. One carries a hairpin structure from *GFP* coding region controlled by a 35S promoter as a positive control, and the other contains *CRP4* coding region controlled by the endogenous *CRP4* promoter. **(B)** Left photograph shows CRP4-GFP signal in synergid cells, shown in white arrow. The picture from right side demonstrates loss of *CRP4-GFP* expression after the CRP4 pro-cds-HP constructwas introduced into CRP4 pro-cds-GFP lines.

The next question was to see whether this *CRP4* knock out transformant would show some severe phenotype, like affected pollination or seed formation. Unfortunately, no seed abortion phenotype was seen. This could be explained by two possibilities; one is that CRP4 protein does not have a function similar to other *CRP* genes that can act as pollen tube attractants in *Torenia fournieri*. On the other hand, although CRP4 might have some function in the female gametophyte, such as pollen tube attractant, some other *CRP* genes act redundantly with CRP4 and complement the phenotype caused by knocking out of *CRP4* gene expression.

The *CRP4* gene is not inducible by flg22 or heat treatment in seedlings

As mentioned before, the *CRP4* gene was classified into the CRP3600 family, which might play a function in plant microbial defense pathways (Silverstein *et al*., 2007). To test whether the *CRP4* gene can be induced by a pathogen protein, I treated seedlings with bacterial flagellin peptide flg22, which is a potent elicitor in plants (Meindl *et al*., 2000). Although I observed induction of *MPK3* (At3g45640), a known

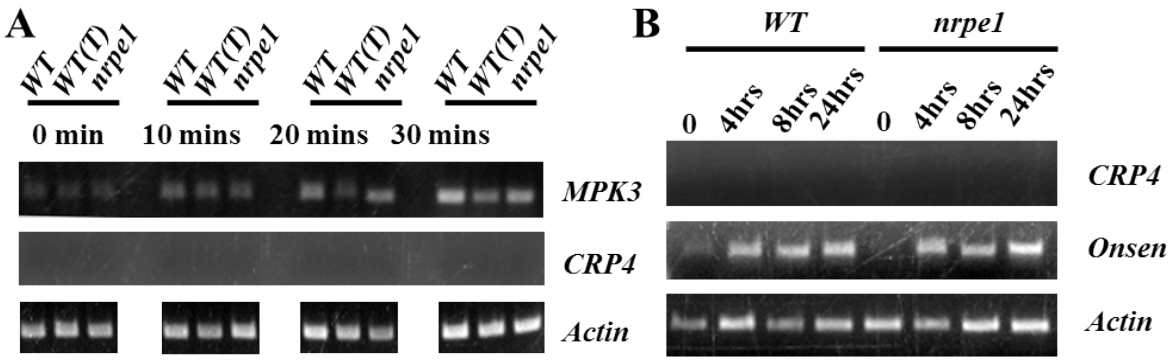

Figure 8 The *CRP4* gene is not induced in seedlings following flg22 treatment or heat stress. Semi-quantitative RT-PCR was used to test possible induction of *CRP4* gene expression in seedlings under different conditions. **(A)** Treatment of seedlings with the bacterial elicitor flg22 for the indicated time periods results in induction of expression of the MAP kinase MPRK3 but not the *CRP4* gene. Similar results were seen in wild-type non-transgenic seedlings (WT), a wild-type transgenic line [WT(T)], and *nrpe1* mutant, which is defective in the largest subunit of Pol V. **(B)** Heat treatment of seedlings for the indicated time periods does not result in induction of *CRP4* gene expression whereas the *Onsen* retrotransposon is induced under the same conditions. In both panels, actin expression is shown as a constitutively expressed control.

flagellin-induced gene, no induction of the *CRP4* gene was observed under the same conditions (Asai *et al*., 2002) (Figure 8A). Similarly, I did not observe induction of the *CRP4* gene after heat stress that was sufficient to activate the *Onsen* retrotransposon (Figure 8B), which requires Pol IV pathway components to restrict trans-generational transposition (Ito *et al*., 2011). Taken together these results suggest that *CRP4* is not a gene involved in anti-microbial defense pathway or in heat stress.

3.2 Discussion of Chapter II

Atypical gene body DNA methylation in *CRP* genes

High throughput sequencing of DNA methylation has revealed that DNA methylation is not only enriched in transposable elements (TEs), repeat sequences and promoter regions, but also occurs in the gene body region of around 30% of protein-coding genes *Arabidopsis* (Zhang *et al*., 2006, Cokus *et al*., 2008, Lister *et al*., 2008). DNA methylation in TEs and repeat sequences is present in all sequence contexts: CG, CHG, CHH (H is A, T or C).By contrast, promoter and gene body methylation occurs primarily in a CG context (Saze and Kakutani, 2011). DNA methylation in TEs and repeat sequences is required to transcriptionally silence these sequences and hence maintain genome integrity and stability (Miura *et al*., 2001, Chan *et al*., 2006). Whereas DNA methylation in promoter regions leads to transcriptional gene silencing (Aufsatz *et al*., 2002), the role of gene body DNA methylation remains unclear (Teixeira and Colot, 2009). Here, we have shown that not only the *CRP4* gene but also several other genes in the CRP 3600 family atypically contain not only CG but also CHG and CHH methylation region, which arises in part from RNA-directed DNA methylation, in their gene body region. The pattern of DNA methylation in *CRP* genes is more similar to that of transposons rather than normal genes. Detection of Pol IV-dependent 24 nt siRNAs that uniquely originate from the *CRP* gene body region also supports this claim. Data on DNA methylation of the *CRP4* gene body region from leaves in RdDM mutants showed a loss of CHH methylation, but a persistence of nearly wild type levels CG and CHG methylation, which indicates the possibility of other methylation mechanisms that contribute to the dense methylation of *CRP* gene bodies in sporophyte tissues. The biological role of the transposon-like methylation pattern of *CRP* gene bodies is not yet clear but it is likely to be related to gene silencing in the sporophyte tissues.

Putative Function of gene body methylation in *CRP* genes

Previous studies showed that the impact of gene body methylation on expression

depends on the sequence context of methylated cytosines. In wild type plants, CG methylation alone in gene bodies does not inhibit transcription and is found at many moderately expressed genes (Cokus *et al*., 2008, Lister *et al*., 2008, Saze and Kakutani, 2011, Zilberman *et al*., 2007). CHG methylation in gene bodies, which is induced in the *ibm1* (increase in *BONSAI* methylation 1) mutant, has variable effects on transcription, with only some target genes showing decreased transcription while others are unchanged or even more highly expressed (Miura *et al*., 2009). In our study, RT-PCR results showed that loss of CHH methylation in the gene body region of *CRP* genes in RdDM mutants does not affect gene expression. Furthermore, the expression of a *CRP4-GFP* fusion gene is correlated with a reduction of CG and CHH methylation in the *CRP4* gene body region in synergid cells (You *et al*., 2012). Taken together, these results suggest that a combination of dense CG and non-CG methylation in gene bodies is associated with gene silencing.

Although gene silencing seems the most likely function and/or consequence of combined CG and non-CG methylation in the *CRP* gene bodies, a possible positive role can also be considered, namely that gene body methylation attunes genes to internal or external cues. This suggestion follows from recent data demonstrating that gene body methylation, whether in a CG or non-CG context, can be dynamic and may render genes sensitive to developmental or environmental signals. For example, a bioinformatics study examining genome-wide gene expression in *Arabidopsis* under many different experimental conditions concluded that CG methylation in gene bodies protects genes from reacting to internal or external stimuli (Aceituno *et al*., 2008). In tomato, a drought-responsive gene that atypically has non-CG methylation in the gene body loses non-CG methylation and gains CG methylation under drought conditions (González *et al*., 2011). Some transcribed genes in human embryonic stem cells loss non-CG methylation in their gene bodies as cells differentiate and re-established during de-differentiation, suggesting a relationship between non-CG methylation in gene bodies and the developmental stage (Lister *et al*., 2009). In view of these findings, it is conceivable that combined CG and non-CG methylation in *CRP* gene bodies allows responsiveness to either developmental cues in the female gametophyte or external factors in the environment. Regarding the latter possibility, we did not observe de-

repression of the methylated *CRP4* gene in leaf tissue following treatment with flg22 or heat stress, but other biotic or abiotic stresses can be tested in the future.

Additionally, our DNA methylation data from sporophyte tissues also showed that loss of CG and CHG methylation in the gene body region of transgenic copy from *CRP4-GFP* fusion gene controlled by 35S promoter results in a translation of fusion protein. By contrast, dense DNA methylation in CG and CHG nucleotide groups was observed in the gene body region of the transgenic copy of the *CRP4 pro-cds-GFP* line. Therefore, these results indicate a putative function of combination of CG and CHG methylation in regulation of *CRP4* expression both in sporophyte and reproductive tissues. However, whether this methylation inhibits transcriptional elongation by Pol II or is the default state when transcription is impaired by other means is not yet known. Also how methylation in *CRP4* gene bodies is lost in synergid cells remains unclear. It may result from active demethylation by DNA glycosylase activities, as observed for imprinted genes in the central cell of the female gametophyte (Bauer and Fischer 2011, Raissig et al., 2011), or a lack of silencing factors in synergid cells or their precursors. Although a recent study demonstrated that genes encoding components of the RNA-directed DNA methylation machinery are expressed in egg cells (Wuest et al., 2010), a similar analysis has not yet been performed for synergid cells.

Transposon like DNA methylation based upon retrogene origin?

Why are the gene bodies of some CRP3600 gene family members methylated like transposons? Although several methylated *CRP* genes are embedded in SAT5 repeats this is not the case for all (for example, *CRP5a* and *CRP5b*), demonstrating that the presence of the flanking SAT5 sequences is not a requirement for methylation. To explain the dense methylation concentrated in the *CRP* coding regions, we propose that some members of the CRP3600 gene family arose through reverse transcription of mature mRNAs (retroposition) and hence can be classified as retroposed genes (retrogenes). Under this hypothesis, an intron-containing progenitor copy would have been highly expressed in the egg cell, the spliced mRNA reverse transcribed, and the

cDNA subsequently integrated into egg cell DNA to become a heritable part of the genome. Through this copying and integration process, *CRP* retrogenes would resemble transposons and be methylated accordingly in their coding regions. Consistent with this hypothesis, only one gene in the CRP3600 family has two introns (At5g53905) while the rest are intron-free and thus candidate retrogenes (Figure 1). The observation that at least some CRP3600 gene family members are probably expressed in the egg cell as well as the neighboring synergid cells is compatible with the retrogene hypothesis (Jones-Rhoades et al., 2007). Interestingly, retrogenes in mammals are expressed mainly the in testes, where transcription is less tightly regulated than in somatic tissues (Vaknin et al., 2009). Whether controls on transcription are similarly relaxed in synergid cells in plants remains to be determined.

However, not all *CRP* gene members of CRP3600 family have a transposon-like methylation pattern in their gene bodies. Some lack methylation entirely and one has CG methylation only (Table 1). Although the source of such differences is presently unknown, one possibility is that recently integrated copies still have the initial transposon-like methylation pattern, which over time degenerates into exclusively CG methylation, ultimately resulting in complete decay of gene body methylation. Because methylated cytosine spontaneously deaminates to produce thymine, methylated *CRP* coding regions may experience accelerated sequence diversification over evolutionary time, thus promoting functional diversification of CRP proteins. Given known or proposed roles of CRPs in pollen tube guidance (Okuda *et al*., 2009), and self-incompatibility (Takayama *et al*., 2001, Marshall *et al*., 2011), such sequence changes and functional diversification may help to drive speciation.

Table 1 Methylation status of members of CRP3600 subgroup.

AGI	Intron Number	Pseudogene	Start Position	End Position	Methylation status	SAT5 embedded
At2g14378	0	F	6112313	6111951	None	No
At2g18042	0	T	7851529	7851897	Transposon	Yes
At2g21905	0	T	9346247	9346557	Transposon	No
At2g24205	0	F	10298498	10298842	CG and CHG	No
At2g27315	0	F	11697031	11696669	CG	No
At3g28985	0	T	10999264	10999576	Transposon	No
At3g42565	0	F	14694396	14694037	Transposon	Yes
At3g48675	0	F	18045177	18045536	None	No
At4g09545	0	F	6037566	6037934	Transposon	Yes
At4g35165	0	F	16735297	16735659	None	No
At5g17340	0	F	5716220	5715738	None	No
At5g36310	0	F	14342288	14341941	None	Yes
At5g36320	0	F	14344487	14344131	Transposon	Yes
At5g36330	0	T	14346684	14346328	Less	Yes
At5g36340	0	F	14348863	14348516	None	Yes
At5g36350	0	F	14351059	14350703	None	Yes
At5g36360	0	F	14353254	14352898	Less	Yes
At5g36370	0	F	14355451	14355095	None	Yes
At5g36380	0	F	14357648	14357292	None	Yes
At5g36390	0	F	14359843	14359487	None	Yes
At5g36400	0	F	14362038	14361682	None	Yes
At5g36410	0	F	14364233	14363877	None	Yes
At5g36420	0	F	14366428	14366072	None	Yes
At5g36430	0	F	14368623	14368267	None	Yes
At5g36440	0	F	14370819	14370463	None	Yes
At5g36450	0	F	14373016	14372660	None	Yes
At5g36460	0	F	14375212	14374856	Less	Yes
At5g36470	0	F	14377409	14377053	None	Yes
At5g36480	0	F	14379606	14379250	Less	Yes
At5g36490	0	F	14381803	14381447	None	Yes
At5g36500	0	F	14384000	14383644	None	Yes
At5g36510	0	T	14386198	14385842	Less	Yes
At5g36520	0	F	14388400	14388044	None	Yes
At5g36530	0	T	14390594	14390238	Less	Yes
At5g36540	0	F	14392787	14392431	None	Yes
At5g36550	0	F	14394980	14394624	None	Yes
At5g36657	0	F	14401406	14401107	CG and CHG	Yes
At5g36658	0	F	14403658	14403302	CG and CHG	Yes
At5g36659	0	F	14405851	14405552	CG	Yes
At5g36661	0	F	14408015	14407659	None	Yes
At5g36662	0	F	14410205	14409849	None	Yes
At5g36738	0	F	14467314	14466958	None	Yes
At5g36739	0	F	14469504	14469148	None	Yes
At5g42895	0	F	17214023	17214391	Transposon	No
At5g44495	0	F	17943937	17943578	Transposon	No
At5g44555	0	T	17962036	17962394	Transposon	No

At5g48210	0	F	19565685	19565996	None	No
At5g51105	0	F	20794858	20795229	None	No
At5g52965	0	F	21496067	21496429	None	No
At5g52975	0	F	21498740	21499105	None	No
At5g53742	0	F	21830897	21831244	Less	No
At5g53905	2	F	21906480	21905374	None	No
At5g54062	0	F	21956083	21956706	None	No
At5g60945	0	F	24543524	24543862	Transposon	No
At5g60964	0	F	24547812	24548150	Transposon	No
At5g60978	0	T	24559725	24560081	Transposon	Yes

There are 57 members of the CRP3600 subgroup. The Arabidopsis Genome Initiative (AGI) numbers, intron numbers, designations as either gene or pseudogene, and genomic coordinates are taken from the study by Silverstein and coworkers (Silverstein *et al*., 2007). The methylation status has been determined from our bisulfite data for At4g09545 and from the publicly available methylome data of Lister and coworkers (Lister *et al*., 2008). A 'Transposon' methylation pattern indicates relatively dense CG, CHG and CHH methylation in the gene body. 'Less' indicates a similar pattern but less methylation. Genes or pseudogenes shown in red letters are mentioned in our study. Those shown in blue letters are present in the repeat block of SAT5 monomers close to the centromere of chromosome 5.

4. Materials and Methods

4.1 Plant materials, growth condition

In this study, all experiments were performed using *Arabidopsis thaliana* accession Col-0. Transgenic plants containing only T locus or T+S loci were used as previously described (Kanno *et al*., 2008; Daxinger *et al*., 2009; Naumann *et al*., 2010; Eun *et al*, 2011; Lorković *et al* 2012; Eun *et al*., 2012). For the mutants defective in RNA-directed DNA methylation the alleles used were *nrpe1-3* (Kanno *et al*., 2010), *drd3-7* (*nrpe1*) (Kanno *et al*., 2005), *rdr2-1* (SAIL_1277H08) (Xie *et al*., 2004), *nrpd1a-7* (Smith *et al*., 2007), *dms4-1* (Kanno *et al*., 2010), *dcl3-1* (SALK_005512) (Xie *et al*., 2004), *dcl2/3/4* (*dcl2-1/dcl3-1/dcl4-2*, SALK_064627.42.50.x/ SALK_005512.38.70.x/ GABI_160G05) (Xie *et al*., 2004; Xie *et al*., 2005), *met1-1* (Kankel *et al*, 2003), *met1-3* (Saze *et al*., 2003), *ddc* (*drm1-2/drm2-2/cmt3-11*) (Chan *et al*., 2006). All the seeds were sterilized with 70% ethanol, 0.05% Triton X-100 solution, 15 min vortexing, afterwards washed with 100% ethanol, followed by ddH_20. Genotyping primers are listed in Tables 4.1 and 4.2. Plants were grown under long day condition (16 hour light/8 hour dark cycle) in either a greenhouse or growth chamber set at ~ 21°C.

Table 4.1: List of primers used for genotyping (Chapter 1)

Primer	Sequence
nrpd1-7 BamHI forward	5'-CCT TTT TGC AGG TTT ATG CTC-3'
nrpd1-7 BamHI reverse	5'-GTT TCT TGT AGA TCC GAG TCC GTG GAT C-3'
nrpe1-3 DraI forward	5'-ACT GGA GAT GCT TAC CGA CAT GTG AAT GA-3'
nrpe1-3 DraI reverse	5'-GGT AGA ACA AAT GGA CAC AAT CAC CTT-3'
Target co-dominant forward	5'-GGA CAC ATC CTA TAG TTC GC-3'
Target co-dominant reverse	5'-GAT CTA CCC ACT AAT CTA CTC G-3'
Target co-dominant internal	5'-GGC GAA CCA AGC CGC TAA TGC-3'
Silencer co-dominant forward	5'-GGT CAT CAC AAA CAT CTC GT-3'
Silencer co-dominant reverse	5'-GTG ATC CAA AGT CAT GGT CT-3'
Silencer co-dominant internal	5'-CAG CCG TCC AAA TGC GGG AT-3'

4.2 Plasmid construction

For the first Chapter, two new plasmids were constructed: 'SD' (siRNAs-DNA interaction) construct, which contains 88bp secondary siRNAs fragment from the target gene (Kanno et al., 2008; Daxinger *et al*., 2009) followed by the ubiquitin promoter fused with dsRED protein (88bps-Ubipro-dsRed) and 'SR' (siRNAs-RNA interaction) construct, which has the same components but in a different order, ubiquitin promoter

Table 4.2: List of primers used for genotyping (Chapter 2)

Primer	Sequence
SALK-LBa1	5'-TGG TTC ACG TAG TGG GCC ATC G-3'
dcl2-1 forward	5'-TGT TAA CAG ATC TCC GCT GCC A-3'
dcl2-1 reverse	5'-TTG GCT GAG ATA CCT CAA GGT GG-3'
dcl3-1 forward	5'-CCT GAA GAG CGT GAA GGA GTG GAA A-3'
dcl3-1 reverse	5'-GAC TAG AAG AGG ACT CAA TGC AAT-3'
GABI 8409A	5'-TCT CCA TAT TGA CCA TCA TAC TCA TT-3'
dcl4-2 forward	5'-TCT AGA TCC GGT CAA AAA CTT GTC-3'
dcl4-2 reverse	5'-TCA GCA AAG GAA TCC AGA ATG CT-3'
rdr2-1 forward	5'-ACA CAT TAG GAC TAA CAA ATT TAC C-3'
rdr2-1 reverse	5'-ATG GTG GTG TCA GAG ACG ACG ACG AAC CGA TCA-3'
SAIL-LB3	5'-TAG CAT CTG AAT TTC ATA ACC AAT CTC GAT ACA C-3'
drd3-7 PstI forward	5'-AAG ATT AAC CCT CTG ATG TG-3'
drd3-7 PstI reverse	5'-CTG CCC TAG ACA TGG ATA AG-3'
*ddm1 Nsi*I forward	5'-ATT TGC TGA TGA CCA GGT CCT-3'
*ddm1 Nsi*I reverse	5'-CAT AAA CCA ATC TCA TGA GGC-3'
nrpd1-7 BamHI forward	5'-CCT TTT TGC AGG TTT ATG CTC-3'
nrpd1-7 BamHI reverse	5'-GTT TCT TGT AGA TCC GAG TCC GTG GAT C-3'
*nrpe1-3 Dra*I forward	5'-ACT GGA GAT GCT TAC CGA CAT GTG AAT GA-3'
*nrpe1-3 Dra*I reverse	5'-GGT AGA ACA AAT GGA CAC AAT CAC CTT-3'
*met1-1 Hae*III forward	5'-ACT TTG GCT TCC CTT CTC GA-3'
*met1-1 Hae*III reverse	5'-TCA CGG TGA TTG GAC GGA-3'
met1-3 forward	5'-TCA GAT TGT GGC TAT ACT TGT G-3'
met1-3 reverse	5'- TGG GAA ACA TCA TAG AAA CAT C-3'
met1-3 LB3	5'-TTT ACA ATT GAA ATG GCT TCC C-3'

followed by 88bp secondary siRNAs fragment fused with dsRED protein (Ubipro-88bps-dsRed). Both of them were modified with restriction enzyme *Spe*I, *Xba*I, *EcoR*V, *Hind*III and *BamH*I sites that are necessary for plasmid construction and ordered from Mr.Gene Company (Regensburg, Germany). Both of these two fragments were inserted into pMPO (Mannopine promoter, Phosphinothricin and Octopine terminator) binary vector by using *Hind*III site. A hairpin structure plasmid containing inverted repeats of secondary siRNAs 88bp fragment downstream of the cauliflower mosaic virus 35S promoter (35spro-88bps HP) modified with restriction enzyme *Sal*I site was also synthesized by Mr.Gene Company (Regensburg, Germany). This inverted repeat fragment was inserted into a new binary vector pPZP 200 series through *Sal*I site (Hajdukiewicz et al., 1994).

For the second Chapter, three constructs were made for expression studies: 35Spromoter-*CRP4-EGFP*, *CRP4*promoter-*CRP4-EGFP*, and *CRP4*promoter-*GUS*. *CRP4* promoter region (including 921bp upstream of ATG site) and *CRP* coding region (369bp) were synthesized by Mr.gene Company (Regensburg, Germany) and modified with restriction enzymes *Eco*RI, *Nco*I, *BamH*I and *Hind*III sites, which are required for the plasmid construction. For construction of CRP4-GFP fusion gene, an enhanced GFP tag was added at C-terminal of CRP4 by *BamH*I and *Hind*III sites (including both promoter and coding region). We also fused the same enhanced GFP tag to the C-terminus of CRP4 coding region CRP4 by *BamH*I and *Hind*III sites, controlled by the cauliflower mosaic virus 35S promoter. Both of these two fragments were inserted into pPZP200 binary vector by using *Eco*RI and *Hind*III sites. To construct β-glucuronidase (GUS) reporter under the control of CRP4 promoter, a GUS coding sequence was inserted downstream of the *CRP4* promoter by using *Nco*I and *BamH*I sites, and then also introduced into pPZP 200 binary vector through *Eco*RI and *Hind*III sites.

Two additional plasmids were constructed to knock down *GFP* and *CRP4* genes *in vivo*. One is containing inverted repeats from *GFP* coding region driven by the 35S promoter (35spro-GFP HP), and the other is including inverted repeats from *CRP4* coding region controlled by *CRP4* own promoter (CRP4pro-CRP4cds HP). Both of them modified with restriction enzyme *Sal*I and *Hind*III sites for cloning were ordered

from Mr.Gene Company (Regensburg, Germany). These two inverted repeats were inserted into MPO binary vector through *Hind*III site.

All the plasmids listed above were transformed into *Agrobacterium tumefaciens* and transgenic plants were obtained by using floral dip method (Matzke and Matzke 1986; Clough and Bent 1998).

4.3 *Agrobacterium tumefaciens*-mediated *Arabidopsis thaliana* floral dip transformation

4.3.1 Tri-parental mating

Tri-parental mating strategy was used for *Agrobacterium tumefaciens* transformation, which is an effective method for moving non-conjugative, but mobilizable plasmids into *Agrobacterium*.

Three initial cultures were grown in 2 ml of LB (Lysogeny broth) medium. One was *Agrobacterium tumefaciens* strain ASE was growing at 28°C for 2 days. The other two were *E. coli* strains containing either pRK2013, which is a helper plasmid to provide the transfer function, or Ti plasmid (target construct) each grown in 2 ml LB medium at 37°C for 1 day. Each culture was supplemented with appropriate antibiotics for selection. From these three pre-cultures 0.5 ml were mixed together and 500 µl were dropped on an LB plate without any antibiotics. The plate was incubated overnight at 28°C. After incubation cells were scratched from the plate and re-suspended in 5 ml of LB medium and aliquots were spread in triplicates on NB (nutrient broth) medium plates containing antibiotics for selection of *Agrobacterium* carrying target construct. The plates was incubated at 28°C for 3 days. Single colonies were picked and inoculated into 2 ml of YEP (Yeast Extract Peptone) liquid medium and cultured over night at 28°C for mini-preps. Isolated DNA was used for PCR amplification of the target construct and subsequent sequencing to confirm that the target plasmid has been successfully transferred to *Agrobacterium.* 500 µl of correct *Agrobacterium* strain carrying target fragment was mixed with 500 µl 40% glycerol and stored as a stock in -

80°C.

4.3.2 *Agrobacterium* mini-prep procedure

1. Collected *Agrobacterium* by centrifugation for 5 min at RT at 13,000 rpm.
2. Re-suspended pellet in 300 µl of solution A, add 100 µl solution B and 100 µl solution C.
3. Extract with phenol-chloroform 3 times (subsequently with 300 µl, 200 µl, 100 µl of phenol-chloroform), spin at 13,000 rpm for 3 min each time.
4. DNA precipitation; add 10 µl 5M NaCl (1/20 Vol) and 400 µl of ethanol to 200 µl supernatant collected from last extraction step and precipitate DNA at -80°C for 10 min; centrifuge for 5 min at 13,000 rpm.
5. Wash the pellet with 70% ethanol and centrifuge for 5 min at 13,000 rpm.
6. Vacuum dry the pellet for 3 min and dissolve the pellet in 100 µl of ddH_2O.

Solution A (100 ml): 1 ml 1M Tris-HCl pH8.0, 200 µl 0.5M EDTA pH8.0, add ddH_2O to 100 ml.

Solution B (50 ml): 0.5 ml 1M Tris-HCl pH8.0, 100 µl 0.5M EDTA pH8.0, 2.5g Sarcosine, add ddH_2O to 50 ml.

Solution C (Pronase Solution 2.5mg/ml) 2.5 mg pronase in 1ml of solution A

4.3.3 Floral dip transformation procedure

1. Grew healthy *Arabidopsis* plants, cut off first bolts to encourage proliferation of many secondary bolts. Optimal plants have many immature flower clusters and not many fertilized siliques.
2. *Agrobacterium* culture; inoculate around 20 µl of glycerol stock into 20 ml YEP medium containing right antibiotics and incubate with shaking at 28°C, 200 rpm for 2 days. Transfer 2.5 ml of this culture into 250 ml YEP medium containing desired antibiotics and incubate overnight at 28°C with shaking (200 rpm).
3. Collected *Agrobacteria* by centrifugation at 4,000 rpm at 15°C for 45 min. Re-suspended pellet in 100 ml of 5% sucrose solution (fresh made) and adjust OD_{600} to be around 1.0 by adding 5% sucrose solution.

Add Silwet L-77 to a final concentration of 0.05% (200 µl Silwet L-77 into 400 ml sucrose solution containing *Agrobacteria*), mix well.

4. Dip above-ground parts of *Arabidopsis* in *Agrobacterium* solution for 1min.
5. Cover dipped plants with foil to keep high humidity and leave them in a dark room overnight and then transfer the plant to green house conditions.

4.4 Bisulfite sequencing analysis

Genomic DNA was isolated from rosette leaves or seedlings using a DNeasy Plant Mini kit (Qiagen). 1.5 µg of genomic DNA was used for bisulfite conversion. Before conversion, genomic DNA was digested overnight with a restriction enzyme that cannot cut the target fragment, *Hind* III was used for instance. Digested DNA was purified with QIAquick PCR Purification Kit (Qiagen). Bisulfite treatment of DNA was conducted by using an EpiTect Bisulphite kit (Qiagen) according to the manufacturer's instructions with the following modification: the conversion PCR programme was changed into: 95^{o}C 2 min, 75^{o}C 2 hr, 95^{o}C 1 min for 9 cycles, and hold at 75^{o}C. The target fragment PCR reactions were performed using Advantage 2 Polymerase Mix from Clontech and the conditions for the amplification of bisulfite-treated DNA were as follows: 95^{o}C for 5 min followed by 39 cycles at 95^{o}C for 30 sec, 30 sec annealing at the temperature depending on the particular primer pair, 72^{o}C for 1 min, and 5 min of final elongation. The PCR was carried out in a total reaction volume of 50 µl. PCR product was gel purified with QIAquick Gel Extraction Kit (Qiagen), ligated into pGEM-T Easy Vector (Promega, Mannheim, Germany), followed by transformation procedure into *E. coli* with blue/white selection plates. Colony PCR was performed in 10 µl with white colonies (containing desired inserts) by using standard M13 primers. After purification with ExoSAP-IT (Affymetrix, Buckinghamshire, UK) PCR products were sent for sequencing. At least 15 clones were used for bisulfite sequencing analysis. Colony PCR conditions were: 95^{o}C 10 min; 95^{o}C 30 sec, 55^{o}C 30 sec, 72^{o}C 1 min for 40 cycles, 72^{o}C 5 min final extension. ExoSAP-IT treatment: add 2 µl of 1:100 dilution of the original solution into 10 µl colony PCR product, incubate at 37^{o}C overnight, inactivate at 80^{o}C for 15 min and store at 4^{o}C until further use. As a control for complete bisulfite

conversion, I used the *PHAVOLUTA* gene: PCR conditions were the same as above except for the first pair of primers I used 40 cycles and for the second pair of primers 26 cycles. Primers used are listed in Tables 4.3 and 4.4.

Table 4.3: List of primers used for bisulfite sequencing (Chapter 1)

Primer	Sequence
Phavoluta primary forward	5'-GTG YAG ATY TGT TTG GAG YTG ATT Y-3'
Phavoluta primary reverse	5'-TTT AAT ATC TAA CAT AAC CAA CCT TT-3'
Phavoluta secondary forward	5'-GGA YYA TAG TGA TGY YAT ATT GTG-3'
Phavoluta secondary reverse	5'-TAT CAT CAA CAA CTT TCC ACA CC-3'
Ubi-88bp-dsred forward	5'-TGG ATG ATG GTA TAT GTA GTA GTT ATA TGT GGA T-3'
Ubi-88bp-dsred reverse	5'-CTC AAA ATA ATA RCC ATT CAC AAA RCC CTC CAT-3'
88bp-ubi-dsred forward	5'-TGA TAG TTT AAA TTG AAG GYG GGA AAY GAT AAT-3'
88bp-ubi-dsred reverse	5'-ACT CAA TTA TCC TTT AAA CCA TAT CTA ACT ATT C-3'
Target forward	5'-GCGGTGTYATYTATGTTAYTAGAT-3'
Target reverse	5'-CTTCTTRATRTTCCATARCTTTCC-3'

4.5 Small RNA isolation and Northern blot analysis

Small RNAs were isolated from pooled *Arabidopsis* inflorescence tissues using the mirVana miRNA isolation kit (Ambion/Applied Biosystems, Brunn am Gebirge, Austria).

4.5.1 Small RNAs isolation procedure:

1. Collected 250 mg inflorescence tissues in 15 ml falcon tube and immediately freeze in liquid nitrogen.
2. Ground tissue into powder by vortexing with 2 grinding spheres for 1 min at maximum speed.
3. Add 2.5 ml lysis/binding buffer and 250 µl Plant RNA isolation aid, vortex and put the tube on ice, added 1/10 volume (250 µl) miRNA homogenate additive

solution, vortex and incubated for 10 min.

Table 4.4: List of primers used for bisulphite sequencing (Chapter 2)

Primer	Sequence
Phavoluta primary forward	5'-GTG YAG ATY TGT TTG GAG YTG ATT Y-3'
Phavoluta primary reverse	5'-TTT AAT ATC TAA CAT AAC CAA CCT TT-3'
Phavoluta secondary forward	5'-GGA YYA TAG TGA TGY YAT ATT GTG-3'
Phavoluta secondary reverse	5'-TAT CAT CAA CAA CTT TCC ACA CC-3'
Chr. 4 SAT5 PART 1 forward	5'-TTT GAA YTY TYA GAG TGG AYT-3'
Chr. 4 SAT5 PART 1 reverse	5'-AAA RAC RAA ACT CCA ART RTA A-3'
Chr. 4 SAT5 PART 2 forward	5'-TGY TTY TTT TAY AYT TGG AGT T-3'
Chr. 4 SAT5 PART 2 reverse	5'-AAC CTA AAT TAC TRA ATT TRC C-3'
Chr. 4 SAT5 PART 3 forward	5'-ATY YGT-GAT-GGA-TAT-TYY-TG-3'
Chr. 4 SAT5 PART 3 reverse	5'-CTC ACA ACT RTC TAT TTA RAA R-3'
CRP4 endocopy forward	5'-TAA GGT TTT TTT ATT TTT TAG TTT TTT ATT ATG-3'
CRP4 endocopy reverse	5'-TAT AAC ATA TAR CAT AAR CTA AAA TTA TAA AAA-3'
CRP4 pro-cds-GFP transgenic copy forward	5'-AAG GTT TTT TTA TTT TTT AGT TTT TTA TTG GAT TT-3'
CRP4 pro-cds-GFP transgenic copy reverse	5'-TCR CCC TTR CTC ACC ATA AAT CTT TT-3'
35spro *CRP4*cds EGFP transgenic copy forward	5'-TTG GAG AGG ATA GGG TAT TYG GGG ATT TAT-3'
35spro *CRP4*cds EGFP transgenic copy reverse	5'-AAC ARC TCC TCR CCC TTR CTC ACC ATA AAT CTT TT-3'
GFP forward	5'-ATG GTG AGT AAG GGY GAG GAG TTG TTT A-3'
GFP reverse	5'-TTA CTT ATA CAR CTC ATC CAT RCC AAA AGT-3'

4. Add 2.5 ml acid phenol-chloroform (Ambion), mix by vortexing for 1 min, centrifuge at 5,000 rpm for 10 min at room temperature (RT) and transfer the supernatant into a new 15 ml falcon tube.

5. Add 1/3 volume of 100% ethanol and vortex thoroughly, apply 700 µl mixture to a filter cartridge in a collection tube and centrifuged at 10,000 rpm for 15 sec at RT, collect flow-through into a new 15 ml falcon tube, repeat with the remaining mixture by using the same filter cartridge and measure the total volume of the filtrate.
6. Added 2/3 volume of 100% ethanol and mix thoroughly, apply 700 µl of mixture to a new filter cartridge with new collection tube, and centrifuge at 10,000 rpm for 15 sec at RT, discard the flow-through and repeated with the remaining mixture.
7. Wash the filter with 700 µl miRNA of wash solution 1, centrifuge at 10,000 rpm for 10 sec at RT.
8. Wash the filter twice with 500 µl of wash solution 2/3, centrifuge at 10,000 rpm for 10 sec for the first time, and 1 min for the second wash at RT.
9. Transfer the filter into a fresh collection tube and elute small RNAs with 100 µl of pre-heated nuclease-free water by incubating at 65^{o}C for 10 min, centrifuged 1 min at 10,000 rpm RT, stored small RNAs at at -80 ^{o}C.

Run small RNA on 15% polyacrylamide gel containing 7% urea for Northern blot hybridization. Sequence of the probe used for secondary siRNAs is: TTC GAT TAT GAA TAA TAA ACA GGC TGC ATC TTC AGG CAT CC.

4.5.2 PAGE and Northern blot procedure

1. Preparation of the 15% polyacrylamide/7M urea gel: dissolve 28.8 g urea in a mixture of 13.5 ml ddH_2O , 22.5 ml 40% acrylamide:bis-acrylamide 37.5:1 (Applichem, Germany), and 1.35 ml 10 X TBE (Tris/Borate/EDTA) solution, heat at 64^{o}C until urea was completely dissolved, filter the solution through miracloth and add 62.5 µl 40% ammonium peroxodisulfate and 40 µl TEMED (Tetramethylethylenediamine), mix well and pour gel (2 mm thick), and then allow the gel to polymerize for at least 2 hr at RT.
2. Preparation of small RNAs samples: mix 50 µl small RNAs sample isolated as described above with 50 µl loading buffer (96% deionized formamide, 20 mM EDTA, 1 mg/ml bromphenol blue), denature at 95^{o}C for 5 min and keep on ice

for 5 min to prevent renaturation.

3. Set up the gel with 0.225 X TBE running buffer (before loading the samples wash the wells with running buffer to get rid of urea from the wells), load the samples to gel and run the gel at 800V, 25mA, 10W until the bromphenol blue reached lower edge of the gel.
4. Stain the gel in 0.5 X TBE containing 1 μg/ml ethidium bromide for 10 min, wash the gel with 0.5 X TBE for 20 min and take a photo under Ultraviolet (UV) light
5. Transfer RNA to Hybond N+ membrane by electroblotting with BioRad semi dry electro blotter (transfer sandwich setup order: from bottom to top, 3 sheets Whatman 3MM, Hybond N+ membrane (Amersham), gel, 3 sheets Whatman 3 MM, all of them were soaked in 0.5 X TBE), transfer RNA to the membrane at 10V for 1 hr in cold room, rinse the membrane briefly in 0.5 X TBE and air dry the membrane.
6. Crosslink RNA to the membrane by using the Stratagene UV-Crosslinker (auto-crosslink function, 1.5 times)
7. Pre-hybridize the membrane with 40 ml Ultrahyb-oligo hybridization buffer (Ambion) at 42 ^{o}C for 3 hr in a hybridization oven.
8. 5' end labelling of the probe with T4 polynucleotide kinase (PNK): 1 μl of 10pMol/ μl oligo, 1 μl T4 PNK 10U/μl (Roche), 2 μl 10 X PNK buffer, 5 μl γ ^{32}P ATP, add 11 μl sterilized water and incubated on ice for 1 hr, stop the reaction by adding 1 μl 0.5M EDTA pH 8.0, add another 79 μl sterilized water and purify the labelled probe with NucAway spin column (Ambion) according to the user's manual.
9. Add purified probe into the hybridization tube containing pre-hybridization buffer and hybridized at 42 ^{o}C in a hybridization oven overnight.
10. Wash the membrane with washing buffer [2X SSC (saline-sodium citrate)/0.5% SDS (sodium dodecyl sulphate)] two times at 42^{o}C for 15 min.
11. Exposed the membrane with sensitive X-ray film (Kodak) at -80 ^{o}C (exposure time depends on probe).

4.6 Chromatin immunoprecipitation assay

Arabidopsis seeds were germinated in soil. Collect 1.5 g of 3-4 week old plant material (do not use roots) per chromatin preparation. It is imperative to avoid contamination with soil as much as possible during harvest.

4.6.1 Chromatin Crosslinking

1. Harvest 1.5 g seedlings and place them into a 50 ml Falcon tube,
2. Rinse seedlings twice with 40 ml bidestilled water. Remove as much water as possible after second wash. This washing step is not necessary if you avoid soil contamination during harvesting.
3. Add 37 ml of 1% formaldehyde solution. Gently submerge seedlings to the bottom of the tube by stuffing the tube with nylon mesh. Screw on cap and poke cap with needle holes. Put in exsicator and draw vacuum for 2 min, then release vacuum slowly and apply again for 8 min.
4. Release vacuum slowly and shake exsicator slightly to remove air bubbles. Seedlings should appear translucent.
5. Add 2.5ml of 2 M glycine to quench crosslinking. Draw vacuum for 5 minutes.
6. Again, release vacuum slowly and shake exsicator slightly to remove air bubbles.
7. Remove nylon mesh, decant supernatant and wash seedlings twice with 40 ml of bidestilled water. After second wash remove as much water as possible and put seedlings between two layers of kitchen paper. Roll up paper layers carefully to remove as much liquid as possible.

At this step plant material can be shock-frozen in liquid nitrogen and stored at -80°C until further use without any affect on reproducibility of results.

4.6.2 Chromatin preparation (Do everything on ice)

1. Pre-cool mortar and pistil with liquid nitrogen. Grind plant material to a fine powder. Alternatively you can use ceramic beads (7 per one tube).

2. Remove ceramic beads and allow powder to cool down, but do not allow thawing. Add 30 ml of Extraction Buffer 1 stored on ice. Vortex to mix and keep on ice until solution is homogenous.
3. Filter extract through Miracloth into a new, ice-cold 50 ml Falcon tube. Wash Miracloth with additional 10-15 ml of Extraction buffer 1. Rigidly press out solid material.
4. Centrifuge extract using the Beckman JS 7.5 rotor (or equivalent) at 2,000 X *g* for 15 minutes at 4°C.
5. Gently pour off supernatant and resuspend pellet in 1ml of Extraction Buffer 2 by pipetting up and down. Transfer solution to Eppendorf tube.
6. Spin in cooled benchtop centrifuge at 2,000 X *g* for 15 min.
7. Remove supernatant and resuspend pellet in 300 µl of Extraction Buffer 2 by pipetting up and down
8. Add 300 µl of Extraction Buffer 3 to fresh Eppendorf tube and carefully layer solution from step 7 onto it.
9. Spin in cooled benchtop centrifuge at 6,000 X *g* for 20 min. In meantime, prepare 10 ml Nuclei Lysis Buffer and 20 ml ChIP Dilution Buffer.
10. Remove supernatant and resuspend pellet in 1 ml of cold Extraction buffer 2 and spin for 15 min at 3,000 X *g*.
11. Resuspend nuclei in 400 µl of Nuclei Lysis Buffer. Resuspend by pipetting up and down and by vortexing.
12. Sonicate 4 X 10 seconds, 40% duty cycle and 20% power (BandelinSonoplus HD 2070 with MS 73 probe). Put on ice for 1 minute between sonication steps.
13. Spin in cooled benchtop centrifuge at 13,000 rpm for 10 minutes. Add supernatant to new Eppendorf tube and measure DNA concentration on Nanodrop. Adjust all samples to same OD values by adding Nuclei Lysis Buffer (10% difference is OK). At this step it is possible to stop procedure and freeze chromatin at -80^{o}C.

4.6.3 Pre-clearing of chromatin and immunoprecipitation (IP)

1. Dilute chromatin with ChIP Dilution Buffer ten times to reduce SDS

concentration to 0.1%.

2. Prepare Protein A agarose beads pre-absorbed with sheared salmon sperm DNA (Upstate, Cat. 16-157) by rinsing 100 µl of beads (per sample) 3 times with 1ml ChIP Dilution Buffer in an Eppendorf tube. Spin in cooled benchtop centrifuge for 30 seconds at 13,000 rpm between the washes to pellet the beads. Add beads to diluted chromatin (in 15 ml Falcon tube) solution and rotate for 2 hour at 4°C.
3. Spin for 30 seconds at 13,000 to pellet the beads. Be careful not to contaminate solution with carryover of beads.
4. Store 1/10 of chromatin solution which you use for each IP at -20°C. This will serve as input control.
5. Separate chromatin solutions into Eppendorf tubes and add appropriate antibodies and Protein A agarose washed with ChIP dilution buffer. I used ~10 µg of RNA Pol II antibody per IP. Also, set up chromatin solution without antibody, which serves as mock IP. Rotate overnight at 4°C.

4.6.4 Collection, washes and elution of immune complexes

1. Spin IP in cooled benchtop centrifuge at 5,000 rpm for 30 sec to collect beads and discard supernatant.
2. Add 1ml of Low Salt Wash Buffer per tube. Rotate for 5 min at 4°C.
3. Spin in cooled benchtop centrifuge at 5,000 rpm for 30 sec to collect beads and discard supernatant.
4. Add 1ml of High Salt Wash Buffer per tube. Rotate for 5 min at 4°C.
5. Spin in cooled benchtop centrifuge at 5,000 rpm for 30 sec to collect beads and discard supernatant.
6. Add 1ml of LiCl Wash Buffer per tube. Rotate for 5 min at 4°C.
7. Spin in cooled benchtop centrifuge at 5,000 rpm for 30 sec to collect beads and discard supernatant.
8. Add 1ml of TE Buffer per tube. Rotate for 5 min at 4°C.
9. Spin in cooled benchtop centrifuge at 5,000 rpm for 30 sec to collect beads and discard supernatant.
10. Repeat TE wash. Spin in cooled benchtop centrifuge at 5,000 rpm for 30 sec to

collect beads and discard supernatant.

11. Elute immune complexes by adding 250 µl of Elution Buffer. Vortex briefly to mix and incubate at 65°C for 15 min. Spin in benchtop centrifuge at 13,000 rpm for 30 sec and transfer supernatant to a fresh Eppendorf tube.
12. Repeat elution and finally combine the two elutes.
13. Add Elution Buffer to the input control aliquoted on Day 1 to obtain 500 µl final volumes.

4.6.5 Reverse crosslinking

Add 20 µl of 5 M NaCl to samples. Incubate overnight at 65°C.

4.6.6 DNA cleanup

1. Add 10 µl of 0.5 M EDTA, 20 µl 1 M Tris-HCl pH 6.5 and 1 µl of 20 mg/ml proteinase K per tube. Incubate for 1 hour at 45°C.
2. Extract samples with 1 volume of phenol-chloroform. Spin in cooled benchtop centrifuge at 13,000 rpm for 15 min and transfer supernatant to 2 ml reaction tube.
3. Precipitate DNA with 1/10 volume 3 M NaOAc pH 5.2 and 3 volumes absolute ethanol. Add 4 µl glycogen (Roche, Cat. 901393) per precipitation. Incubate at -20°C for at least 1 hour.
4. Spin in cooled benchtop centrifuge at 13,000 rpm for 15 min. Wash pellet with 1 ml of 70% ethanol and spin again. Discard supernatant and vacuum-dry pellet for 10 minutes. Dissolve DNA in 50 µl 10 mM Tris-HCl pH7.5. Proceed to PCR reactions.

4.6.7 Materials and Reagents for ChIP

Extraction Buffer 1: 0.4 M Sucrose, 10 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$,5 mM β-mercaptoethanol, with Protease Inhibitors

Extraction Buffer 2: 0.25 M Sucrose, 10 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, 1% Triton X-100, 5 mM β-mercaptoethanol, with Protease Inhibitors

Extraction Buffer 3: 1.7 M Sucrose, 10 mM Tris-HCl, pH 8.0, 2 mM $MgCl_2$, 0.15% Triton X-100, 5 mM β-mercaptoethanol, with Protease Inhibitors
Nuclei Lysis Buffer: 50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 1% SDS, with Protease Inhibitors
ChIP Dilution Buffer: 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl, pH 8.0, 167 mM NaCl
Elution Buffer: 1% SDS, 0.1 M $NaHCO_3$
Low Salt Wash Buffer: 150 mM NaCl, 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.0
High Salt Wash Buffer: 500 mM NaCl, 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.0
LiCl Wash Buffer: 0.25 M LiCl, 1% Nonidet P-40, 1% sodium deoxycholate, 1 mM EDTA, 10 mM Tris-HCl, pH 8.0
TE Buffer: 10 mM Tris-HCl, pH 8.0, 1 mM EDTA
Protease Inhibitors: 1 tablet (Complete Mini Protease Inhibitor Cocktail tablets (Roche)) per 50ml of solution.

4.7 Real-time PCR

For performing the real-time PCR, first dilute input from ChIP into 1:10, 1:100, 1:1000, 1:10000, and also diluted mock and Pol II precipitated DNA into 1:10. Use 10 µl of diluted DNA as a template plus 10 µl of PCR mix (SensiFAST™ SYBR & Fluorescein Kit from Bioline) and 0.8 µl primer mix. PCR program as follows, step 1: 95°C for 5 min, step 2: 95°C for 30 sec and 60°C for 1 min, followed by 40 cycles, ran on BioRad iCylcer iQ5 Real-time PCR Thermal cycler. Primers used for real-time PCR are listed in Table 4.5.

4.8 GUS assays and GFP visualization

GUS assays were performed as described previously (Jones-Rhoades MW *et al*.,

2007). Briefly, ovary was dissected from unopened flowers buds (emasculated 24 hr before) in a small petri dish containing GUS staining solution (50 mM sodium phosphate buffer pH 7.0, 10 mM EDTA, 0.1% Triton X-100, 2 mM potassium ferrocyanide, 2 mM potassium ferricyanide, and 1 mg/ml X-Gluc) on ice. Ovaries were incubated at 37 ˚C for 30 min, cleared with 20% methanol/4% concentrated HCl solution at 55˚C for 15 min, and then transferred into a solution containing 60% ethanol/1.8 M NaOH at 25 ˚C for 10 min. Samples were washed with 30% ethanol and

Table 4.5: Primers used for real-time and semi-quantitative PCR (Chapter 1)

Primer	Sequence
Ubi ChIP short forward	5'- TTG ACA ACA GGA CTC TAC AG-3'
Ubi ChIP short reverse	5'-AAA CCA TTA ACC CTA AAC C-3'
Ubi ChIP long forward	5'- CAC CAA CCA GCG AAC CAG CA-3'
Ubi ChIP long reverse	5'-ATT TCT GGA TGC CGA CAG-3'
Actin for ChIP Forward	5'-CGT TTC GCT TTC CTT AGT GTT AGC T-3'
Actin for ChIP Reverse	5'-AGC GAA CGG ATC TAG AGA CTC ACC TTG-3'
IGN5 for ChIP Forward	5'-TCC CGA GAA GAG TAG AAC AAA TGC TAA AA-3'
IGN5 for ChIP Reverse	5'-CTG AGG TAT TCC ATA GCC CCT GAT CC-3'
88bp-dsRED RT forward	5'-AAT AAT AAA CAG GCT GCA TC-3'
88bp-dsRED RT reverse	5'-CTA CTG GGA GCC GGA GTG GC-3'
dsRED RT forward	5'-ATG GAC AAC ACC GAG GAC GT-3'
dsRED RT reverse	5'- CTA CTG GGA GCC GGA GTG GC-3'
RT forward	5'-CGA CAA TCT GAT CCA CTA-3'
RT reverse	5'-AAA CCA TTA ACC CTA AAC C-3'
Ubipro oligo	5'-ATT TCT GGA TGC CGA CAG-3'
At3g18780 (*act2*) forward	5'-GCC ATC CAA GCT GTT CTC TC-3'
At3g18780 (*act2*) reverse	5'-GGG CAT CTG AAT CTC TCA GC-3'

10% ethanol and mounted onto slides with a 50% glycerol solution. Ovules were dissected from ovary, and images were acquired using a Leica stereo fluorescence microscope MZFLIII.

To visualize GFP, ovules were dissected ovaries isolated from unopened flower buds, and localization of the GFP signal was visualised by using Leica stereo fluorescence microscope MZFLIII.

4.9 Treatment with flg22 and heat stress

Dried seeds were sterilised with 70% ethanol/ 0.05% TritonX-100 and germinated on Murashige and Skoog (MS) medium. One week after germination, small seedlings were transferred to liquid MS medium supplied with 1 µM Flagellin22 (flg22) peptide (kindly provided by Elke Logemann and Imre Somssich, Max Planck Institute for Plant Breeding Research, Cologne) (4 seedlings per 400 µl of medium). Batches of seedlings were collected after 10 min, 20 min, and 30 min of flg22 treatment followed by total RNA isolation and semi-quantitative RT-PCR to detect expression of the *CRP4* and *MPK3,* which is a positive control for flg22 treatment.

For heat stress, approximately one week old seedlings were transferred into liquid MS medium at 4˚C 24 hr, followed by 4, 8, and 12 hours incubation at 37 ˚C. Stressed seedlings were allowed to recover at 21 ˚C for 3 days, and then used for RNA isolation and semi-quantitative RT-PCR to detect expression of the *CRP4* gene and Onsen retrotransposon.

4.10 Semi-quantitative RT-PCR

Total RNA was extracted from 3 week-old seedlings (100 mg) or mixed floral inflorescence tissues (100 mg) by using TRIzol® Reagent (Invitrogen).

4.10.1 Total RNA isolation procedure:

All material used (i.e. tubes, pipette tips) were RNase free.

1. About 100 mg of rosette leaves or flower tissues were harvested into 15 ml falcon tubes and frozen in liquid nitrogen immediately to avoid RNA degradation.
2. Frozen tissue was disrupted with two grinding spheres by vortexing at maximum speed for 1 min; 1 ml of TRIzol was added, vortexed and the tubes were left at RT for 5 min.
3. Solution was transferred into a 2 ml Eppendorf tube, 200 µl of chloroform were added, vortexed thoroughly for 1 min and centrifuged at 14,000 rpm at 4ºC for 15 min.
4. The aqueous phase was transferred into a new 1.5 ml Eppendorf tube. RNA was precipitated by addition of the same volume of isopropanol, incubation at RT for 10 minutes and centrifugation at 13,000 rpm for 10 min at 4ºC.
5. The supernatant was discarded. The pellet was washed with 1 ml of 75% ethanol, followed by centrifugation at 13,000 rpm at 4ºC for 5 min.
6. Subsequently, the pellet was air-dried and dissolved in 150 µl RNase-free water
7. Add same volume of phenol-chloroform-isoamyl alcohol (25:24:1) to RNA solution, vortex at maximum speed for 1 min, centrifuge at 13,000 rpm for 10 min at 4ºC
8. Transfer the supernatant into a new 1.5 ml Eppendorf tube, add same volume of chloroform-isoamyl alcohol (24:1), vortex at maximum speed for 1 min, centrifuge at 13,000 rpm for10 min at 4ºC
9. Transferred the supernatant into a new 1.5ml Eppendorf tube, add 1/10 3M pH5.2 sodium acetate, and 2.5 volumes of 100% ethanol. Precipitate RNA at -20 ºC for at least 1 hr. Centrifuged at 13,000 rpm for 10 min at 4ºC
10. Wash the pellet with 700 µl of 75% ethanol (gently mixed by inversion) centrifuge at 13,000 rpm for 5 min at 4ºC
11. Dissolved the pellet in 150 µl RNase-free water
12. Repeat steps 7-10, and finally dissolve RNA in 20 µl RNase-free water
13. Quantify the RNA sample with Nano drop and store RNA samples in -80 ºC.

One μg of total RNA was used for reverse transcription by using RevertAid™ H Minus First Strand cDNA Synthesis Kit (Fermantas) following the manufacturer's instructions. One μL cDNA was used for semi-quantitative reverse transcriptase–mediated (RT) PCR analysis. The PCR conditions were 95°C for 5 min followed by 23 (actin), 25 (MPK3), 40 (CRP genes) amplification cycles (95°C for 30 sec, 55°C for 30 sec, and 72°C for 1 min). Actin was used as an internal control. Primers used for semi-quantitative RT-PCR are listed in Tables 4.5 and 4.6.

4.11 Protein extraction and Western blotting

Total proteins were extracted from 3 week old seedlings grown on Murashige and Skoog (MS) medium. Approximately 50 mg seedlings were harvested, ground in liquid nitrogen and suspended in 100 μl of extraction buffer (50 mM HEPES-KOH pH 7.9, 400 mM KCl, 2.5 mM $MgCl_2$, 1 mM EDTA, 1 mM DTT, 0.1% Triton X-100) supplemented with EDTA-free protease inhibitor cocktail (Roche, Vienna, Austria). After centrifugation for 15 min at 4°C, the supernatant was mixed with the same volume of extraction buffer without KCl. Proteins were mixed with the same volume of 2 X loading dye and after denaturing for 5 min at 95^{o}C loaded onto a 10% SDS–PAGE and transferred to a PVDF membrane. Membranes were probed with GFP antibody (1:1,000; mouse monoclonal, Roche, Vienna, Austria) and tubulin antibody (1:1,000; mouse monoclonal, Sigma-Aldrich, Vienna, Austria) followed by horseradish peroxidase conjugated goat anti- mouse secondary antibody (1:10,000; Biorad, Vienna, Austria). The blots were developed using an enhanced chemoluminescence ECL kit (GE Healthcare, Vienna, Austria), and exposed to X-ray film (GE Healthcare, Vienna, Austria).

Table 4.6: Primers used for semi-quantitative PCR (Chapter 2)

Primer	Sequence
At2g18042 forward	5'-TAT GCT TTC ACT CCT AGC GG-3'
At2g18042 reverse	5'-AAC AAT GCT TTA CAA CAA GAT G-3'
At3g42565 forward	5'-TGT GTT TTC CCT TCT ATA TG-3'
At3g42565 reverse	5'-GTA TGC AAT TAA CTT TGG CA-3'
At4g09545 (*CRP4*) forward	5'-TTC TGT AAA TGC TCA GTT ACC T-3'
At4g09545 (*CRP4*) reverse	5'-GGA GGG AAA AGC GGA ATA AG-3'
At5G60978 forward	5'-TCT CCG TAA ATG CTC AAC TA-3'
At5G60978 reverse	5'-GCA TGC AAT TGG ATT CGA CA-3'
Onsen forward	5'-TCT CTG TCT AAT CTA TCA A-3'
Onsen reverse	5'-AAG TTG CTC TAT TGT CAT AGC-3'
At3g45640 (*ATMPK3*) forward	5'-GGA ATC GTT TGC TCT GTG TTG-3'
At3g45640 (*ATMPK3*) reverse	5'-CAT GAT CAA GAT GAC GAA GAA GC-3'
At3g18780 (*act2*) forward	5'-GCC ATC CAA GCT GTT CTC TC-3'
At3g18780 (*act2*) reverse	5'-GGG CAT CTG AAT CTC TCA GC-3'
CRP4 endocopy RT forward-1	5'-TAT TTC TCA GTT TCT CAT TAT GC-3'
CRP4 endocopy RT forward-2	5'-AAG GCT TTT TTA TTT CTC AGT-3'
35spro *CRP4* transgenic copy RT forward	5'-ACA GGG TAC CCG GGG ATC CAT-3'
MYB98 RT forward	5′-CTC CTT TCC AAA ACA ATG GAG AAT TTC GTC-3′
MYB98 RT reverse	5′-GTC CTC TTC TTC ACT CCA TGT TTC TTT CTT A-3′

5. References

Aceituno FF, Moseyko N, Rhee SY, Gutiérrez RA. 2008. The rules of gene expression in plants: organ identify and gene body methylation are key factors for regulation of gene expression inArabidopsis thaliana. BMC Genomics. 9: 438.

Allen E, Xie Z, Gustafson AM, Carrington JC. 2005. microRNA-Directed Phasing during *Trans*-Acting siRNA Biogenesis in Plants. Cell. 121: 207-221.

Alvarez-Venegas R, Pien S, Sadder M, Witmer X, Grossniklaus U, Avramova Z. 2003. ATX1, an *Arabidopsis* homolog of trithorax, activates flower homeotic genes. Current Biology. 13: 627-37.

Asai T, Tena G, Plotnikova J, Willmann MR, Chiu WL, Gomez Gomez L, Boller T, Ausubel FM, Sheen J. 2002. MAP kinase signalling cascade in Arabidopsisinnate immunity. Nature. 415: 977-983.

Aufsatz W, Mette MF, van der Winden J, Matzke AJ, Matzke M. 2002. RNA-directed DNA methylation in *Arabidopsis*. Proc. Natl. Acad. Sci. USA. 99: (Suppl 4) 16499-16506.

Aufsatz W, Mette MF, van der Winden J, Matzke M, Matzke AJ. 2002. HDA6, a putative histone deacetylase needed to enhance DNA methylation induced by double-stranded RNA. EMBO J. 21: 6832-6841.

Ausin I, Mockler TC, Chory J, Jacobsen SE. 2009. IDN1 and IDN2 are required for *de novo* DNA methylation in *Arabidopsis thaliana*. Nature Struct. Mol. Biol.16: 1325-1327.

Bauer MJ, Fischer RL. 2011. Genome demethylation and imprinting in the endosperm. Curr. Opin. Plant Biol. 14: 162-167.

Baumberger N, Baulcombe DC. 2005. *Arabidopsis* ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. Proc. Natl. Acad. Sci. USA. 102: 11928-11933.

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: 649-654.

Borsani O, Zhu JH, Verslues PE, Sunkar R, Zhu JK. 2005. Endogenous siRNAs derived from a pair of natural *cis*-antisense transcripts regulate salt tolerance in *Arabidopsis*. Cell. 123:1279-1291.

Cao X, Jacobsen SE. 2002. Role of the *Arabidopsis* DRM methyltransferases in *de novo* DNA methylation and gene silencing. Current Biology. 12: 1138-1144.

Chan SW, Zhang X, Bernatavichute YV, Jacobsen SE. 2006. Two-step recruitment of RNA-directed DNA methylation to tandem repeats. PLoS Biol. 4: e363.

Chan SW, Henderson IR, Zhang X, Shah G, Chien JS, Jacobsen SE. 2006. RNAi, DRD1, and histone methylation actively target developmentally important non-CG DNA methylation in *Arabidopsis*. PLoS Genetics. 2: e83.

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

Chen X. 2009. Small RNAs and their roles in plant development. Annu. Rev. Cell Dev. Biol. 25: 21-44.

Choi Y, Gehring M, Johnson L, Hannon M, Harada JJ, Goldberg RB, Jacobsen SE, Fischer RL. 2002. DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in *Arabidopsis*. Cell. 110: 33-42.

Clough SJ, Bent AF 1998. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant J. 16: 735-743.

Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE. 2008. Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. Nature. 452: 215-219.

Dalmay T, Hamilton A, Rudd S, Angell S, Baulcombe DC. 2000. An RNA-dependent

RNA polymerase gene in *Arabidopsis* is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. Cell 101: 543-553.

Daxinger L. Ph.D Dissertation. 2008.

Daxinger L, Kanno T, Bucher E, van der Winden J, Naumann U, Matzke AJM, Matzke M. 2009. A stepwise pathway for biogenesis of 24-nt secondary siRNAs and spreading of DNA methylation. EMBO J. 28:48-57.

Dowen RH, Pelizzola M, Schmitz RJ, Lister R, Dowen JM, Nery JR, Dixon JE, Ecker JR. 2012. Widespread dynamic DNA methylation in response to biotic stress. Proc. Natl. Acad. Sci. USA. 109 pp. E2183-E2191

Earley K, Lawrence RJ, Pontes O, Reuther R, Enciso AJ, Silva M, Neves N, Gross M, Viegas W, Pikaard CS. 2006. Erasure of histone acetylation by *Arabidopsis* HDA6 mediates large-scale gene silencing in nucleolar dominance. Genes & Development. 20: 1283-1293.

Ebbs ML, Bartee L, Bender J. 2005. H3 lysine 9 methylation is maintained on a transcribed inverted repeat by combined action of SUVH6 and SUVH4 methyltransferases. Molecular and Cellular Biology. 25: 10507-10515.

Ebbs ML, Bender J. 2006. Locus-specific control of DNA methylation by the *Arabidopsis* SUVH5 histone methyltransferase. Plant Cell. 18: 1166-1176.

El-Shami M, Pontier D, Lahmy S, Braun L, Picart C, Vega D, Hakimi MA, Jacobsen SE, Cooke R, Lagrange T. 2007. Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. Genes & Development. 21: 2539-2544.

Eun C, Lorkovic Z, Naumann U, Long Q, Havecker E, Simon SA, Meyers BC, Matzke AJM, Matzke M. 2011. AGO6 functions in RNA-mediated transcriptional gene silencing in shoot and root meristems in *Arabidopsis thaliana*. PLoS One. 6:e25730.

Eun C, Lorkovic ZJ, Sasaki T, Naumann U, Matzke AJ, Matzke M. 2012. Use of Forward Genetic Screens to Identify Genes Required for RNA-Directed DNA

Methylation in Arabidopsis thaliana. Cold Spring Harb. Symp. Quant. Biol. Epub ahead of print.

Feng Q, Wang H, Ng HH, Erdjument-Bromage H, Tempst P, Struhl K, Zhang Y.2002. Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. Current Biology. 12: 1052-1058.

Finnegan EJ, Margis R, Waterhouse PM. 2003. Posttranscriptional gene silencing is not compromised in the *Arabidopsis CARPEL FACTORY (DICER-LIKE1)* mutant, a homolog of Dicer-1 from *Drosophila*. Current Biology. 13: 236-240.

Gao Z, Liu HL, Daxinger L, Pontes O, He X, Qian W, Lin H, Xie M,Lorkovic ZJ, Zhang S, Miki D, Zhan X, Pontier D, Lagrange T, Jin H, Matzke AJ, Matzke M, Pikaard CS, Zhu JK. 2010. An RNA polymerase II- and AGO4-associated protein acts in RNA-directed DNA methylation. Nature. 465: 106-109.

Gong Z, Morales-Ruiz T, Ariza RR, Roldán-Arjona T, David L, Zhu JK. 2002. ROS1, a repressor of transcriptional gene silencing in *Arabidopsis*, encodes a DNA glycosylase/lyase. Cell 111: 803-14.

González RM, Ricardi MM, Iusem ND. 2011. Atypical epigenetic mark in an atypical location: cytosine methylation at asymmetric (CNN) sites within the body of a non-repetitive tomato gene. BMC Plant Biology 11: 94.

Goodrich J, Puangsomlee P, Martin M, Long D, Meyerowitz EM, Coupland G. 1997. A Polycombgroup gene regulates homeotic gene expression in *Arabidopsis*. Nature. 386: 44-51.

Grossniklaus U, Vielle-Calzada JP, Hoeppner MA, Gagliano WB. 1998. Maternal control of embryogenesis by MEDEA, a polycomb group gene in *Arabidopsis*. Science. 280: 446-450.

Grossniklaus U. 2011. Plant germline development: a tale of cross-talk, signaling and cellular interactions. Sex Plant Reprod. 24: 91-95.

Haag JR, Pikaard CS. 2011. Multisubunit RNA polymerases IV and V: purveyors of

non-coding RNA for plant gene silencing. Nature Rev. Mol. Cell Biol. 12: 483-492

Hajdukiewicz P, Svab Z, Maliga P. 1994. The small versatile pPZP family of *Agrobacterium* binary vectors for plant transformation. Plant Mol. Biol. 25: 989–994.

He XJ, Hsu YF, Zhu S, Wierzbicki AT, Pontes O, Pikaard CS, Liu HL,Wang CS, Jin H, Zhu JK. 2009. An effector of RNA-directed DNA methylation in *Arabidopsis* is an ARGONAUTE 4- and RNA-binding protein. Cell. 137: 498-508.

He XJ, Hsu YF, Zhu S, Liu HL, Pontes O, Zhu J, Cui X, Wang CS, Zhu JK. 2009. A conserved transcriptional regulator is required for RNA-directed DNA methylation and plant development. Genes & Development. 23: 2717-2722.

Henderson IR, Dean C. 2004. Control of *Arabidopsis* flowering: the chill before the bloom. Development. 131: 3829-38.

Henderson IR, Zhang X, Lu C, Johnson L, Meyers BC, Green PJ, Jacobsen SE. 2006. Dissecting *Arabidopsis thaliana* DICER function in small RNA processing, gene silencing and DNA methylation patterning. Nature Genetics 38:721-725.

Herr AJ, Jensen MB, Dalmay T, Baulcombe DC. 2005. RNA polymerase IV directs silencing of endogenous DNA. Science 308:118-120.

Ibarra CA, Feng X, Schoft VK, Hsieh TF, Uzawa R, Rodrigues JA, Zemach A, Chumak N, Machlicova A, Nishimura T, Rojas D, Fischer R, Tamaru H, Zilberman D. 2012. Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. Science 337: 1360-1364.

Ito H, Gaubert H, Bucher E, Mirouze M, Vaillant I, Paszkowski J. 2011. A siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. Nature. 472: 115-119.

Jackson JP, Lindroth AM, Cao X, Jacobsen SE. 2002. Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature .416: 556-560.

Jacob Y, Feng S, Leblanc CA, Bernatavichute YV, Stroud H, Cokus S, Johnson LM,

Pellegrini M, Jacobsen SE, Michaels SD. 2009. ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing. Nature Struct. Mol. Biol. 16: 763-768.

Jeddeloh JA, Stokes TL, Richards EJ. 1999. Maintenance of genomic methylation requires a SWI2/SNF2-like protein. Nature Genetics. 22: 94-97.

Jenuwein T, Laible G, Dorn R, Reuter G. 1998. SET domain proteins modulate chromatin domains in eu- and heterochromatin. Cell Mol Life Sci. 54: 80-93.

Johnson L, Cao X, Jacobsen S. 2002. Interplay between two epigenetic marks. DNA methylation and histone H3 lysine 9 methylation. Current Biology. 12: 1360-67.

Jones L, Hamilton AJ, Voinnet O, Thomas CL, Maule AJ, Baulcombe DC. 1999. RNA–DNA interactions and DNA methylation in post-transcriptional gene silencing. Plant Cell. 11: 2291-2301.

Jones-Rhoades MW, Borevitz JO, Preuss D. 2007. Genome-wide expression profiling of the *Arabidopsis* female gametophyte identifies families of small, secreted proteins. PLoS Genetics. 3: 1848-1861.

Jullien PE, Katz A, Oliva M, Ohad N, Berger F. 2006. Polycomb group complexes self-regulate imprinting of the Polycomb group gene MEDEA in *Arabidopsis*. Current Biology. 16: 486-92.

Jurka J. 2006. ATSAT5: A recently expanded satellite from chromosome 5. Repbase Reports. 6(2): 50-50.

Kankel M, Ramsey D, Stokes T, Flowers S, Haag J, Jeddeloh J, Riddle N, Verbsky M, Richards E. 2003. *Arabidopsis* MET1 cytosine methyltransferase mutants. Genetics 163: 1109-1122.

Kanno T, Mette MF, Kreil DP, Aufsatz W, Matzke M, Matzke AJ. 2004. Involvement of putative SNF2 chromatin remodeling protein DRD1 in RNA-directed DNA methylation. Current Biology. 14: 801-805.

Kanno T, Aufsatz W, Jaligot E, Mette MF, Matzke M, Matzke AJ. 2005. A SNF2-like protein facilitates dynamic control of DNA methylation. EMBO Reports 6: 649-655.

Kanno T, Huettel B, Mette MF, Aufsatz W, Jaligot E, Daxinger L, Kreil DP, Matzke M, Matzke AJ. 2005. Atypical RNA polymerase subunits required for RNA-directed DNA methylation. Nature Genetics. 37: 761-765.

Kanno T, Bucher E, Daxinger L, Huettel B, Bohmdorfer G, Gregor W, Kreil DP, Matzke M, Matzke AJ. 2008. A structural-maintenance-of-chromosomes hinge domain-containing protein is required for RNA-directed DNA methylation. Nature Genetics. 40: 670-675.

Kanno T, Bucher E, Daxinger L, Huettel B, Kreil DP, Breinig F, Lind M, Schmitt M, Simon SA, Gurazada SG, Meyers BC, Lorkovic ZJ, Matzke AJM, Matzke M. 2010. RNA-directed DNA methylation and plant development require an IWR1-type transcription factor. EMBO Reports. 11: 65-71.

Kasahara R. D., Portereiko M. F., Sandaklie-Nikolova L., Rabiger D. S., Drews G. N. 2005. MYB98 is required for pollen tube guidance and synergid cell differentiation in Arabidopsis. Plant Cell. 17: 2981-2992.

Katiyar-Agarwal S, Morgan R, Dahlbeck D, Borsani O, Villegas A, Zhu JK, Staskawicz BJ, Jin HL. 2006. A pathogen-inducible endogenous siRNA in plant immunity. Proc Natl Acad Sci USA. 103: 18002-18007.

Katz A, Oliva M, Mosquna A, Hakim O, Ohad N. 2004. FIE and CURLY LEAF polycomb proteins interact in the regulation of homeobox gene expression during sporophyte development. Plant Journal. 37 (5): 707-719.

Kohler C, Hennig L, Bouveret R, Gheyselinck J, Grossniklaus U, Gruissem W. 2003. *Arabidopsis* MSI1 is a component of the MEA/FIE Polycomb group complex and required for seed development. EMBO J. 22: 4804-14.

Law JA, Jacobsen SE. 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nature Review Genetics 11:204-220.

Law JA, Ausin I, Johnson LM, Vashisht AA, Zhu JK, Wohlschlegel JA, JacobsenSE. 2 010. A protein complex required for polymerase V transcripts and RNA-directed DNA methylation in *Arabidopsis*. Current Biology 20: 951-956.

Li CF, Pontes O, El-Shami M, Henderson IR, Bernatavichute YV, Chan SW, Lagrange T, Pikaard CS, Jacobsen SE. 2006. An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in *Arabidopsis thaliana*. Cell. 126: 93-106.

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR 2008. Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. Cell. 133: 523-536.

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, Ecker JR. 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature. 462: 315-322.

Liu C, Lu F, Cui X, Cao X. 2010. Histone methylation in higher plants. Annu. Rev. Plant Biol. 61: 395-420.

Lorković Z, Naumann U, Matzke AJM, Matzke M. 2012. Involvement of a GHKL ATPase in RNA-directed DNA methylation in *Arabidopsis thaliana*. Current Biology. 22:933-938.

Lu F, Cui X, Zhang S, Jenuwein T, Cao X. 2011. *Arabidopsis* REF6 is a histone H3 lysine 27 demethylase. Nature Genetics. 43: 715-719.

Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature. 389: 251-260.

Malagnac F, Bartee L, Bender J. 2002. An *Arabidopsis* SET domain protein required for maintenance but not establishment of DNA methylation. EMBO J. 21: 6842-52.

Marshall E, Costa LM, Gutierrez-Marcos J 2011. Cysteine-rich peptides (CRPs) mediate diverse aspects of cell-to-cell communication in plant reproduction and

development. J Ex Bot. 62: 1677-1686.

Mathieu O, Probst AV, Paszkowski J. 2005. Distinct regulation of histone H3 methylation at lysines 27 and 9 by CpG methylation in *Arabidopsis*. EMBO J. 24: 2783-91.

Matzke AJM, Matzke MA1986. A set of novel Ti plasmid-derived vectors for the production of transgenic plants. Plant Mol. Biol. 7:357-365.

Matzke MA, Birchler JA. 2005. RNAi-mediated pathways in the nucleus. Nature Review Genetics 6: 24-35.

Meindl T, Boller T, Felix G. 2000. The bacterial elicitor flagellin activates its receptor in tomato cells according to the address-message concept. Plant Cell. 12:1783-1794.

Mette MF, van der Winden J, Matzke MA, Matzke AJM. 1999. Production of aberrant promoter transcripts contributes to methylation and silencing of unlinked homologous promoters in *trans*. EMBO J. 18: 241.

Mette MF, Aufsatz W, van der Winden J, Matzke MA, Matzke AJ. 2000. Transcriptional silencing and promoter methylation triggered by double-stranded RNA. EMBO J. 19: 5194-5201.

Miura A, Yonebayashi S, Watanabe K, Toyama T, Shimada H, Kakutani T. 2001. Mobilization of transposons by a mutation abolishing full DNA methylation in *Arabidopsis*. Nature. 411: 212-214.

Miura A, Nakamura M, Inagaki S, Kobayashi A, Saze H, Kakutani T. 2009. An *Arabidopsis* jmjC domain protein protects transcribed genes from DNA methylation at CHG sites. EMBO J. 28:1078-86.

Mourrain P, Béclin C, Elmayan T, Feuerbach F, Godon C, Morel J-B, Nicic S, Picault N, Rémoué K, Sanial M, Vo T-A, Vaucheret H. 2000. *Arabidopsis SGS2* and *SGS3* genes are required for posttranscriptional gene silencing and natural virus resistance. Cell 101: 533-542.

Naumann U, Daxinger L, Kanno T, Eun C, Long Q, Lorkovic ZJ, Matzke M, Matzke AJ. 2011. Genetic evidence that DNA methyltransferase DRM2 has a direct catalytic role in RNA-directed DNA methylation in *Arabidopsis thaliana*. Genetics 187:977-979.

Jullien PE, Mosquna A, Ingouff M, Sakata T, Ohad N, Berger F. 2008. Retinoblastoma and its binding partner MSI1 control imprinting in *Arabidopsis*. PLoS Biology 6: e194

Noh B, Lee SH, Kim HJ, Yi G, Shin EA, Lee M, Jung KJ, Doyle MR, Amasino RM, Noh YS. 2004. Divergent roles of a pair of homologous jumonji/zinc-finger-class transcription factor proteins in the regulation of *Arabidopsis* flowering time. Plant Cell 16: 2601-2613.

Okuda S, Tsutsui H, Shiina K, Sprunck S, Takeuchi H, Yui R, Kasahara RD, Hamamura Y, Mizukami A, Susaki D, Kawano N, Sakakibara T, Namiki S, Itoh K, Otsuka K, Matsuzaki M, Nozaki H, Kuroiwa T, Nakano A, Kanaoka MM, Dresselhaus T, Sasaki N, Hihashiyama T. 2009. Defensin-like polypeptide LUREs are pollen tube attractants secreted from synergid cells. Nature. 458: 357-361.

Onodera Y, Haag JR, Ream T, Nunes PC, Pontes O, Pikaard CS. 2005. Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. Cell 120:613-622.

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*. Current Biology 12:1484-1495.

Penterman J, Zilberman D, Huh JH, Ballinger T, Henikoff S, Fischer RL. 2007. DNA demethylation in the *Arabidopsis* genome. Proc. Natl. Acad. Sci. USA 104: 6752-57

Peragine A, Yoshikawa M, Wu G, Albrecht HL, Poethig RS. 2004. SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of *trans*-acting siRNAs in *Arabidopsis*. Genes & Development 18: 2368-2379.

Pélissier T, Wassenegger M. 2000. A DNA target of 30 bp is sufficient for RNA-directed DNA methylation. RNA 6: 55-65.

Pontier D, Yahubyan G, Vega D, Bulski A, Saez-Vasquez J, Hakimi MA, Lerbs-Mache S, Colot V, 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: 2030-2040.

Pontier D, Picart C, Roudier F, Garcia D, Lahmy S, Azevedo J, Alart E, Laudié M, Karlowski WM, Cooke R, Colot V, Voinnet O, Lagrange T. 2012. NERD, a Plant-Specific GW Protein, Defines an Additional RNAi-Dependent Chromatin-Based Pathway in *Arabidopsis*. Mol. Cell. 48: 121-132.

Qi Y, He X, Wang XJ, Kohany O, Jurka J, Hannon GJ. 2006. Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. Nature 443: 1008-1012.

Qian W, Miki D, Zhang H, Liu Y, Zhang X, Tang K, Kan Y, La H, Li X, Li S, Zhu X, Shi X, Zhang K, Pontes O, Chen X, Liu R, Gong Z, Zhu JK.2012. A histone acetyltransferase regulates active DNA demethylation in *Arabidopsis*. Science. 336: 1445-1448.

Raissig MT, Baroux C, Grossniklaus U. 2011. Regulation and flexibility of genomic imprinting during seed development. Plant Cell 23: 16-26.

Riggs AD, Martienssen RA, Russo VEA. 1996. Introduction. In Epigenetic mechanisms of gene regulation (ed.V.E.A. Russo *et al*.), p. 1. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Saleh A, Alvarez-Venegas R, Yilmaz M, Le O, Hou G, Sadder M, AlAbdallat A, Xia Y, Lu G, Ladunga I, Avramova Z. 2008. The highly similar *Arabidopsis* homologs of trithorax ATX1 and ATX2 encode proteins with divergent biochemical functions. Plant Cell 20: 568-579.

Saze H, Mittelsten Scheid O, Paszkowski J. 2003. Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis Nature Genetics 34: 65-69.

Saze H, Shiraishi A, Miura A, Kakutani T. 2008. Control of genic DNA methylation by a jmjC domaincontaining protein in *Arabidopsis thaliana*. Science 319: 462-65.

Saze H, Kakutani T. 2011. Differentiation of epigenetic modifications between transposons and genes. Curr. Opin. Plant Biol.14: 81-87.

Schoft VK, Chumak N, Choi Y, Hannon M, Garcia-Aguilar M, Machlicova A, Slusarz L, Mosiolek M, Park JS, Park GT, Fischer RL, Tamaru H. 2011. Function of the *DEMETER* DNA glycosylase in the *Arabidopsis thaliana* male gametophyte. Proc. Natl. Acad. Sci. 108: 8024-8047.

Schubert D, Primavesi L, Bishopp A, Roberts G, Doonan J, Jenuwein T, GoodrichJ. 200 6. Silencing by plant Polycomb-group genes requires dispersed trimethylation of histone H3 at lysine 27. EMBO J. 25:4638-49.

Silverstein KA, Moskal WA, Wu HC, Underwood BA, Graham MA, Town CD, VandenBosch KA. 2007. Small cysteine-rich peptides resembling antimicrobial peptides have been under-predicted in plants. Plant J. 51:262-280.

Simmer F, Buscaino A, Kos-Braun IC, Kagansky A, Boukaba A, Urano T, Kerr AR, Allshire RC. 2010. Hairpin RNA induces secondary small interfering RNA synthesis and silencing in *trans* in fission yeast. EMBO Reports 11: 112-118.

Slotkin RK, Vaughn M, Borges F, Tanurdzic M, Becker JD, Feijo JA, Martienssen RA. 2009. Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell. 136: 461-472.

Smith L, Pontes O, Searle I, Yelina N, Yousafzai FK, Herr AJ, Pikaard CS, Baulcombe DC. 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.

Soppe WJ, Jasencakova Z, Houben A, Kakutani T, Meister A, Huang MS, Jacobsen SE, Schubert I, Fransz PF. 2002. DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in *Arabidopsis*. EMBO J. 21: 6549-6559.

Sridhar VV, Kapoor A, Zhang K, Zhu J, Zhou T, Hasegawa PM, Bressan RA, Zhu JK.

2007. Control of DNA methylation and heterochromatic silencing by histone H2B deubiquitination. Nature 447: 735-738.

Takayama S, Shimosato H, Shiba H, Funato M, Che FS, Watanabe M, Iwano M, Isogai A. 2001. Direct ligand-receptor complex interaction controls *Brassica* self-incompatibility. Nature 413: 534-538.

Tariq M, Saze H, Probst AV, Lichota J, Habu Y, Paszkowski J. 2003. Erasure of CpG methylation in *Arabidopsis* alters patterns of histone H3 methylation in heterochromatin. Proc. Natl. Acad. Sci. USA 100: 8823-27.

Teixeira FK, Colot V. 2009. Gene body DNA methylation in plants: a means to an end or an end to a means? EMBO J. 28: 997-998.

Vaknin K, Goren A, Ast G. 2009. TEs or not TEs? That is the evolutionary question. J Biol. 8: 83.

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. Genes & Development 18: 1187-1197.

Vaucheret H. 2005. RNA polymerase IV and transcriptional silencing. Nature Genetics. 37: 659-60

Vazquez F, Vaucheret H, Rajagopalan R, Lepers C, Gasciolli V, Mallory AC, Hilbert JL, Bartel DP, Crete P. 2004. Endogenous trans-acting siRNAs regulate the accumulationof *Arabidopsis* mRNAs. Molecular Cell 16: 69-79.

Waddington CH. 1942. Endeavour 1, 18-20.

Wassenegger M, Heimes S, Riedel L, Sänger HL. 1994. RNA-directed de novo methylation of genomic sequences in plants. Cell. 76: 567-576.

Wassenegger M. 2000. RNA-directed DNA methylation. Plant Mol. Biol.43: 203-220.

Wheeler JI, Irving HR. 2010. Evolutionary advantages of secreted peptide signaling molecules in plants. Functional Plant Biology 37: 382-394.

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: 635-648.

Wierzbicki AT, Ream TS, Haag JR, Pikaard CS. 2009. RNA polymerase V transcription guides ARGONAUTE4 to chromatin. Nature Genetics 41: 630-634.

Wu L., Mao L., Qi Y. 2012. Roles of DICER-LIKE and ARGONAUTE proteins in TAS-derived siRNAs triggered DNA methylation. Plant Physiology 160. 2: 990-99

Wuest SE, Vijverberg K, Schmidt A, Weiss M, Gheyselinck J, Lohr M, Wellmer F, Rahnenführer J, von Mering C, Grossniklaus U. 2010. *Arabidopsis* female gametophyte gene expression map reveals similarities between plant and animal gametes. Current Biology 20: 506-512.

Xie Z, Kasschau KD, Carrington JC. 2003. Negative feedback regulation of *Dicer-like1* (*DCL1*) in *Arabidopsis* by microRNA-guided mRNA degradation. Current Biology 13: 784-789.

Xie Z, Johansen LK, Gustafson AM, Kasschau KD, Lellis AD, Zilberman D, Jacobsen SE, Carrington JC. 2004. Genetic and functional diversification of small RNA pathways in plants. PLoS Biology 2: E104.

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. USA 102: 12984-12989.

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

You W, Tyczewska A, Spencer M, Daxinger L, Schmid M, Grossniklaus U, Simon SA, Meyers BC, Matzke AJ, Matzke M. 2012. Atypical DNA methylation of genes encoding cysteine-rich peptides in *Arabidopsis thaliana*. BMC Plant Biol. 12: 51

Yu B, Yang Z, Li J, Minakhina S, Yang M, Padgett RW, Steward R, Chen X. 2005. Methylation as a crucial step in plant microRNA biogenesis. Science. 307: 932-935.

Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE, Ecker JR. 2006. Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. Cell 126: 1189-1201.

Zheng B, Wang Z, Li S, Yu B, Liu J, 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: 2850-2860.

Zilberman D, Cao X, Jacobsen SE. 2003. ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. Science 299: 716719.

Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S. 2007. Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an inter-dependence between methylation and transcription. Nature Genetics. 39: 61-69.

List of abbreviations

bp	base pair
ChIP	Chromatin Immunoprecipitation
CRP	Cysteine Rich Peptides
dsRNAs	double strand RNAs
HP	hairpin structure
miRNAs	microRNAs
nt	nucleotide
RdDM	RNA-directed DNA Methylation
siRNAs	small interfering RNAs
ssRNAs	single strand RNAs
MSAP	Methylation-Sensitive Amplification Polymorphism
'SD'	construct indicate potential siRNAs-DNA interaction
'SR'	construct indicate potential siRNAs-RNA interaction
T line	transgenic plant carrying original Target locus
T+S line	transgenic plant carrying both original Target and Silencer loci
TT line	transgenic plant carrying original target, silencer and 'SD' or 'SR'

Curriculum Vitae

Personal information

Name: Wanhui You
Date of Birth: 1982-01-02
Place of Birth: Dan Dong, Liaoning Province, China
Nationality: Chinese
E-mail: youwanhui@gmail.com

Education and Research Experience

Since 07/2009	Ph.D study, Supervisor: Dr. Marjori Matzke Gregor Mendel Institute of Molecular Plant Biology, Vienna, Austria
09/2006-06/2009	Post-graduate study (Master's degree in 06/2009), Institute of Genetics and Cytology, Northeast Normal University, Changchun, Jilin Province, China
03/2008-06/2008	Visiting student, Dr. Daniel Schubert's lab Institute of Genetics, Heinrich-Heine-University, Düsseldorf, Germany
09/2004-08/2006	Research assistant Institute of Genetics and Cytology, Northeast Normal University Changchun, Jilin Province, China
09/2000-06/2004	Undergraduate study (Bachelor's degree in 06/2004) School of Life Science, Northeast Normal University Changchun, Jilin Province, China

Conferences

The EMBO meeting 2011, Vienna, Austria, Volunteer

International conference on Arabidopsis, 2012 Vienna, Austria

Poster presentation
Title: Atypical DNA methylation of genes encoding cysteine-rich peptides in *Arabidopsis thaliana*

Cold Spring Harbor Asia conference: Plant Epigenetics, Stress and Evolution, Suzhou, China
Poster Presentation
Title: Pol IV/RDR2-dependent secdondary siRNAs can act in *trans* to induce DNA methylation at target regions transcribed by Pol II

Publications

Atypical DNA methylation of genes encoding cysteine-rich peptides in *Arabidopsis thaliana*. **You W**, Tyczewska A, Spencer M, Daxinger L, Schmid M, Grossniklaus U, Simon SA, Meyers BC, Matzke AJ, Matzke M. BMC Plant Biol. 2012. 12(1): 51

RNA helicase A acts as a bridging factor linking nuclear beta-actin with RNA polymerase II. Tang W, **You W**, Shi F, Qi T, Wang L, Djouder Z, Liu W, Zeng X.Biochem J. 2009 420(3): 421-8