CHARACTERIZATION OF AtSUVR3 FUNCTIONS IN *Arabidopsis thaliana* USING RNA INTERFERENCE

A Dissertation

by

TAO WANG

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2007

Major Subject: Biology

CHARACTERIZATION OF AtSUVR3 FUNCTIONS IN *Arabidopsis thaliana* USING RNA INTERFERENCE

A Dissertation

by

TAO WANG

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Timothy C. Hall
Committee Members,	Jeffrey Chen
	Thomas McKnight
	Alan E. Pepper
Head of Department,	Vincent M. Cassone

August 2007

Major Subject: Biology

ABSTRACT

Characterization of AtSUVR3 Functions in *Arabidopsis thaliana*

Using RNA Interference.

(August 2007)

Tao Wang, B.S., University of Science and Technology of China

Chair of Advisory Committee: Dr. Timothy C. Hall

Variability of transgene expression levels resulting from gene silencing is considered as a hindrance to the successful application of plant genetic engineering. Towards alleviating gene silencing, I decided to screen for novel genes involved in transgene silencing and to investigate how these genes regulate plant development. Genes encoding putative chromatin remodeling factors, especially those including a SET domain, were selected as candidate targets. A bioinformatic analysis of the *Arabidopsis* SET genes (*AtSET*) was performed and these genes were classified into 6 groups based on the domain architecture.

RNA interference (RNAi) vectors were constructed for ~ 20 *AtSET* genes and were introduced into both wild type lines and transgenic lines silenced for a *GFP* reporter gene. Surprisingly, altered developmental phenotypes were only observed for three constructs, raising questions as to the effectiveness of the RNAi approach for the chosen *Arabidopsis* system. To assess this situation, I targeted a phytoene desaturase (*PDS*) gene using the same RNAi approach. Inactivation of *PDS* renders plant a readily identifiable phenotype. Whereas the RNAi penetrance in *Arabidopsis* can be very high, the expressivity of RNAi in various tissues and among different plants can vary dramatically. Contradictory to previous reports,

I found that there is correlation between transcript level and silencing phenotype. Possible reasons for this discrepancy are discussed. No apparent correlation between transgene copy number and RNAi phenotypes was observed.

Among the three RNAi constructs that caused an abnormal development in *Arabidopsis*, K-23 which targets *SuvR3* has the highest expressivity and could reactivate a silenced *GFP* locus. *SuvR3* RNAi lines were selfed for six generations and were screened for morphological phenotypes. Abnormal number of flower organs, loss of viability of male gametophytes, and decreased seedling germination percentage were found in *SuvR3* RNAi lines. A progressive increase in both severity and frequency of abnormal phenotypes were seen in subsequent generations, suggesting an epigenetic regulatory mechanism involved with *SuvR3*. Alternative splicing of *SuvR3* was also observed in most of *Arabidopsis* tissues. One of the protein isoforms, SuvR3α, lacks 16 amino acids within the highly conserved SET domain. Possible effects of isoform interaction are proposed.

DEDICATION

This dissertation is dedicated to my grandparents and parents for their love and support.

ACKNOWLEDGMENTS

I would like to express my gratitude and appreciation to my supervisor, Timothy C. Hall, for his constant encouragement, patient guidance and support during the development of this dissertation. I am indebted to him for his valuable advice, patience and insightful comments. In addition to his academic support, Dr. Hall has provided me a friendly and welcoming home where we enjoyed dinners and games together.

I would also like to thank all my committee members, Dr. Jeffrey Chen, Dr. Alan Pepper, Dr. Thomas McKnight, and Dr. Wayne Versaw, for all of their advice, ideas, and critical comments on this dissertation. My gratitude also goes to Mark Zoran, Debby Siegele, Michael Benedik, and Kay Goldman from the Graduate Advisory Office for their wonderful help in answering my questions concerning my study.

I also thank Laks Iyer for sharing his invaluable insight on bioinformatic analysis of *Arabidopsis* SET proteins, Rodolfo Aramayo for providing *Neurospora* stains and chemicals for our *SuvR3 - DIM5* complementation experiments, and Peter Waterhouse for offering the pHannibal plasmid. I am especially grateful to Sunee Kertbundit with whom I have had so many pleasant working experiences.

This dissertation would not be finished without all my labmates and student helpers, Wang Kit Ng, Xiangyu Shi, Yiming Jiang, Xin Zhou, Jinjiang Dong, Guojun Yang, Mahesh Chandershakeran, Rakesh Pancholy, Joshua Wollam, Dan Grunspan, and Dmitry Verkhoturov. I am indebted for their valuable discussions, ideas, help, and comments.

I am also thankful for all the support I got from personnel in IDMB, Caitlin

Donovan, Charlie Harris and Satira O'Connell-Miller, Leslie Ungerer and Larry Harris-Haller and Ginger Stuessy for her help in plant care and providing me a humidifier that enabled the pollen germination experiment.

Last but not least, I would like to thank the Introductory Biology Program for providing me a Teaching Assistantship when financial support was needed.

This work was supported by National Science Foundation grant MCB-014704.

TABLE OF CONTENTS

CHAPTER Page

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

A fundamental and remarkable fact about complex organisms is that, while the nucleus of each somatic cell of a given organism contains an identical complement of genetic information, developmental regulatory events yield tissues or arrays of cells that have differentiated forms and functions. Although great insight to the mechanisms involved has come from molecular genetics, it is now evident that epigenetic events, which are largely histone-based, are essential contributors (Strahl and Allis, 2000).

In general, actively expressing genes are typically found in euchromatic chromosomal regions which are characterized by relatively loose interactions between histones and DNA. This open chromatin architecture can change rapidly to a closed, inactive heterochromatic structure in response to specific modifications of histone residues, especially methylation and acetylation. Chemical modifications of DNA also profoundly affect gene expression and hence development. A crucial insight to the connection between transcription factor-mediated regulation and chromatin-mediated regulation was the discovery that methylation of certain histone lysine residues, e.g. lysine 9 of histone H3(H3K9), can signal downstream processes leading to methylation of cytosine residues in DNA (Tamaru and Selker, 2001; Jackson et al., 2002). This modification frequently results in gene inactivation or gene silencing.

This dissertation follows the style of Planta.

Transcriptional and post-transcriptional gene silencing

Genetic engineering techniques have opened a wide range of opportunities to study various fundamental aspects of plant biology. However, crop improvement through transformation encountered a substantial obstacle in the unpredicted phenomenon of gene silencing, when genes that are expected to be active are inactivated. Such features are not acceptable for commercialization of a genetically engineered crop. Transcription factor interactions, epigenetic events, chromatin structure and many other phenomena determine when, where, and how a specific gene is to be expressed in the plant. Consequently, studying the effects of associations between modified histones, chromatin architecture and gene function is likely to provide novel insight to development processes, including gene silencing.

Gene silencing can occur both transcriptionally (TGS) or post-transcriptionally (PTGS). Both transgenes and endogenous genes are subjected to these two regulatory mechanisms. TGS is usually associated with heterochromatinization in which alteration of chromatin conformation renders targeted genes inaccessible to transcriptional machineries. In contrast, transcripts can be produced in PTGS but they are degraded by RNA-dependent silencing complexes. DNA methylation is usually associated with promoter regions and coding regions in TGS and PTGS respectively (Jones et al., 1999; Morel et al., 2000). Despite these differences between TGS and PTGS, these two silencing pathways are mechanistically interlinked. For example, double-stranded RNA (dsRNA), an intermediate product in PTGS processes, can be a trigger of TGS (Morel et al., 2000). A single transgene locus, *271*, can trigger both TGS and PTGS by simultaneously producing dsRNA

corresponding to both promoter and transcribed sequences (Mourrain et al., 2007). Additionally, mutation of *Argonaute,* a gene that is involved in PTGS, can profoundly affect heterochromatin formation (Martienssen et al., 2005; Kim et al., 2006). In summary, it is evident that both PTGS and TGS processes are intimately involved in the regulation of gene expression (Sijen et al., 2001).

Histone code hypothesis

The fundamental repeating unit of eukaryotic chromatin is the nucleosome, in which a core composed of an octamer of histones (two copies each of H2A, H2B, H3, and H4) around which are wrapped 1.75 turns (~146 bp) of double stranded DNA in a left-handed superhelix. A fifth histone, H1, is typically associated with ~ 50 bp double stranded DNA to form a linker between two adjacent nucleosomes (Luger et al., 1997). Nucleosomes not only provide the first level of compaction of genomic DNA within the nucleus, they are also carriers of two types of epigenetic information: histone modifications and DNA methylation. Numerous experiments have demonstrated that both nucleosome structure and positioning (Li et al., 1998; Lia et al., 2006; Ng et al., 2006) play important roles in regulating gene expression.

The N-terminal tails (16 - 44 amino acid residues) of histones are subjected to multiple types of post translational modification. These include acetylation, methylation, phosphorylation, ubiquitination, glycosylation, sumoylation and ADP-ribosylation (Jenuwein and Allis, 2001; Fuchs et al., 2006). From the electrostatic point of view, it was

generally conceived that histone acetylation could neutralize the positive charge of histones to decondense the chromatin structure, whereas histone methylation would have the opposite effect. Consequently, histone methylation and deacetylation are associated with gene repression and expression, respectively. Exceptions to this situation, such as H4K12 acetylation in transcriptionally silent regions in yeast (Braunstein et al., 1996) and the discovery of a spectrum of diverse histone modifications, has led to a more comprehensive model termed the Histone Code (Strahl and Allis, 2000; Jenuwein and Allis, 2001). According to this model, various chromatin modifications, primarily on the histone N-terminal domains and usually determined by upstream events can interact with each other to compose a coded histone language. The histone modifications often occur in sequential orders and the resulting codes can be deciphered by other proteins or protein domains to determine specific downstream events. This model is now well accepted and is supported by multiple lines of evidence.

Among various histone modifications, acetylation is the most extensively characterized. Two antagonistic enzymes, histone acetyl transferase (HAT) and histone deacetylase (HDAC), reversibly and dynamically control the status of acetylation in histones (Tian and Chen, 2001). In contrast, methylation on the histones is more stable and is suggested to be involved in epigenetic cellular memory (Volkel and Angrand, 2006).

Histone methylation

It has been known for over forty years that nitrogen atoms of lysines and arginines on the side chains of histones can be methylated (Allfrey et al., 1964; Murray, 1964). The arginine residue in the histone can be mono- or di- methylated by protein arginine methyltransferase (PRMT) in a symmetric (both side chain amino groups are methylated) or asymmetric (one side chain amino group) manner. Histone arginine methylation can be associated with either gene activation (Strahl et al., 2001; Bauer et al., 2002) or gene repression (Pal et al., 2004). However, such modifications in *Arabidopsis* have not been identified.

Compared to arginine methylation, histone lysine methylation is more systematically studied in both animals and plants and is now thought to play a central role in epigenetic regulation of gene expression. Lysine residues in the histone tails, including residues 4, 9, 27, and 36 in histone 3 (H3), and residue 20 in histone 4 (H4) can be mono-, di-, or tri-methylated [for review, see Lachner and Jenuwein (2002)]. Another lysine residue in the globular domain, K79 of H3, can be also methylated (Ng et al., 2003). Generally, methylation at K4, K36, and K79 is associated with gene activation, whilst methylation at H3K9, H3K79, and H4K20 is related to gene repression (Cheng et al., 2005). It is noteworthy that the methylation status at H3K9 shows interesting, but distinct, correlation with chromatin states in various species. In animals and fungi, euchromatin regions are usually marked with H3K9 mono- or di- methylation and heterochromatin regions with tri-methylation. In contrast, heterochromatin regions in plants show H3K9 dimethylation. Little,

if any, H3K9 trimethylation is present in *Arabidopsis* chromatin (Jackson et al., 2004).

SET domain proteins and histone methylation

When studying the *Drosophila* polycomb-group (PcG) gene *Enhancer of zeste* [E(z)], Jones and Gelbart (1993) found that E(z) contains a C-terminus (C-ter) region (~ 130 aa) with high sequence similarity to two previously identified Trithorax-group (TrxG) proteins: trithorax (Trx, *Drosophila*) (Mazo et al., 1990) and acute lymphoblastic lymphocytic 1 (ALL-1/Hrx, Human) (Gu et al., 1992; Tkachuk et al., 1992). PcG proteins are generally transcriptional activators whereas TrxG proteins repress transcription. The presence of this conserved region in two proteins with antagonistic functions led the authors to surmise that this region may comprise a domain that interacts with common nucleic acid or protein targets. The opposite effects of these two proteins on gene transcription are conjectured to be regulated by other regions of the proteins. More recently, a suppressor of position-effect of variegation, *Su(var)3-9* (Tschiersch et al., 1994), was found to encode a protein that also contains the C-ter domain shared by E(z) and Trx. This conserved domain was then named as SET domain, after three founding *Drosophila* proteins containing this conserved region: **S**uppressor of variegation 3-9, **E**nhancer of zeste, and **T**rithorax. Subsequently, numerous proteins containing the SET domain were found from various species and a small portion of them have been functionally characterized. As of April, 2007, 1026 entries for SET proteins from various species are cataloged in Pfam sequence alignment database (this

contains duplicate entries for some SET proteins), and studies of over 40 of them have been reported.

The first functional analysis of the SET domain was done by Rea et al. (2000) in human suppressor of variegation 3-9 (SUV39H1). The protein was shown to be a H3K9 histone methyltransferase (HMT) and the evolutionarily conserved SET domain to be the catalytic motif. Similar results were subsequently obtained in fungal and plant SET proteins. The first characterized *Arabidopsis* SET domain HMT was KRYPTONITE (KYP), which specifically di-methylates H3K9. Thus far, all HMTs that modify histone tails contain the SET domain. A non-SET domain HMT, DOT1, was found to methylate H3K79, a residue within the histone globular domain. In addition to being the catalytic motif, the SET domain in some proteins can direct protein-protein interactions. For example, the PR motif within the SET domain interacts with dual specificity phosphatase (dsPTP) to modulate cell growth (Cui et al., 1998a). SET1 and SET2 contain a motif named as Single-stranded Nucleic Acid Binding Linked to SET (SSBLS) near the SET domain boundaries, indicating these SET proteins could interact with RNAi machineries such as RNAi Induced Silencing Complex (RISC) through binding to small interfering RNA (siRNA) (Krajewski et al., 2005).

Interestingly, SET domains were found in viruses and prokaryotic organisms that do not have histone proteins. This indicates SET domain proteins may originate from the common ancestor of prokaryotes and eukaryotes. Accordingly, the substrates of the ancestor SET proteins may not be histones. Supportive evidence for this hypothesis is the discovery of SET domain-containing rubisco large subunit methyltransferase (RuBisCo LSMT or

LSMT). Alternatively, the SET domain could have originated from eukaryotes but was introduced into prokaryotes through horizontal gene transfer. This theory is favored by the fact that all bacteria containing SET domain proteins are pathogenic bacteria (Aravind and Iyer, 2003). Only one viral SET domain protein (vSET) has been identified, and its origin remains unknown (Manzur et al., 2003). In solutions, vSET can form dimers to specifically di-methylate H3K27. Cumulative evidence indicates that vSET is involved in repression of host gene transcription upon virus infection (Qian et al., 2006; Yamada et al., 2006).

Protein crystallization has been accomplished for a few fungal and animal SET proteins, plant LSMT, and vSET. Conversely, structural analysis revealed that SET domain proteins contain 5 regions within the SET domain and its vicinities. They are, from the amino terminal to the carboxyl terminal, N flanking, SET-N, SET-I, SET-C, and C flanking, respectively. In some SET proteins, N and C flanking regions are conserved motifs named as Pre-SET and Post-SET, respectively. Each of these regions may contribute to a distinct aspect of protein function. The N flanking region could interplay with SET-N to stabilize the tertiary structure of the SET domain by a range of different surface interactions (Marmorstein, 2003). The SET-I region is less conserved than SET-N, and SET-C regions and may contribute to substrate specificity. SET-C forms a topologically unusual "pseudo-knot" structure that contains the catalytic site (NHS motif). The C flanking region in some proteins can form a channel against the SET domain to provide binding surface for both

cofactor and protein substrate (Xiao et al., 2003). The C flanking region posses catalytic activity (Esteve, 2005).

Histone demethylation

For a long time, histone methylation was thought to be a permanent covalent modifications to the histones. This hypothesis theory proved incorrect with the discoveries of two types of histone demethylase (HDM). The first type is lysine-specific demethylase 1 (LSD1). This type of histone demethylase is a flavin-dependent monoamine oxidase that can remove mono- and di- methyl groups from H3K4 through an oxidation reaction, in which a methylated lysine or arginine residue is converted to its non methylation status and a formaldehyde (Shi et al., 2004; Forneris et al., 2005). Interestingly, functional interaction between LSD1 and HDAC complexes has been demonstrated (Lee et al., 2005; Lee et al., 2006). The other type of HDM is jumonji domain containing histone demethylase (JmjC). This type of histone demethylase can target mono-, di-, and tri-methylated H3K9 or H3K36, and demethylation is achieved through hydoxylation (Trewick et al., 2005; Tsukada et al., 2006; Whetstine et al., 2006). In *Arabidopsis*, no histone demethylase has been experimentally identified. However, JmjC domain, the histone demethylase signature motif (Tsukada et al., 2006), has been found in 26 proteins (Pfam database), among which, EARLY FLOWERING 6 and RELATIVE OF EARLY FLOWERING 6 were proposed to have HDM activity. The presence of these histone demethylases suggests that histone

methylation is dynamically regulated and the histone code can be readily reset during developmental transitions.

Histone methylation and DNA methylation

Both histone methylation and DNA methylation are epigenetic information carriers, and they often interact with each other. Recently, two alternative interaction mechanisms have been observed. The first model places histone methylation preceding DNA methylation. The first evidence to support this model came from *Neurospora*, a filamentous fungus. Loss of *DIM*-5, a histone methyltransferase gene, can result in total loss of DNA methylation (Tamaru and Selker, 2001). *Arabidopsis* KRYPTONITE (KYP, or SuvH4) control DNA methylation (Jackson et al., 2002). KYP is a SET domain containing HMT belonging to the SuvH class. In *kyp* mutants, methylation on the CpNpG sites in the *SUPERMAN* locus and other retrotransposon loci are lost. Both KYP and a DNA methyltransferase, CHROMOMETHYLASE 3 (CMT3), form physical interactions through HETEROCHROMATIN ASSOCIATED PROTEIN 1 (HP1). Additionally, H3K27 methylation is recognized by CMT3 (Lindroth et al., 2001). Thus, a model was proposed in which methylated histones (K9 and K27) will be first recognized by HP1, which will then recruit CMT3 to methylate DNA.

In the second model, DNA methylation is the prerequisite for histone methylation. H3K9 methylation was drastically reduced at a tumor suppressor gene in cells that were

deficient in DNA methyltransferases (Bachman et al., 2003). Similarly, in *Arabidopsis met1* mutants that are defective for METHYLASE 1, H3K9 methylation was greatly reduced (Tariq et al., 2003). Furthermore, SETDB1 interacts with methyl binding domain-containing proteins, a group of proteins that bind to methylated DNA (Sarraf and Stancheva, 2004). These observations place histone methylation as a downstream event triggered by DNA methylation.

The discrepancies in various reports may reflect the diverse interaction mechanisms between different epigenetic information carriers. Collectively, the symmetric CG methylation in DNA, maintained in mitosis by MET1, will recruit MBD proteins that interact with HMTs. In turn, methylation in the histones will induce DNA methylation on CpNpG and other non-symmetric sites.

RNA interference

One of the best ways to characterize the function of an unknown gene is to debilitate this gene in the experimental organism and observe the phenotype displayed in the loss-of-function mutants. The potential function of the unknown gene can then be deduced from the mutant phenotypes. There are many different approaches for gene inactivation, such as gene replacement by homologous recombination (HR), Targeting Induced Local Lesions IN Genomes (TILLING), T-DNA tag, and RNAi.

Among these approaches, gene replacement by HR should be most effective in silencing a gene as it can completely remove the gene itself from an organism.

Unfortunately, HR has not yet been successful in plants thus far. Only one example in plants has been reported, and it is speculated that plants are inefficient in homologous recombination as they do not have double stranded break repair genes (Kempin et al., 1997). TILLING is very successful in many organisms, including plants. However, the cost of this technology is still too high ($1500/gene) and can not be used on characterization of genes on a large scale.

RNA interference (RNAi) is proving to be a powerful approach for gene characterization. In RNAi, the presence of dsRNA complementary to a gene of interest is recognized by DICER and degraded into 21-25 nt small interfering (siRNA) fragments that are incorporated into a RNAi-induced silencing complex (RISC) that guides sequence-specific degradation of the target RNA, thereby crippling or silencing the target gene.

Large-scale analyses of gene function using RNAi have been demonstrated in various organisms, including yeast (Giaever *et al.*, 2002), *C. elegans* (Kamath *et al.*, 2003), *Drosophila* (Boutros *et al.*, 2004) and mammalian cell lines (Berns *et al.*, 2004). However, such a high-throughput analysis has not been reported in plants. Although efficient RNAi vector construction approaches, especially those based on Gateway recombination technologies, are available, lack of phenotypes in RNAi mutants and dependence on *Agrobacterium*-mediated stable transformation still limit the scale of RNAi study in plants.

Compared with the other gene inactivation strategies for plant functional genomics, RNAi has several advantages. (1) It can precisely target the gene of interest without the establishment of a tagged library, which is required for a T-DNA tag approach. (2) Various

degrees of silencing are usually observed among transformant populations; this is highly desirable when characterizing gene function. (3) RNAi can be inducible chemically (Guo *et al.*, 2003) and physically (Masclaux *et al.*, 2004), making it possible to discover genes whose loss of function will lead to lethality. (4) Genes present as more than one copy can be silenced (Matthew, 2004).

There are some disadvantages for RNAi compared with T-DNA inactivation approaches. (1) RNAi usually results in "Knock-Down" instead of true "Knock-Out (KO)", thereby increases the difficulty in detection of mutant phenotypes and necessitates molecular characterization of transcript levels in individual transgenic plants. (2) RNAi is based on computational annotation of genes, which still has many errors (Galperin and Koonin, 1998; Andrade *et al.*, 1999; Iliopoulos *et al.*, 2003). (3) A minimum of one transformation is required per RNAi-targeted gene. This apparently renders genomic-scale functional analysis by RNAi relatively costly and tedious.

CHAPTER II

BIOINFORMATIC ANALYSIS OF *Arabidopsis thaliana* SET DOMAIN GENES

Introduction

During the past decade, the SET (Suppressor of variegation, Enhancer of zeste, and Trithorax) domain has been recognized to play an important role in epigenetic regulation of gene expression. This domain, usually composed of ~ 130 amino acid residues and often localized at the C-terminus (C-ter), is evolutionarily conserved (Jenuwein et al., 1998). Proteins containing a SET domain (abbreviated as SET proteins) can be found in organisms ranging from virus to all three domains of life (Bacteria, Archaea, and Eukaryota). SET proteins have protein lysine methyltransferase (LMT) activity and have the ability to transfer one or multiple methyl groups to the ϵ-Nitrogen of specific lysine residues in histones, rubisco bicarboxyl phosphorylase large subunit, and cytochrome c (Aravind and Iyer, 2003). In various SET proteins, the SET domain can either carry catalytic activity of LMT (Tamaru and Selker, 2001) or mediate protein-protein interactions (Manzur et al., 2003). SET proteins have been found to be involved in various molecular and developmental aspects of life, including cell cycle (Raynaud et al., 2006), growth control (Cui et al., 1998b), genomic imprinting (Vielle-Calzada et al., 1999), and reproduction (Makarevich et al., 2006b).

AtSET proteins have been generally classified into four classes according to their SET domain similarity hierarchy: 1) ASH1 homologs and related; 2) Enhancer of zeste *(E)z* homologs; 3) Trithorax (trx) homologs and related; and 4) Suppressor of variegation

(suv(var)) homologs and related (Baumbusch et al., 2001; Alvarez-Venegas and Avramova, 2002; Springer et al., 2002). This classification method is of value since it is based on evolutionary relatedness. However, we have observed cases where some of the proteins that were classified within the same sub-group based on their general evolutionary relationship have radically different domain composition. For example, the architecture of ASHR1, an ASH1-related polypeptide, resembles that of the ATXR2. Similarly, the architecture of ASHR3 resembles that of ATXR5 and ATXR6. Given that the function of any polypeptide can reasonably be predicted on the basis of its domain composition, polypeptides having a similar protein domain composition are likely to have related functions irrespective of their evolutionary relatedness. Based on this rationale, a reclassification of the *Arabidopsis* SET domain-containing proteins according to their domain architecture is presented in this chapter.

Materials and methods

Examination of alternative splicing in AtSET

The online Alternative Splicing in *Arabidopsis* (ASIP, Wang and Brendel 2006) and UniPro data bases were used to scan all 47 AtSET genes and proteins, respectively. A total number of 16 AtSET were found to be alternatively spliced in ASIP database and UniProt revealed that one more AtSET (At1g77300 - SDG8/EFS/ASSH2)) undergoes alternative splicing but was missed by the ASIP database. The alternative splicing mechanisms for the 16 AtSET

have been provided by the online server. For ASSH2 alternative splicing, cDNA isoforms were obtained from NCBI and were analyzed using Vector NTI to reveal its alternative splicing mechanism. Alternatively spliced transcripts were then conceptually translated using Gene Construction Kit to determine its effect on proteins sequences and whether the SET domain is altered by alternative splicing.

Analysis of presence of antisense transcript

In scanning AtSET genes in the ASIP database , the orientation of each EST sequences for each gene was examined. The presence of one or more natural antisense transcripts was recorded and compared with the published or annotated gene structure to determine the position of the antisense transcript.

Results

Classification of *AtSET* genes according to protein domain architecture

77 entries for *Arabidopsis* SET proteins were retrieved from the Pfam database version 20.0 (Finn et al., 2006). The amino acid sequences were extracted from these proteins and were used in BLASTP search against *Arabidopsis* genome to determine the AGI locus for each AtSET gene. Fragmented and exact duplicated entries were removed. Partial overlapping

entries were analyzed for possible alternative spliced isoforms. 45 AtSET genes obtained using this approach were then compared with the published list of AtSET genes in the literature (Baumbusch et al., 2001) and in the online ChromDB server (http://chromdb.org/). Two more AtSET (At2g19640 - *ASHR2*, and At5g06610 - *ATXR5*) were found in this comparison. Thus, we concluded that there are at least 47 AtSET genes in *Arabidopsis*. The domain architectures for each of these AtSET proteins were obtained from Pfam, SMART, and Conserved Domain database (CDD). The presence of domains in the proteins were visually inspected and summarized in Table 2.1.

Group 1 is composed of Enhancer of Zeste homologs. This group has three members with protein lengths ranging from 689 to 902 amino acids (aa). The SET domain is located at the C-terminal end and is usually preceded by conserved domains of unknown function (e.g., Pfam-B_14655, Pfam-B_2595, and Pfam-B_53073). In addition to the C-terminally located SET domain, a conserved SANT (SWI3, ADA2, N-CoR and TFIIIB") DNA-binding domain is located in members of this group through the SMART annotation. This domain can also be found in rice OsiEZ1 and maize MEZ1, MEZ2 and MEZ3 E(z) homologs (Springer et al., 2002). Therefore, it is highly likely that this domain is important for the SET protein function within the PcG complex although its presence was considered to be a false positive by the conserved domain database (CDD) and not shown in Figure 2.1.

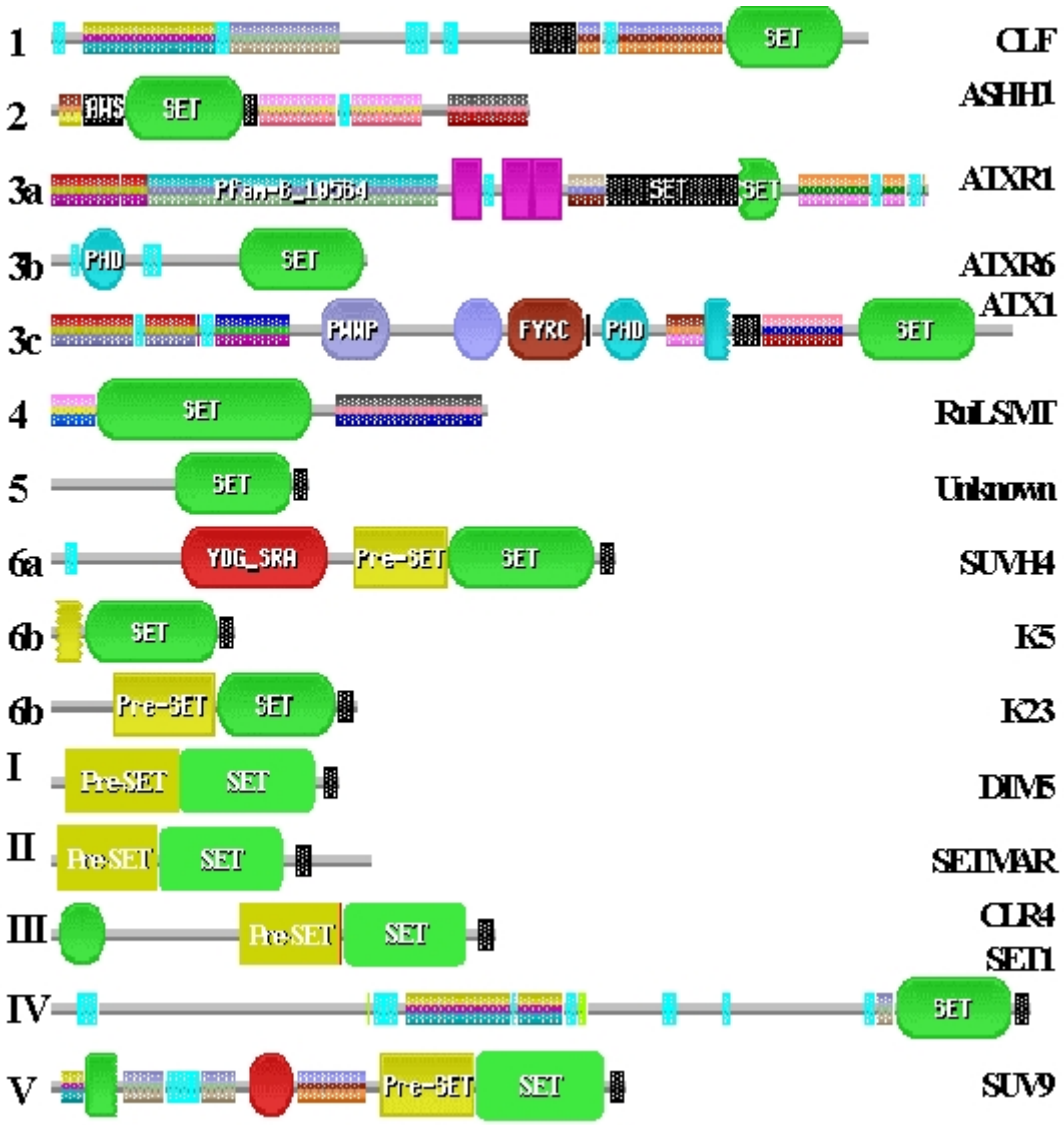

Fig. 2.1 Architecture of representative members of six classes of *Arabidopsis* SET proteins and five SET proteins from other species. Six classes of *Arabidopsis* SET proteins are: **1**, Enhancer of zeste homologs; **2**, ASH1 homologs and related, **3**, Trx homologs and related; **4**, Non histone protein methyltransferase; **5**, unknown; **6** Suppressor of variegation 3-9 homologs and related. Five representative members of SET proteins from species other than *Arabidopsis* are: **I**, *Neurospora* DIM5; **II**, human SETMAR; **III**, fission yeast CLR4; **IV**, baker's yeast SET1; **V**, fruit fly Suv9. The names for each representative member are listed at the right side of the diagram. Black boxes containing white dots denote domains predicted by SMART. Pfam-A database-predicted domains are represented by boxes of various shapes and colors. Pfam-B database-predicted domains are represented by boxes filled with two colors and white spots. Conserved domains are labeled: **AWS**, a conserved subdomain found in the Pre-SET; **FYRC**,: F/Y rich C-terminal region; **Pfam-B-10564**, automatically annotated domain 10564 that may contain low complexity regions; **PHD**: nt homeodomain that can fold into an interleaved type of Zn-finger that chelates two Zn ions; **PRE-SET**, a Zn-binding domain containing 9 conserved cysteines that coordinate three Zn ions; **PWWP**, a domain named after its Pro-Trp-Trp-Pro motif but of unknown function; **YDG_SRA**, a domain named after its conserved YDG motif but of unknown function.

Table 2.1 SET domain-containing proteins in *Arabidopsis thaliana*

Group	Name	Synonyms	aar	Domains		Locus	GI
				Pfam-A	SMART		
1	CLF	ICU1, PIF1, PIF2, SDG1, SET1	902	SET	SANT	At2g23380	30912630
1	EZA1	SWN, SDG10, SET10	856	SET	SANT	At4g02020	30913136
1	MEDEA	FIS1, MEDEA, SDG5, SET5	689	SET	SANT	At1g02580	30913012
2	ASHH1	SDG26, SET26	492	SET	AWS, Post-SET	At1g76710	75243465
2	ASHH2	EFS, SDG8, SET8	1759	ZF-CW, SET	AWS, Post-SET	At1g77300	947071110
2	ASHH3	SDG7, SET7	363	SET	AWS, Post-SET	At2g44150	947071125
2	ASHH4	SDG24, SET24	352	SET	AWS, Post-SET	At3g59960	75264575
3a	ATXR1	T24P13.14	969	TPR1, SET	SET	At1g26760	75335344
3a	ATXR3	SDG2, SET2	2351	SET	-	At4g15180	75219219
3a	ATXR7	-	1421	SET	-	At5g42400	75262514
3b	ASHR3	SDG4, SET4, SIML	497	SET	PHD, Post-SET	At4g30860	75164864
3b	ATXR5	SDG15, SET15	352	PHD, SET	-	At5g09790	75248529
3b	ATXR6	SDG34, SET34	349	PHD, SET	-	At5g24330	75262758
3c	ATX1	SDG27, SET27, TRX1	1052	PWWP, FYRN, FYRC, 2 PHD, SET	FYRC, C1, PHD, Post-SET	At2g31650	95147534
3c	ATX2	SDG30, SET30	1193	PWWP, FYRN, FYRC, 2 PHD, SET	PHD, Post-SET	At1g05830	75191912
3c	ATX3	SDG14, SET14	902	PWWP, 2PHD, SET	SET, Post-SET	At3g61740	75264636
3c	ATX4	SDG16, SET16, TX4	912	PWWP, SET	PHD, SET	At4g27910	947071310
3c	ATX5	SDG29, SET29	1043	PWWP, 2 PHD, SET	PHD, Post-SET	At5g53430	75244456
4	Unknown protein	Low similarity to RuBisCO LSMT	572	SET	-	At1g01920	42561601
4	RBCMT	RuBisCO LSMT	482	SET	-	At1g14030	17369870
4	Q9FYK3	F21J9.27 protein	476	SET	-	At1g24610	75263218
4	At2g18850		543	SET	-	At2g18850	75251247
4	RBCMT	RuBisCO SSMT	504	SET	-	At3g07670	75265363
4	At3g58080		463	SET	-	At3g58080	75251252
4	-	Hypothetical protein T5P19_220	531	SET	-	At3g56570	75264360
4	RBCMT	RuBisCO LSMT	483	SET	-	At4g20130	42566980
4	At5g14260	hypothetical protein	514	SET	-	At5g14260	75248518
4	SDG40	SET40	491	SET	-	At5g17240	75271674
5	ASHR1	SDG37, SET37	480	ZF-MYND, SET	-	At2g17900	947071144
5	ASHR2	SDG39, SET39	398	-	SET	At2g19640	947071155
5	ATXR2	SDG36, SET36	473	ZF-MYND, SET	-	At3g21820	75251251
5	ATXR4	SDG38, SET38	258	-	SET	At5g06620	75170339
6a	SUVH1	SDG32, SET32	670	YDG_SRA, Pre-SET, SET	Post-SET	At5g04940	30580528
6a	SUVH2	SDG3, SET3	651	YDG_SRA, Pre-SET, SET	-	At2g33290	30580518
6a	SUVH3	SDG19, SET19	669	YDG_SRA, Pre-SET, SET	Post-SET	At1g73100	30580525
6a	SUVH4	KYP, SDG33, SET33	624	YDG_SRA, Pre-SET, SET	Post-SET	At5g13960	30580520
6a	SUVH5	SDG9, SET9	794	YDG_SRA, Pre-SET, SET	Post-SET	At2g35160	30580519
6a	SUVH6	SDG23, SET23	790	YDG_SRA, Pre-SET, SET	Post-SET	At2g22740	30580521
6a	SUVH7	SDG17, SET17	693	YDG_SRA, Pre-SET, SET	AT_Hook, Post-SET	At1g17770	30580524
6a	SUVH8	SDG21, SET21	755	YDG_SRA, Pre-SET, SET	2 AT_Hooks, Post-SET	At2g24740	30580523
6a	SUVH9	SDG22, SET22	650	YDG_SRA, Pre-SET, SET	-	At4g13460	30580529
6a	SUVH10	SDG11, SET11	312	YDG_SRA, Pre-SET, SET	-	At2g05900	947030579
6b	SUVR1	SDG13, SET13	630	Pre-SET, SET	Post-SET	At1g04050	75249638
6b	SUVR3	SDG20, SET20	338	Pre-SET, SET	Post-SET	At3g03750	947030581
6b	SUVR4	SDG31, SET31	492	Pre-SET, SET	-	At3g04380	947030582
6b	SUVR5	SDG6, SET6	203	Pre-SET, SET	Post-SET	At2g23740	75219857

Key to Table 2.1

AT Hook = DNA binding domain with preference for A/T rich region
AWS = associated with SET domain
ZF-CW = a zinc finger with conserved cysteine and tryptophan residues
FYRC = F/Y rich C-terminus
FYRN = F/Y-rich N-terminus
PHD = Plant homeodomain
PWWP = domain named after a conserved Pro-Trp-Trp-Pro motif
SANT = SANT SWI3, ADA2, N-CoR and TFIIIB" DNA-binding domain
SRA = SET and Ring finger Associated
TPR = Tetratricopeptide repeat
YDG_SRA = SRA domain that contains a conserved YDG motif
Zf-MYND = Zinc finger MYND domain (myeloid, Nervy, and DEAF-1)

Group 2 is composed of ASH1 homologs and related proteins. This group has four members with protein lengths ranging from 352 to 1759-aa. The SET domain is usually centrally located and is invariantly preceded by an AWS SMART domain, a sub-domain of the pre-SET domain. In addition to the canonical domains present in this group of proteins, ASHH2 contains a CW (cysteine and tryptophan conserved) domain (Table 2.1) that can be found in at least five other protein families in higher plants (Perry and Zhao, 2003).

Group 3 consists of Trithorax homologs. This group has 11 members and can be further classified into three sub-groups: 3a, 3b and 3c.The 3a sub-group contains ATXR3 and ATXR7 with protein length of 2,351 and 1423-aa respectively. Both members have C-terminal located SET domains.

The 3b sub-group includes three members with protein lengths ranging from 349 to 497-aa. Raynaud et al. (2006) recently revealed that *Arabidopsis* ATXR5 and ATXR6 are involved in cell cycle regulation or DNA replication through interactions with proliferating cell nuclear antigen (PCNA). The presence of a distinct plant homeodomain (PHD), which is classical of nuclear proteins and believed to be involved in chromatin regulation (Aasland et al., 1995; Bienz, 2006), further support their role in cell cycle regulation. Interestingly, the PHD finger is missing in other ATX or ATX-related proteins that are classified into group 3a and 5. In addition, sequence alignment of SET domain proteins from maize and *Arabidopsis* revealed that certain conserved amino acid residues important for histone lysine methyltransferase activity are different in ATXR5 and ATXR6 (Springer et al., 2003). This further supported that classification based on domain architecture may supplement sequence homology based-classification and provide an alternative means of proteins classification

within the same gene family. ASHR3 is the only ASH-related protein that is classified into this group of proteins containing trithorax homologs. Although its functions remain to be determine, the presence of a PHD domain suggests that it may share functions similar to ATXR5 and ATXR6.

The complex 3c sub-group includes five members with protein lengths ranging from 902 to 1,193-aa. Similar to the 3a and 3b sub-groups, all these proteins have a C-terminal SET domain. In contrast, members of this last sub-group contain several highly conserved protein domains (e.g., Pfam PWWP, Pfam FYRN, Pfam FYRC, and Pfam PHD). With the exception of Pfam FYRN and Pfam FYRC, that are only present in two sub-members, the Pfam PWWP and Pfam PHD domains are present in all polypeptides. The Pfam PHD domain is sometimes present in more than one copy per protein.

Group 4 is likely to be composed of rubisco methyltransferase like proteins. This group has 10 members with protein lengths ranging from 463 to 572-aa. Among these SET genes, three are rubisco methyltransferase (RBCMT) and other unnamed/ hypothetical proteins were classified into this group because they contain domain architecture resembling that of RBCMT. Members in this group usually have a SET domain that is N-terminally located. It is worthy to note that the SET domain of this group of proteins is usually bigger than those found in the other groups (250 aa vs. 120-150 aa, respectively), due to the presence of long SET-I regions.

Group 5 has two and three proteins from the ASH-related and ATX-related group respectively. Proteins within this group have sizes ranging from 258 to 969-aa and they all possess a truncated or non-canonical SET domains.

Group 6 is Suppressor of variegation homologs. This group has 15 members that can be sub-divided into two sub-groups: 6a and 6b. The 6a sub-group contains nine members with protein lengths ranging from 624 to 794-aa. In all these proteins the Pfam Pre-SET, and the Pfam SET domains always follow the highly conserved Pfam YDG_SRA domain. The presence of other, less conserved Pfam-B domains (e.g., Pfam-B 18882), sometimes precedes the Pfam YDG_SRA domain. This sub-group is best exemplified by the KRYPTONITE H3-K9 methyl-transferase. The highly conserved, plant-specific protein domain distribution of these proteins suggests that this subgroup of proteins share a recent, plant-specific common ancestor. The 6b sub-group is composed of 5 members with protein ranging from 203 to 734-aa. In contrast to the 6a sub-group, proteins belonging to the 6b class all lack the Pfam YDG_SRA domain. However, as for the 6a sub-group, most members of the 6b sub-group have a variety of Pfam-B domains preceding the Pfam Pre-SET, and the Pfam SET domains. The distinct architecture of this sub-group of proteins is clearly more related to that of the founder member of the H3-K9 methyl-transferases, DIM-5. In fact, SET domain-containing proteins having similar architecture can be found in all metazoa. Amazingly, a clear example of the versatility of this domain combination can be seen in the fusion of a Pfam Pre-SET and Pfam SET domains to a transposase in the genomes of human beings and dogs. It seems very likely that this combination facilitated the insertion of these repeated elements into recombinationally silent regions of the genome.

Natural antisense transcript is present for nine *AtSET* genes

Interestingly, naturally present antisense transcript was found in ASIP database ((Wang and Brendel, 2006)) and *Arabidopsis* Cis-NAT pairs database (Wang et al., 2005b) for 9 AtSET genes (Table 2.1). Natural antisense transcripts (NAT), present in both prokaryotes and eukaryotes, are involved in epigenetic regulation of gene expression such as genomic imprinting (Moore et al., 1997) and X chromosome inactivation (Lee et al., 1999) through either siRNA-mediated RNA interference or miRNA-mediated translational inhibition. It is of vital importance to understand how the NAT can regulate the expression of AtSET genes. NAT could facilitate a sensitive, rapid and dynamic control over AtSET expression under different environmental cues. Determination of the spatial and temporal expression patterns of these NAT and their corresponding genes will provide some clues to the role of AtSET NAT in regulating plant development. Alternatively, NAT could assist the permanent imprinting of AtSET genes. Except for MEDAE , a self regulated imprinting SET gene (Kinoshita et al., 1999; Gehring et al., 2006; Jullien et al., 2006) devoid of antisense transcript, no AtSET has been reported to be imprinted. These 9 AtSET genes could be the candidate genes that are regulated by genomic imprinting.

Alternative splicing of AtSET genes

Although alternative splicing in AtSET genes has not been reported, scanning of all 47 AtSET genes (Table 2.2) against alternative splicing in *Arabidopsis* database (ASIP, (Wang

et al., 2002)) and UniProt consortium revealed that 18 (38%) AtSET genes undergo alternative splicing (Table 2.2), a percentage higher than the overall percentage (21.8%) of *Arabidopsis* genes that have transcript isoforms. A diverse alternative splicing pattern and possible consequences of alternative splicing are found in AtSET. In *ATX2*, *SuvH1*, and an uncharacterized SET gene (At5g14260), alternative splicing occurs in the 3'UTR region which could influence protein expression levels (Mendrysa et al., 2001; Fetherson et al., 2006). In contrast, in five other SET genes (*At1g01920*, *SuvR3*, *At3g55080*, *AtxR5*, and *At5g17240*), alternative splicing seems to occur within regions encoding the conserved SET domains. In the remaining 9 AtSET genes, alternative splicing is in regions encoding other parts of the SET proteins. These observed alternative splicing events further enrich the complexity of SET genes in *Arabidopsis*. Interestingly, a search in the ASIP reveals the presence of conserved alternative splicing mechanisms between *Arabidopsis* and rice for three pairs of orthologous SET genes (At5g14260 and Os02g36740, At5g17240 and Os07g28840, and, At2g19640 and Os08g10470 [AshR2]).

Table 2.2 Arabidopsis SET genes that undergo alternative splicing

AGI locus	AS pattern	Protein	Possible effects on proteins
At1g01920	AltA (2); IntronR (3);	NA	affect SET domain
At1g04050	IntronR(1)	SuvR1	affects 5'UTR and start codon selection
At1g05830	AltA (2);	ATX2	affects 3'UTR
At1g76710	AltA (2);	AshH1	affect SET C-flanking
At1g77300	AltA (2); ExonS (1);	ASHH2	no effect on SET domain
At2g19640	IntronR (1);	AshR2	no effect on SET domain, affect N-Terminus
At2g31650	AltD (2);	ATX1	3 nt difference close to the SET domain region
At2g35160	AltA (2);	SuvH5	use a different start codon to produce a shorter protein, may cause NMD
At3g03750	AltP (2);	SuvR3	affect SET domain, SET-N and SET-I
At3g04380	IntronR (1);	SuvR4	affect N-side of the PreSET domain
At3g55080	AltA (6); ExonS (1);	NA	affect SET domain by frame shift
At3g61740	IntronR (2);	ATX3	two extra introns, may affect SET-C flanking
At4g30860	IntronR (2);	ASHR3	no effect on SET domain
At5g04940	IntronR (1);	SuvH1	affect 3'UTR
At5g09790	AltA (2); ExonS (1);	ATXR5	affect SET domain
At5g14260	AltD (3); IntronR (3);	NA	affect 3'UTR
At5g17240	IntronR (1);	NA	affect SET domain
At5g43990	AltA (2);	SuvR2	use a different start codon

The possible effects on proteins are summarized according to the conceptual translation of alternatively spliced transcripts. NA denotes *Arabidopsis* SET proteins that have a similar domain architecture to RuBisCo methyltransferase but have **n**o **a**ssigned common name. Abbreviations for alternative splicing mechanisms are as in Fig 1. AltA, alternative acceptor (3' side of introns); AltD, alternative donor (5' side of introns); AltP, alternative positions (both 5' and 3' side of introns); IntronR, intron retention; ExonS, exon skipping.

Discussion

Domain architecture can be used to classify proteins

Protein domain architecture based analysis can supplement the conventional sequence alignment approaches to classify the protein relationship. Since domain identification is based on protein sequences, two approaches often gave identical classification results, as exemplified in this study for Group 1 (E[z]), Group 2 (ASHH), Group3c (ATX1), Group 6a (SUVH), and Group 6b (SUVR) SET domain proteins. However, while sequence alignment is focusing on the evolutionary origin and relatedness among various proteins, the domain based method places more emphasis on the protein functions. The intuition that proteins that share similar structure also have similar functions is the basis for the latter approach. Discrepancies between two classification approaches are found in Group 3a, Group 3b, Group4, and Group 5. In our approach, the seven ATXR members defined by Baubumsch et al. (2001) are dismantled and reshuffled with members from other groups to form 3 small groups: 3a, 3b and 5. We believe this is more likely to reflect their functions. For example, the two ATXR proteins we placed into Group 5 have only truncated SET domain and thus probably lost their HMT activity. We also predict that ASHR3 may not have HMT activity but could be involved in cell cycle control, as this protein shares identical domain structure with ATXR5 and ATXR6, two previously characterized SET proteins that do not have HMT activity and are involved in cell cycle regulation (Raynaud et al., 2006).

A disadvantage of domain based classification is the insufficient information about domains and misannotation of domains (Finn et al., 2006). The addition or deletion of a single domain may affect the classification result drastically.

Alternative splicing in AtSET genes

Alternative splicing can produce two or more forms of mature mRNA from a precursor mRNA. For various genes, alternative splicing can occur in either 5' or 3' UTRs, or in coding sequences. Overall, alternative splicing in the non-translated regions can have an impact on protein expression levels whereas alternative splicing in coding regions can alter protein structure and functions. In extreme cases, alternative splicing can even lead to production of two proteins with antagonistic functions (Mumberg et al., 1991). It is estimated that 60-80 % of human genes undergo alternative splicing (Lee and Wang, 2005) which may contribute to human genome complexity. A smaller percentage of plant genes, ~22% in *Arabidopsis* and 10% in rice (Wang and Brendel, 2006), also undergo alternative splicing, probably partially due to the fact that fewer plant EST or cDNA sequences have been identified. While a portion of these alternative splicing events are results of experimental artifacts such as sequencing errors, or spliceosomal errors in which the aberrant transcripts are subjected to nonsense-mediated decay (NMD)(Wang et al., 2002), some alternative splicing events may be biologically important.

Alternative splicing in SET genes has been documented in various species with effects on protein function ranging from no apparent consequence to completely abolishing or changing the protein activity. For example, *Drosophila Su(var)3-9* can express two distinct transcripts (2.4 kb and 2.0 kb) which encode two proteins only sharing the first 80 aa at the N-ter. While the 2.4 kb transcript is translated to a SET domain HMT, the 2.0 kb transcript encodes the gamma subunit of the eukaryotic translation initiation factor 2 (eIF2) which does not have a SET domain (Wang and Brendel, 2006). Similarly, maize SET protein Mez2 has three isoforms resulted from alternative splicing: Mez2, Mez2^{AS1}, and Mez2 ^{AS2}. Only Mez2 has the C-ter localized SET protein, the other two isoforms have no SET domain protein due to either frameshift or in-frame deletion (Springer et al., 2002). Alternative splicing in SET proteins can also produce two or more proteins with possibly duplicate functions. An example is zebrafish *SmyD1* gene. Inactivation of either isoform of this gene causes no morphological phenotype. Conversely, inactivation of both isoforms simultaneously had severe effects on myofibril organization. Other SET genes that generate spliced variants include human *G9a* (Brown et al., 2001) and *EZH1* (Abel et al., 1996), mouse *ESET* (Blackburn et al., 2003), *C. elegans SET2* (Yu et al., 2004), and Drosophila *WHSC1* (Stec et al., 1998) and *ASH2L* (Wang et al., 2001). It is interesting to know that the non-SET-domain-containing HMT, mouse *mDot1* (Zhang et al., 2004), also undergoes alternative splicing.

Although there is no experimental evidence demonstrating that the alternative splicing in AtSET results in protein isoforms and some isoform transcripts may be subjected

to NMD, there may exist different protein isoforms for some AtSET. The presence of SET protein isoforms may place an extra level of regulation over SET protein functions.

Possible dimer formation of SET domain proteins.

Virus SET proteins can form homodimers in solution through interactions between domain II tethering (Manzur et al., 2003; Qian et al., 2006). Similarly, homodimerization has also been observed for human ALL-1/MLL (a Trithorax homolog), G9a, and GLP, and *Drosophila* Ash1 and Trithorax1 through SET-SET interactions (Rozovskaia et al., 2000), although a different dimerization mechanism may be involved in these two *Drosophila* proteins (Rozovskaia et al. 2000). Heterodimer formation was also found for human G9a and GLP, and Drosophila Ash1 and Trithorax1. If plant SET proteins form dimers, alternative splicing in AtSET could generate various homo- and hetero-dimers, which may have distinct biological activity. A good candidate for SET protein dimers in *Arabidopsis* is SuvR3 protein isoforms (see Chapter V).

CHAPTER III

ASSESSMENT OF PENETRANCE AND EXPRESSIVITY OF RNAI-MEDIATED SILENCING OF THE *Arabidopsis thaliana PHYTOENE DESATURASE* GENE*

Introduction

An exciting challenge for modern biology is how to decipher the vast amount of raw information from genome sequencing so that individual genes can be identified and their biological function revealed. Among the various gain- or loss-of-function approaches available for interpreting gene function, RNA interference (RNAi) is especially powerful and well-suited for functional analysis of *Arabidopsis* and rice, plants for which physical sequencing of the genome is essentially complete.

This double-stranded RNA (dsRNA)-induced gene-silencing phenomenon, conserved among many organisms, including animals and plants, has several advantages over other approaches. In contrast to virus-induced (Baulcombe, 1999) and agroinfiltration-mediated (Schob et al., 1997) systems for transient gene silencing in plants, RNAi silencing is stable, allowing its effects to be studied in progeny (Carthew, 2001). Unlike other mutagenesis methods such as T-DNA insertion (Sallaud et al., 2003), transposon tagging (Brutnell, 2002) and TILLING (McCallum et al., 2000), RNAi silencing can be made inducible and reversible

---

* This chapter is reproduced with permission from Wang, et al. (2005). "Assessment of penetrance and expressivity of RNAi-mediated silencing of the *Arabidopsis phytoene desaturase* gene." New Phytol. **167**(3): 751-760.

(Guo et al., 2003), attributes that are especially useful in studying genes crucial to early development. A single RNAi construct can silence duplicated genes or genes sharing coding regions of sequence identity, making it possible to characterize genes with redundant copies in the genome (Waterhouse and Helliwell, 2003). Given that some 60% of known genes in the small *Arabidopsis* genome are duplicated (Blanc et al., 2000), this is an important consideration. In 2002, The AGRIKOLA project (*Arabidopsis* Genomic RNAi Knock-out Line Analysis) was initiated to study the function of 25,000 *Arabidopsis* genes using RNAi (Hilson et al., 2004).

Despite its many attributes, the value of RNAi for gene discovery and characterization is diminished where debilitation of gene function fails to produce a visible phenotype. Another caveat to its use is that the effects may vary for individual transformants. For example, Wesley et al. (2001) found that transformation of *Arabidopsis* and rice plants with constructs that generated hairpin-RNA (hpRNA) yielded a series of independent lines with various phenotypes and degrees of target mRNA reduction. Indeed, levels of the targeted mRNA have been reported to range from wild type to undetectable (Kerschen et al., 2004). We encountered a similar situation in the use of RNAi-induced silencing to determine the function of a series of SET domain-containing genes in *Arabidopsis* as an altered morphological phenotype was detected for only three of some 20 different constructs. In such cases, quantitative determination of transcript reduction in the silenced population is essential to confirm functionality of the RNAi construct. For the present studies, RT-PCR was chosen for estimation of transcript abundance as it is more sensitive than are typical genomic RNA blots and requires much less tissue, permitting large-scale analysis of many

plants. This technique has been successfully used in several studies to measure transcript abundance, usually expressed as a percentage of transcript depletion, defined in relation to controls for the specific investigation. The *PDS* gene that encodes phytoene desaturase (PDS) was chosen as the target since its silencing results in photobleached leaves (Goodwin, 1988), a readily visible phenotype. Although silencing of *PDS* has been used as a qualitative reporter of RNAi vector-based silencing in various plants, quantitative analysis of RNAi silencing in these studies is limited to the objective of establishment of efficiency of the proposed RNAi construct or reporter system under constitutive (Miki and Shimamoto, 2004) or inducible systems (Guo et al., 2003). Since *PDS* silencing has been used by several investigators, our results can be compared with their work to provide a general guide to assessment of expectations for RNAi-mediated silencing.

However, none of these studies have focussed on the issue that, using the same RNAi vector system, a uniform population of plants exhibiting equal level of silencing is rarely obtained. This situation is not likely to be unique to PDS and, therefore, has implications for all RNAi-mediated silencing studies in plants. Thus, in this study, we have used the terms "penetrance" and "expressivity" in an attempt to address the issue of variable silencing effects and their quantitative assessment in a more global connotation. These terms are commonly used in population genetics and general studies (Zlotogora, 2003). Classically, penetrance is defined as the percentage of individuals with a given genotype that exhibit the phenotype associated with that genotype, whereas expressivity measures the extent to which a given genotype is expressed in an individual at the phenotypic level (Griffiths, 1996). Thus, for *PDS* silencing, the percentage of transgenic plants displaying an identifiable

photobleached phenotype represents penetrance and the percentage depletion of endogenous *PDS* mRNA defines expressivity.

Materials and methods

Plant material

Arabidopsis thaliana (ecotype Columbia) seeds were germinated in soil (Redi Earth, Scotts) and, following vernalization at 4°C for 48 h in the dark, grown at 22°C under a 16/8 h light/dark cycle. Transformants were selected on Murashige and Skoog (MS) medium (GIBCO™ Invitrogen Corporation, Carlsbad, CA) containing 100 μg ml^{-1} Timentin (ticarcillin disodium and clavulanate potassium, SmithKline Beecham Pharmaceuticals, UK), 50 μg ml^{-1} kanamycin (Sigma, St. Louis, MO), 3% (w/v) sucrose and 0.2% (w/v) phytagel (Sigma). After two weeks, resistant plants were transferred to soil and grown in a growth chamber. Pictures of plants were taken 21 d after transfer to soil using an Olympus C-3040ZOOM digital camera (Olympus, Melville, NY).

Plasmid construction and transformation

Two primer sets with different restriction enzyme recognition site overhangs were used to amplify a 179 bp region (spanning exons 8 and 9) of *PDS*. One set, with *Xho*I and *Kpn*I overhangs, was inserted in a sense orientation into pHannibal (Wesley et al., 2001); the

other, with *Bam*HI and *Cla*I overhangs, was inserted in an antisense orientation. In the second step, a fragment containing *35S*:*PDS-s*:intron:*PDS-as*:*ocs* was released by *Not*I and inserted into the binary vector pArt27 (Gleave, 1992) to form the RNAi construct K-1. Primer sequences were:

*Bam*HI: 5'-GTCAGTGGATCCCATGGTTCCAAGATGGCATTC-3'

*Cla*I: 5'-ACGGACATCGATAGCTTCAGGATATCGACTGGAGCG-3'

*Xho*I: 5'-GTCAGTCTCGAGCATGGTTCCAAGATGGCATTC-3'

*Kpn*I: 5'-ACGGACGGTACCAGCTTCAGGATATCGACTGGAGCG-3'

Thermocycling conditions were 94°C for 5 min, followed by 35 cycles of 94°C for 1 min, 53°C for 1 min and 72°C for 2 min, with a final polymerization step at 72°C for 10 min. The K-1 construct was transformed into *Agrobacterium* (GV3101) using electroporation with a Gene Pulser (Bio-Rad, Hercules, CA). *Arabidopsis* plants were transformed using vacuum infiltration (Bechtold and Pelletier, 1998).

Genomic DNA blot analysis

Genomic DNA was extracted from four to five leaves automatically with an AutoGenprep 850alpha (Autogen, Holliston, MA). Genomic DNA (500 ng) was digested with 20 units of *Bam*HI for 17 h. After electrophoretic separation in a 0.7% agarose gel, the DNA fragments were transferred to Hybond-N+ membrane (Amersham). DNA probes were labeled using a DECAprime II kit (Ambion, Austin, TX). Hybridizations were performed using ULTRAhyb solution (Ambion) according to the manufacturer's recommendations.

Hybridization signals were detected by exposure to a PhosphorImager (Fuji, Stamford, CT) and quantitated using the public domain NIH ImageJ program (developed at the U.S. National Institutes of Health and available on the Internet at http://rsb.info.nih.gov/ij//). The densitometry ratios for the intact transgene to the endogenous *PDS* gene were calculated and the copy number of the transgene expressed as an integer relative to the plant having the lowest ratio. The copy number of rearranged transgenes was estimated by counting the aberrant transgenic bands.

Reverse transcription-PCR (RT-PCR)

Leaves exhibiting a similar PDS silencing phenotype were used for RNA extraction. If more than one leaf on a plant exhibited the phenotype, the leaves were pooled. However, leaves from different plants were never pooled. RNA was extracted using TRIzol (Invitrogen, San Diego, CA) according to the manufacturer's instructions. RNA (1 μg) was treated with DNaseI (1 u μl^{-1}, Invitrogen) and RT-PCR was carried using the QIAGEN OneStep RT-PCR Kit (Qiagen, Valencia, CA) to amplify a PDS coding region upstream of the RNAi target region. Reactions contained 250 ng RNA, 6 μM of each PDS primer and 0.08 μM of each EF1α primer in a final volume of 25 μl. Primer sequences for PDS were: 5'-GTATGAGACTGGTTTACATATTTTCT-3' and 5'-CCGCAAAATAGCCCAAATACC-3'. Primer sequences for the internal control EF1α were: 5'-TGCTGTCCTTATCATTGACTCCACCAC-3' and 5'-TTGGAGTACTTGGGGGTAGTGGCATC-3'. Thermocycling conditions were: reverse-

transcription at 50°C for 30 min, 95°C for 15 min, followed by 25 cycles of 94°C for 1 min, 55°C for 1 min and 72°C for 1 min, with a final polymerization step at 72°C for 10 min. The products of the RT-PCR amplification were subjected to electrophoresis through a 2.0% agarose gel, followed by staining with ethidium bromide (100 ng ml^{-1}). The gel was then digitally imaged and was analyzed using ImageJ.

Results

PDS is encoded by a single copy gene in *Arabidopsis*

That the *Arabidopsis* genome contains a single copy of PDS was validated by a BLAST search (Altschul et al., 1997) using Accession NM117498.2, the original full length cDNA sequence (Scolnik and Bartley, 1994), against the *Arabidopsis* genome. Although two additional cDNAs were identified, they differed from the original PDS cDNA by only one or two nucleotides. Further examination revealed that PDS3 (Accession Dl3154c) is a C to G correction at position 42 of the 5'-UTR and the other (Accession NM202816.1) is a splice variant of the original cDNA that results in a difference of two amino acid residues (GV to AI, encoded by exons 7 and 8). Nevertheless, all three cDNAs (4344 bp) originated from the same 4837 bp gene locus (At4g14210) on chromosome 4. Thus, only one copy of *PDS* is present in the *Arabidopsis* genome.

High penetrance of RNAi-mediated silencing

To inactivate the *Arabidopsis PDS* gene, an RNAi construct (K-1, Fig. 3.1) containing sense- and antisense-orientations of a fragment spanning exons 8 and 9 of the *PDS* coding region flanking the pHannibal intron, was transformed into *A. thaliana* ecotype Columbia. A series of three replicate transformations generated 485 kanamycin-resistant (Kan^r) T_1 plants. Only 5% of the Kan^r plants lacked visible phenotype. Genomic DNA blot analysis revealed that 6 of 7 randomly selected P0 plants contained at least one copy of transgene. To simplify the calculation, all P0 Kan^r were counted as *bona fide* transgenic plants. Thus the penetrance of the K-1 RNAi construct was 95%. This demonstrates the sensitivity of the PDS system for evaluating the efficacy of silencing induced by the RNAi construct.

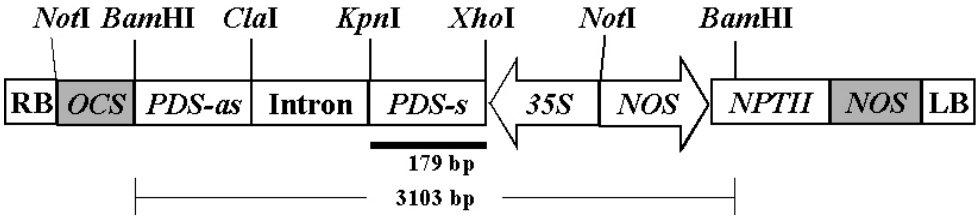

Fig. 3.1 Organization of the T-DNA region of RNAi vector K-1 used to target *PDS*. RB and LB: T-DNA right and left border; *35S*, cauliflower mosaic virus *35S* promoter; *PDS-s* and *PDS-as*: sense and antisense orientation, respectively, of the targeted *PDS* fragment. *NPT*II, neomycin phosphotransferase II. The shaded boxes marked *OCS* and *NOS* denote terminators of *Agrobacterium octopine synthase* and *nopaline synthase* genes, respectively; the arrow labeled *NOS* denotes a *nopaline synthase* promoter. The *Kpn*I–*Xho*I region (179 bp, thick bar) corresponds to the PDS coding sequence used to generate a probe for genomic DNA blot analysis. The presence of a 3103 bp *Bam*HI-*Bam*HI fragment in genomic blots was used to confirm the presence of the intact RNAi construct.

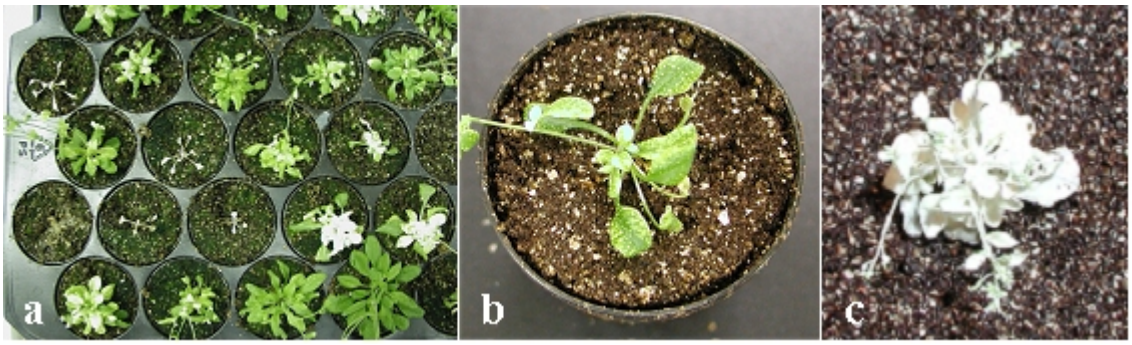

Fig. 3.2 Wide range of RNAi-induced *PDS* phenotypes in T_1 plants. (a) A representative tray containing a variety of silencing phenotypes (21 d post transfer). Representative plants exhibiting a mild (P1) silencing phenotype and a severe (P4) phenotype are shown in (b) and (c), respectively. A classification of phenotypes is given in Table 3.1.

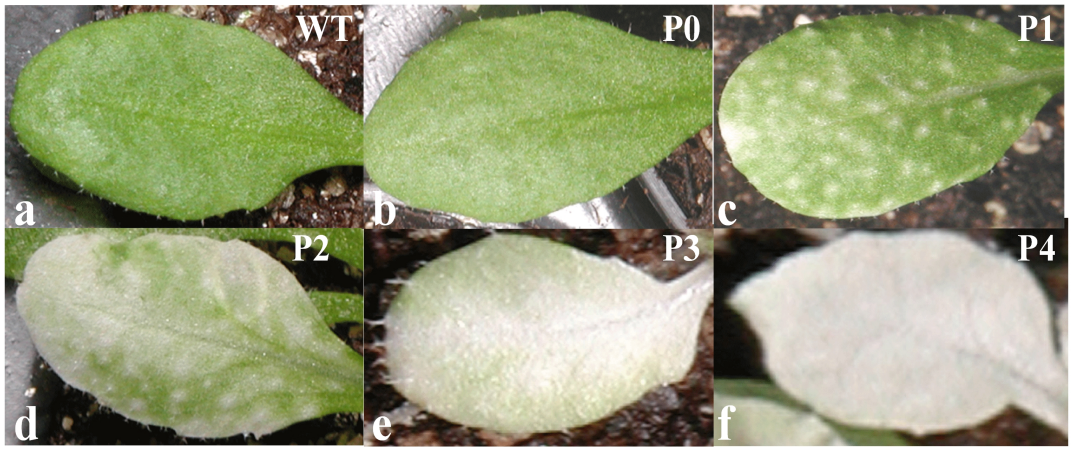

Fig. 3.3 Range of phenotypes seen in PDS-silenced leaves. (a) A wild type leaf. (b-f) Representative leaves showing P0 to P4 phenotypes, respectively. A classification of phenotypes is given in Table 3.1.

Relationship between expressivity and phenotype

The wide range of photobleached phenotypes present in T_1 progeny as a result of *PDS* silencing (Fig. 3.2) indicated that the expressivity of the K-1 construct can be dramatically

different in individual plants and often for different parts of a plant. To characterize the efficacy of RNAi-mediated *PDS* silencing, the plants were grouped into six classes (P0-P5), based on their phenotypes (Fig. 3.3 and Table 3.1, P5 plants were not shown in Fig.3.3 because they were already dead before the photos were taken).

Approximately 20% of the transformants displayed readily discernable mixed phenotypes; these were classified according to their severest phenotypes. An example is shown in Fig. 3.4a, in which six relatively old rosette leaves of a P2 plant had only very small (< 1 mm) white patches, a typical P1 phenotype, whereas the cauline leaf and three relatively young rosette leaves showed a P2 phenotype, with bigger white patches (3-5 mm). Similarly, the T_1 plant in Fig. 3.4B displayed a P2 phenotype in old rosette leaves

Fig. 3.4 Expressivity of RNAi-mediated silencing in different plant parts**.** Representative plants exhibiting different *PDS* silencing phenotypes in various tissues are shown. (a) Plant displaying a mixture of P1 and P2 phenotypes. Arrow 1 indicates a cauline leaf showing a P2 phenotype; arrows 2 and 3 indicate rosette leaves exhibiting P1 and P2 phenotypes, respectively. (b) Plant showing P3 (in older rosette leaves) and P4 (in younger rosette leaves) phenotypes. (c) A wild-type plant. Rare phenotypes (d) showing leaves variegated for P0 and P4 and (e) a plant with P0 and P2 rosette leaves and P4 cauline leaves, stems and flowers.

Table 3.1 High penetrance and various expressivities of RNAi-mediated PDS silencing in T_1 plants

Class	# of T1 plants	% of T1 plants	Phenotype description	Expre-ssivity
P0	25	5	no visible phenotype	4%
P1	74	15	mild symptoms, ~1 mm white patches on the leaves	21%
P2	158	33	3-5 mm white patches on the leaves	48%
P3	27	6	leaves predominatly white, with green patches	64%
P4	22	5	white leaves	85%
P5	179	37	white leaves, very stunted, died within 14 dpt.	nd

The penetrance of RNAi-mediated silencing was calculated as the percentage of kanamycin-resistant T_1 plants displaying an identifiable *PDS* silencing phenotype (P1-P5: see text and Fig. 3.3). The *PDS* transcript depletion level was used as a measure of expressivity for the K-1 construct (Fig. 3.1). nd: not determined.

and a P2 phenotype in young rosette and cauline leaves. Two T_1 plants (out of 485 plants) displayed rare phenotypes. In one plant (Fig. 3.4a), about 70% of rosette and cauline leaves variously displayed P0, P2 and P2 phenotypes and the other 30% were variegated, having one side (25-50%) of the leaf area completely white and the other completely green.

While it is tempting to think that these effects reflect the spread of silencing induced by RNAi, current evidence does not support this possibility. For example, Vestige et al. (2002) have shown that *Arabidopsis PDS* mRNA cannot be used as a template by the RNA-dependent RNA polymerase SDE1. The production of dsRNA that could trigger the spread of PDS silencing (Himber et al., 2003) is lacking . The plant shown in Fig. 3.4a would have

been valuable for studying the possibility that systemic gene silencing of PDS occurred. Unfortunately, it was sterile and two attempts to regenerate it through young silique culture were not successful. Similarly, the plant shown in Fig. 3.4e, with rosette leaves that displayed no or mild silencing while the rest of the shoot system was completely white, was also sterile.

Penetrance and expressivity of RNAi-mediated silencing in T_2 progeny

Compared with agroinfiltration (Schob et al., 1997) and other transient methods for expressing transgenes, *Agrobacterium*-mediated stable transformation has the advantage that transgenic progeny can be obtained. Opportunity was taken of this advantage to inspect whether RNAi-induced silencing was heritable and, if so, to determine the penetrance of K-1 in T_2 progeny. Flowers and small siliques developed and T_2 seeds were obtained from six P0, ten P1, ten P2 and five P2 plants (P2 plants were sterile and P5 plants were dead before

Table 3.2 Penetrance of RNAi-mediated silencing in T_2 progeny

T1 plant phenotype	Kanamycin resistant T2 progeny No.	P0	P1	P2	P3	P4	RNAi penetrance in T2 progeny
P0	62	62	0	0	0	0	0%
P1	72	39	33	0	0	0	46%
P2	70	30	27	9	3	1	57%
P3	53	9	14	20	6	4	83%
Overall	257	140	74	29	9	5	46%

reaching a reproductive stage). A portion of seed (20 to 30 from each plant) was germinated under kanamycin selection. T_2 progeny derived from a given class of T_1 parental plants were grouped together and screened for *PDS* silencing (Table 3.2). The penetrance of K-1 in T_2 progeny of P0 plants was 0 as none displayed detectable *PDS* silencing. T_2 progeny of P1 plants showed both P0 and P1 phenotypes, with a penetrance of ~46%. For P2 and P3 plants, *PDS* silencing decreased in most T_2 progeny but enhanced in a few (< 8%). Overall, penetrance of the K-1 construct in T_2 progeny dropped sharply to 46% from 95% in T_1 plants (Table 3.2).

Relationship between transgene copy number and severity of RNAi-mediated silencing

The number of transgene copies present can be either positively or negatively associated with the level of transgene expression. Usually, increased expression corresponds with higher transgene copy number if the copies of the transgene are intact and the copy number is below a threshold whose value is dependent on the transgene itself (Hobbs et al., 1993; Lechtenberg et al., 2003). However, transcript levels that are above this threshold or aberrant transcripts can trigger posttranscriptional gene silencing (PTGS) (Garrick et al., 1998). Therefore we considered the possibility that the level of hairpin RNA (hpRNA), the RNAi silencing trigger generated from the K-1 construct (Wesley et al. 2001) , was related to transgene copy number.

To determine the K-1 transgene copy number, genomic DNA was isolated from more than 70 randomly selected kanamycin-resistant plants, digested with *Bam*HI and subjected

to Southern (Southern, 1975) analysis. A single 7083 bp fragment corresponding to the endogenous *PDS* gene was detected in all of the transformants tested. Fig. 3.5 is a representative blot for 12 independently transformed plants, hybridized with a *PDS* probe (Fig. 3.1). One to three copies of a 3103 bp fragment, representing the intact transgene, was detected in all Kan^r plants except for one P0 plant (Fig. 3.5, lane 4), in which only a rearranged copy of the transgene (~ 5.5 kb) was found. One to four partial or rearranged transgene fragments were present in many transformants. Phenotypes P0 to P2 were represented in this sample, but P2 and P5 plants were excluded as they provided insufficient plant material.

Figure 3.5 shows the phenotype and copy number for each of the plants for which Southern analysis was conducted. From these data, it appears that a single intact transgene is not necessarily associated with a severe silencing phenotype. For example, it can be seen in Fig. 3.5 that single intact copy transgenic plants displayed P0 (lane 2), P1 (lanes 5,11), P2 (lanes 6, 7 and 9) and P2 (lane 10) phenotypes. A range of phenotypes was also evident from the data shown in Fig. 3.5 for plants containing two or more copies of an intact transgene: P0, lane 1; P1, lanes 8 and 12; P2, lane 3). Indeed, the copy number of rearranged transgenes also showed little correlation with the severity of the phenotype as plants with 3 or 4 copies of rearranged transgene displayed P0 (Fig. 3.5, lanes 1 and 2); P1 (Fig. 3.5, lane 12), P2

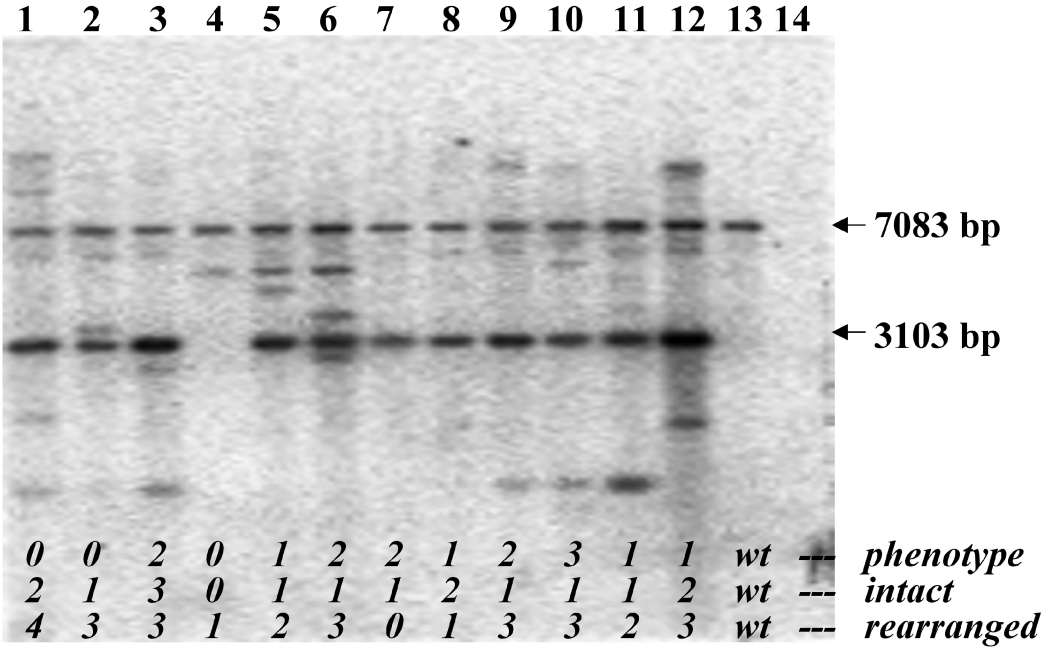

Fig. 3.5 Transgene copy number shows little correlation with phenotypic severity. DNA extracted from twelve T_1 plants (lanes 1-12) randomly selected from a total of 485 plants transgenic for the K-1 construct, and a wild-type plant (wt, lane 13), was digested with *Bam*HI and subjected to genomic DNA blot analysis. The PDS phenotypes (see Table 3.1) and number of intact and rearranged *PDS* transgene fragments are indicated at the bottom of each lane. The probe used for hybridization corresponded to the PCR amplicon (*Xho*I-*Kpn*I fragment, Fig. 3.1) employed in the construction of the K-1 vector. Intact transgene copy numbers were calculated as described in Materials and methods with the signal ratio of transgene to endogenous PDS for lane 10 set at 1. The positions of the endogenous gene (7083 bp) and the transgene (3103 bp) are indicated by arrows. Lane 14 contained a 1 kb DNA ladder (New England Biolabs).

(Fig. 3.5, lanes 3, 6, and 9) and P2 (Fig. 3.5, lane 10) phenotypes. Another blot of *Xho*I-digested DNA from a different set of 12 independent transformants showed similar results (data not shown). Taken together, the results from these plants do not provide any evidence that a correlation exists between transgene copy number and severity of RNAi-mediated silencing.

Correlation between *PDS* transcript level and *PDS* silencing phenotype

Since the photobleaching phenotype in *PDS* plants is caused by reduction of endogenous *PDS* mRNA level by the K-1 RNAi construct, we conjectured that the depletion level of *PDS* mRNA is in positive correlation with *PDS* silencing phenotypes. The level of *PDS* transcripts in *Arabidopsis* has been reported to be below the detectable limit of Northern blot analysis (von Lintig et al., 1997; Wetzel and Rodermel, 1998). Therefore, a semi-quantitative relative RT-PCR technique (Dean et al., 2002) was used. In this technique, the gene of interest is co-amplified with an internal control gene to determine the relative abundance of endogenous *PDS* transcripts in each class of *PDS* plants. EF-1α was chosen as the internal control because it produces stable transcripts and its amplification remains

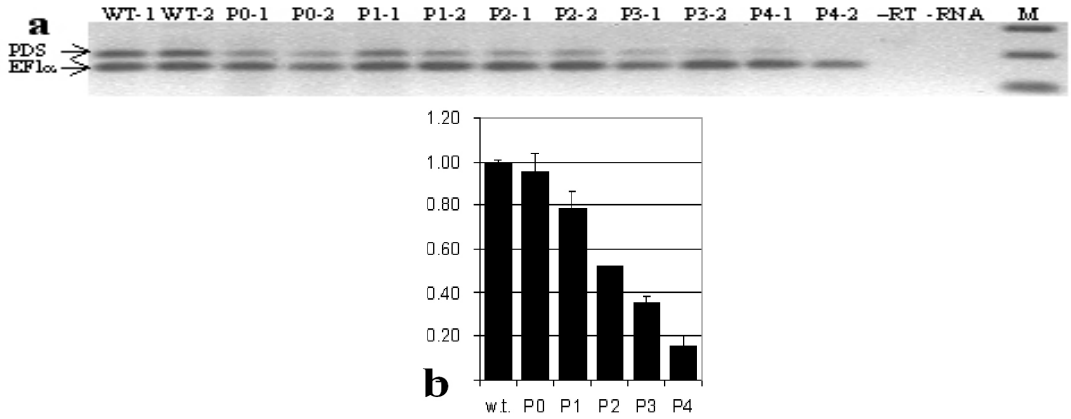

Fig. 3.6 *PDS* transcript levels diminish in correspondence with severity of photobleaching. (a) Relative RT-PCR analysis in wild type and transgenic plants displaying various degrees of bleaching (P0 to P4: see text and Table 1). Arrows denote the predicted position of *PDS* amplicon and the control (EF-1α) amplicon. (b) Normalized *PDS* transcript levels for the various phenotypes. RDI: relative densitometric intensities (pixels·mm^{-2}), normalized relative to *EF-1α*, was obtained using MacBAS v2.5 software (Fuji, Tokyo, Japan). The RDI for wild-type plants was set as 1.0. Error bars denote standard error of the mean.

in a log-linear stage at the same optimal conditions as those for *PDS*. The amplified products were subjected to DNA gel analysis followed by densitometry quantitation and the relative *PDS* : *EF-1α* expression ratio was calculated in wild-type and *PDS* plants. As expected, a close relationship between severity of phenotype and depression of *PDS* transcript level was displayed (Fig. 3.6). In plants displaying no- (P0) or mild-silencing (P1) *PDS* plants, *PDS* transcripts level averaged 96% and 79% of wild-type plants, respectively. This number had dropped to ~ 40% in medium-silencing *PDS* plants (P2 and P2) and only 15% in severe-silencing plants (P2). Thus, the depletion level of *PDS* transcript, a direct result form K-1 expression, was consistent with the silencing phenotype and was used as a measure of expressivity of K-1 construct (Table 3.1).

Discussion

The apparent penetrance of RNAi inactivation is influenced by phenotype

Using the intron-containing vector pHannibal (Wesley et al., 2001), high RNAi penetrance (95%) was observed (Table 3.1) that may be attributed to the ease of identification of the silencing phenotype, permitting the detection of even a mild degree of *PDS* silencing. For example, if P1 and P2 plants could not be visually identified, RNAi penetrance would drop to 47%, less than half of the original level. High RNAi penetrance is unlikely to be limited to *PDS* and may be routinely achieved if optimal target regions are used in the construction of RNAi vectors. Conversely, RNAi penetrance could be greatly underestimated for genes

whose loss-of-function mutants result in little or no visible phenotype. In regard to this, an important consideration is that expression of the targeted gene may be limited to specific tissues or certain developmental stages and/or certain environmental conditions. In such cases, phenotypic changes resulting from down-regulation may only be detectable in the relevant tissues and conditions. These caveats complicate gene discovery or characterization by RNAi-mediated gene silencing as biochemical, rather than phenotypic, analysis may be required. However, for genes whose function is predicted but not proven, RNAi remains a valuable discovery tool as it permits a guided analysis of the predicted function.

Endogenous gene expression level and RNAi efficacy

RNAi phenotype or RNAi efficiency may be related to the nature of the target gene. For example, in *C. elegans*, RNAi phenotypes were shown to be more scoreable for highly expressed genes than for genes expressed at low levels (Cutter et al., 2003). However, although Kerschen *et al*. (2004) found that transcript levels were effectively reduced by RNAi in *Arabidopsis* for several moderately and highly expressed genes, they found that RNAi was effective for *HDA9*, *HDT4* and *SGA1*, genes normally expressed at low levels. Similarly, despite the low endogenous level of *PDS* mRNA expression in *Arabidopsis* (Wetzel and Rodermel, 1998), very high RNAi penetrance (95%; Table 1) and expressivity (Fig. 3.3f) were observed, suggesting that RNAi efficiency and the endogenous transcription level of the targeted gene are not necessarily related.

Transgene copy number and variability of silencing

Several studies on RNAi-mediated gene silencing have shown a wide variability for individual plants (Chuang and Meyerowitz, 2000; Levin et al., 2000; Stoutjesdijk et al., 2002). Kerschen *et al.* (2004) reported little variability in target transcript reduction for single copy RNAi lines and implicated that the presence of multiple copies of the RNAi construct were a major cause of variability. Evidence supporting this view included the observation that no multi-copy line depleted the target transcript more than single-copy lines. However, their data show variability in transcript depletion for both single-copy and multi-copy lines targeting *HDA2*, *HAG5* and *CHR2*, with the greatest variability for *HAG5* being among single copy lines. For *PDS*, silencing was considerably more effective for some multi-copy transgenic plants than for single copy transgenic plants and no correlation was found between copy number and silencing severity (Fig. 3.5). Another consideration is that Kerschen *et al.* (2004) used pooled RNA from several seedlings, thus obtaining an average transcript level that would mask any plant to plant variation.

The establishment of single copy transgenic lines is usually preferred over multicopy lines because they are more readily taken to homozygosity. In general, transgene expression levels from single copy lines are more stable than from multi-copy lines. However, this probably reflects the organization, rather than the copy number, of the transgene (Lechtenberg et al., 2003). This follows from the finding that gene silencing typically arises as a posttranscriptional event incited by aberrant RNA transcribed from the rearranged insert rather than as a homology-dependent event (Mette et al., 2000; Matzke et al., 2002).

Consequently, the presence of multiple intact copies of the RNAi-generating transgene may be beneficial, as they have the potential to provide higher RNAi transcript levels than can single copy inserts.

High-throughput characterization of plant genes by RNAi

While fabrication of RNAi constructs is rarely a limiting step in high-throughput identification of gene function using RNAi, delivery and analysis can be constraining. In *C. elegans*, highly efficient delivery of small interfering RNA (siRNA) occurs by ingestion, and phenotypic analysis of function is straightforward (Kamath et al., 2003). In *Drosophila* and mammals, the establishment of an *in vitro* system and the availability of numerous cell lines have simplified the delivery of siRNA or long double-stranded RNA (dsRNA), and greatly facilitated screening of genes involved in particular pathways at the cellular level (Boutros et al., 2004; Foley and O'Farrell, 2004; Paddison et al., 2004).

In plants, agroinfiltration and virus induced gene silencing (Waterhouse and Helliwell, 2003) can provide approaches for large scale temporary and non-heritable gene silencing. However, the understanding of gene function at the organismal level requires *Agrobacterium*-mediated stable transformation. For this, the development of intron-containing hairpin RNA constructs (Smith et al., 2000) and Gateway recombination-based cloning technology (Wesley et al., 2001) have facilitated high throughput construction of RNAi vectors.

For gene discovery using RNAi, many primary transformants need to be screened. Transcriptional inactivation and an associated phenotype can be expected to range from little to complete and from no phenotype to extreme phenotype or lethality. Lines having a visible phenotype are retained for further study. When no phenotype is apparent, the functionality of the RNAi construct needs to be demonstrated by analysis of transcript depletion. The complete functional inactivation of some genes can be predicted to be lethal. Since the cause of lethality, rather than lethality *per se* is of interest, advantage can be taken of lines that express RNAi weakly. Even more valuable is the ability to use RNAi expressed from an inducible promoter (Guo et al., 2003) since this provides flexibility for the timing and degree of gene inactivation and has the potential for reversal of silencing by withdrawal of the inducer (Gupta et al., 2004). If effective depletion is not substantiated, the use of alternative target or promoter sequences for the RNAi vector is indicated. Estimation of transgene copy number will identify single copy transformants. If it is assumed that single copy lines are always more effective in transcript depletion, then multicopy lines will be discarded. However, in contrast to the studies of Kerschen *et al.* (2004), our data show that multicopy lines can have higher expressivity than some single copy lines. In such situations, transcript depletion is a more meaningful selection criterion than is copy number. Even for single copy lines, variation in expressivity can be expected and screening of these lines for those showing greatest transcript depletion is still desirable. For detailed functional investigation of the newly identified gene, well-defined stably-expressing RNAi lines need to be established. Clearly, whatever protocol is followed, functional analysis will be time-consuming.

CHAPTER IV

CHARACTERIZATION OF SUPPRESSOR OF VARIEGATION RELATED 3, AN *Arabidopsis thaliana* SET DOMAIN GENE

Introduction

In an emerging model for transgene silencing (Mutskov and Felsenfeld, 2004), the initial event is recruitment of histone deacetylase (HDAC), resulting in loss of histone acetylation and inactivation of transcription. Although plants use an HD2-type HDAC that differs from other HDAC classes and requires phosphorylation for activity (Lusser et al., 2001), it appears to play a similar role in transgene silencing. While it is likely that several different stimuli incite recruitment of HDAC to a genetic locus, various forms of RNA appear to be major factors. These include dsRNA that acts as a trigger for silencing by RNA interference (RNAi) pathways (Hamilton and Baulcombe, 1999; Hamilton et al., 2002), aberrant RNA (Mette et al., 2000) and small RNAs (Grewal and Rice, 2004).

Methylation of H3K4, H3K36 and of H3K79 is correlated with transcriptional activation, whereas methylation of H3K9, H3K27 and H4K20 are characteristic of repressive chromatin (Lachner et al., 2003; Schotta et al., 2004). The occurrence of mono-, di- or trimethylation (Tamaru and Selker, 2001) provides an opportunity for additional epigenetic signaling. In transgene silencing, loss of di- or tri-methylation at H3K4 occurs concurrently with histone deacetylation. Silencing of the transgene is subsequently reinforced and stabilized as a result of methylation of H3K9 by histone methyltransferase (HMT) and of

CpG by DNA methyltransferase (DNMT). Jackson et al. (2004), and Malagnac et al. (2002), studying the SET domain-containing gene *KRYPTONITE* (*KYP,* also known as *SuvH4*), found a major difference between plants and other organisms (animals, *Neurospora*) in that H3K9m^2, rather than H3K9m^3, marks silent loci in *Arabidopsis*. Indeed, little if any H3K9m^3 was found in bulk chromatin of *Arabidopsis* (Jackson et al., 2004).

Whereas developmental decisions regarding gene expression and differentiation are complete at an early stage of animal development, plants can switch their developmental fate throughout their life cycle, especially in response to environmental stimuli (Kohler and Spatz, 2002). From an analysis of phylogenetic relationships of 37 SET domain proteins from *Arabidopsis*, Baumbusch et al. (2001) concluded that there are 7 classes of SET domain proteins in plants: Enhancer of zeste homologs (E[z]), Trithorx homologs and related (ATXH and ATXR), Suppressor of variegation homologs and related (SUVH and SUVR), and Aabsent, small or homeotic discs1 homologs and related (ASHH and ASHR). Thus far, members from five of these groups (all except groups SUVR and ASHR) have been experimentally characterized.

CURLY LEAF (*CLF*) and its close relative SWINGER (SWN) are (E)z type Polycomb group (PcG) proteins with H3K27 HMT activity (Goodrich et al., 1997; Makarevich et al., 2006a). They have redundant functions in controlling leaf and flower morphology as well as flowering time through repression of the floral homeotic gene (Katz et al., 2004; Schubert et al., 2006). The third member of in E(z) group, MEDEA (MEA) is also an H3K27 HMT but is involved in seed development and can maintain its own imprinting during endosperm development (Chanvivattana et al., 2004; Makarevich et al.,

2006a; Schubert et al., 2006). In ASHH group, ASHH2 has H3K36 HMT activity and is involved in flowering control through histone methylation at the *FLOWERING LOCUS C* (*FLC*). Loss of *ASHH2* (*SDG8*) function resulted in early flowering (Zhao et al., 2005). In ATXH and ATXR groups, three members have been studied. ATX1 was the first confirmed H3K4 methyltransferase in plants and it is involved in floral development (Alvarez-Venegas et al., 2003). The other two reported ATX proteins, ATXR5 and ATXR6, have only a PHD domain and a truncated SET domain. They were found to regulate the cell cycle or DNA replication through interactions with proliferating cell nuclear antigen (Raynaud et al., 2006). Four members of the AtSUVH SET proteins have been functionally characterized as having H3K9 dimethylation specificity. While SuvH4 (also known as as KYP), SuvH5 and SuvH6 are involved in locus specific control of H3K9 dimethylation (Jackson et al., 2002)(Ebbs, 2006), SuvH2 is involved in overall heterochromatin formation (Naumann et al., 2005).

Acceptance of histone methylation as a major regulatory event in the regulation of eukaryotic gene expression required that it existed in all major model species. It is of interest, therefore, that *Dim-5*, the first and, thus far the only, HMT identified in *Neurospora* has a similar domain architecture to that of the plant SUVR proteins. In the *Neurospora Dim-5* mutant, DNA methylation is abolished at nearly all genomic loci. This mutant also shows a slow growth rate and produces few spores, most of which are not viable (Tamaru and Selker, 2001). Because both *Neurospora* Dim5 and *Arabidopsis* SuvR proteins have intact SET, Pre-SET and Post-SET domains but are devoid of other domains, it is a reasonable speculation that SUVR SET proteins may have HMT activity and could thus be involved in gene silencing. We conjectured that depletion of SUVR transcript would

interfere with the silencing, resulting in restoration of wild type functions. To evaluate this possibility, *SUVR3* was cloned into the RNAi vector pHANNIBAL (Wesley et al., 2001) and used for *Arabidopsis* transformation.

Materials and methods

RNAi vector construction

Two primer sets with different restriction enzyme recognition site overhangs were used to amplify a 400 bp coding region (spanning boh exons 1 and 2) of *SuvR3*. The PCR product digested with *Xho*I and *Kpn*I was inserted in a sense orientation into pHannibal (Wesley et al., 2001); the same PCR product digested with *BamH*I and *Cla*I was inserted in an antisense orientation. In the second step, a fragment containing *35S*:*SuvR3-s*:intron:*SuvR3-as*:*ocs* was released by *Not*I digestion and inserted into the binary vector *pArt27* (Gleave, 1992) to form the RNAi construct K-23. Primer sequences were:

*BamH*I and *Xho*I: 5'-gtc agt gga tcc ctc gag ctc aac gat acg cgt act tc-3'

*Cla*I and *Kpn*I: 5'-acg gac atc gat ggt acc gca tat tca cag atg aat tgg c-3'

Thermocycling conditions were 94°C for 5 min, followed by 35 cycles of 94°C for 1 min, 53°C for 1 min and 72°C for 2 min, with a final polymerization step at 72°C for 10 min. The K-1 construct was transformed into *Agrobacterium* (GV3101) using electroporation with a Gene Pulser (Bio-Rad, Hercules, CA). *Arabidopsis* plants were transformed using vacuum infiltration (Bechtold and Pelletier, 1998).

Reverse transcription PCR (RT-PCR)

RNA was extracted using TRIzol (Invitrogen, San Diego, CA) according to the manufacturer's instructions. RNA (1 μg) was treated with DNaseI (1 u μl^{-1}, Invitrogen) and RT-PCR was carried out using the QIAGEN OneStep RT-PCR Kit (Qiagen, Valencia, CA) to amplify a SuvR3 coding region upstream of the RNAi target region. Reactions contained 250 ng RNA, 6 μM of each SuvR3 primer and 0.08 μM of each EF1α primer in a final volume of 25 μl. Primer sequences for SuvR3 were:

Forward: 5'-CTCAACGATACGCGTACTTC-3' and

Reverse: 5'-GCATATTCACAGATGAATTGGC-3'.

Primer sequences for the internal control EF1α were:

Forward: 5'-TGCTGTCCTTATCATTGACTCCACCAC-3' and

Reverse: 5'-TTGGAGTACTTGGGGGTAGTGGCATC-3'.

Thermocycling conditions were: reverse-transcription at 50°C for 30 min, 95°C for 15 min, followed by 25 cycles of 94°C for 1 min, 55°C for 1 min and 72°C for 1 min, with a final polymerization step at 72°C for 10 min. The products of the RT-PCR amplification were subjected to electrophoresis through a 2.0% agarose gel, followed by staining with ethidium bromide (100 ng ml^{-1}). The gel was then digitally imaged and was analysed using ImageJ.

Plant genomic DNA blot

Genomic DNA was extracted from four to five leaves using an AutoGenprep 850alpha (Autogen, Holliston, MA). Genomic DNA (500 ng) was digested with 20 units of *Bam*HI for 17 h. After electrophoretic separation in a 0.7% agarose gel, the DNA fragments were transferred to Hybond-N+ membrane (Amersham). DNA probes were labeled using a DECAprime II kit (Ambion, Austin, TX). Hybridizations were performed using ULTRAhyb solution (Ambion) according to the manufacturer's recommendations. Hybridization signals were detected by exposure to a PhosphorImager (Fuji, Stamford, CT) and quantitated using the public domain NIH ImageJ program (developed at the U.S. National Institutes of Health and available on the Internet at http://rsb.info.nih.gov/ij//).

Plant transformation

Arabidopsis thaliana (ecotype Columbia) seeds were germinated in soil (Redi Earth, Scotts) and, following vernalization at 4°C for 48 h in the dark, grown at 22°C under a 16/8 h light/dark cycle. Transformants were selected on Murashige and Skoog (MS) medium (GIBCO™ Invitrogen Corporation, Carlsbad, CA) containing 100 μg ml^{-1} Timentin (ticarcillin disodium and clavulanate potassium, SmithKline Beecham Pharmaceuticals, UK), 50 μg ml^{-1} kanamycin (Sigma, St. Louis, MO), 3% (w/v) sucrose and 0.2% (w/v)

phytagel (Sigma). After two weeks, resistant plants were transferred to soil and grown in a growth chamber.

Microscopy analysis

For fluorescence microscopy, plants bearing a GFP transgene were screened for GFP expression using a Zeiss SV-11 fluorescence microscope equipped with suitable filters: excitation 450 nm; emission 500 nm (filters out red autofluorescence) or 525 nm. Pictures were taken using AxioCam HRc camera (Zeiss, Germany). The electron microscopy data were obtained using a ESEM microscope purchased under National Science Foundation grant No. ECS-9214314.

In vitro pollen germination

The procedure for *in vitro* pollen germination was essentially as described in Fan et al. (2001). In each experiment, 6 to 12 randomly selected flowers bearing freshly dehisced anthers (stage 13, (Sanders et al., 2000)) were used as pollen grains have the highest germination percentage at this developmental stage (Fan et al., 2001). Stamen were

collected carefully, so as not to damage the anther from individual flowers and the anthers were dipped into liquid pollen germination medium containing 5 mM MES (pH 5.8 adjusted with TRIS), 1 mM KCl, 10 mM CaCl2, 0.8 mM MgSO4, 1.5 mM boric acid, 16.6% (w/v) sucrose, 3.65% (w/v) sorbitol, and 10 μg ml^{-1} myo-inositol. The anthers were then gently rubbed onto the surface of solid pollen germination medium (the liquid medium plus 1% (w/v) agar) in a 24-well flat-bottom agar plate. Both liquid and solid media were prepared in deionized water and heated to 100°C for 2 min. Following pollen application, the agar plates were transferred to a growth chamber with continuous light and 100% humidity. The number of pollen grains with or without germination tubes were counted 24 h after the transfer to determine the pollen germination percentage. In one experiment, over 100 pollen grains from a single flower were counted and the experiment was repeated three times with different flowers from the same plants.

Results

RNAi-mediated knockdown of AtSET genes

Through a PSI-BLAST search in the then-incompletely sequenced *Arabidopsis* genome using SET domains present in yeast SET proteins SET1 and Clr4, and human SET protein SUV39H1, we identified 20 *Arabidopsis* SET domain genes, most of which were annotated as either hypothetical or unknown proteins. RNAi vectors targeting these SET genes were constructed and were introduced into *Arabidopsis* (ecotype Columbia) via *Agrobacterium*-mediated transformation. The resulting T_1 plants and their progeny were examined for morphological phenotypes. RNAi constructs that yielded abnormal phenotypes were *K-16*, *K-23*, and *K-24. These constructs respectively targeted ASH1 Homolog 3* (*ASHH3*), *SUPPRESSOR of VARIEGATION 3* (*SUVR3*), and *Trithorax 4* (*ATX4*). Among the three constructs, *K-23* (Fig. 4.1) that targets *SUVR3* showed the highest and inheritable RNAi

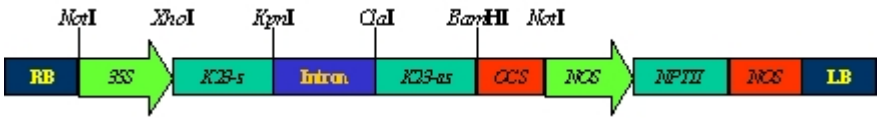

Fig 4.1 Architecture of T-DNA in *SuvR3*-targeting RNAi construct K-23. RB and LB: T-DNA right and left border; *35S*, cauliflower mosaic virus *35S* promoter; *K23-s* and *K23-as*: sense and antisense orientation, respectively, of the targeted *SuvR3* fragment. *NPT*II, neomycin phosphotransferase II. *OCS* and *NOS* denote terminators of *Agrobacterium* octopine synthase and nopaline synthase genes, respectively; the arrow labeled *NOS* denotes a nopaline synthase promoter. Relative positions are shown for the cleavage sites for the indicated restriction endonucleases used in the construction of the RNAi vectors.

expressivity (Wang et al., 2005a) and hence *SUVR3* was selected as the primary SET gene for further study (Fig. 4.2).

Inactivation of *SUVR3* can reactivate a silenced GFP reporter gene

To establish lines suitable for gene reactivation, *A. thaliana* ecotype Columbia was transformed [Bechtold, 1998 #7910] with *35S/mt-mGFP5/nos::nos/bar/nos*. Some 10,000 seeds were germinated on MS medium containing 10 μM bialaphos. Herbicide-resistant T_1

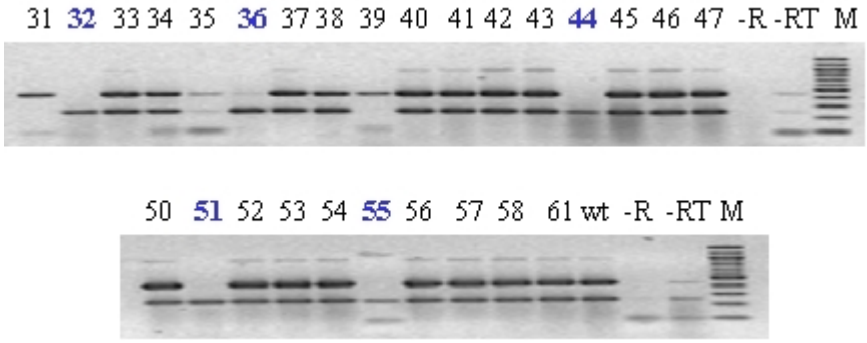

Fig. 4.2 Transcript depletion analysis in T_2 progeny of K23-8. Relative RT-PCR (40 cycles) was performed in wt and T_2 progeny of K23-8. Blue numbers represent K23 progeny lines exhibiting severe transcript knock-down. M: 100 bp DNA ladder, -R: no RNA control in RT-PCR, -RT: no reverse transcriptase control in RT-PCR.

seedlings were screened for GFP expression. Of the 100 T_1 progeny examined, 30 non-fluorescing, presumably silenced, lines were obtained and selfed; T_2 progeny seedlings were again selected on bialaphos and examined for fluorescence (Fig. 4.3). All T_2 progeny of six of the T_1 lines showed no fluorescence (Fig. 4.3) and were considered candidate silenced (SI)

lines. T_3 seedlings of these six lines were again screened and all progeny of one of these lines (SI-19-35) were non-fluorescent but were bialaphos resistant. Genomic DNA blot analysis showed that the parental line was a multi-copy, single locus homozygous plant (data not shown). This parental plant was chosen as the founder of homozygous silenced lines used in reactivation experiments.

RNAi construct *K-23* was supertransformed (ST) into T_4 progeny of selfed SI-19-35 and the resulting double transformants were screened for reactivation of *GFP* in roots, stems, leaves, and flowers. Reactivation was detected in segments of the roots in 4 out of 20 ST_1 progeny (Fig 4.4) but was not observed in other parts of all supertransformants including flowers, where endogenous *SUVR3* has the highest expression level. As an important control, parallel supertransformation of *GFP* silenced line SI-19-35 was performed with an empty RNAi vector that does not target any gene, or RNAi constructs that target genes encoding a methyl-binding domain protein, or eight other SET domain proteins. For each supertransformation, at least 20 plants were carefully examined for *GFP* expression and no reactivation of *GFP* was observed in any tissue.

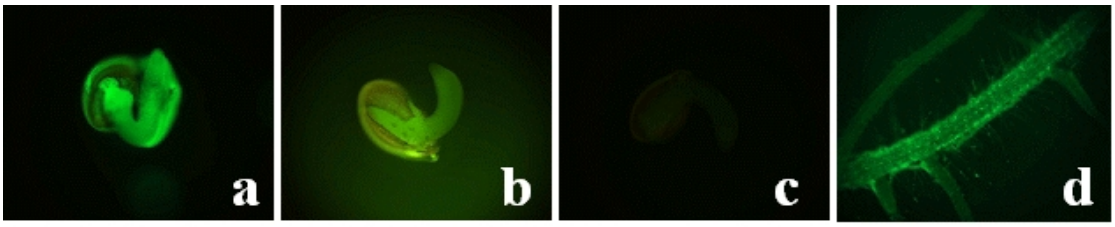

Fig. 4.3 Establishment of *Arabidopsis* silenced for *35S-GFP*. (a) Germinating T_2 seedling showing bright fluorescence at short (1,600 ms) exposure. (b) Silenced T_2 seedling showing fluorescence at long (20,566 ms) exposure but essentially none at short (1,600 ms) exposure (c). (d) Roots of a progeny plant of silenced line SI-19-35 reactivated by germination in MS medium containing 5-azaC (50 μM).

Fig. 4.4 Reactivation of GFP. A. Segment of root from *Arabidopsis* line SI-19-35-n (silenced for GFP) supertransformed with pArt27 (vector only control), representative of 14 independent lines. B. Root segment from line SI-19-35-n supertransformed with pK-23 (RNAi construct for *K-23* knockdown) showing reactivation of GFP, representative of 4 of 20 independent lines. Exposure was for 6092 ms. Both plants were 2 wk old.

Epigenetic regulation of flower development by *SuvR3*

As for many SET-domain genes, the database annotates *SuvR3* (At3g03750) only as a hypothetical protein, leaving many opportunities to learn its function. Through relative RT-PCR using RNA extracted from various organs of wild type (wt) plants as template, we found that wt *SuvR3* endogenous transcript was undetectable in siliques, present in moderate amounts in roots, rosette and cauline leaves, and stems; high levels were found in flowers (data not shown). This pattern suggested that *SuvR3* may play a role in flower development.

Thus, our phenotypic observations of *SuvR3* RNAi lines are mainly focused on, but not limited to, flower development.

Initially, no distinguishable developmental phenotypes were observed in virtually all T_1 and T_2 plants bearing the *SuvR3* RNAi construct. Subtle morphological changes observed were: delayed flowering and decreased plant size for a few (~ 2%) individual plants. However, these phenotypic aberrancies were not correlated with *SuvR3* transcript depletion level. In the T_3 generation, one plant (K23-8-36-18), out of 36 plants examined, displayed an interesting phenotype (Fig. 4.5): five flowers from one branch had 5 petals and sepals instead of 4 of each in the regular flowers. Stamen and carpel morphology in these aberrant flowers was normal. The aberrant petal phenotype is heritable as 10% (5 out of 50) of T_4 progeny and 30% of T_5 progeny of K23-8-36-18 displayed similar abnormal flowers. It was

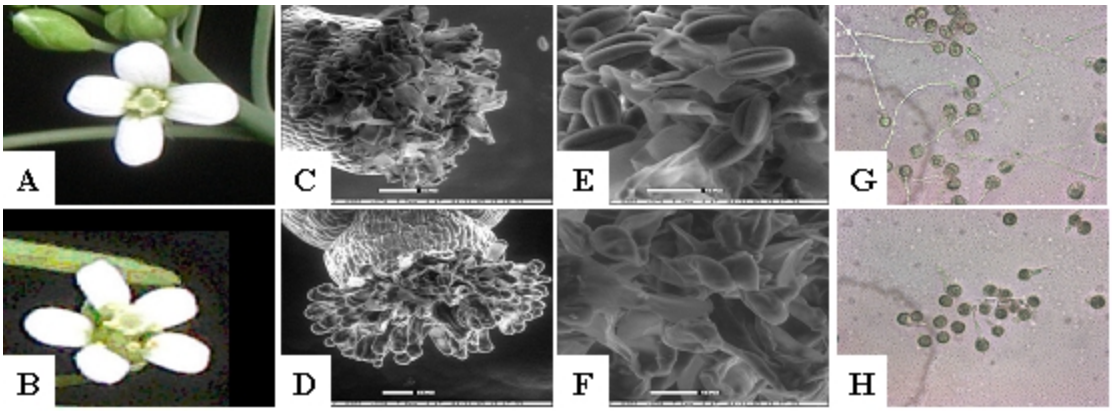

Fig. 4.5 Aberrant flower morphology in *SuvR3* RNAi lines. A: a typical flower of wt plants. B: a representative *SuvR3* aberrant flower. C-F: scanning electron micrograph (SEM) of a wt flower and an aberrant flower from *SuvR3* RNAi lines. C and D: stigma of wt and *SuvR3* 5-petal flower at 200 X magnification, respectively. E and F: stigma of wt and *SuvR3* 5-petal flower at 900 X magnification, respectively. SEM graphs were taken using Electroscan ESEM E-3. G: pollen grains from a wt plant. H. pollen grains from a representative flower (normal number of petals) of a *SuvR3* RNAi plant.

later found that another independent T_1 plant (K23-58) and 6% (3 out 45) of its T_2 progeny also had similar aberrant flowers. It is unlikely that this phenotype is caused by T-DNA inactivation or autonomous mutation of another gene since this phenotype (5 petals) was not observed in over 2000 plants that were transformed with either control plasmids and RNAi constructs (except K-16) that target other AtSET genes.

The severity and frequency of aberrant phenotypes was found to increase in the progeny of abnormal RNAi transformants. While 5 petals were seen in the T_3 generation of K23-8, petal numbers were found to vary between 3 and 6 in T_5 progeny of this plant. Carpels in two flowers from two different T_5 progeny of K23-8 remain unfused. Although the frequency of aberrant carpel development is very low, representing only 0.2% (2/~1000) of all flowers in K23-8 T_5 progeny that were screened, this phenotype may still be associated with *SUVR3* inactivation, as a similarly low frequency (~0.2%) of rare silencing phenotypes was observed in our earlier study of RNAi-mediated knockdown of the *phytoene desaturase* gene (Wang et al., 2005a).

All flowers that have an aberrant number of petals are sterile. To investigate the cause of infertility, the aberrant flowers in K23-8-36-18 progeny were examined using scanning electronic microscopy (SEM). Whereas the morphology of sepal, petal, stamen, and stigma appeared normal, virtually no pollen grains could be found on the stigmas of 5-petal flowers in the *SuvR3* RNAi lines (Fig. 4.5 D and F), even though the coordination of filament and style elongation appeared to be unaffected. To further study the parental origin of the sterility, a two-way cross fertilization between *SuvR*3 5-petal flowers and wt flowers was performed. Whereas ten crosses using wt stamens and *SuvR3* stigmas generated ~ 100

F1 seeds, the same number of crosses between wt stigmas and *SuvR3* stamens did not produce any F1 seed, indicating the sterility in *SuvR3* 5-petal flowers is due to a defective male gametophyte.

Morphologically normal (4-petal) flowers in *SuvR3* RNAi lines were also examined. A total number of 116 flowers randomly selected from progeny of four *SuvR3* RNAi lines were dissected to examine the flower organ morphology. Over 40% (47/116) of the flowers have decreased number of stamens ranging from zero to five (six in wt). The average number of stamen in these flowers from *SuvR3* RNAi lines was 5.39 ± 0.09 (Table 4.1). In contrast, six stamen were found in all 21 wt flowers examined. No differences in the number and morphology of stigmas and sepals were detected in these flowers from *SuvR3* RNAi lines.

Table 4.1 Effect of RNAi-mediated knockdown on flower development and pollen germination

Plant lines		WT	SuvR3 RNAi line
No. of flower organs	Total No. of flowers examined	21	116
	petals	4 ± 0	4 ± 0
	sepals	4 ± 0	3.98 ± 0.01
	stamen	6 ± 0	5.39 ± 0.09
	stigma	1 ± 0	1 ± 0
Pollen germination	Total No. of pollen grains examined	1374	1564
	% of germinated pollen grains	52 ± 7	24 ± 8

Number of flower organs and pollen germination percentage were recorded as mean ± standard

The viability of male gametophytes also tested. Since *in vivo* pollen germination in *Arabidopsis* is difficult to track, an *in vitro* approach (Fan et al., 2001) was used to study the fertility of pollen grains. Under our experimental conditions (see Materials and methods), approximately 50% of wt pollen grains germinated in vitro (Fig.4.5 G). In contrast, only ~ 20% of pollen grains from *SuvR3* RNAi lines germinated (Fig. 4.5 H).

Questions arising from the above observations include: Why is there a delay of several generations prior to the appearance of the phenotype and why does it occur at such a low frequency? Explanations include the possibility that insufficient SuvR3 protein is produced, leading (over generations) to the progressive loss of histone and DNA methylation in certain chromatin regions. Only when the methylation level is below a certain threshold are the phenotypes displayed. Moreover, the phenotype (e.g. flower aberrancy) may not be uniform in all parts of the plants (e.g. flower branches) since RNAi expressivity can be dramatically different in various parts of the plants (Wang et al., 2005a). Additionally, it has been observed that only about 20% of the flowers in homozygous *ATX1* T-DNA insertion lines displayed an aberrant phenotype in androecium and gynoecium development (Alvarez-Venegas et al., 2003).

SuvR3 is essential in seedling germination

While maintaining the plant lines silenced for *SuvR3*, we found that fewer seeds from high *SuvR3* RNAi expressivity lines could germinate even in media without antibiotic selection. For T_3 progeny of T_1 line K23-8 that expressed detectable levels of the *SuvR3* transcript,

>50% of the seed germinated on medium without any selection reagent whereas <20% germinated from K23-8 progeny lines for which RNAi expressivity was high and the endogenous *SuvR3* transcript was undetectable. In contrast, seeds from wt plants can have virtually 100% seedling gemination frequency under the same culture conditions. Seedlings derived from plants showing high RNAi expressivity were grown to produce progeny generations. Under our growth conditions, a wt plant usually produces ~ 2500 (100 µl) seeds. However, when the *SuvR3* RNAi lines were selfed to the T_6 generation, 3 out of 36 plants examined became sterile, and one plant produced only a few (~50) seeds.

While we analyzing the *SuvR3* RNAi lines, we learned that a T-DNA insertion line (SALK_063174) for *SuvR3* is available from *Arabidopsis* Biological Resource Center (ABRC). According to the flanking tag sequencing results (available at http://signal.salk.edu/cgi-bin/tdnaexpress), the insertion in this line is at the second exon, ~130 nt upstream from the stop codon. Seeds for this T-DNA insertion line were obtained from ABRC and were tested for the germination percentage. A total of 137 seeds, acquired in three batches, were tested in media without any antibiotic selection. Only six seeds germinated. DNA were extracted from the derived plants and were used in three-primer PCR to verify the T-DNA insertion locus. None of these six plants had a T-DNA in the specified position. DNA from those seeds that were unable to germinate was also extracted and was used in the same PCR procedure to verify the insertion locus. Thus far, we have not been able to verify the presence of T-DNA in the designated position in these non-viable seeds.

Alternative splicing of *SuvR3*

Alternative splicing of transcripts is thought to contribute to genome complexity. While it has been estimated that it occurs for 35 to 60% of all human genes (Brett et al., 2000) and for 22% of *Arabidopsis* genes (Wang and Brendel, 2006), functional characterization of isoforms and their interactions is poorly described. Using RT-PCR, we have recently confirmed that two isoform SuvR3 RNAs can be reproducibly identified in *Arabidopsis* tissues (Fig. 4.6). Sequencing of the transcripts revealed that both sequences are present in

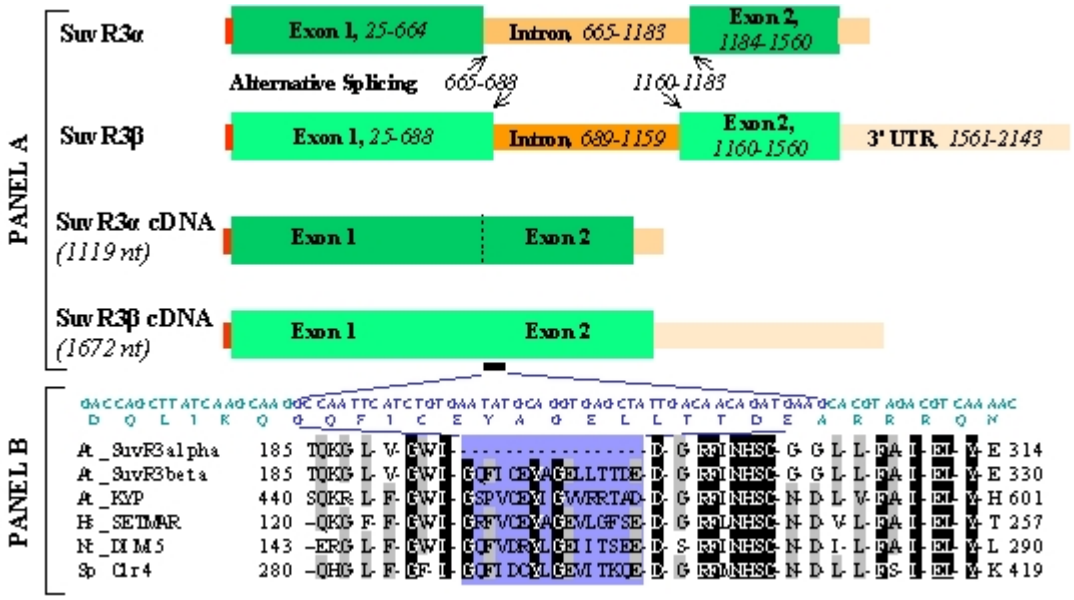

Fig. 4.6 Structures of *SUVR3*α and *SUVR3*β isoforms. Panel A: The red, green, orange and boxes denote 5' UTR, exon, intron and 3' UTR regions, respectively. Their positions in *SUVR3*α and *SUVR3*β are indicated. The 48-nt fragment represented by the black bar, which belongs to the exonic regions in *SUVR3*β but belongs to the intronic region in *SUVR3*α, and its corresponding 16-aar are shown in blue letters. Panel B: Sequence alignment of the SET domain of three *Arabidopsis* SET proteins: SUVR3α, SUVR3β, *Arabidopsis* KRYPTONITE (At_KYP, AAK28969); human SET domain mariner transposase (Hs_SETMAR, AAH11635); *N. crassa* (*Nc*) DIM-5 (CAF06044) and *S. pombe* (*Sp*) CLR4 (NP_595186). The aar highlighted are invariant (white on black) or conserved (white on gray) among all six SET proteins.

cDNA and EST libraries (Asamizu et al., 2000; Yamada et al., 2003), each comprising two exons and one intron. The shorter sequence is designated here as SuvR3α and the longer one as SuvR3β. SuvR3α (NM_111246) is 1119 nt in length, including a 24 nt 5' UTR and a 78 nt 3'UTR; it encodes a protein of 338 amino acid residues (aa) with a calculated pI value of 7.076. SuvR3β (NM_202483) consists of 2143 nt and includes the same 24 nt 5' UTR but bears a much longer 3' UTR (583 nt) and encodes a predicted protein of 354 aa with a calculated pI value of 5.797.

Examination of the transcripts from these isoforms revealed that SuvR3α has the unusual intron border sequence GC–AG (Brown et al., 2002) instead of the normal GU–AG that is in SuvR3β. Compared to SuvR3β, the 5' border of the intron in SuvR3α is 24 nt upstream of that in SuvR3β and 3' border is 24 nt downstream of that in SuvR3β. Thus, the total coding sequence length of SuvR3α is 48 nt shorter than that in SuvR3β, resulting in a 16 aa difference between the two isoforms. Protein sequence alignment of both SuvR3 isoforms with KYP (Jackson et al., 2002), Dim-5 (Tamaru and Selker, 2001) and other SET proteins revealed that these 16 aa contain 3 invariant and 3 conserved residues among almost all members of the SUV39 family. Moreover, of the 12 conserved strands within the SET domain defined by Zhang et al. (2002), two (strands 8 and 9) are lost due to the absence of these 16 aa. This indicates that the *SuvR3* isoforms may differ in functions.

Discussion

Regulation of transgene silencing by *SuvR3*

Chromatin remodeling genes are known to regulate gene expression through either changes in chromatin structure or altered nucleosome positioning. SET proteins are key components of such gene expression regulatory systems. Members from the SuvH group of SET genes are known to control heterochromatin formation and to suppress the expression of retrotransposons. Here we provide evidence for the involvement of a SuvR SET gene in transgene silencing. Both SuvR and SuvH have Pre-SET, SET, and Post-SET domains. However, a SET and ring finger associated region (the SRA/YDG domain), is absent in SuvR but present in SuvH.

The SRA/YDG domain is important in regulating the interaction between the SET domain proteins and histones, especially H3 (Citterio et al., 2004). Studies in AtSuvH2 show that the YDG domain is important in directing DNA methylation to the target sequences (not limited to histone 3), a prerequisite for histone methylation mediated by AtSuvH2 (Naumann et al., 2005). Examination of the YDG domain revealed that this domain can be found in 80 proteins from plants, animals and green algae. In animals, the YDG domain is always associated with a PHD (plant homeo domain) domain. The association of SET domain and YDG domain seems to have originated in the last common ancestor of green algae and land plants (~ 700 million years ago, Heckman et al. 2001)), as this unique combination was only found in these two types of organisms.

Using the Pfam database, we found that there are 17 non-SET proteins in *Arabidopsis* containing the YDG domain. Members of these proteins, or even a SuvH SET protein, may interact with SuvR3 to methylate certain histone residues and hence change chromatin structure.

A question rises regarding the lack of *GFP* reactivation in flower tissues since the endogenous *SuvR3* transcripts are most abundant in flowers. There could be two possible factors contributing to this phenomenon. First, the reporter *GFP* gene is driven by cauliflower mosaic virus *35S* promoter. Although this promoter is widely considered to be a constitutive promoter, several investigations have shown it displays variable spatial and temporal expression patterns (Benfey et al., 1989; Zhou et al., 2005)(Battraw and Hall 1990). In several *Arabidopsis* lines transgenic for *GFP* we have observed that the double-enhanced *35S* promoter we have is highly expressed in roots, but only moderately expressed in leaves and stems. Thus, an alleviation of gene silencing in the CaMV *35S* promoter may be discerned in roots but not in other organs. Second, chlorophylls, present in both leaves and flowers, have been shown to quench the fluorescence emitted by *GFP* (Zhou et al., 2005).

SuvR3 regulates flower development in an epigenetic manner

The delayed onset of aberrant flower morphology in *SuvR3* RNAi lines is intriguing. Few aberrant flowers were seen in the T_1 generation, even in lines with high RNAi expressivities, suggesting that reduction of *SuvR3* expression by the RNAi construct in one generation has little impact on some as yet unidentified downstream element that regulates flower

development. Therefore, several generations are required before the mutant phenotype can be readily discerned. Similar phenomena have been observed for other genes. In a telomerase reverse transcriptase mutant (Riha et al., 2001), loss of function did not result in an identifiable phenotype until the 6th generation. Some morphological phenotypes became progressively more severe in later generations for *ddm1* mutants (Kakutani et al., 1996). In this study, RNAi-mediated knock down of *SuvR3* could direct progressive loss of histone and DNA methylation in some chromatin region. This is supported by the observation that the percentage of plants that have aberrant phenotypes increased from 1.5% in the T_1 generation to 30% in the T_5 generation. Moreover, the severity of the aberrancy increases over the generations. Initially, only gain of one petal were seen in T_1 to T_3 generations while other parts of flowers remained normal. In contrast, in T_5 and T_6 generations, both loss of one petal and gain of multiple petals were seen, in company with abnormal carpel development. Recent additional evidence is our finding that the number of stamen in individual flowers (six in T_1 to T_3 plants), was also reduced in T_6 plants, regardless of petal numbers. To summarize, our data show that *SuvR3* regulates flower development through as yet unidentified epigenetic mechanisms.

Pivotal epigenetic role of *SuvR3* in early seedling development

Seedling germination tests showed that most seeds derived from *SuvR3* high expressivity lines failed to germinate, even without antibiotic selection. Since this could signal an

important finding, we decided to purchase T-DNA insertion lines from ABRC. Only six out of 167 seeds germinated; all six grew like wild type. These results were exciting as they suggested that normal expression of *SuvR3* is crucial in seedling development. However, the possibility existed that some mistake occurred during handling or labeling of the insertion line seeds. Verification that the T-DNA insertion was at the designated site was sought through PCR as this would provide unequivocal support for the importance of *SuvR3* function in seedling development. Unfortunately, no T-DNA insertion was detected at the *SuvR3* locus in ten independent assays (representing 20 non-viable seeds).

SuvR3 alternative splicing and possible dimer formation between isoform proteins

Alternative splicing of *SuvR3* results in two SET proteins with 95% homology and differing in only 16 aa in their SET domains (Fig. 4.6). This 16 aa region spans the junction of SET-N and SET-I with a majority of aa (12) in the SET-I. To our knowledge, this is the first example of AS in SET genes that results in two isoform proteins that are identical in all domains except a short region within the SET domain. Although the HMT catalytic motifs within the SET domain, SET-C and SET C flanking region, are not affected in both isoforms, the observation that 6 of these 16 aa are highly conserved in all characterized SET domain proteins with HMT activity led to the speculation that the shorter isoform, SuvR3α, may have no or reduced HMT activity by itself or have lost its ability of substrate

specification, a function assigned to SET-I region. Alternatively, SuvR3α may retain its binding ability but have lost its catalytic function and hence serve as a competitive inhibitor.

It is also possible that SuvR3α and SuvR3β function as homodimers or heterodimers. Virus SET proteins can form homodimers in solution by domain II tethering interactions (Manzur et al., 2003; Qian et al., 2006). Similarly, homodimerization has also been observed for human ALL-1/MLL (a Trithorax homolog), G9a, and GLP, and Drosophila Ash1 and Trithorax1 through SET-SET interactions (Rozovskaia et al., 2000). Heterodimer formation was also found for human G9a and GLP, and *Drosophila Ash1* and Trithorax1. If such dimerization exists for SuvR3 isoform proteins, two alternative hypotheses can be envisaged concerning the mechanism by which dimerization affects biological activity. First, the heterodimer SuvR3αβ could be the active complex and the homodimer SuvR3ββ the inactive protein complex. Alternatively, the SuvR3αβ heterodimer could be inactive and the SuvR3ββ homodimer be the active complex. In the second model, the synthesis of relatively small amounts of SuvR3α may be needed to achieve the same goal. We favor the second hypothesis as it provides the system with greater sensitivity and flexibility. As SuvR3α lacks two conserved motifs within its SET domain (Fig. 4.7), it is unlikely to have HMT activity by itself, or as a homodimer. Further in vitro experiments with purified SuvR3 isoform proteins are necessary to distinguish these two hypotheses.

CHAPTER V

SUMMARY

In this dissertation, I used domain architecture as an important criterion to study the relationship among various proteins. Whereas the conventional sequence alignment-based classification has been very valuable in illustrating the possible origin and evolutionary relatedness of proteins, my approach focuses on the function of the proteins. Proteins with the same origin from various species may have diverged too far to retain a close sequence resemblance but may still possess the same domain architecture. Alternatively, proteins from distinct origins could have reached the same domain architecture but do not have high sequence similarity. The structure-based classification approach, but not the sequence alignment-based approach, can predict whether or not the proteins being compared share similar functions. However, protein cataloging using the intuitive domain architecture approach is highly dependent upon the correct prediction of domains.

The domain architecture for 47 *Arabidopsis* SET (AtSET) proteins was obtained from Pfam, SMART and UniProt databases and was used to classify AtSET proteins. Comparison of the two classification approached for AtSET proteins resulted in identical results for most subgroups but discrepancies were found for four groups. In particular, we found that ASHR3, classified as an ASH-related group according to its SET domain sequence similarity to that of other ASSH proteins, is more related to two cycle regulators, ATXR5 and ATXR6 (Raynaud et al., 2006), because they share identical domain

architecture. Therefore, a regulatory role of ASHR3 in cell cycle is predicted. Another distinction is that I found that the ATXR group can be further classified into 3 small groups, with one of them (Group 5) possibly devoid of HMT activity.

While searching the EST and cDNA databases for these 47 AtSET genes, I found that natural antisense transcripts (NAT) were present for nine AtSET genes, indicating that their expression could be profoundly regulated by RNA. I also found that 17 (36%) of AtSET genes undergo alternative splicing. The high frequency of alternative splicing in AtSET suggested that alternative splicing may be important for their function. One possibility is that alternative splicing could result in two (or more) protein isoforms that can interact with each other, with one component acting as a regulatory element and the other as catalytic unit. Alternatively, the proteins derived from transcript isoforms may have unrelated functions or antagonistic functions.

To study the functions of AtSET genes, I constructed RNAi vectors targeting 20 AtSET genes but found that introduction of most of these RNAi plasmids into plants did not result in readily discernable phenotypes. Relative RT-PCR in the T_1 transformants revealed that transcript depletion levels can be different among various transformants. These two observations prompted us to evaluate the efficacy of the RNAi-mediated approach for inactivation of gene function in plants using a *phytoene desaturase* (*PDS*) gene, inactivation of which results in a readily identifiable photobleached phenotype. In a population of nearly 500 T_1 transformants, we found that vector-based RNAi in plants have a high penetrance. However, the expressivity of such RNAi constructs can vary from subtle to extreme.

Moreover, I studied the relationship between the RNA depletion level and severity of photobleaching. A strong correlation between transcript depletion level and morphological phenotypes was observed for PDS silencing, rendering it a valuable method for determining a quantitative measurement of RNAi expressivity. Previous studies have not reported this correlation. Readons for this include the use of pooled plant samples and the failure to appreciate that RNAi expressivity can differ dramatically between different tissues (e.g. flower and rosette leaf) and even within the same tissue. In addition, we found that RNAi expressivity is decreases from generation to generation. Furthermore, we demonstrated that there is no correlation between transgene copy number and transgene expressivity, supporting an earlier observation that transgene organization rather than copy number is the trigger for gene silencing (Yang et al., 2005)

One of the AtSET genes, *SuvR3* was further studied for its role in plant development and transgene silencing. Tissue specific RT-PCR showed that the highest expression levels of *SuvR3* occurred in flowers. Indeed, I showed that this gene may epigenetically regulate flower development, especially anther development. In two independent RNAi lines exhibiting high RNAi expressivity, a small portion of flowers had an aberrant number of petals. Interestingly, both the percentage of aberrant flowers and the severity of phenotypes can increase over generations, suggesting that an epigenetic regulatory pathway is present in the control of flower development by *SuvR3*. Selfed *SuvR3* RNAi lines showed loss of fertility, probably due to a defective male gametophyte and loss of pollen grain viability. Seedling germination tests of RNAi lines showed that *SuvR3* may have a crucial function in

early plant development. This follows from the observation that seeds derived from high RNAi expressivity lines have a much lower germination rate than those from low RNAi expressivity lines or wt lines. Additionally, *SuvR3* may be involved in transgene silencing as inactivation of *SuvR3* was shown to reactivate a previously silenced *GFP* reporter gene in the roots.

Interestingly, *SuvR3* can undergo alternative splicing through an alternative positioning mechanism (Wang and Brendel, 2006). I confirmed the presence of two AS transcripts in several tissues of *Arabidopsis* by RT-PCR. Conceptually translated protein isoforms differ in their SET-I region with SuvR3α lacking 16 amino acids, six of which were shown to be highly conserved among characterized H3K9 histone methyltransferase. The absence of conserved residues in SuvR3α suggests that it may not be an active enzyme. Lastly, the possible interplay between the two *SuvR3* isoform proteins was discussed.

REFERENCES

Aasland, R., Gibson, T.J., and Stewart, A.F. (1995). The PHD finger: implications for chromatin-mediated transcriptional regulation. Trends Biochem. Sci. 20, 56-59.

Abel, K.J., Brody, L.C., Valdes, J.M., Erdos, M.R., McKinley, D.R., Castilla, L.H., Merajver, S.D., Couch, F.J., Friedman, L.S., Ostermeyer, E.A., Lynch, E.D., King, M.C., Welcsh, P.L., Osborne-Lawrence, S., Spillman, M., Bowcock, A.M., Collins, F.S., and Weber, B.L. (1996). Characterization of *EZH1*, a human homolog of *Drosophila Enhancer of zeste* near *BRCA1*. Genomics 37, 161-171.

Allfrey, V.G., Faulkner, R., and Mirsky, A.E. (1964). Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. Proc. Natl. Acad. Sci. U.S.A. 51, 786-894.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389-3402.

Alvarez-Venegas, R., and Avramova, Z. (2002). SET-domain proteins of the Su(var)3-9, E(z) and trithorax families. Gene 285, 25-37.

Alvarez-Venegas, R., Pien, S., Sadder, M., Witmer, X., Grossniklaus, U., and Avramova, Z. (2003). ATX-1, an *Arabidopsis* homolog of trithorax, activates flower homeotic genes. Curr Biol 13, 627-637.

Aravind, L., and Iyer, L.M. (2003). Provenance of SET-domain histone methyltransferases through duplication of a simple structural unit. Cell Cycle 2, 369-376.

Asamizu, E., Nakamura, Y., Sato, S., and Tabata, S. (2000). A large scale analysis of cDNA in *Arabidopsis thaliana*: generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries. DNA Res. 7, 175-180.

Bachman, K.E., Park, B.H., Rhee, I., Rajagopalan, H., Herman, J.G., Baylin, S.B., Kinzler, K.W., and Vogelstein, B. (2003). Histone modifications and silencing prior to DNA methylation of a tumor suppressor gene. Cancer Cell 3, 89-95.

Battraw, M.J., and Hall, T.C. (1990). Histochemical analysis of *CaMV 35S* promoter-β-glucuronidase gene expression in transgenic rice plants. Plant Mol. Biol. 15, 527-538.

Bauer, U.M., Daujat, S., Nielsen, S.J., Nightingale, K., and Kouzarides, T. (2002). Methylation at arginine 17 of histone H3 is linked to gene activation. EMBO Rep. 3, 39-44.

Baulcombe, D.C. (1999). Fast forward genetics based on virus-induced gene silencing. Curr. Opin. Plant. Biol 2, 109-113.

Baumbusch, L.O., Thorstensen, T., Krauss, V., Fischer, A., Naumann, K., Assalkhou, R., Schulz, I., Reuter, G., and Aalen, R.B. (2001). The *Arabidopsis thaliana* genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. Nucleic Acids Res 29, 4319-4333.

Benfey, P.N., Ren, L., and Chua, N.H. (1989). The CaMV *35S* enhancer contains at least two domains which can confer different developmental and tissue-specific expression patterns. EMBO J. 8, 2195-2202.

Bienz, M. (2006). The PHD finger, a nuclear protein-interaction domain. Trends Biochem. Sci. 31, 35-40.

Blackburn, M.L., Chansky, H.A., Zielinska-Kwiatkowska, A., Matsui, Y., and Yang, L. (2003). Genomic structure and expression of the mouse *ESET* gene encoding an ERG-associated histone methyltransferase with a SET domain. Biochim. Biophys. Acta. 1629, 8-14.

Blanc, G., Barakat, A., Guyot, R., Cooke, R., and Delseny, M. (2000). Extensive duplication and reshuffling in the *Arabidopsis* genome. Plant Cell 12, 1093-1101.

Boutros, M., Kiger, A.A., Armknecht, S., Kerr, K., Hild, M., Koch, B., Haas, S.A., Consortium, H.F., Paro, R., and Perrimon, N. (2004). Genome-wide RNAi analysis of growth and viability in *Drosophila* cells. Science 303, 832-835.

Braunstein, M., Sobel, R.E., Allis, C.D., Turner, B.M., and Broach, J.R. (1996). Efficient transcriptional silencing in *Saccharomyces cerevisiae* requires a heterochromatin histone acetylation pattern. Mol. Cell Biol 16, 4349-4356.

Brett, D., Hanke, J., Lehmann, G., Haase, S., Delbruck, S., Krueger, S., Reich, J., and Bork, P. (2000). EST comparison indicates 38% of human mRNAs contain possible alternative splice forms. FEBS Lett 474, 83-86.

Brown, J.W., Simpson, C.G., Thow, G., Clark, G.P., Jennings, S.N., Medina-Escobar, N., Haupt, S., Chapman, S.C., and Oparka, K.J. (2002). Splicing signals and factors in plant intron removal. Biochem Soc Trans 30, 146-149.

Brown, S.E., Campbell, R.D., and Sanderson, C.M. (2001). Novel NG36/G9a gene products encoded within the human and mouse MHC class III regions. Mamm. Genome. 12, 916-924.

Brutnell, T.P. (2002). Transposon tagging in maize. Funct. Integr. Genomics 2, 4-12.

Carthew, R.W. (2001). Gene silencing by double-stranded RNA. Curr. Opin. Cell Biol. 13, 244-248.

Chanvivattana, Y., Bishopp, A., Schubert, D., Stock, C., Moon, Y.H., Sung, Z.R., and Goodrich, J. (2004). Interaction of Polycomb-group proteins controlling flowering in *Arabidopsis*. Development 131, 5263-5276.

Cheng, X., Collins, R.E., and Zhang, X. (2005). Structural and sequence motifs of protein (histone) methylation enzymes. Annu. Rev. Biophys. Biomol. Struct. 34, 267-294.

Chuang, C.F., and Meyerowitz, E.M. (2000). Specific and heritable genetic interference by double-stranded RNA in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. U.S.A. 97, 4985-4990.

Citterio, E., Papait, R., Nicassio, F., Vecchi, M., Gomiero, P., Mantovani, R., Di Fiore, P.P., and Bonapace, I.M. (2004). Np95 is a histone-binding protein endowed with ubiquitin ligase activity. Mol. Cell. Biol. 24, 2526-2535.

Cui, X., De Vivo, I., Slany, R., Miyamoto, A., Firestein, R., and Cleary, M.L. (1998a). Association of SET domain and myotubularin-related proteins modulates growth control. Nat. Genet. 18, 331-337.

Cui, X., De Vivo, I., Slany, R., Miyamoto, A., Firestein, R., and Cleary, M.L. (1998b). Association of SET domain and myotubularin-related proteins modulates growth control. Nat. Genet. 18, 331-337.

Cutter, A.D., Payseur, B.A., Salcedo, T., Estes, A.M., Good, J.M., Wood, E., Hartl, T., Maughan, H., Strempel, J., Wang, B., Bryan, A.C., and Dellos, M. (2003). Molecular correlates of genes exhibiting RNAi phenotypes in *Caenorhabditis elegans*. Genome Res. 13, 2651-2657.

Dean, J.D., Goodwin, P.H., and T., H. (2002). Comparison of relative RT-PCR and Northern blot analyses to measure expression of β-1,3-Glucanase in *Nicotiana benthamiana* infected with *Colltotrichum destrictivum*. Plant Mol. Biol. Rep. 20, 347-356.

Fan, L.M., Wang, Y.F., Wang, H., and Wu, W.H. (2001). *In vitro Arabidopsis* pollen germination and characterization of the inward potassium currents in *Arabidopsis* pollen grain protoplasts. J. Exp. Bot. 52, 1603-1614.

Fetherson, R.A., Strock, S.B., White, K.N., and Vaughn, J.C. (2006). Alternative pre-mRNA splicing in *Drosophila* spliceosomal assembly factor RNP-4F during development. Gene 371, 234-245.

Finn, R.D., Mistry, J., Schuster-Bockler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., Moxon, S., Marshall, M., Khanna, A., Durbin, R., Eddy, S.R., Sonnhammer, E.L., and Bateman, A. (2006). PFAM: clans, web tools and services. Nucleic Acids Res. 34, D247-251.

Foley, E., and O'Farrell, P.H. (2004). Functional dissection of an innate immune response by a genome-wide RNAi screen. PLoS Biol. 2, E203.

Forneris, F., Binda, C., Vanoni, M.A., Mattevi, A., and Battaglioli, E. (2005). Histone demethylation catalysed by LSD1 is a flavin-dependent oxidative process. FEBS Lett. 579, 2203-2207.

Fuchs, J., Demidov, D., Houben, A., and Schubert, I. (2006). Chromosomal histone modification patterns--from conservation to diversity. Trends Plant Sci. 11, 199-208.

Garrick, D., Fiering, S., Martin, D.I., and Whitelaw, E. (1998). Repeat-induced gene silencing in mammals. Nature Genetics 18, 56-59.

Gehring, M., Huh, J.H., Hsieh, T.F., Penterman, J., Choi, Y., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (2006). DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. Cell 124, 495-506.

Gleave, A.P. (1992). A versatile binary vector system with a T-DNA organisational structure conducive to efficient integration of cloned DNA into the plant genome. Plant Mol.Biol. 120, 1203-1207.

Goodrich, J., Puangsomlee, P., Martin, M., Long, D., Meyerowitz, E.M., and Coupland, G. (1997). A Polycomb-group gene regulates homeotic gene expression in *Arabidopsis*. Nature 386, 44-51.

Goodwin, T.W. (1988). Plant pigments. (San Diego, CA: Academic Press).

Grewal, S.I., and Rice, J.C. (2004). Regulation of heterochromatin by histone methylation and small RNAs. Curr. Opin. Cell. Biol. 16, 230-238.

Griffiths, A.J.F. (1996). An introduction to genetic analysis. (New York: W.H. Freeman).

Gu, Y., Nakamura, T., Alder, H., Prasad, R., Canaani, O., Cimino, G., Croce, C.M., and Canaani, E. (1992). The t(4;11) chromosome translocation of human acute leukemias fuses the *ALL-1* gene, related to *Drosophila trithorax*, to the *AF-4* gene. Cell 71, 701-708.

Guo, H.S., Fei, J.F., Xie, Q., and Chua, N.H. (2003). A chemical-regulated inducible RNAi system in plants. Plant J. 34, 383-392.

Gupta, S., Schoer, R.A., Egan, J.E., Hannon, G.J., and Mittal, V. (2004). Inducible, reversible, and stable RNA interference in mammalian cells. Proc. Natl. Acad. Sci. U.S.A. 101, 1927-1932.

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

Hamilton, A.J., and Baulcombe, D.C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. Science 286, 950-952.

Hilson, P., Allemeersch, J., Altmann, T., Aubourg, S., Avon, A., Beynon, J., Bhalerao, R.P., Bitton, F., Caboche, M., Cannoot, B., Chardakov, V., Cognet-Holliger, C., Colot, V., Crowe, M., Darimont, C., Durinck, S., Eickhoff, H., de Longevialle, A.F., Farmer, E.E., Grant, M., Kuiper, M.T., Lehrach, H., Leon, C., Leyva, A., Lundeberg, J., Lurin, C., Moreau, Y., Nietfeld, W., Paz-Ares, J., Reymond, P., Rouze, P., Sandberg, G., Segura, M.D., Serizet, C., Tabrett, A., Taconnat, L., Thareau, V., Van Hummelen, P., Vercruysse, S., Vuylsteke, M., Weingartner, M., Weisbeek, P.J., Wirta, V., Wittink, F.R., Zabeau, M., and Small, I. (2004). Versatile gene-specific

sequence tags for *Arabidopsis* functional genomics: transcript profiling and reverse genetics applications. Genome Res. 14, 2176-2189.

Himber, C., Dunoyer, P., Moissiard, G., Ritzenthaler, C., and Voinnet, O. (2003). Transitivity-dependent and -independent cell-to-cell movement of RNA silencing. EMBO J. 22, 4523-4533.

Hobbs, S.L., Warkentin, T.D., and DeLong, C.M. (1993). Transgene copy number can be positively or negatively associated with transgene expression. Plant Mol. Biol. 21, 17-26.

Jackson, J.P., Lindroth, A.M., Cao, X., and Jacobsen, S.E. (2002). Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature 416, 556-560.

Jackson, J.P., Johnson, L., Jasencakova, Z., Zhang, X., PerezBurgos, L., Singh, P.B., Cheng, X., Schubert, I., Jenuwein, T., and Jacobsen, S.E. (2004). Dimethylation of histone H3 lysine 9 is a critical mark for DNA methylation and gene silencing in *Arabidopsis thaliana*. Chromosoma 112, 308-315.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science 293, 1074-1080.

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

Jones, L., Hamilton, A.J., Voinnet, O., Thomas, C.L., Maule, A.J., and Baulcombe, D.C. (1999). RNA-DNA interactions and DNA methylation in post-transcriptional gene silencing. Plant Cell 11, 2291-2302.

Jones, R.S., and Gelbart, W.M. (1993). The *Drosophila* Polycomb-group gene *Enhancer of zeste* contains a region with sequence similarity to *trithorax*. Mol. Cell. Biol. 13, 6357-6366.

Jullien, P.E., Katz, A., Oliva, M., Ohad, N., and Berger, F. (2006). Polycomb group complexes self-regulate imprinting of the Polycomb group gene *MEDEA* in *Arabidopsis*. Curr. Biol. 16, 486-492.

Kakutani, T., Jeddeloh, J.A., Flowers, S.K., Munakata, K., and Richards, E.J. (1996). Developmental abnormalities and epimutations associated with DNA hypomethylation mutations. Proc. Natl. Acad. Sci. U.S.A. 93, 12406-12411.

Kamath, R.S., Fraser, A.G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M., Welchman, D.P., Zipperlen, P., and Ahringer, J. (2003). Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. Nature 421, 231-237.

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

Kempin, S.A., Liljegren, S.J., Block, L.M., Rounsley, S.D., Yanofsky, M.F., and Lam, E. (1997). Targeted disruption in *Arabidopsis*. Nature 389, 802-823.

Kerschen, A., Napoli, C.A., Jorgensen, R.A., and Muller, A.E. (2004). Effectiveness of RNA interference in transgenic plants. FEBS Lett. 566, 223-228.

Kim, D.H., Villeneuve, L.M., Morris, K.V., and Rossi, J.J. (2006). Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. Nat. Struct. Mol. Biol. 13, 793-797.

Kinoshita, T., Yadegari, R., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (1999). Imprinting of the *MEDEA* polycomb gene in the *Arabidopsis* endosperm. Plant Cell 11, 1945-1952.

Kohler, L., and Spatz, H.C. (2002). Micromechanics of plant tissues beyond the linear-elastic range. Planta 215, 33-40.

Krajewski, W.A., Nakamura, T., Mazo, A., and Canaani, E. (2005). A motif within SET-domain proteins binds single-stranded nucleic acids and transcribed and supercoiled DNAs and can interfere with assembly of nucleosomes. Mol. Cell. Biol. 25, 1891-1899.

Lachner, M., O'Sullivan, R.J., and Jenuwein, T. (2003). An epigenetic road map for histone lysine methylation. J. Cell. Sci. 116, 2117-2124.

Lechtenberg, B., Schubert, D., Forsbach, A., Gils, M., and Schmidt, R. (2003). Neither inverted repeat T-DNA configurations nor arrangements of tandemly repeated transgenes are sufficient to trigger transgene silencing. Plant J. 34, 507-517.

Lee, C., and Wang, Q. (2005). Bioinformatics analysis of alternative splicing. Brief Bioinform. 6, 23-33.

Lee, J.T., Davidow, L.S., and Warshawsky, D. (1999). *Tsix*, a gene antisense to Xist at the X-inactivation centre. Nat. Genet. 21, 400-404.

Lee, M.G., Wynder, C., Cooch, N., and Shiekhattar, R. (2005). An essential role for *CoREST* in nucleosomal histone 3 lysine 4 demethylation. Nature 437, 432-435.

Lee, M.G., Wynder, C., Bochar, D.A., Hakimi, M.A., Cooch, N., and Shiekhattar, R. (2006). Functional interplay between histone demethylase and deacetylase enzymes. Mol. Cell. Biol. 26, 6395-6402.

Levin, J.Z., de Framond, A.J., Tuttle, A., Bauer, M.W., and Heifetz, P.B. (2000). Methods of double-stranded RNA-mediated gene inactivation in *Arabidopsis* and their use to define an essential gene in methionine biosynthesis. Plant Mol. Biol. 44, 759-775.

Li, G., Chandler, S.P., Wolffe, A.P., and Hall, T.C. (1998). Architectural specificity in chromatin structure at the TATA box *in vivo*: Nucleosome displacement upon β-phaseolin gene activation. Proc. Natl. Acad. Sci. U.S.A. 95, 4772-4777.

Lia, G., Praly, E., Ferreira, H., Stockdale, C., Tse-Dinh, Y.C., Dunlap, D., Croquette, V., Bensimon, D., and Owen-Hughes, T. (2006). Direct observation of DNA distortion by the RSC complex. Mol. Cell 21, 417-425.

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

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389, 251-260.

Lusser, A., Kolle, D., and Loidl, P. (2001). Histone acetylation: lessons from the plant kingdom. Trends Plant Sci. 6, 59-65.

Makarevich, G., Leroy, O., Akinci, U., Schubert, D., Clarenz, O., Goodrich, J., Grossniklaus, U., and Kohler, C. (2006a). Different Polycomb group complexes regulate common target genes in *Arabidopsis*. EMBO Rep.

Makarevich, G., Leroy, O., Akinci, U., Schubert, D., Clarenz, O., Goodrich, J., Grossniklaus, U., and Kohler, C. (2006b). Different Polycomb group complexes regulate common target genes in *Arabidopsis*. EMBO Rep. 7, 947-952.

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

Manzur, K.L., Farooq, A., Zeng, L., Plotnikova, O., Koch, A.W., Sachchidanand, and Zhou, M.M. (2003). A dimeric viral SET domain methyltransferase specific to Lys27 of histone H3. Nat. Struct. Biol. 10, 187-196.

Marmorstein, R. (2003). Structure of SET domain proteins: a new twist on histone methylation. Trends. Biochem. Sci. 28, 59-62.

Martienssen, R.A., Zaratiegui, M., and Goto, D.B. (2005). RNA interference and heterochromatin in the fission yeast *Schizosaccharomyces pombe*. Trends Genet. 21, 450-456.

Matzke, M.A., Aufsatz, W., Kanno, T., Mette, M.F., and Matzke, A.J. (2002). Homology-dependent gene silencing and host defense in plants. Adv. Genet. 46, 235-275.

Mazo, A.M., Huang, D.H., Mozer, B.A., and Dawid, I.B. (1990). The *trithorax* gene, a trans-acting regulator of the bithorax complex in *Drosophila*, encodes a protein with zinc-binding domains. Proc. Natl. Acad. Sci. U.S.A. 87, 2112-2116.

McCallum, C.M., Comai, L., Greene, E.A., and Henikoff, S. (2000). Targeting induced local lesions IN genomes (TILLING) for plant functional genomics. Plant Physiol. 123, 439-442.

Mendrysa, S.M., McElwee, M.K., and Perry, M.E. (2001). Characterization of the 5' and 3' untranslated regions in murine *mdm2* mRNAs. Gene 264, 139-146.

Mette, M.F., Aufsatz, W., van Der Winden, J., Matzke, M.A., and Matzke, A.J. (2000). Transcriptional silencing and promoter methylation triggered by double-stranded RNA. EMBO J. 19, 5194-5201.

Miki, D., and Shimamoto, K. (2004). Simple RNAi vectors for stable and transient suppression of gene function in rice. Plant Cell Physiol. 45, 490-495.

Moore, T., Constancia, M., Zubair, M., Bailleul, B., Feil, R., Sasaki, H., and Reik, W. (1997). Multiple imprinted sense and antisense transcripts, differential methylation and tandem repeats in a putative imprinting control region upstream of mouse *Igf2*. Proc. Natl. Acad. Sci. U.S.A. 94, 12509-12514.

Morel, J.B., Mourrain, P., Beclin, C., and Vaucheret, H. (2000). DNA methylation and chromatin structure affect transcriptional and post-transcriptional transgene silencing in *Arabidopsis*. Curr. Biol. 10, 1591-1594.

Mourrain, P., van Blokland, R., Kooter, J.M., and Vaucheret, H. (2007). A single transgene locus triggers both transcriptional and post-transcriptional silencing through double-stranded RNA production. Planta 225, 365-379.

Mumberg, D., Lucibello, F.C., Schuermann, M., and Muller, R. (1991). Alternative splicing of *fosB* transcripts results in differentially expressed mRNAs encoding functionally antagonistic proteins. Genes Dev. 5, 1212-1223.

Murray, K. (1964). The occurrence of epsilon-N-methyl lysine in histones. Biochemistry 3, 10-15.

Mutskov, V., and Felsenfeld, G. (2004). Silencing of transgene transcription precedes methylation of promoter DNA and histone H3 lysine 9. EMBO J. 23, 138-149.

Naumann, K., Fischer, A., Hofmann, I., Krauss, V., Phalke, S., Irmler, K., Hause, G., Aurich, A.C., Dorn, R., Jenuwein, T., and Reuter, G. (2005). Pivotal role of AtSUVH2 in heterochromatic histone methylation and gene silencing in *Arabidopsis*. EMBO J. 24, 1418-1429.

Ng, D.W., Chandrasekharan, M.B., and Hall, T.C. (2006). Ordered histone modifications are associated with transcriptional poising and activation of the phaseolin promoter. Plant Cell 18, 119-132.

Ng, H.H., Ciccone, D.N., Morshead, K.B., Oettinger, M.A., and Struhl, K. (2003). Lysine-79 of histone H3 is hypomethylated at silenced loci in yeast and mammalian cells: a potential mechanism for position-effect variegation. Proc. Natl. Acad. Sci. U.S.A. 100, 1820-1825.

Paddison, P.J., Caudy, A.A., Sachidanandam, R., and Hannon, G.J. (2004). Short hairpin activated gene silencing in mammalian cells. Methods Mol. Biol. 265, 85-100.

Pal, S., Vishwanath, S.N., Erdjument-Bromage, H., Tempst, P., and Sif, S. (2004). Human SWI/SNF-associated PRMT5 methylates histone H3 arginine 8 and negatively

regulates expression of ST7 and NM23 tumor suppressor genes. Mol. Cell. Biol. 24, 9630-9645.

Perry, J., and Zhao, Y. (2003). The CW domain, a structural module shared amongst vertebrates, vertebrate-infecting parasites and higher plants. Trends Biochem. Sci. 28, 576-580.

Qian, C., Wang, X., Manzur, K., Sachchidanand, Farooq, A., Zeng, L., Wang, R., and Zhou, M.M. (2006). Structural insights of the specificity and catalysis of a viral histone H3 lysine 27 methyltransferase. J. Mol. Biol. 359, 86-96.

Raynaud, C., Sozzani, R., Glab, N., Domenichini, S., Perennes, C., Cella, R., Kondorosi, E., and Bergounioux, C. (2006). Two cell-cycle regulated SET-domain proteins interact with proliferating cell nuclear antigen (PCNA) in *Arabidopsis*. Plant J 47, 395-407.

Riha, K., McKnight, T.D., Griffing, L.R., and Shippen, D.E. (2001). Living with genome instability: plant responses to telomere dysfunction. Science 291, 1797-1800.

Rozovskaia, T., Rozenblatt-Rosen, O., Sedkov, Y., Burakov, D., Yano, T., Nakamura, T., Petruck, S., Ben-Simchon, L., Croce, C.M., Mazo, A., and Canaani, E. (2000). Self-association of the SET domains of human ALL-1 and of *Drosophila* TRITHORAX and ASH1 proteins. Oncogene 19, 351-357.

Sallaud, C., Meynard, D., van Boxtel, J., Gay, C., Bes, M., Brizard, J.P., Larmande, P., Ortega, D., Raynal, M., Portefaix, M., Ouwerkerk, P.B., Rueb, S., Delseny, M., and Guiderdoni, E. (2003). Highly efficient production and characterization of T-DNA plants for rice (*Oryza sativa L.*) functional genomics. Theor. Appl. Genet. 106, 1396-1408.

Sanders, P.M., Lee, P.Y., Biesgen, C., Boone, J.D., Beals, T.P., Weiler, E.W., and Goldberg, R.B. (2000). The *Arabidopsis DELAYED DEHISCENCE1* gene encodes an enzyme in the jasmonic acid synthesis pathway. Plant Cell 12, 1041-1061.

Sarraf, S.A., and Stancheva, I. (2004). Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. Mol. Cell 15, 595-605.

Schob, H., Kunz, C., and Meins, F., Jr. (1997). Silencing of transgenes introduced into leaves by agroinfiltration: a simple, rapid method for investigating sequence requirements for gene silencing. Mol .Gen. Genet. 256, 581-585.

Schotta, G., Lachner, M., Sarma, K., Ebert, A., Sengupta, R., Reuter, G., Reinberg, D., and Jenuwein, T. (2004). A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. Genes Dev. 18, 1251-1262.

Schubert, D., Primavesi, L., Bishopp, A., Roberts, G., Doonan, J., Jenuwein, T., and Goodrich, J. (2006). Silencing by plant Polycomb group genes requires dispersed trimethylation of histone H3 at lysine 27. EMBO J. 25, 4638-4649.

Scolnik, P.A., and Bartley, G.E. (1994). Nucleotide sequence of an *Arabidopsis* cDNA for phytoene synthase. Plant Physiol. 104, 1471-1472.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., and Casero, R.A. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell 119, 941-953.

Sijen, T., Vijn, I., Rebocho, A., van Blokland, R., Roelofs, D., Mol, J.N., and Kooter, J.M. (2001). Transcriptional and posttranscriptional gene silencing are mechanistically related. Curr. Biol. 11, 436-440.

Smith, N.A., Singh, S.P., Wang, M.B., Stoutjesdijk, P.A., Green, A.G., and Waterhouse, P.M. (2000). Total silencing by intron-spliced hairpin RNAs. Nature 407, 319-320.

Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98, 503-517.

Springer, N.M., Danilevskaya, O.N., Hermon, P., Helentjaris, T.G., Phillips, R.L., Kaeppler, H.F., and Kaeppler, S.M. (2002). Sequence relationships, conserved domains, and expression patterns for maize homologs of the polycomb group genes E(z), esc, and E(Pc). Plant Physiol 128, 1332-1345.

Springer, N.M., Napoli, C.A., Selinger, D.A., Pandey, R., Cone, K.C., Chandler, V.L., Kaeppler, H.F., and Kaeppler, S.M. (2003). Comparative analysis of SET domain proteins in maize and *Arabidopsis* reveals multiple duplications preceding the divergence of monocots and dicots. Plant Physiol 132, 907-925.

Stec, I., Wright, T.J., van Ommen, G.J., de Boer, P.A., van Haeringen, A., Moorman, A.F., Altherr, M.R., and den Dunnen, J.T. (1998). *WHSC1*, a 90 kb SET domain-containing gene, expressed in early development and homologous to a *Drosophila* dysmorphy gene maps in the Wolf-Hirschhorn syndrome critical region and is fused to *IgH* in t(4;14) multiple myeloma. Hum. Mol. Genet. 7, 1071-1082.

Stoutjesdijk, P.A., Singh, S.P., Liu, Q., Hurlstone, C.J., Waterhouse, P.A., and Green, A.G. (2002). hpRNA-mediated targeting of the *Arabidopsis FAD2* gene gives highly efficient and stable silencing. Plant Physiol. 129, 1723-1731.

Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature 403, 41-45.

Strahl, B.D., Briggs, S.D., Brame, C.J., Caldwell, J.A., Koh, S.S., Ma, H., Cook, R.G., Shabanowitz, J., Hunt, D.F., Stallcup, M.R., and Allis, C.D. (2001). Methylation of histone H4 at arginine 3 occurs in vivo and is mediated by the nuclear receptor coactivator PRMT1. Curr. Biol. 11, 996-1000.

Tamaru, H., and Selker, E.U. (2001). A histone H3 methyltransferase controls DNA methylation in *Neurospora crassa.* Nature 414, 277-283.

Tariq, M., Saze, H., Probst, A.V., Lichota, J., Habu, Y., and Paszkowski, J. (2003). Erasure of CpG methylation in *Arabidopsis* alters patterns of histone H3 methylation in heterochromatin. Proc. Natl. Acad. Sci. U.S.A. 100, 8823-8827.

Tian, L., and Chen, Z.J. (2001). Blocking histone deacetylation in *Arabidopsis* induces pleiotropic effects on plant gene regulation and development. Proc. Natl. Acad. Sci. U.S.A. 98, 200-205.

Tkachuk, D.C., Kohler, S., and Cleary, M.L. (1992). Involvement of a homolog of *Drosophila trithorax* by 11q23 chromosomal translocations in acute leukemias. Cell 71, 691-700.

Trewick, S.C., McLaughlin, P.J., and Allshire, R.C. (2005). Methylation: lost in hydroxylation? EMBO Rep. 6, 315-320.

Tschiersch, B., Hofmann, A., Krauss, V., Dorn, R., Korge, G., and Reuter, G. (1994). The protein encoded by the *Drosophila* position-effect variegation suppressor gene *Su(var)3-9* combines domains of antagonistic regulators of homeotic gene complexes. EMBO J. 13, 3822-3831.

Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M.E., Borchers, C.H., Tempst, P., and Zhang, Y. (2006). Histone demethylation by a family of JmjC domain-containing proteins. Nature 439, 811-816.

Vaistij, F.E., Jones, L., and Baulcombe, D.C. (2002). Spreading of RNA targeting and DNA methylation in RNA silencing requires transcription of the target gene and a putative RNA-dependent RNA polymerase. Plant Cell 14, 857-867.

Vielle-Calzada, J.P., Thomas, J., Spillane, C., Coluccio, A., Hoeppner, M.A., and Grossniklaus, U. (1999). Maintenance of genomic imprinting at the *Arabidopsis* medea locus requires zygotic DDM1 activity. Genes Dev. 13, 2971-2982.

Volkel, P., and Angrand, P.O. (2006). The control of histone lysine methylation in epigenetic regulation. Biochimie. 89, 1-20.

von Lintig, J., Welsch, R., Bonk, M., Giuliano, G., Batschauer, A., and Kleinig, H. (1997). Light-dependent regulation of carotenoid biosynthesis occurs at the level of phytoene synthase expression and is mediated by phytochrome in *Sinapis alba* and *Arabidopsis thaliana* seedlings. Plant J. 12, 625-634.

Wang, B.B., and Brendel, V. (2006). Genomewide comparative analysis of alternative splicing in plants. Proc. Natl. Acad. Sci. U.S.A. 103, 7175-7180.

Wang, J., Chang, Y.F., Hamilton, J.I., and Wilkinson, M.F. (2002). Nonsense-associated altered splicing: a frame-dependent response distinct from nonsense-mediated decay. Mol. Cell. 10, 951-957.

Wang, J., Zhou, Y., Yin, B., Du, G., Huang, X., Li, G., Shen, Y., Yuan, J., and Qiang, B. (2001). *ASH2L*: alternative splicing and downregulation during induced megakaryocytic differentiation of multipotential leukemia cell lines. J. Mol. Med. 79, 399-405.

Wang, T., Iyer, L.M., Pancholy, R., Shi, X., and Hall, T.C. (2005a). Assessment of penetrance and expressivity of RNAi-mediated silencing of the *Arabidopsis* phytoene desaturase gene. New Phytol. 167, 751-760.

Wang, X.J., Gaasterland, T., and Chua, N.H. (2005b). Genome-wide prediction and identification of cis-natural antisense transcripts in *Arabidopsis thaliana*. Genome Biol. 6, R30.

Waterhouse, P.M., and Helliwell, C.A. (2003). Exploring plant genomes by RNA-induced gene silencing. Nat. Rev. Genet. 4, 29-38.

Wesley, S.V., Helliwell, C.A., Smith, N.A., Wang, M.B., Rouse, D.T., Liu, Q., Gooding, P.S., Singh, S.P., Abbott, D., Stoutjesdijk, P.A., Robinson, S.P., Gleave, A.P., Green, A.G., and Waterhouse, P.M. (2001). Construct design for efficient, effective and high-throughput gene silencing in plants. Plant J. 27, 581-590.

Wetzel, C.M., and Rodermel, S.R. (1998). Regulation of phytoene desaturase expression is independent of leaf pigment content in *Arabidopsis thaliana*. Plant Mol. Biol. 37, 1045-1053.

Whetstine, J.R., Nottke, A., Lan, F., Huarte, M., Smolikov, S., Chen, Z., Spooner, E., Li, E., Zhang, G., Colaiacovo, M., and Shi, Y. (2006). Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. Cell 125, 467-481.

Xiao, B., Wilson, J.R., and Gamblin, S.J. (2003). SET domains and histone methylation. Curr. Opin. Struct. Biol. 13, 699-705.

Yamada, K., Lim, J., Dale, J.M., Chen, H., Shinn, P., Palm, C.J., Southwick, A.M., Wu, H.C., Kim, C., Nguyen, M., Pham, P., Cheuk, R., Karlin-Newmann, G., Liu, S.X., Lam, B., Sakano, H., Wu, T., Yu, G., Miranda, M., Quach, H.L., Tripp, M., Chang, C.H., Lee, J.M., Toriumi, M., Chan, M.M., Tang, C.C., Onodera, C.S., Deng, J.M., Akiyama, K., Ansari, Y., Arakawa, T., Banh, J., Banno, F., Bowser, L., Brooks, S., Carninci, P., Chao, Q., Choy, N., Enju, A., Goldsmith, A.D., Gurjal, M., Hansen, N.F., Hayashizaki, Y., Johnson-Hopson, C., Hsuan, V.W., Iida, K., Karnes, M., Khan, S., Koesema, E., Ishida, J., Jiang, P.X., Jones, T., Kawai, J., Kamiya, A., Meyers, C., Nakajima, M., Narusaka, M., Seki, M., Sakurai, T., Satou, M., Tamse, R., Vaysberg, M., Wallender, E.K., Wong, C., Yamamura, Y., Yuan, S., Shinozaki, K., Davis, R.W., Theologis, A., and Ecker, J.R. (2003). Empirical analysis of transcriptional activity in the *Arabidopsis* genome. Science 302, 842-846.

Yamada, T., Onimatsu, H., and Van Etten, J.L. (2006). *Chlorella* viruses. Adv. Virus Res. 66, 293-336.

Yang, G., Lee, Y.H., Jiang, Y., Kumpatla, S.P., and Hall, T.C. (2005). Organization, not duplication, triggers silencing in a complex transgene locus in rice. Plant Mol. Biol. 58, 351-366.

Yu, Y., Dong, A., and Shen, W.H. (2004). Molecular characterization of the tobacco SET domain protein NtSET1 unravels its role in histone methylation, chromatin binding, and segregation. Plant J 40, 699-711.

Zhang, W., Hayashizaki, Y., and Kone, B.C. (2004). Structure and regulation of the *mDot1* gene, a mouse histone H3 methyltransferase. Biochem. J. 377, 641-651.

Zhao, Z., Yu, Y., Meyer, D., Wu, C., and Shen, W.H. (2005). Prevention of early flowering by expression of FLOWERING LOCUS C requires methylation of histone H3 K36. Nat Cell Biol 7, 1256-1260.

Zhou, X., Carranco, R., Vitha, S., and Hall, T.C. (2005). The dark side of green fluorescent protein. New Phytol. 168, 313-322.

Zlotogora, J. (2003). Penetrance and expressivity in the molecular age. Genet. Med. 5, 347-352.

VITA

Name Tao Wang

Address C/O Dr. Timothy C. Hall, Department of Biology, Texas A&M University,, College Station, TX 77843

Email address twang@idmb.tamu.edu

Education B.S., Biology, University of Science and Technology of China, 1997

Ph.D., Biology, Texas A&M University, 2007