Identification and characterization of factors involved in DNA demethylation and anti-silencing in Arabidopsis

Zhaobo Lang

Purdue University

Follow this and additional works at: https://docs.lib.purdue.edu/open_access_dissertations

Part of the Biology Commons, Botany Commons, Genetics Commons, and the Horticulture Commons

Recommended Citation

https://docs.lib.purdue.edu/open_access_dissertations/315

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.
PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Zhaobo Lang

Entitled IDENTIFICATION AND CHARACTERIZATION OF FACTORS INVOLVED IN
DNA DEMETHYLATION AND ANTI-SILENCING IN ARABIDOPSIS

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

Jian-Kang Zhu

Ray A. Bressan

Stephen C. Weller

Larry Murdock

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication delay, and Certification/Disclaimer (Graduate School Form 32), this thesis/dissertation
Purdue University’s “Policy on Integrity in Research” and the use of copyrighted material.

Approved by Major Professor(s): Jian-Kang Zhu

Approved by: Robert J. Joly 10/10/2014

Head of the Graduate Program Date
IDENTIFICATION AND CHARACTERIZATION OF FACTORS INVOLVED IN DNA DEMETHYLATION AND ANTI-SILENCING IN *ARABIDOPSIS*

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Zhaobo Lang

In Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

December 2014

Purdue University

West Lafayette, Indiana
This work is dedicated to my father, Zhongwu Lang, my mother, Xiumei Li, and my beloved husband, Yinfeng Qin, for their continuous support, encouragement, and love.
ACKNOWLEDGMENTS

First and foremost, I am very grateful to my advisor, Prof. Jian-Kang Zhu, for the five exciting years working together. I thank him for both his guidance in scientific research and his generous sharing of life wisdom. It is extremely fortunate to be able to work with such a great scientist.

Secondly, I would like to thank my committee advisors: Prof. Larry Murdock, Prof. Steve Weller, and Prof. Ray Bressan, for all their insights and suggestions that inspired and improved my work.

My gratitude also goes to my most helpful colleagues: Huiming Zhang, Xingang Wang, Kai Tang, Bang-Shing Wang, and Heng Zhang. Their support is indispensable in my research progress.

Lastly, I would like to express my sincere gratitude to my parents in China, without whose love I would not have delivered this work. I would also like to thank my husband, Yinfeng Qin, for his support and understanding: we had to live in different cities for the past five years for pursuing my Ph.D. Knowing their being proud of me is the best reward for my years’ work.
TABLE OF CONTENTS

| LIST OF TABLES                          | vii    |
| LIST OF FIGURES                        | viii   |
| ABBREVIATIONS                          | xi     |
| GLOSSARY                               | xiv    |
| ABSTRACT                               | xv     |
| 1 Introduction                         | 1      |
| 1.1 DNA methylation mechanism          | 1      |
| 1.2 DNA demethylation mechanism        | 4      |
| 1.3 Genetic systems for studying DNA demethylation | 5 |
| 1.4 References                         | 6      |
| 2 A methyl-CpG-binding protein prevents the spread of DNA methylation and transcriptional gene silencing by recruiting factors for active DNA demethylation | 11    |
| 2.1 Abstract                           | 11     |
| 2.2 Introduction                       | 12     |
| 2.3 Results                            | 14     |
| 2.3.1 Mutation in MBD7 causes transgene silencing | 14 |
| 2.3.2 DNA hypermethylation and repressive histone modification accompany the silencing of transgenes | 16 |
| 2.3.3 The mbd7 mutation causes genome-wide DNA hypermethylation | 17 |
| 2.3.4 MBD7-binding positively correlates with mCG methylation density | 18 |
| 2.3.5 MBD7 binding is associated with the presence of mCHG and mCHH | 19 |
| 2.3.6 MBD7 binding influences DNA methylation level | 20 |
2.3.7 MBD7 interacts with other anti-silencing factors ........ 21
2.3.8 MBD7 affects a subset of genomic regions targeted by active DNA demethylation .................................................. 22
2.3.9 MBD7 promotes the expression of endogenous loci ........ 24
2.4 Discussion ................................................. 24
2.5 Material and methods ...................................... 27
  2.5.1 Plant materials and growth conditions ................. 27
  2.5.2 Mutant screening ....................................... 28
  2.5.3 Map-based cloning ..................................... 28
  2.5.4 Complementation experiments .......................... 29
  2.5.5 Northern blot, RT-PCR, and RT-qPCR ................. 29
  2.5.6 Whole-genome bisulfite sequencing ..................... 30
  2.5.7 Chop-PCR ............................................. 30
  2.5.8 Chromatin immune-precipitation (ChIP) assay .......... 30
  2.5.9 ChIP-seq ............................................ 31
  2.5.10 RNA-seq ........................................... 31
  2.5.11 MethylC-Seq data analysis ............................. 31
  2.5.12 MBD7 ChIP-Seq Analysis .............................. 32
2.6 References ................................................. 33
2.7 Figures and tables ........................................ 38
3 Identification of a molecular regulator involved in the active DNA demethylation pathway based on genetic screening in Arabidopsis ............ 92
  3.1 Abstract .................................................. 92
  3.2 Introduction ............................................. 92
  3.3 Results ................................................... 94
    3.3.1 Isolation of ros3 enhancer mutants .................... 94
    3.3.2 rte1 is a DNA hypermethylated mutant ................ 95
    3.3.3 Isolation of the rte1 single mutant ................... 96
    3.3.4 The methylome of the rte1 single mutant ............. 97
3.3.5 Genetic mapping of *rte1* ................................................... 98
3.4 Discussion .................................................................................. 98
3.5 Material and methods ................................................................. 100
  3.5.1 Plant materials and growth conditions ..................................... 100
  3.5.2 Mutant screen ......................................................................... 101
  3.5.3 Real-Time PCR ........................................................................ 101
  3.5.4 Chop-PCR .............................................................................. 102
  3.5.5 Individual-locus bisulfite sequencing ...................................... 102
  3.5.6 Whole-genome bisulfite sequencing ........................................ 102
3.6 References .................................................................................... 103
3.7 Figures and tables ......................................................................... 103
4 Summary ......................................................................................... 115
VITA ................................................................................................... 118
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Results of immunoprecipitation against MBD7-4xMYC followed by LC-MS.</td>
<td>91</td>
</tr>
<tr>
<td>3.1 Rough mapping of the \textit{rte1} mutant.</td>
<td>104</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Isolation of the 91-1 mutant based on its long root phenotype.</td>
<td>38</td>
</tr>
<tr>
<td>2.2</td>
<td>Hygromycin and kanamycin sensitivity of 91-1 mutants.</td>
<td>39</td>
</tr>
<tr>
<td>2.3</td>
<td>Expression level of transgenes in 91-1 mutant plants.</td>
<td>40</td>
</tr>
<tr>
<td>2.4</td>
<td>Identification of the mutation in the 91-1 mutant that is responsible for transgene silencing.</td>
<td>41</td>
</tr>
<tr>
<td>2.5</td>
<td>Genotyping of mbd7-2 containing 35S::SUC2 transgene.</td>
<td>42</td>
</tr>
<tr>
<td>2.6</td>
<td>Root phenotype of mbd7-2.</td>
<td>43</td>
</tr>
<tr>
<td>2.7</td>
<td>Functional complementation of the mbd7-1 mutant.</td>
<td>44</td>
</tr>
<tr>
<td>2.8</td>
<td>Determination of Pol II occupancy at 35S promoters.</td>
<td>45</td>
</tr>
<tr>
<td>2.9</td>
<td>Determination of H3K9me2 level at the 35S promoter.</td>
<td>46</td>
</tr>
<tr>
<td>2.10</td>
<td>Determination of H3K18ac level at the 35S promoter.</td>
<td>47</td>
</tr>
<tr>
<td>2.11</td>
<td>Determination of H3K23ac level at the 35S promoter.</td>
<td>48</td>
</tr>
<tr>
<td>2.12</td>
<td>DNA methylation status of the 35S::SUC2 transgene in the wild type and mbd7-1.</td>
<td>49</td>
</tr>
<tr>
<td>2.13</td>
<td>Hyper- and hypo-DMRs (differently methylated regions) in mbd7-1.</td>
<td>50</td>
</tr>
<tr>
<td>2.14</td>
<td>Determination of DNA hypermethylated loci in mbd7-1 using Chop-PCR.</td>
<td>51</td>
</tr>
<tr>
<td>2.15</td>
<td>Examples of hypermethylated regions in mbd7-1.</td>
<td>52</td>
</tr>
<tr>
<td>2.16</td>
<td>Composition of hyper-DMRs in mbd7-1.</td>
<td>53</td>
</tr>
<tr>
<td>2.17</td>
<td>Average methylation levels in hyper-DMR-associated TE bodies and flanking 2-kb regions.</td>
<td>54</td>
</tr>
<tr>
<td>2.18</td>
<td>Compositions of total TEs and TEs with hyper mCHH in mbd7-1.</td>
<td>55</td>
</tr>
<tr>
<td>2.19</td>
<td>Average methylation level of hyper-DMR-associated TEs with different lengths.</td>
<td>56</td>
</tr>
<tr>
<td>2.20</td>
<td>Chromosomal distribution of hyper-DMRs in mbd7-1.</td>
<td>57</td>
</tr>
<tr>
<td>2.21</td>
<td>Position preference of hyper-DMR-associated TEs in mbd7.</td>
<td>58</td>
</tr>
</tbody>
</table>
Figure | Page
--- | ---
2.22 Recombinant MBD7 protein expression in *MBD7-4xMYC/mbd7-1.* | 59
2.23 Correlation of MBD7 enrichment with DNA methylation level. | 60
2.24 Correlation of MBD7 enrichment with DNA methylation density. | 61
2.25 Genome-wide relationship between MBD7 enrichment and DNA methylation densities. | 62
2.26 Screenshots showing endogenous MBD7-binding targets. | 63
2.27 Composition of endogenous MBD7 binding targets based on analysis of 2452 peaks of ChIP-seq signals. | 64
2.28 Comparison of mCG methylation density between genes and TE s. | 65
2.29 Comparison of MBD7 enrichment between 1-kb gene and TE windows with similar mCG methylation densities. | 66
2.30 Comparison of MBD7 enrichment between 1-kb gene and TE windows with similar mCG methylation density. | 67
2.31 Comparison of MBD7 enrichment between windows with and without non-CG methylation. | 68
2.32 Comparison of MBD7 enrichment between windows with and without non-CG methylation. | 69
2.33 Correlation of MBD7 enrichment with DNA hypermethylation in *mbd7-1* mutant plants. | 70
2.34 Correlation of MBD7 enrichment with hyper mCG, mCHG, and mCHH in *mbd7-1* mutant plants. | 71
2.35 The effect of MBD7 binding on DNA methylation. | 72
2.36 MBD7 binding at the 35S transgene promoter. | 73
2.37 Examination of MBD7 enrichment at transgene promoter using individual locus ChIP assay. | 74
2.38 Screenshots showing that DNA methylation level is increased in MBD7-binding regions. | 75
2.39 MBD7 physically interacts with IDM2 and IDM3. | 76
2.40 Interaction of MBD7 with IDM2 and IDM3 using split luciferase transient expression assays. | 77
2.41 Diagram showing structures of full MBD7 and truncated MBD7 used in Y2H assay. | 78
<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.42 Y2H assay testing interactions between IDM2 and truncated MBD7 proteins.</td>
<td>79</td>
</tr>
<tr>
<td>2.43 Y2H assay testing protein interactions between IDM3 and truncated MBD7 proteins.</td>
<td>80</td>
</tr>
<tr>
<td>2.44 Complementation of the mbd7-1 mutant with a truncated MBD7 construct.</td>
<td>81</td>
</tr>
<tr>
<td>2.45 Levels of DNA methylation and MBD7 chromatin enrichment at shared and not shared hyper-DMRs in the mbd7-2, idm1-1, idm2-1 and ros1-4 mutants.</td>
<td>82</td>
</tr>
<tr>
<td>2.46 Overlapping between mbd7-2 DMRs and DMRs in idm1-1 or idm2-1.</td>
<td>83</td>
</tr>
<tr>
<td>2.47 Box plots displaying mCG density (Y-axis) in different groups of DMRs (X-axis).</td>
<td>84</td>
</tr>
<tr>
<td>2.48 Box plots displaying MBD7 chromatin enrichment (Y-axis) in different groups of DMRs (X-axis).</td>
<td>85</td>
</tr>
<tr>
<td>2.49 Heat-induced expression of ONSEN genes in mbd7 mutants.</td>
<td>86</td>
</tr>
<tr>
<td>2.50 Heat-induced expression of TSI genes in mbd7 mutants.</td>
<td>87</td>
</tr>
<tr>
<td>2.51 Methylation status and MBD7 enrichment of the ONSEN gene.</td>
<td>88</td>
</tr>
<tr>
<td>2.52 Methylation status and MBD7 enrichment of the TSI gene.</td>
<td>89</td>
</tr>
<tr>
<td>2.53 Working model of MBD7 function in anti-silencing.</td>
<td>90</td>
</tr>
<tr>
<td>3.1 Luminescence phenotype of the rte1 mutant.</td>
<td>105</td>
</tr>
<tr>
<td>3.2 Silencing of RD29A promoter-driven transgenic and endogenous genes in the rte1 mutant.</td>
<td>106</td>
</tr>
<tr>
<td>3.3 Silencing of the 35S promoter-driven NPTII transgene in the rte1 mutant.</td>
<td>107</td>
</tr>
<tr>
<td>3.4 Examination of the rte1 methylation status using Chop-PCR.</td>
<td>108</td>
</tr>
<tr>
<td>3.5 Methylation level of the RD29A promoter region in the rte1 mutant.</td>
<td>109</td>
</tr>
<tr>
<td>3.6 Isolation of rte1 single mutants.</td>
<td>110</td>
</tr>
<tr>
<td>3.7 Kanamycin-resistance phenotype of the rte1 single mutant.</td>
<td>111</td>
</tr>
<tr>
<td>3.8 Methylation status of the RD29A promoter in the rte1 single mutant #12.</td>
<td>112</td>
</tr>
<tr>
<td>3.9 Genome-wide DNA methylation status in the rte1 single mutant #12.</td>
<td>113</td>
</tr>
<tr>
<td>3.10 Endogenous hypermethylated regions in the rte1 single mutant #12.</td>
<td>114</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>5mC</td>
<td>5-methylcytosine</td>
</tr>
<tr>
<td>RdDM</td>
<td>RNA-directed DNA methylation</td>
</tr>
<tr>
<td>Pol IV</td>
<td>DNA-dependent RNA polymerase IV</td>
</tr>
<tr>
<td>DTF1</td>
<td>DNA-binding transcription factor</td>
</tr>
<tr>
<td>SHH1</td>
<td>Sawadee homeodomain homolog 1</td>
</tr>
<tr>
<td>CLSY1</td>
<td>CLASSY 1</td>
</tr>
<tr>
<td>RDR2</td>
<td>RNA-dependent RNA polymerase II</td>
</tr>
<tr>
<td>DCL3</td>
<td>Dicer-like 3</td>
</tr>
<tr>
<td>HEN1</td>
<td>Hua enhancer 1</td>
</tr>
<tr>
<td>DRD1</td>
<td>defective in RNA-directed DNA methylation 1</td>
</tr>
<tr>
<td>DMS3</td>
<td>defective in meristem silencing 3</td>
</tr>
<tr>
<td>RDM1</td>
<td>RNA-directed DNA methylation 1</td>
</tr>
<tr>
<td>KTF1</td>
<td>KOW domain-containing transcription factor 1</td>
</tr>
<tr>
<td>ZOP1</td>
<td>zinc finger and OCRE domain containing Protein 1</td>
</tr>
<tr>
<td>MET1</td>
<td>DNA methyltransferase I</td>
</tr>
<tr>
<td>VIM1</td>
<td>Variant in methylation 1</td>
</tr>
<tr>
<td>DDM1</td>
<td>decrease in DNA methylation 1</td>
</tr>
<tr>
<td>CMT3</td>
<td>Chromomethylase 3</td>
</tr>
<tr>
<td>KYP</td>
<td>KRYPTONITE</td>
</tr>
<tr>
<td>H3K9</td>
<td>histone H3 lysine 9</td>
</tr>
<tr>
<td>H3K18</td>
<td>histone H3 lysine 18</td>
</tr>
<tr>
<td>H3K23</td>
<td>histone H3 lysine 23</td>
</tr>
<tr>
<td>IBM1</td>
<td>Increase in Bonsai Methylation 1</td>
</tr>
<tr>
<td>CMT2</td>
<td>Chromomethylase 2</td>
</tr>
</tbody>
</table>
ROS1/DML1  Repressor of silencer 1/Demeter like 1
DME      Demeter
DML2     Demeter like 2
DML3     Demeter like 3
ZDP      Arabidopsis thaliana zinc finger DNA 3’ phosphoesterase
ROS3     Repressor of Silencing 3
WT       wild-type
IDM1     increased DNA methylation 1
IDM2     increased DNA methylation 2
IDM3     increased DNA methylation 3
SUC2     sucrose transporter 2
MBD7     methyl-CpG binding protein 7
MBD      Methyl-CpG- Binding Domain
TE       transposable elements
PRMT11   Protein arginine methyltransferase 11
FT       flowering locus T
SOC1     suppressor of constant 1
FLC      flowering locus C
StkC     sticky-c
HDA6     histone deacetyltransferase 6
SSLP     simple sequence length polymorphisms
SUC2     sucrose transporter 2
DMR      differentially methylated regions
DMC      differentially methylated cytosine
HMC      hypermethylated cytosines
IP       immuno-precipitation
LC-MS    liquid chromatography-mass spectrometry
Y2H      Yeast two-hybrid
EMS      ethyl methanesulfonate
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCD</td>
<td>Charged couple device</td>
</tr>
<tr>
<td>RTE1</td>
<td>Ros Three Enhancer mutant 1</td>
</tr>
<tr>
<td>Leu</td>
<td>leucine</td>
</tr>
<tr>
<td>Trp</td>
<td>tryptophan</td>
</tr>
<tr>
<td>His</td>
<td>histidine</td>
</tr>
<tr>
<td>Ade</td>
<td>adenine</td>
</tr>
</tbody>
</table>
GLOSSARY

mC  methylation of 5-position of cytosine
mCG  5-methylcytosine in CpG
mCHG  5-methylcytosine in CHG (H=A, C, or T)
mCHH  5-methylcytosine in CHH (H=A, C, or T)
H3K9me2  histone H3 lysine 9 dimethylation
H3K18ac  histone H3 lysine 18 acetylation
H3K23ac  histone H3 lysine 23 acetylation
ChIP-seq  chromatin immunoprecipitation-sequencing
5-aza  5-Aza-2’-deoxycytidine, a DNA methylation inhibitor
NPTII  gene  kanamycin resistance gene
HPTII  gene  hygromycin resistance gene
minus-three media  yeast media lacking leucine, tryptophan, and histidine
minus-four media  yeast media lacking leucine, tryptophan, histidine and adenine
HpaII and HaeIII  restriction enzymes, enzyme activity can be blocked by CpG methylation at cut sites.
ABSTRACT

Lang, Zhaobo Ph.D., Purdue University, December 2014. Identification and characterization of factors involved in DNA demethylation and anti-silencing in *Arabidopsis*. Major Professor: Jian-Kang Zhu Professor.

DNA methylation is a key epigenetic mark for transcriptional gene silencing in many eukaryotes. DNA methylation status can be dynamically controlled by methylation and active demethylation processes. Compared to the well-known mechanisms of DNA methylation, the mechanisms of DNA demethylation and its regulation are poorly understood. In order to better understand the DNA demethylation pathway, we developed two genetic screening systems in *Arabidopsis* to identify new components involved in the DNA demethylation. In the first system, which is based on 35S promoter driven *SUC2* (sucrose transporter 2) transgene, a mutant (91-1) was isolated and map-based cloning identified the responsible gene as *MBD7* (methyl-CpG binding protein 7). As important interpreters of DNA methylation, Methyl-CpG-Binding Domain (MBD) proteins are known to be involved in gene silencing through recruitment of transcriptional repressors and protein factors for heterochromatin formation. Our data show that Arabidopsis MBD7 functions as a cellular anti-silencing factor. MBD7 is required for the expression of transgenes by preventing DNA hypermethylation at the transgene promoter. Analysis of the methylome of *mbd7* mutant plants revealed that MBD7 prevents DNA hypermethylation at thousands of genomic loci. MBD7 have been characterized previously as methyl-CpG-binding proteins *in vitro*, but its *in vivo* binding targets are not known. We performed chromatin immunoprecipitation-sequencing (ChIP-seq) experiments to characterize the genome-wide binding targets of MBD7 in plants. We found that MBD7 preferentially binds to loci with dense and high levels of mCG. This binding is associated with the role
of MBD7 in preventing the spread of DNA methylation. Protein interaction assays revealed that MBD7 is physically associated with other anti-silencing factors such as IDM1 (Increased DNA methylation 1), IDM2, and IDM3. In fact, MBD7 interacts directly with the alpha-crystallin domain proteins IDM2 and IDM3. IDM1 (a histone H3 acetyltransferase) and IDM2 are known to function as regulatory factors for ROS1, a methylcytosine DNA glycosylase/lyase critical for active DNA demethylation. Our results suggest that MBD7 binds to genomic regions with dense and highly methylated CGs and prevents the spread of DNA methylation by recruiting other anti-silencing factors such as regulators and enzymes for active DNA demethylation. The second system is based on the \textit{RD29A} promoter-driven \textit{LUC} (Luciferase) transgene. Previous studies in our lab have identified two important mutants from this system, including \textit{ros} (repressor of silencing) 1 and \textit{ros3}. ROS1 is the first DNA demethylase identified, and ROS3 is a regulatory component in the active DNA demethylation pathway. In the current study, we carried out an enhancer mutant screen in the \textit{ros3} background and isolated and characterized a mutant, \textit{rte1} (ros three enhancer 1). Our results suggest that RTE1 is a novel factor involved in the regulation of active DNA demethylation.
1. INTRODUCTION

1.1 DNA methylation mechanism

Epigenetics refers to any changes in chromatin status that leads to a modification in gene expression without alteration in DNA sequence [1]. As one of the most important epigenetic marks, DNA methylation is a biochemical process that involves the addition of a methyl group to the 5th position of the cytosine pyrimidine ring, forming 5-methylcytosine (5mC) [2]. DNA methylation is often considered a repressive epigenetic mark [2]. In mammals and plants, epigenetic regulation is essential for many biological processes, such as gene silencing, imprinting, and X chromosome inactivation [2]. The investigation of epigenetic modification also has important applied value. In humans, for instance, aberrant DNA methylation occurs in tumor cells, and DNA methylation markers have been developed for detecting cancer and for monitoring the development of cancer [3,4]. Moreover, the reversible modification of DNA methylation provides an opportunity for epigenetic cancer therapy [5,6]. In plants, epigenetic modifications play important roles in development, stress response, polyploidy, and hybrid vigor [7,8]. The manipulation of epigenetic modification may enable researchers to improve plant productivity and adaptability to the environment [7–9].

Based on DNA sequence contexts, methylation of cytosine can be classified into mCG, mCHG, and mCHH (H is C, A, or T). Although mammalian DNA methylation predominantly occurs in CG context, recent studies have identified some non-CG methylation in stem cells and mouse brains [10–12]. In contrast, plant cytosine methylation occurs in all three DNA sequence contexts. In Arabidopsis thaliana, the most abundant type of methylation is mCG, which represents 55% of the total methylated cytosines, while CHG and CHH represent 23 and 22%, respectively [13]. Like CHG
and CHH methylation, CG methylation preferentially targets transposons and repetitive DNA sequences, although CG methylation is also found to be enriched in some gene coding regions [13, 14].

In plants, the DNA methylation status is dynamically controlled by DNA methylation and DNA demethylation processes [15]. DNA methylation occurs through two processes: the establishment of DNA methylation and the maintenance of DNA methylation. The establishment of DNA methylation is mediated by small interfering RNAs (siRNAs), and the process is designated as RNA-directed DNA methylation (RdDM) [16]. In the RdDM pathway, small interfering RNAs guide de novo DNA methylation to complementary genetic loci. Biogenesis of most siRNAs depends on the plant-specific DNA-dependent RNA polymerase IV (Pol IV) [16]. DTF1 (DNA-binding transcription factor) / SHH1 (Sawadee homeodomain homolog 1) interacts with Pol IV in vivo [17, 18], and the abolishment of Pol IV occupancy in the dtf1 mutant indicated a role of DTF1 in targeting Pol IV [18]. Another chromatin remodeling protein, CLSY1 (CLASSY I), functions with Pol IV and DTF1, and is also required for siRNA biogenesis [18, 19]. Then, RDR2 (RNA-dependent RNA polymerase II) synthesizes double-stranded RNAs, which are cleaved into 24-nt siRNAs by DCL3 (Dicer-like 3) [20]. After modification by HEN1 (HUA ENHANCER 1), the siRNA turns into mature siRNA [21]. Mature siRNAs can be loaded into AGO4 and its closely related agonaute proteins, forming agonaute-siRNA complexes, which are later recruited to RdDM loci through base pairing between siRNAs and non-coding scaffold RNA transcripts. Recent studies showed that Pol II and another plant-specific DNA-dependent RNA polymerase, Pol V, can generate long, non-coding RNA transcripts at RdDM target loci, which may function as scaffolds for recruiting complementary siRNAs [22, 23]. A protein complex comprised of DRD1 (defective in RNA-directed DNA methylation 1), DMS3 (defective in meristem silencing 3), and RDM1 (RNA-directed DNA methylation 1) is involved in the production of Pol V-dependent non-coding RNA transcripts [16, 24]. After siRNA biogenesis and siRNA-scaffold RNA pairing, DRM2 is recruited to catalyze de novo DNA methylation. KTF1 (KOW domain-
containing transcription factor 1) is capable of binding both the scaffold RNAs and AGO4, thereby reinforcing the recruitment of the AGO-siRNA complexes to RdDM target loci \[25\]. The pre-mRNA splicing factor ZOP1 (zinc finger and OCRE domain containing Protein 1) is required for silencing some of the RdDM target loci and for accumulating some Pol IV-dependent 24-nt siRNAs, although ZOP1 is not directly involved in siRNA biogenesis. Analysis of siRNA accumulation in the \(zop1\) mutant suggested that ZOP1 might function downstream of the siRNA biogenesis step in the RdDM pathway \[26\].

Once established, the DNA methylation pattern must be maintained after DNA replication during cell division. The methylation in CG and CHG can be maintained because CG and CHG sequences are symmetrical. The methylation of CHH, however, cannot be maintained and must be regenerated after each DNA replication because the CHH sequence is asymmetrical. In plants, the maintenance methylation at CG context is catalyzed by DNA METHYLTRANSFERASE I (MET1) \[27\]. After a round of DNA replication, one fully methylated CpG site becomes two hemimethylated CpG sites, which can serve as substrate for MET1. Rather than fully- or un-methylated CpG, the hemimethylated CpG site is recognized and bound by VIM1 (VARIANT IN METHYLATION 1), and VIM1 has been shown to be essential for maintaining CpG methylation \textit{in vivo} \[28,29\]. CG maintenance methylation is also influenced by DDM1 (decrease in DNA methylation 1), a chromatin remodeling protein, and the absence of DDM1 results in hypomethylation not only in CG context but also in CHG and CHH contexts \[30,31\]. DDM1 is thought to provide an environment that makes chromatin accessible to DNA methyltransferases \[31\].

Maintenance of CHG methylation is catalyzed by CMT3 (CHROMOMETHYLASE 3) \[32\]. According to recent research, two other proteins regulate the methylation at CHG sites through their histone-modification activities. One is KRYPTONITE (KYP/SUVH4), a histone H3 lysine 9 (H3K9) methyltransferase. CHG methylation by CMT3 relies on the physical interaction between CMT3 and H3K9me2, and CHG methylation is decreased in the \(kyp\) mutant \[33\]. The other protein known
to affect CHG methylation levels is the jmjC domain-containing protein IBM1 (Increase in Bonsai Methylation 1). IBM1 can negatively regulate CHG methylation through its H3K9 demethylase activity [34].

As noted, the DNA methylation in the CHH context cannot be maintained because of its asymmetric nature [2]. To date, two pathways responsible for CHH methylation have been identified: the RNA-dependent-DNA methylation (RdDM) pathway and the RdDM-independent pathway. In contrast to RdDM, which is guided by siRNAs and catalyzed by DRM2, RdDM-independent CHH methylation is guided by specific chromatin environment and is catalyzed by CMT2 (CHROMOMETHYLASE 2) [31]. The recruitment of CMT2 to its target loci depends on the chromatin environment provided by DDM1. This RdDM-independent CMT2-mediated CHH methylation pathway is thought to account for most instances of CHH methylation [31].

1.2 DNA demethylation mechanism

The DNA methylation status is dynamically regulated during biological processes [2] and in addition to being established and maintained, it can also be removed. In contrast to the well-established DNA methylation pathway, the mechanism of DNA demethylation is less understood. DNA demethylation can be either passive, active, or a combination of both. Passive DNA demethylation is caused by inhibition or lack of DNA maintenance activities during cycles of replication, whereas active DNA demethylation requires specific enzymatic reactions [2]. A class of glycosylase/lyases has been identified as DNA demethylases, and these include ROS1/DML1 (REPRESSOR OF SILENCER 1/DEMETER LIKE 1), DME (DEMETER), DML2 (DEMETER LIKE 2), and DML3 (DEMETER LIKE 3). Based on the activities of these enzymes, researchers initially inferred that active DNA demethylation relies on the base excision repair mechanism [35]. The first DNA demethylase identified was ROS1 [36]. Mutations in ROS1 caused transcriptional silencing of a transgene and its homologous endogenous gene. Biochemical evidence was later presented that
the 5-methylcytosine, instead of only the methyl group, was removed from DNA through the glycosylase activity of ROS1 [37]. Three other enzymes (DME, DML2, and DML3) in the same family as ROS1 were determined to be also involved in DNA demethylation [2]. In particular, DME plays an important role in gene imprinting in plants. Expression of DME in the central cell before double fertilization is required for maternal allele demethylation and gene imprinting in the endosperm [38,39]. A recent study found that DME is expressed in the vegetative cell but not in sperm cells in the male gametophyte and that DME is important for male fertility [40]. After methylcytosine is removed from DNA and the lyase activity of ROS1 cleaves the DNA background at the abasic site, the single-nucleotide gap must be filled with an unmethylated cytosine. ZDP (Arabidopsis thaliana zinc finger DNA 3’ phosphoesterase), a DNA 3’ phosphatase, removes the phosphate group from the 3’-end of the gap, resulting in a 3’-hydroxyl terminus that is necessary for the subsequent DNA repair process [41]. In addition to these enzymes, ROS3 (Repressor of Silencing 3) was identified as an essential regulator of DNA demethylation [42]. ROS3 contains an RNA-recognition motif and is a small RNA-binding protein. Co-localization data and analysis of zdp/rosl and rosl/ros3 double mutants suggested that ZDP, ROS1, and ROS3 might function in the same genetic pathway [41,42].

1.3 Genetic systems for studying DNA demethylation

Compared to the well-known mechanisms of DNA methylation and its regulation, the mechanisms of DNA demethylation and its regulation are poorly understood and require more research. To identify new components involved in the DNA demethylation pathway, we developed two genetic screening systems in Arabidopsis.

The first system was developed using the Col-0 ecotype of Arabidopsis. Three transgenes exist in the wild-type (WT) transgenic plants, including the cauliflower mosaic virus 35S promoter-driven sucrose transporter 2 (SUC2), the double 35S promoter-driven neomycin phosphotransferase II (NPTII), and the double 35S promoter-
driven hygromycin phosphotransferase II (HPTII). The transgenic plants are very sensitive to sucrose and show a short-root phenotype on sucrose medium due to the over expression of the SUC2 transgene. The expression of the SUC2 transgene can be silenced by DNA hypermethylation at the 35S promoter region [43,44]. To identify factors involved in DNA demethylation, we mutagenized WT transgenic plants with EMS and screened for mutants in the M2 population with a long-root phenotype, which might be caused by silencing of the SUC2 transgene. In this study, a mutant (91-1) was isolated that displayed the long-root phenotype on sucrose growth medium. Through genetic mapping, this mutant was identified as a new recessive allele of MBD7 (methyl-CpG binding protein 7) and will be described in Chapter 2.

The second genetic screening system was developed using the C24 ecotype of Arabidopsis. Two transgenes were introduced into C24 plants: the RD29A promoter-driven firefly luciferase reporter gene (proRD29A::LUC), and the 35S promoter-driven NPTII gene (35S::NPTII). The stress-responsive RD29A promoter can be induced by salt, cold, or ABA treatments. This transgenic line has been very stable for many generations over the last 17 years. This system has been used to screen for mutants with higher or lower levels of stress-induced expression of the proRD29A-LUC transgene. ros1 and ros3 mutants were both screened from this system [36,42].

In the current project, we initiated a screen for genetic enhancers in the ros3 mutant background. Based on luminescence phenotype, a ros3 enhancer mutant, rte1 (ros three enhancer 1), was identified and characterized as described in Chapter 3.

1.4 References


[4] Ewa Brzezińska, Agata Dutkowska, and Adam Antczak. The significance of epi-


[8] Xiangsong Chen and Dao-Xiu Zhou. Rice epigenomics and epigenetics: chal-


[11] Wei Xie, Cathy L Barr, Audrey Kim, Feng Yue, Ah Young Lee, James Eu-


2. A METHYL-CPG-BINDING PROTEIN PREVENTS THE SPREAD OF DNA METHYLATION AND TRANSCRIPTIONAL GENE SILENCING BY RECRUITING FACTORS FOR ACTIVE DNA DEMETHYLATION

2.1 Abstract

DNA methylation is a key epigenetic mark for transcriptional gene silencing in many eukaryotes. As important interpreters of DNA methylation, Methyl-CpG-Binding Domain (MBD) proteins are known to be involved in gene silencing through recruitment of transcriptional repressors and protein factors for heterochromatin formation. Here, we report the identification of *Arabidopsis* MBD7 as a cellular anti-silencing factor. MBD7 is required for the expression of transgenes by preventing DNA hypermethylation at the transgene promoter. Analysis of the methylome of *mbd7* mutant plants revealed that MBD7 prevents DNA hypermethylation at thousands of transposable elements (TEs) and other genomic loci. In *mbd7* mutant plants, hypermethylation occurs mainly in CHG and CHH sequence contexts (H=A, C, or T). Among the 13 members of the MBD family of proteins in *Arabidopsis*, MBD5, MBD6, and MBD7 have been characterized previously as methyl-CpG-binding proteins *in vitro*, but their *in vivo* binding targets are not known. We performed chromatin immunoprecipitation-sequencing (ChIP-seq) experiments to characterize the genome-wide binding targets of MBD7 in plants. We found that MBD7 preferentially binds to loci with dense and high levels of mCG. This binding is associated with the role of MBD7 in preventing the spread of DNA methylation. Protein interaction assays revealed that MBD7 is physically associated with other anti-silencing factors such as IDM1, IDM2, and IDM3. In fact, MBD7 interacts directly with the alpha-
crystallin domain proteins IDM2 and IDM3. IDM1 (a histone H3 acetyltransferase) and IDM2 are known to function as regulatory factors for ROS1, a methylcytosine DNA glycosylase/lyase critical for active DNA demethylation. Our results suggest that MBD7 binds to genomic regions with dense and highly methylated CGs and prevents the spread of DNA methylation by recruiting other anti-silencing factors such as regulators and enzymes for active DNA demethylation.

2.2 Introduction

In mammals and plants, DNA methylation is important for many epigenetic processes, such as gene silencing, imprinting, and X chromosome inactivation [1]. DNA methylation occurs in three contexts: CG, CHG, and CHH (where H is A, C, or T) in plants [2]. In addition to DNA methylation, histone modifications such as ubiquitination, methylation, acetylation, sumoylation, and phosphorylation are also important epigenetic marks [3,4]. Histone modifications may create a chromatin environment that is more or less accessible for transcription. Methylation and acetylation of lysines in histones are two well-characterized modifications. Previous studies showed that lysine acetylation in histones H3 and H4 correlates with active transcription [4–7]. Researchers have identified some histone acetyltransferases and deacetylases that catalyze this reversible modification. Identified histone acetyltransferases included HAG1, HAG3, HAM1, HAC1, and TAF1 [8]. Histone deactylases (HDACs) include three types of enzymes: RPD3-like superfamily HDACs, HD2-like HDACs, and SIR2 family HDACs [8–10]. IDM1/ROS4, a recently identified histone acetyltransferase, specifically modifies H3K18 and K3K23 in Arabidopsis. In the idm1 mutant, a transgene is silenced because of a lack of H3K18ac and H3K23ac in the transgene promoter regions, which leads to DNA hypermethylation of the promoter [7,11]. In contrast to histone acetylation, methylation of histones can either activate or repress transcription. Methylation of H3K4, H3K36, and H3K79 are epigenetic marks for active genes, whereas methylation of H3K9, H3K27, and H3K20 are marks for inactive genes [3,12].
DNA methylation and histone modifications are interconnected. For instance, methylation of CHG by CMT3 relies on the binding of CMT3 to H3K9me2 [13]. MET1, the enzyme catalyzing CG methylation, interacts with the histone deacetyltransferase HDA6 in vivo [14]. A histone-binding protein, DTF1/SHH1, specifically recognizes H3 with methylated K9 and unmethylated K4, and recruits PolIV to initiate RNA-directed DNA methylation (RdDM) [15–17]. DNA methylation can also be negatively regulated by H3K4 methylases such as JMJ14, LDL, and LDL2, and by the histone acetyltransferase IDM1 [7,11,18].

The epigenetic marks are interpreted by specific proteins to cause downstream effects. The methyl-CpG-binding domain (MBD) proteins are important interpreters of DNA methylation. Plant MBD proteins were first characterized by Zemach et al. in 2003 [19]. This protein family contains 13 putative MBD proteins based on bioinformatics analysis [19–22]. MBD8, MBD9, and MBD11 were suggested to be important for normal development in Arabidopsis [22–24]. MBD11 knockdown mutants displayed aberrant morphological and developmental phenotypes, such as aerial rosettes, serrated leaves, and reduced fertility [22]. The mutation of MBD8 resulted in a late flowering phenotype due to down regulation of two major promoters of flowering, FT (flowering locus T) and SOC1 (suppressor of constant 1). The late flowering phenotype of the mbd8 mutant was observed in the C24 ecotype but not in the Col-0 ecotype [23]. In contrast to MBD8, MBD9 is a negative regulator of flowering. MBD9 directly controls flowering time by its histone acetyltransferase activity. In the mbd9 mutant the acetylation level of histone H3 and H4 on FLC (flowering locus C) chromatin is dramatically reduced, leading to repressed expression of the flowering repressor FLC [24,25]. In addition to MBD9, MBD6 is another member that is associated with histone modification in vitro [19]. After incubation with Arabidopsis leaf nuclear extract, the complex immuno-precipitated by GST-MBD6 possessed a histone deacetylase activity [19]. Although all of these proteins contain a methyl-DNA-binding domain, only MBD5, MBD6, and MBD7 are capable of specifically binding to methylated DNA [19,21,26]. Further investigations showed
that all three members are able to bind to mCpG sites but not to mCHG sites in vitro and that MBD5 can also bind to mCHH sites. In addition, MBD11 has a DNA-binding ability independent of DNA methylation status [19]. MBD5 and MBD6 are capable of binding to DDM1 in vitro [27].

MBD7 is distinguished from all other MBD members by the presence of three MBD domains [22]. MBD7 contains four functional domains, including the three MBD domains and a StkC (sticky-c) domain at the C terminus [28]. By using 35S::MBD7-GFP transiently expressed in protoplasts, researchers determined that MBD7 localizes to all chromocenters [29]. Localization of truncated MBD7 proteins suggested that the three MBD motifs are sufficient for chromocenter localization and that the number of MBD motifs influences MBD7 localization [29]. Although the StkC domain of MBD7 is not involved in protein localization, it has strong chromatin-binding activity and may help MBD7 fasten to chromosomal mCpG sites [28]. MBD7 is a substrate for PRMT11 in vitro, but PRMT11 does not influence MBD7 localization [30].

In this study, we discovered that MBD7 has a critical role in protecting transgenes from transcriptional silencing. MBD7 binds to densely methylated CpG sites and prevents the spreading of DNA methylation to neighboring regions. We found that MBD7 interacts with other cellular anti-silencing factors such as IDM3, IDM2 and IDM1 [7,31]. These latter factors are regulators of the DNA demethylase ROS1, and are thus necessary for active DNA demethylation at many genomic loci. Therefore, MBD7 interprets DNA methylation not to cause silencing but to prevent DNA methylation spreading to avoid silencing.

2.3 Results

2.3.1 Mutation in MBD7 causes transgene silencing

We developed a very effective system for screening for Arabidopsis mutants with enhanced silencing based on root-length phenotype [32,33]. In this system, a 35S::SUC2
transgene is introduced into Col-0 wild-type (WT) plants. Because of over-expression of \textit{SUC2}, the transgenic plants are very sensitive to sucrose, resulting in a short-root phenotype when grown on sucrose-containing media (Fig. 2.1). This system also contains double 35S promoter-driven \textit{NPTII} and \textit{HPTII}. If a gene is required for anti-silencing, its loss-of-function mutation will cause 35S::\textit{SUC2} transgene silencing. To identify anti-silencing factors, we mutagenized WT plants with EMS and screened the M2 population for mutants with long-root phenotype, a phenotype that indicates possible silencing of the \textit{SUC2} transgene. Our mutant screen has identified several novel anti-silencing factors, such as Anti-silencing 1 (ASI1), Enhanced Downy Mildew 2 (EDM2). In addition, we identified new mutant alleles of known anti-silencing factors, such as IDM1 and ROS1 [32, 33].

In this study, we characterized a mutant, \textit{91-1}, which developed long roots on sucrose growth medium (Fig. 2.1). In addition to having long roots, mutant \textit{91-1} is sensitive to both hygromycin and kanamycin (Fig. 2.2), indicating that all three transgenes might be silenced in this mutant. Results from both Northern blot and RT-qPCR showed that transcript levels of all three transgenes are lower in \textit{91-1} than in WT plants (Fig. 2.3). Through map-based cloning, we identified the mutant \textit{91-1} as a recessive allele of \textit{mbd7} (AT5g59800), hereafter named \textit{mbd7-1}. A C-to-T point mutation changed the 18th amino acid of MBD7 to a stop codon (Fig. 2.4). To confirm that the silencing of transgenes was caused by the mutation of \textit{MBD7}, we crossed WT with a T-DNA insertion allele of \textit{MBD7} (CS876032), \textit{mbd7-2}, and found that plants with 35S::\textit{SUC2} and homozygous T-DNA insertion in \textit{MBD7} also showed the long-root phenotype (Fig. 2.5 and 2.6). In addition, the root phenotype of \textit{mbd7-1} could be complemented by native promoter-driven \textit{MBD7} full-length genomic DNA fused with either an MYC tag or a GFP tag (Fig. 2.7A). RT-PCR indicated that \textit{proMBD7::gMBD7-4xMYC} is able to rescue the suppressed expression of 35S::\textit{SUC2} and 35S::\textit{HPTII} transgenes (Fig. 2.7B). Together, these results demonstrate that mutations in \textit{MBD7} cause transgene silencing, thereby revealing an anti-silencing role for MBD7.
2.3.2 DNA hypermethylation and repressive histone modification accompany the silencing of transgenes

To determine whether silencing of these transgenes occurs at transcriptional levels in *mbd7-1*, we examined RNA Polymerase II (Pol II) occupancy on the 35S transgene promoters in *mbd7-1*. Chromatin immunoprecipitation (ChIP) assays showed that Pol II occupancy at both 35S::SUC2 and 35S::HPTII promoters is reduced in *mbd7-1* (Fig. 2.8), suggesting that the reduced transgene RNA levels are caused by decreased Pol II transcription. Previous results indicated that silencing of 35S::SUC2 is correlated with alterations in the levels of several histone modifications, including H3K9me2, H3K18ac, and H3K23ac [7, 11, 32, 33]. Consistent with the decreased Pol II transcription, ChIP assays showed that, at 35S promoter regions, the repressive histone mark H3K9me2 is increased in *mbd7-1* (Fig. 2.9), while the active histone marks H3K18ac and H3K23ac are reduced in *mbd7-1* (Fig.2.10 and 2.11).

In addition to histone modifications, DNA methylation is another important epigenetic mark that influences chromatin status and can confer transcriptional silencing. To investigate whether DNA methylation plays a role in transgene silencing in *mbd7-1*, we determined the effect of 5-aza-2'-deoxycytidine (5-aza), a DNA methylation inhibitor, on the kanamycin resistance of *mbd7-1*. In the absence of 5-aza treatment, *mbd7-1* was more sensitive than the WT to kanamycin (Fig. 2.2). The difference in kanamycin resistance between the WT and *mbd7-1*, however, was reduced by 5-aza treatment, indicating that DNA methylation plays an important role in silencing the NPTII transgene (Fig.2.2). To examine whether the *mbd7-1* mutation affects DNA methylation, we performed whole-genome bisulfite sequencing for both the WT and *mbd7-1*. The DNA methylation level at 35S promoters was greater in *mbd7-1* than in the WT even though the WT already has a high level of DNA methylation at 35S promoters (Fig.2.12). Together, these results demonstrate that MBD7 dysfunction causes transcriptional silencing through epigenetic modifications.
2.3.3 The *mbd7* mutation causes genome-wide DNA hypermethylation

Whole-genome bisulfite sequencing identified 7561 DMRs (differentially methylated regions) in *mbd7-1*, among which 5694 are hypermethylated, while only 1867 are hypomethylated. In CHG and CHH contexts, the hypermethylated regions are significantly more abundant than the hypomethylated regions (Fig.2.13). Most hypermethylated loci are rich in the mCHH context (Fig.2.13). The much greater number of hypermethylated regions suggested that MBD7 mainly antagonizes the DNA methylation pathway.

Ten hypermethylated cytosines (HMC) were selected for validation by Chop-PCR. Genomic DNA of the WT and *mbd7-1* was extracted and digested by methylation-sensitive endonucleases, such as Hpa II. Subsequently, the digested DNA was used as template for PCR amplification that covered the restriction sites. The results confirmed DNA hypermethylation in *mbd7-1* and were therefore consistent with the whole-genome bisulfite sequencing results (Fig.2.14).

For most of the hypermethylated loci, mCHG and/or mCHH were increased (see the two screenshots on the left of Fig.2.15). mCG methylation was also increased in some regions (see the 3rd screenshot in Fig.2.15).

We found that about 67% of the hyper-DMRs are located in transposable elements, 22% are located in intergenic regions, and only 11% are in genes and pseudogenes (Fig.2.16). Among these DMR-associated TEs, the mCG level was not changed while the mCHG level was slightly increased, and the mCHH level was obviously increased in the *mbd7-1* mutant (Fig.2.17). We then determined whether specific types of TEs are preferentially influenced. However, a comparison of the compositions of total TEs in the WT and of MBD7-affected TEs revealed no obvious preference for any specific type of TE (Fig.2.18). At TEs affected by MBD7, changes of DNA methylation levels are independent of TE length (Fig.2.19).

In the *mbd7-1* mutant, hyper DMRs tend to be distributed near centromeric and pericentromeric regions (Fig.2.20), indicating that MBD7 might preferentially
regulate TEs located at pericentromeric regions. To test this possibility, we defined hypermethylated TE density as the number of hyper DMR-associated TEs divided by the total number of TEs. Hypermethylated TE density was calculated in consecutive 50-kb windows. Hypermethylated TE density distribution across five chromosomes showed that there was no obvious preference for centromeric or pericentromeric TEs (Fig. 2.21), suggesting that hypermethylated TEs aggregate around pericentromeric regions simply because of the high TE density in the regions and not because of a preference of MBD7 for pericentromeric TEs.

### 2.3.4 MBD7-binding positively correlates with mCG methylation density

MBD7 was previously shown as a methyl DNA-binding protein [19]. To date, there has been no report of genome-wide binding targets of MBD7 or any other MBD proteins in plants. *proMBD7::gMBD7-4xMYC* transgenic plants were used for ChIP assays. Western blot analysis showed that the tagged MBD7 protein was expressed, and the mutant phenotype was complemented, suggesting that the tagged protein was functional (Fig. 2.22 and 2.7). To characterize endogenous MBD7 binding targets, we performed ChIP against MBD7-4xMYC followed by high throughput sequencing (ChIP-seq).

We examined the correlation between MBD7 binding and DNA methylation, including CG methylation, CHG methylation, and CHH methylation. First, the genome was divided into 1-kb regions, and MBD7 enrichment was calculated for each region. The top 1% of the MBD7-enriched regions was selected for further analysis. To evaluate the influences of DNA methylation level on MBD7 enrichment, we ranked these regions by CG, CHG, and CHH methylation level, and generated a heat map of MBD7 enrichment (Fig. 2.23). However, no pattern of MBD7 enrichment was evident in all three ranks, suggesting that methylation level does not determine MBD7 binding.

To consider not only methylation level but also the methyl cytosine density of one region, we calculated the methylation density for 1-kb windows. Methylation
density is a value reflecting both methylation level and methyl cytosine frequency for one region, and it is equal to the sum of methylation percentages of individual cytosines normalized by region length [34]. When these regions were ranked by mCG, mCHG, and mCHH methylation density, the heat map pattern indicated that MBD7 enrichment is associated with CG methylation density rather than with CHG and CHH methylation density (Fig. 2.24), which is consistent with the previous finding that MBD7 can bind to mCG sites but not to mCHG sites in vitro [19]. The analysis of methylation level and methylation density suggested that regions with dense and highly methylated mCGs are preferred by MBD7. The tight relationship between MBD7 enrichment and CG methylation density was also supported by genome-wide analysis. We ranked genomic 1-kb windows by CG, CHG, and CHH methylation density, and found that MBD7 enrichment was positively correlated with CG methylation density but not with CHG or CHH methylation density (Fig. 2.25). Two regions were selected to show that the MBD7 could specifically bind to regions with high CG methylation density (Fig. 2.26).

2.3.5 MBD7 binding is associated with the presence of mCHG and mCHH

We identified 2452 MBD7-binding sites through two biological replicates of ChIP-seq. Of 2452 loci, 1930 (78.7%) are TE regions, 94 (3.8%) are intergenic regions, 401 (16.4%) are genes, and 27 (1.1%) are pseudogenes or others (Fig. 2.27). We analyzed the mCG methylation density of all TE-associated 1-kb windows (TE windows) and all gene-associated 1-kb windows (gene windows), and found that TE windows generally have higher CG methylation density than gene windows (Fig. 2.28). To test whether the TE-preferred binding of MBD7 is due to the higher CG methylation density of TE or due to other TE-specific features, we selected TE windows and gene windows with comparable CG methylation density and then compared their MBD7 enrichment (Fig. 2.29). Analysis of MBD7 enrichment revealed that MBD7 binding was more associated with high mCG density in TEs than in genes (Fig. 2.29); this
pattern was consistent in all tested ranges of mCG methylation density (Fig.2.30). We therefore hypothesized that MBD7 may not bind to all regions with high mCG density but instead binds to regions with high mCG density in specific chromatin environments.

Although MBD7 binding did not correlate well with CHG or CHH methylation density, all MBD7 enriched regions appeared to have CHG or CHH methylation (Fig.2.25). We therefore compared MBD7 enrichment in two types of gene windows with the same mCG density. Type I windows lacked mCHG and mCHH, and type II windows had mCHG or mCHH. At comparable mCG methylation densities, MBD7 enrichment was greater in type II windows than in type I windows (Fig.2.31 and 2.32), indicating that MBD7 enrichment is associated with the presence of non-CG methylation.

In summary, MBD7 enrichment is positively correlated with CG methylation density but MBD7 binding is also associated with the presence of non-CG methylation.

### 2.3.6 MBD7 binding influences DNA methylation level

To investigate the correlation of MBD7 binding with its influence on DNA methylation, we calculated MBD7 enrichment and the number of hyper differentially methylated cytosine (DMC) in the *mbd7-1* mutant for each 50-kb region. Hyper DMCs were positively correlated with MBD7 enrichment (Fig.2.33), indicating that DNA hypermethylation in *mbd7-1* is tightly related to the absence of MBD7 around that region. A similar correlation exists between MBD7 enrichment and numbers of hyper mCG, mCHG, and mCHH (Fig.2.34).

To further understand the influence of MBD7 on DNA methylation at its binding loci, we compared the methylation level of binding loci in both the WT and *mbd7-1*. Two groups of control regions were randomly selected, and their methylation levels were also calculated. In *mbd7-1*, mCHG and mCHH methylation levels were elevated
at MBD7-binding regions but not in control regions (Fig. 2.35). The mCG methylation level increased slightly at MBD7-binding regions.

These findings were supported by a closer examination of transgene regions. At the 35S::SUC2 locus, ChIP-seq results showed that MBD7 proteins were enriched in the 35S promoter, which was also hypermethylated in the mbd7-1 mutant (Fig. 2.36). Binding of MBD7 to the 35S promoter was also confirmed by individual MBD7 ChIP-qPCR (Fig. 2.37). Four endogenous MBD7-binding regions were selected to show that DNA methylation at MBD7-binding loci was increased in the mbd7-1 mutant (Fig. 2.38).

### 2.3.7 MBD7 interacts with other anti-silencing factors

To identify MBD7-interacting proteins, we used proMBD7::MBD7-4xMYC transgenic plants and performed immuno-precipitation (IP) followed by LC-MS (liquid chromatography-mass spectrometry). WT plants without the proMBD7::MBD7-4xMYC transgene served as controls. A total of 249 proteins were co-precipitated with MBD7-4xMYC in transgenic plants, but only 79 of the 249 were absent in control samples. Of the 79 proteins, three were of particular interest: IDM1, IDM2 and IDM3 (Table 2.1). These three genes are known to be involved in the DNA demethylation pathway [7, 31]. We also isolated mutants of these three genes from the same genetic screen of mbd7-1 [31], indicating that they also function in anti-silencing.

Yeast two-hybrid (Y2H) assays were used to determine whether MBD7 interacts directly with IDM1, IDM2, and IDM3. Full-length MBD7 was transferred into yeast together with these genes to test their interactions. Yeast cells expressing MBD7 and IDM3 grew on both minus-three media (media lacking leucine, tryptophan, and Histidine) and minus-four media (media lacking leucine, tryptophan, histidine and adenine) (Fig. 2.39), suggesting a strong interaction between MBD7 and IDM3. MBD7 and IDM2 co-transferred yeast cells grew on minus-three media but not on minus-four media (Fig. 2.39), indicating moderate binding between MBD7 and IDM2. MBD7
was not observed to interact with IDM1 in the Y2H assay (data not shown). The interaction between MBD7 and IDM2 or IDM3 was confirmed in a split luciferase transient expression assay in protoplasts. Results in protoplasts were consistent with those of the Y2H assays. IDM3 and MBD7 have a strong interaction, while IDM2 and MBD7 have a weaker interaction (Fig.2.40).

MBD7 contains four functional domains, including three MBD domains and one C-terminal domain (StkC) (Fig.2.41) [28,29]. Y2H assay indicated that both IDM2 and IDM3 bind to the StkC domain of MBD7 rather than to the three MBD domains (Fig.2.42 and 2.43). To test whether StkC, the protein-interacting domain, is important for the anti-silencing function of MBD7, we separately transferred the native promoter-driven full MBD7 cDNA (proMBD7::cMBD7-4xMYC) and native promoter-driven cDNA of the three MBD domains (proMBD7::c3MBD-4xMYC) into mbd7-1. The three MBD domains (lacking the StkC domain) could not rescue the root phenotype of mbd7-1 (Fig.2.44), indicating that the StkC domain is required for the anti-silencing function of MBD7, at least at the 35S::SUC2 transgene locus.

2.3.8 MBD7 affects a subset of genomic regions targeted by active DNA demethylation

IDM1 and IDM2 have been shown to be required for preventing the hypermethylation of hundreds of genomic regions that are a subset of the targets of active demethylation by ROS1, DML2 and DML3 [7,31]. We previously determined the methylomes of the idm1-1, idm2-1 and rdd (ros1dml2dml3) mutants, which were T-DNA insertion mutants in the Col-0 background. The 35S::SUC2 genetic background has a big impact on the DNA methylome (data not shown). In order to compare the effect of mbd7 with those of idm1, idm2 and rdd mutations on the DNA methylome, we sequenced the DNA methylome of the mbd7-2 mutant, which is a T-DNA insertion allele in the Col-0 background. To analyze the overlaps of hyper-DMRs between the mutants, we generated lists of hyper-DMRs regardless of CG, CHG and CHH
sequence contexts using a method that we previously employed [7, 31]. The analysis found 2094 DMRs with increased DNA methylation (hyper-DMRs) in *mbd7-2*. Approximately half (1049) of the hyper-DMRs in *mbd7-2* overlaps with the hyper-DMRs in *rdd* mutants, and 33% (690) overlaps with the hyper-DMRs in the *ros1-4* mutant. Of the hyper-DMRs in *idm1-1* and *idm2-1* mutants, 30.3% (661) and 29.2% (608), respectively, are also hypermethylated in *mbd7-2*. Several examples shared and not shared hyper-DMRs in the *mbd7-2*, *idm1*, *idm2* and *rdd* mutants are shown in Fig.2.45, which also display ChIP-seq results showing MBD7 enrichment levels in these regions.

We noticed that about half of the hyper-DMRs in *idm1-1* or *idm2-1* have DNA hypermethylation at only CGs, while another half have hypermethylation in both CG and non-CG contexts. In contrast, only about 8% (161) of the hyper-DMRs in *mbd7-2* are hypermethylated at only CGs. Out of the hyper-DMRs in *idm1-1* and *idm2-1* that overlap with those in *mbd7-2*, 83% (495) and 81% (429), respectively, have hypermethylation in both CG and non-CG contexts (Fig.2.46). This suggests that that IDM1, IDM2 and MBD7 tend to work together at regions where both CG and non-CG hypermethylation need to be prevented. Our results suggest that MBD7 affects only a subset of genomic regions that require IDM1 or IDM2 for demethylation. To understand how MBD7 may distinguish these regions from others, we analyzed the CG methylation level and CG methylation density of the different groups of hyper-DMRs. In both *idm1* and *idm2* mutants, the hyper-DMRs that overlap with those in *mbd7-2* have higher CG methylation density than the non-overlapping hyper-DMRs. Simulations with randomly selected regions do not show a significant difference (Fig.2.47). The results suggest that MBD7 functions together with IDM1 and IDM2 preferentially at genomic regions with a high CG methylation density, consistent with the requirement of high mCG density for MBD7 binding. Indeed, analyses of the MBD7 ChIP-seq data showed that MBD7 is more enriched at the overlapping than the non-overlapping hyper-DMRs. No significant difference was found using the control ChIP-seq data from WT plants (Fig.2.48).
2.3.9 MBD7 promotes the expression of endogenous loci

Although TEs represented most of the DMRs and MBD7-binding loci, under normal conditions TE transcripts are not abundant; therefore, detecting enhanced silencing of TEs in the mbd7 mutant is difficult. Expression of repeats, like ONSEN and TSI, can be induced by prolonged heat treatment [35]. Heat-induced expression of repeats enabled us to detect changes in TE transcripts between the WT and mbd7-1. After 36 h at 37°C, RNA from the WT and mbd7 mutants was extracted and subjected to RT-qPCR. In two alleles of MBD7, mbd7-1 and mbd7-2, heat-induced expression of ONSEN and TSI was lower compared to that in the WT (Fig. 2.49 and 2.50). Because the primers used for ONSEN and TSI detected transcripts from all copies of ONSEN or TSI, we designed two pairs of primers for specific detection of transcripts of AT1G21945 and AT1G58140, which are two copies of ONSEN genes. In both mbd7-1 and mbd7-2, heat-induced expression of AT1G21945 and AT1G58140 was compromised, which is consistent with the results obtained for all ONSEN copies (Fig. 2.49). HSP70 was used as a control to show effective heat treatment (Fig. 2.50). ChIP-seq of MBD7-4xMYC revealed an enriched MBD7 signal at ONSEN and TSI regions, and increased DNA methylation at ONSEN and TSI was observed in mbd7-1 (Fig. 2.51 and 2.52). The results suggest that physical binding of MBD7 at these loci may antagonize TE silencing by preventing DNA hypermethylation.

2.4 Discussion

In this study, we discovered a new cellular anti-silencing factor, MBD7, required for stopping DNA methylation spread or avoiding hypermethylation to prevent transcriptional gene silencing. DNA methylation is a conserved epigenetic mark for the silencing of TEs and other invasive elements [36–38]. Many plant and animal genomes are occupied mainly by transposable elements (TEs), such that the genes are islands in the sea of TEs [39–41]. When the TEs are silenced by DNA methylation, the methylation may spread and the spread would cause silencing of nearby genes [40,41].
In order for the many TE proximal genes to be properly expressed during development or in response to environment changes, mechanisms must exist to prevent DNA methylation spread from TEs. Active DNA demethylation is one such mechanism. For example, the EPF2 gene that controls the size of stomatal stem cell population is close to a methylated TE in Arabidopsis, and active DNA demethylation is required for preventing methylation spread from the TE and therefore for preventing transcriptional silencing of EPF2 [42]. In the DNA demethylase ros1 mutants, EPF2 is silenced by DNA methylation that is spread from the proximal TE, resulting in an over-production of stomatal lineage cells [42]. Many imprinted genes in plants require the DNA demethylase DME for DNA demethylation and expression in the endosperm because the imprinted genes evolved to have TEs in or near their regulatory sequences [43, 44]. Active DNA demethylation is also necessary for pruning the DNA methylation landscape of many TEs and is thus important for preventing over-silencing of TEs [45]. In addition, transgenes are often subjected to transcriptional as well as posttranscriptional silencing in plants. It has long been known that DNA methylation can cause transcriptional silencing of transgenes in plants [46]. It is important to understand how transcriptional silencing of transgenes is avoided or prevented in order to keep transgenic traits stable in the agricultural biotechnology industry.

The key enzymes for active DNA demethylation have been identified in recent years. In mammals, active DNA demethylation starts from oxidation of 5mC by TETs, followed by the DNA glycosylase TDG [47], while in plants, active DNA demethylation is initiated by the ROS1/DME family of 5mC DNA glycosylases [1]. The mechanisms for targeting the active DNA demethylation enzymes to specific genomic loci are poorly understood. In plants, the histone acetyltransferase IDM1 is required for targeting ROS1 to a subset of genomic loci for demethylation [7]. Although it is not known how H3K18 and H3K23 acetylation marks created by IDM1 affect ROS1 targeting, such a regulation may be considered as an “acetylation switch”, analogous to what has been proposed for the targeting of the chromatin remodeling
complex SWR1 in yeast [48]. Like IDM1, IDM2 regulates the demethylation of a similar subset of genomic loci targeted by the ROS1 family of demethylases [31]. IDM2 is a nuclear alpha-crystallin domain (ACD) protein that mainly affects the H3K18 acetylation activity of IDM1. IDM2 interacts with IDM1 in vitro and in vivo [31], and thus may be considered as a partner protein of IDM1. Interestingly, we found in this study that IDM3, an ACD protein that is closely related to IDM2, is also required for preventing DNA hypermethylation and gene silencing.

MBD7 is a novel MBD protein that contains three MBD domains and a C-terminal Stkc domain [28,29]. Consistent with its capacity to bind methylated CpGs in vitro [19], our ChIP-seq assays found that it binds to thousands of genomic regions, and the binding correlates with the density of methylated CG but not methylated CHG or CHH sites. This binding is coincident with the role of MBD7 in preventing DNA hypermethylation. Therefore, instead of reading the DNA methylation signal to cause silencing, MBD7 interprets the DNA methylation signal to avoid DNA methylation spread or hypermethylation to prevent silencing. Many genes have high levels of CG methylation in their gene bodies, but this type of CG methylation does not cause silencing [2,49]. Our results show that MBD7 binding also requires the presence of non-CG methylation, so it does not bind to these genes since their methylation is exclusively mCG. It will be of interest to determine how MBD7 distinguishes CG methylation with the presence of non-CG methylation from exclusive CG methylation. It is possible that specific chromatin features associated with the presence of non-CG methylation influences MBD7 binding.

MBD7 interacts directly with both IDM2 and IDM3, with the interaction with IDM3 being stronger. Like IDM2, IDM3 also interacts with IDM1. In addition, IDM2 and IDM3 interact with each other. Our IP-LC-MS results show that MBD7 is associated with IDM1 as well as with IDM2 and IDM3 in vivo. The IP-LC-MS results indicated that the amounts of IDM1 and IDM2 that were pulled down with MBD7 were low, compared to the amounts of IDM3 (Table 2.1). The results suggest that while MBD7 and IDM3 may exist in a tight complex, IDM2 and IDM1 are loosely
associated with the complex or in the complex only part of the time. In addition to having three MBD motifs, MBD7 also contains an Stkc domain that is conserved in plant MBD7 orthologs but is not found in other proteins. Our results show that the Stkc domain is necessary for MBD7 function in anti-silencing in plants, consistent with the Stkc domain being responsible for mediating the interaction of MBD7 with IDM2 and IDM3.

Our results suggest a model in which MBD7 binds to methylated TEs and other repeats through its MBD motifs, and uses its Stkc domain to bind to IDM3 and IDM2, to bring IDM1 to the methylated DNA (Fig. 2.53). The H3K18ac and H3K23ac marks created by IDM1 then allow ROS1 and related DNA demethylases to be recruited to stop methylation spread or to prevent hypermethylation by active demethylation (Fig. 2.53). Consistent with this model, H3K18ac and H3K23ac marks are reduced in mbd7 mutant plants, as in idm1 plants at the tested loci. It is interesting that IDM1 also contains an MBD domain that can bind methylated CG \textit{in vitro} [7]. It seems that the single MBD domain of IDM1 is not sufficient to bring IDM1 to some of its target genomic regions, and the targeting requires MBD7 that contains multiple MBD domains. Nevertheless, the single MBD domain may help anchor IDM1 to the genomic sites once it is recruited by MBD7. MBD7 may also help recruit histone modification enzymes other than IDM1 to help target the DNA demethylases since IDM1 only affects a subset of genomic regions demethylated by ROS1 and related DNA glycosylases. On the other hand, MBD7 affects only a subset of genomics regions subjected to active DNA demethylation, indicating also the presence of MBD7-independent mechanisms of targeting the DNA demethylases.

2.5 Material and methods

2.5.1 Plant materials and growth conditions

All plants were \textit{Arabidopsis} in Col-0 genetic background unless stated otherwise. Wild-type (WT) plants contained three transgenes: 35S promoter-driven \textit{SUC2}, dou-
ble 35s promoter-driven \textit{HPTII}, and double 35s promoter-driven \textit{NPTII}. Seeds were disinfested in 50\% (vol/vol) bleach for 15 minutes and then washed five times with sterilized water. After stratification at 4C for 48 h, the seeds were plated on $\frac{1}{2}$MS medium. Plants were grown at 23C in a growth chamber with 16 h of light and 8 h of darkness. The medium used for examining root phenotype was $\frac{1}{2}$MS with 2\% sucrose and 1\% agar. Otherwise, 1\% glucose was added to the $\frac{1}{2}$MS medium to avoid root growth suppression by sucrose. \textit{mbd7-2} (CS876032) was ordered from the Arabidopsis Biological Resource Center (http://www.arabidopsis.org). \textit{mbd7-2} was crossed to the WT containing the \textit{35S::SUC2} transgene. In the F2 population, homozygous of \textit{mbd7-2} containing \textit{35S::SUC2} transgene was isolated and examined for root-length phenotype.

\subsection*{2.5.2 Mutant screening}

An EMS-mutagenized pool of plants was generated and screened for mutants with a long-root phenotype. M2 seedlings were grown vertically on $\frac{1}{2}$MS plates with 2\% sucrose and 1\% agar. WT plants exhibit severely suppressed root growth on this medium, and we screened for mutants with long-root phenotype among 10-day-old seedlings. \textit{mbd7-1} was obtained from this screening.

\subsection*{2.5.3 Map-based cloning}

Genetic mapping was performed as described previously [32]. Briefly, a mapping population was generated from selfed F1 of \textit{mbd7-1} x \textit{Ler}. On $\frac{1}{2}$MS plates with 2\% sucrose and 1\% agar, F2 plants with a long-root phenotype were selected for genetic mapping. For rough mapping, 96 plants and 25 simple sequence length polymorphisms (SSLP) markers evenly distributed across five chromosomes were used to locate the mutation. Through rough mapping, the mutation was determined to be located between Chr.5 22800000 and Chr.5 24724000. For fine mapping with 708 additional plants and 7 SSLP markers, the mutation was narrowed to Chr. 5:
Whole-genome sequencing of the *mbd7-1* mutant was performed to find the mutated gene.

### 2.5.4 Complementation experiments

To generate complementation lines, *MBD7* native promoter-driven full-length *MBD7* genomic DNA was amplified and inserted into the pENTRY- D-topo vector (Invitrogen, *cat.* K2400-20). Through LR recombination reaction, pro*MBD7::gMBD7* was inserted into the pGWB16 vector (with a 4xMYC tag at the C terminal) and into the pGWB4 vector (with a GFP tag at the C terminal). *pGWB16-proMBD7:gMBD7-4xMYC* and *pGWB4-proMBD7:gMBD7:GFP* were transferred into *Agrobacterium tumefaciens* GV3101, and then flowering *mbd7-1* plants were subjected to flower dip transformation. Transgenic plants were genotyped by PCR amplification of transgenes, and transgene expression was examined by western blot using 12-day-old seedlings. Unsegregated T3 plants were identified as homozygous complementation lines and were used for MBD7 ChIP assay.

### 2.5.5 Northern blot, RT-PCR, and RT-qPCR

According to the manufacturer instructions, total RNA was extracted with Trizol reagent (Ambion, 15596-026) from 12-day-old seedlings, except that heat-treated seedlings were only 8 days old. Turbo DNase (Ambion, AM2238) was used to remove DNA contamination from total RNA. A 10 μg quantity of total RNA was treated with 1 μl of DNase I in a 20 μl reaction for 30 min at 37°C; the reaction was stopped by adding 2 μl of inactivation reagent and incubating the preparation at room temperature for 5 min. For reverse transcription, 1 μg of RNA and oligo dT primers were used to synthesize cDNA using the qScript cDNA SuperMix kit (Quanta, 95048-025). Then, 1 μl of cDNA was used as template in each reaction for both RT-PCR and RT-qPCR. Real-time PCR was carried out on a BIO-RAD machine with PerfeCTa SYBR Green FastMix (Quanta, 95072-250). UBC28 was used as an internal control in
heat treatment-related reactions; for all other reactions, either ACTIN2 or ACTIN7 was used as an internal control.

For Northern Blot assay, 10 μg of RNA was run on a 1% formaldehyde Mops agarose gel, and the RNA was transferred to a Hybond-N+ nylon membrane (GE Healthcare) overnight [50]. Probes were synthesized by gene-specific PCR amplifications and labeled using the PCR DIG Probe Synthesis Kit (Roche, 11636090910). Hybridization, washing, and detection were performed following the protocol provided with the DIG-High Prime DNA Labeling and Detection Starter Kit II (Roche, 11585614910).

2.5.6 Whole-genome bisulfite sequencing

Fourteen-day-old mbd7-1 and WT seedlings were used for genomic DNA extraction. Then, bisulfite conversion, library construction, and deep sequencing were performed by the Beijing Genomics Institute (BGI, in Shenzhen, China).

2.5.7 Chop-PCR

Genomic DNA was extracted from 12-day-old mbd7-1 and WT seedlings. After RNase (invitrogen) treatment at 37°C for 1 h, DNA was purified with phenochloroform-isopropyl (Thermo, 17909) and precipitated by 2V ethanol and 1/10V NaOAc (3 M, pH 5.2). In a 30-μl reaction, methylation-sensitive enzymes were used to digest 500 ng of DNA for 16 h, and 2 μl was used for PCR amplification. The HpaII used in Chop-PCR were obtained from New England Biolab (NEB) company.

2.5.8 Chromatin immune-precipitation (ChIP) assay

The ChIP assay was performed were as previously described [51]. In brief, 3 g of 14-day-old seedlings was collected, placed in 35 ml of 1% formaldehyde PBS buffer, and then fixed by vacuum for 2 min followed by another 8 min. After glycine was
added, samples were subjected to a vacuum for 5 min to stop the fixation. Samples were washed three times with sterile water and then ground into powder with liquid nitrogen. Nuclei were extracted with Honda Buffer. The nuclei solution was sonicated to break genomic DNA into 500-bp fragments. After centrifugation, supernatants were removed and used for immuno-precipitation. Dynabeads (Invitrogen, 10003D) were used for pre-clear and antibody binding. The antibodies were: anti-H3K9me2 (Abcam), anti-H3K18ac (Abcam, ab1191), anti-H3K23ac (Millipore, 07-355), anti-MYC (Millipore, 05-724), and anti-Pol II (Abcam, ab5408). Precipitated DNA was dissolved in 30 μl of nuclease-free water, and 2 μl was used for qPCR amplification.

2.5.9 ChIP-seq

*proMBD7::gMBD7: 4xMYC* transgenetic plants were used for ChIP-seq, with WT plants as controls. ChIP was performed according to a previously published protocol [51]. ChIP samples were sent to the Genomics Core Facilities of the Shanghai Center for Plant Stress Biology, SIBS, CAS (Shanghai, China) for library construction and Illumina sequencing.

2.5.10 RNA-seq

RNA was isolated from 14-day-old seedlings of *mbd7-1* and the WT, and the method for extraction of total RNA was the same as described above. Oligo (dT)-conjugated magnetic beads (Qiagen) were used to enrich mRNAs, and then samples were sent to BGI, where library construction and Illumina sequencing were performed.

2.5.11 MethylC-Seq data analysis

For data analysis, low quality sequences (q<20) were trimmed using trim in BRAT-BW [52], and clean reads were mapped to the TAIR10 genome using BRAT-BW and allowing two mismatches. To remove potential PCR duplicates, the remove-
dupl command of BRAT-BW was used. Differentially methylated regions (DMRs) were identified as described previously [53] with minor modifications. The DMRs were separately identified for each of the three contexts (CG, CHG, and CHH). For each context, the *Arabidopsis* genome (TAIR10) was divided into 100-bp windows, and the number of called Cs and Ts were compared. Windows with an absolute methylation difference of 0.4, 0.2, and 0.1 for CG, CHG, and CHH, respectively, were selected as candidate DMRs. Finally, windows with Benjamini-Hochberg corrected FDR ≤ 0.01 (Fisher exact test) and at least 4 cytosines in the corresponding context that are each covered by at least 4 reads in both mutant and wild-type were retained.

2.5.12 MBD7 ChIP-Seq Analysis

The quality of the sequencing data was checked with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The paired-end reads were mapped to the unmasked TAIR10 genome of *Arabidopsis thaliana* by means of bowtie2, allowing the mapping of up to 10 different genomic sites with parameter “-k 10 –no-mixed –no-discordant -I 0 -X 1000” [54]. To remove potential PCR duplicates, rmdup from SAMtools [55] was used. Peaks were called with MACS2 callpeak function [56], comparing the two replicates of MBD7-Myc ChIP samples with the two replicates of wild-type control. The genome size (-g) was set at 1.2e8, and the -keep-dup parameter was set as “all”.

Relationship between MBD7 ChIP Enrichments and mCG identity

The *Arabidopsis* genome was divided into 1-kb bins with an overlap of 500 bp. The MBD7 protein enrichment definition was similar as that of [34]. They were calculated as follows:

\[
\text{Enrichment} = \log_2(8 + n_{\text{ChIP}}) - \log_2(8 + n_{\text{Input}} \times N_{\text{ChIP}}/N_{\text{Input}})
\]

where \( n_{\text{ChIP}} \) and \( n_{\text{Input}} \) represent the number of mapped ChIP and input tags, respectively, in the corresponding 1-kb bin. \( N_{\text{ChIP}} \) and \( N_{\text{Input}} \) are the sum of all mapped tags. Eight pseudo counts were added to overcome sampling noise. The program featureCounts [57] was used to obtain \( n_{\text{ChIP}} \) and \( n_{\text{Input}} \). mCG(CHG/CHH) density was calculated as sum of CG(CHG/CHH) methylation level in the corre-
sponding 1kb bin and normalized to 100 bp. Only those 1-kb bins that had at least 80% CG(CHG/CHH) covered by ≥ 4 reads were used for downstream analysis.

2.6 References


2.7 Figures and tables

Fig. 2.1. Isolation of the 91-1 mutant based on its long root phenotype. Wild-type (WT) plants with the 35S::SUC2 (sucrose transporter 2 gene driven by the CaMV 35S promoter) transgene display a short root phenotype on medium containing sucrose, while mbd7 mutant plants have long roots. The root growth of Col-0, WT, and 91-1 on glucose medium served as controls.
Fig. 2.2. Hygromycin and kanamycin sensitivity of 91-1 mutants. $2\times35S::HPTII$ and $2\times35S::NPTII$ also exist in WT plants, leading to both hygromycin and kanamycin resistance. Comparison of Col-0, the WT, and the 91-1 mutant showed that the 91-1 mutant was sensitive to both hygromycin and kanamycin; with growth on medium without antibiotics served as the control. However, the kanamycin sensitivity of 91-1 mutants was compromised by treatment with the DNA methylation inhibitor 5'-Aza.
Fig. 2.3. Expression level of transgenes in 91-1 mutant plants.
(A) Transcript levels of \textit{SUC2} and \textit{HPTII} in 91-1 mutant plants. Northern blot was used to determine the expression level of transgenes. Compared to the WT, 91-1 mutant plants had reduced levels of \textit{SUC2} and \textit{HPTII} transcripts. The level of \textit{SKP1} transcripts was used as the loading control.
(B) Expression levels of \textit{SUC2}, \textit{HPTII}, and \textit{NPTII} in 91-1 mutant plants were examined using RT-qPCR.
Fig. 2.4. Identification of the mutation in the 91-1 mutant that is responsible for transgene silencing.

(A) Map-based cloning of 91-1. The top line is the diagram of chromosome 5 of *Arabidopsis*. The lower line is the diagram of chromosome region containing *MBD7* (AT5G59800). Positions of molecular markers used for mapping are shown below the lower line. Ratios on the bottom of the diagram are the numbers of recombinants over the total numbers of mapped chromosomes.

(B) Diagram showing the EMS mutation site in the *mbd7-1* (91-1) mutant and the T-DNA insertion position in *mbd7-2*. 

41
Fig. 2.5. Genotyping of \textit{mbd7-2} containing \textit{35S::SUC2} transgene. The diagram shows positions of primers used in genotyping the \textit{mbd7-2} mutant containing the \textit{35S::SUC2} transgene. Wild-type plants were used as controls. Full-length genomic DNA of MBD7 was amplified with \textit{MBD7F} and \textit{MBD7R} primers. T-DNA insertion was genotyped using Lb1 and \textit{MBD7R}. Primers of 35sF and SUC2 R were used to amplify the \textit{35S::SUC2} transgene.
Fig. 2.6. Root phenotype of mbd7-2. 35S::SUC2 was introduced into another allele of mbd7 mutant, mbd7-2 (CS876032). On a medium containing sucrose, mbd7-2 has the same long root phenotype as mbd7-1.
Fig. 2.7. Functional complementation of the $mbd7-1$ mutant. (A) Complementation of the $mbd7-1$ mutant restored the short root phenotype. Expression of native promoter driven genomic DNA of $MBD7$ could rescue the root phenotype of $mbd7-1$ mutants. (B) Restored expression level of $SUC2$ and $HPTII$ in $MBD7-4xMYC/mbd7-1$ transgenic plants. $Actin2$ expression level is shown as a control.
Fig. 2.8. Determination of Pol II occupancy at 35S promoters.
(A) Diagram of the 35S::SUC2 transgene and labeling of examined regions in the ChIP assay.
(B) Pol II occupancy at the transgene promoter was tested in the wild type and mbd7-1. A ChIP assay used 12-day-old seedlings and antibody against Pol II. Actin2 was used as the control region.
Fig. 2.9. Determination of H3K9me2 level at the 35S promoter.
(A) Diagram of the 35S::SUC2 transgene, and labeling of examined regions in the ChIP assay.
(B) H3K9me2 level at the 35S promoter and control region. ChIP was performed in wild-type and mbd7-1 plants with anti-H3K9me2 antibody. Actin2 was used as the control region.
Fig. 2.10. Determination of H3K18ac level at the 35S promoter.
(A) Diagram of the 35S::SUC2 transgene, and labeling of examined regions in the ChIP assay.
(B) H3K18ac level at the 35S promoter and control region. ChIP was performed in wild-type, mbd7-1, and idm1-9 plants. Actin7 was used as the control region.
Fig. 2.11. Determination of H3K23ac level at the 35S promoter.
(A) Diagram of the 35S::SUC2 transgene, and labeling of examined regions in ChIP assay.
(B) H3K23ac level was examined at the 35S promoter and control region. ChIP assay was performed in wild-type, mbd7-1, and idm1-9 plants.
Fig. 2.12. DNA methylation status of the 35S::SUC2 transgene in the wild type and mbd7-1. Bisulfite sequencing results are displayed as a screenshot of IGB tracks. The upper lane indicates the methylation status in the WT, which served as the control. The lower lane indicates the methylation status in mbd7-1. Each orange bar represents a methyl cytosine, and the height of the bar indicates the methylation level of the single cytosine. The 35S::SUC2 transgene diagram is shown above the screenshot.
Whole genome methylation status

- Number of Hypermethylated regions
- Number of Hypomethylated regions

Fig. 2.13. Hyper- and hypo-DMRs (differently methylated regions) in *mbd7-1*. Whole-genome bisulfide sequencing was performed using 12-day-old seedlings. With wild-type plants as controls, hyper- and hypo methylated regions in *mbd7-1* were identified in CG, CHG, and CHH contexts.
Fig. 2.14. Determination of DNA hypermethylated loci in \textit{mbd7-1} using Chop-PCR. Ten loci were selected to validate whole-genome bisulfite sequencing results. Chop-PCR was performed with two biological replicates. Methylation-sensitive restriction enzymes were used to digest genomic DNA of the wild type and \textit{mbd7-1} mutants.
Fig. 2.15. Examples of hypermethylated regions in \textit{mbd7-1}. The methylation status of total mC, mCG, mCHG, and mCHH in the wild type and \textit{mbd7-1} is shown. For most DMRs, either the amount or level of mCHH and mCHG were increased in \textit{mbd7-1}, as indicated in the two regions to the left. For some regions, the mCG level also changed, as shown in the third screenshot.
Fig. 2.16. Composition of hyper-DMRs in mbd7-1. Compositions of hypermethylated regions in the mbd7-1 mutant were analyzed based on CG, CHG, and CHH contexts.
Fig. 2.17. Average methylation levels in hyper-DMR-associated TE bodies and flanking 2 kb regions. Each TE was aligned from start to end and divided into 20 equal bins. Upstream and downstream 2 kb regions were each also divided into 20 equal bins. Methylation level was calculated for each of the 60 bins across all of the corresponding regions and plotted. Only cytosines with > four-fold coverage were used for this analysis.
Fig. 2.18. Compositions of total TEs and TEs with hyper mCHH in *mbd7-1*. TE categories are labeled with different colors.
Fig. 2.19. Average methylation level of hyper-DMR-associated TEs with different lengths. DMR associated-TEs were classified by length. The range of lengths is indicated above the figure.
Fig. 2.20. Chromosomal distribution of hyper-DMRs in mbd7-1.
Fig. 2.21. Position preference of hyper-DMR-associated TEs in mbd7. Number of total TEs and mbd7-1 hyper-DMR-associated TEs was calculated in every 500-kb window of the genome. The distribution of TE number/500 kb along five chromosomes is shown in the first and second lanes. The ratios of hypermethylated TEs to total TEs per 500 kb were calculated and displayed as in the bottom figure. The trend of hypermethylated TE ratio is represented by the black dashed line.
Fig. 2.22. Recombinant MBD7 protein expression in $MBD7$-$4xMYC/mbd7$-1. Western blot was used to detect recombinant MBD7 protein expression in transgenic plants. WT was used as the control. Positions of molecular mass markers are on the left.
Fig. 2.23. Correlation of MBD7 enrichment with DNA methylation level. The top 1% of MBD7-enriched regions was selected for analysis. For each region, MBD7 enrichment was represented by color from light yellow (lowest) to red (highest). Windows were ranked by mCG, mCHG, and mCHH methylation levels.
Fig. 2.24. Correlation of MBD7 enrichment with DNA methylation density. The top 1% of MBD7-enriched regions was selected for analysis. For each region, MBD7 enrichment was represented by color from light yellow to red. Windows were ranked by mCG, mCHG, and mCHH methylation density.
Fig. 2.25. Genome-wide relationship between MBD7 enrichment and DNA methylation densities. Genomic 1-kb bins were used in this analysis. ChIP-seq of the wild type served as the control. All 1-kb bins were ranked by mCG, mCHG, or mCHH methylation density. The blue dashed line indicates the DNA methylation density, and red line indicates MBD7 enrichment.
Fig. 2.26. Screenshots showing endogenous MBD7-binding targets. Total DNA methylation of the wild type is labeled by orange bars in the first row, and CG methylation of the wild type is labeled by yellow bars in the second row. The 3rd and 4th rows are MBD7 enrichment in WT and MBD7-4xMYC transgenic plants, respectively. MBD7 enrichment is labeled by blue bars. Gene or TE annotations are below the screenshot.
Fig. 2.27. Composition of endogenous MBD7 binding targets based on analysis of 2452 peaks of ChIP-seq signals.
Fig. 2.28. Comparison of mCG methylation density between genes and TEs. Total 1-kb gene windows and 1-kb TE windows were used in this analysis. Methylation density of gene windows is indicated by purple, and methylation density of TE windows is indicated by green. Gene windows were 1-kb windows that had at least a 500-bp overlap with genes but not with TEs. TE windows were 1-kb windows that had at least a 500-bp overlap with TEs.
Fig. 2.29. Comparison of MBD7 enrichment between 1-kb gene and TE windows with similar mCG methylation densities. One-kb gene and TE windows with 1.8-2.2 mCG methylation density were used. mCG methylation densities of gene windows and TE windows are displayed on the left of the figure, and their corresponding MBD7 enrichment is shown on the right of the figure.
Fig. 2.30. Comparison of MBD7 enrichment between 1-kb gene and TE windows with similar mCG methylation density. This analysis was similar to that shown in Figure 28, except that three mCG methylation density ranges were selected: 0.8-1.0, 1.0-1.4, and 1.4-1.8. Gene and TE windows within the same range were used to compare their MBD7 enrichment. For each mCG methylation density range, comparable mCG methylation density between gene and TE windows is shown on the left, and their MBD7 enrichment is shown on the right.
Fig. 2.31. Comparison of MBD7 enrichment between windows with and without non-CG methylation. One-kb windows with 1.8-2.2 mCG methylation density were used. Windows were sorted by absence (type I) or presence (type II) of non-CG methylation, including mCHG and mCHH. mCG methylation densities of type I and type II windows are displayed on the left, and their MBD7 enrichment is shown on the right. The cutoff of methylation level for mCHG and mCHH is 5%.
Fig. 2.32. Comparison of MBD7 enrichment between windows with and without non-CG methylation. This analysis is similar to that of Figure 31, except that three mCG methylation density ranges were used: the 1-kb windows had methylation densities of 0.8-1.0, 1.0-1.4, and 1.4-1.8. 1.8-2.2. Windows were classified by absence (type I) or presence (type II) of non-CG methylation. mCG methylation densities of type I and type II windows are displayed on the left, and their MBD7 enrichment is shown on the right.
Fig. 2.33. Correlation of MBD7 enrichment with DNA hypermethylation in mbd7-1 mutant plants. For each 50-kb genomic window, MBD7 enrichment and the number of hypermethylated cytosines [hyper-differentially methylated cytosines (hyper-DMCs)] in mbd7-1 were calculated. Hyper-DMC was defined as cytosines that showed a 1.5X increase in methylation level in mbd7-1 compared to WT.
Fig. 2.34. Correlation of MBD7 enrichment with hyper mCG, mCHG, and mCHH in \textit{mbd7-1} mutant plants. Hypermethylated cytosines in \textit{mbd7-1} were classified by mCG, mCHG, and mCHH contexts. For each 50-kb genomic window, MBD7 enrichment and the numbers of hyper mCG, mCHG, and mCHH were calculated.
Fig. 2.35. The effect of MBD7 binding on DNA methylation. The average methylation levels of 2452 peaks were calculated. The methylation levels of mCG, mCHG, mCHH, and total mC are displayed separately. Random regions were selected through computational simulation based on the lengths of the 2452 peaks. The purple line indicates the methylation level in mbd7-1, and the green line indicates the methylation level in the wild type.
Fig. 2.36. MBD7 binding at the 35S transgene promoter. Diagram for 35S::SUC2 transgene is on the top. Upper two lanes are DNA methylation status of 35S::SUC2 transgene in wild type and mbd7-1 respectively. MBD7 enrichment at the transgene is shown in the lower two lanes.
Fig. 2.37. Examination of MBD7 enrichment at transgene promoter using individual locus ChIP assay. Examined regions are as indicated above the figure. Promoter of ACTIN7 was used as the control.
Fig. 2.38. Screenshots showing that DNA methylation level is increased in MBD7-binding regions. In the wild type and mbd7-1, methylation levels of total mC, mCG, mCHG, and mCHH are shown separately. The last two rows indicate MBD7 enrichment in the WT and MBD7-4xMYC transgenic plants. MBD7 enrichment is labeled by blue bars. Gene or TE annotations are below the screenshot.
Fig. 2.39. MBD7 physically interacts with IDM2 and IDM3. IDM2 and IDM3 are two closely related alpha-crystallin domain proteins that also function in anti-silencing. Protein interactions were tested using the yeast two-hybrid assay. Yeast growth on media without Leu (leucine), Trp (tryptophan), and His (Histidine) indicates an interaction between two proteins.
Fig. 2.40. Interaction of MBD7 with IDM2 and IDM3 using split luciferase transient expression assays. Tested proteins were transiently expressed in *Arabidopsis* protoplasts. The co-transferred *GUS* transgene was used to standardize protoplast transfer efficiency. Protoplasts expressing the tested protein and irrelevant proteins including AHGI (ABA-hypersensitive germination 1) and PYL13 (PYR1-like protein 13) served as negative controls.
Fig. 2.41. Diagram showing structures of full MBD7 and truncated MBD7 used in Y2H assay. The full MBD7 contains three MBD domains and one STKC domain, as shown on the first row. The truncated MBD7 protein labeled 3MBD in the second row contains three MBD domains (1-232 aa). The truncated MBD7 protein labeled as STKC in the bottom row contains only the STKC domain (232-306 aa).
Fig. 2.42. Y2H assay testing interactions between IDM2 and truncated MBD7 proteins. BD-fused 3MBD, BD-fused STKC, and AD-fused IDM2 were used in this Y2H assay. Media lacking Leu, Trp, and His were used to detect moderate protein interaction. Media lacking Leu, Trp, His, and Ade (adenine) were used to detect strong protein interaction.
Fig. 2.43. Y2H assay testing protein interactions between IDM3 and truncated MBD7 proteins. BD-fused 3MBD, BD-fused STKC, and AD-fused IDM3 were used in this Y2H assay. Media lacking Leu, Trp, and His were used to detect moderate protein interaction. Media lacking Leu, Trp, His, and Ade were used to detect strong protein interaction.
Fig. 2.44. Complementation of the mbd7-1 mutant with a truncated MBD7 construct. Native promoter-driven full MBD7 CDS (proMBD7::cMBD7-4xMYC) and native promoter-driven 3MBD CDS (proMBD7::c3MBD-4xMYC) were transferred to mbd7-1. Root phenotypes of the two transgenic lines were tested on a sucrose-containing medium.
Fig. 2.45. Three examples of shared DMRs are shown in the first three (from left to right) dashed boxes. The last three dashed boxes show examples of hyper-DMRs that are independent of MBD7.
Fig. 2.46. Within each DMR, hyper-methylated cytosines are classified into two types, hyper-mCG and hyper-non-mCG.
Fig. 2.47. Simulation with randomly selected regions was shown on the right for comparison. Mann-Whitney test was performed for statistical analyses. P-values < 0.01 indicate statistical significance.
Fig. 2.48. Simulation with randomly selected regions was shown on the right for comparison. Mann-Whitney test was performed for statistical analyses. P-values $< 0.01$ indicate statistical significance.
Fig. 2.49. Examination of heat-induced expression of \textit{ONSEN} genes in \textit{mbd7} mutants. Expression was tested in both \textit{mbd7-1} and \textit{mbd7-2} mutants, with WT and Col-0 as controls. Eight-day-old plants were treated at 37°C for 36 h. The gene expression levels in \textit{mbd7-1} and in \textit{mbd7-2} are relative to those in the WT and Col-0, respectively. Primers used to detect \textit{ONSEN} could amplify cDNA from multiple copies of \textit{ONSEN}. Primers for \textit{AT1G58140} and \textit{AT1G21945} were designed to specifically amplify only one copy of the \textit{ONSEN} gene.
Fig. 2.50. Examination of heat-induced expression of *TSI* genes in *mbd7* mutants. Heat induction of *HSP70* served as a control. Eight-day-old plants were treated at 37°C for 36 h. The gene expression levels in *mbd7-1* and in *mbd7-2* are relative to those in the WT and Col-0, respectively. The primers used to detect *TSI* could amplify cDNA from multiple copies of *TSI*. 
Fig. 2.51. Methylation status and MBD7 enrichment of the \textit{ONSEN} gene. \textit{AT1G21945}, one of the \textit{ONSEN} genes, is displayed as an example. The first eight lanes show the methylation status of \textit{AT1G21945} in both the WT and \textit{mbd7-1}. Total mC, mCG, mCHG, and mCHH are indicated. The last two lanes show MBD7 enrichment at \textit{AT1G21945}. \textit{AT1G21945} is labeled with a black arrow below the screenshot.
Fig. 2.52. Methylation status and MBD7 enrichment of the TSI gene. AT1G38360, one of the TSI genes, is displayed as an example. The first eight lanes show the methylation status of AT1G38360 in both the WT and mbd7-1. Total mC, mCG, mCHG, and mCHH are displayed. The last two lanes show MBD7 enrichment at AT1G38360. AT1G38360 is labeled with a black arrow below the screenshot. The methylation status of the framed region (Chr. 1: 14437062-14446605) is shown separately.
Fig. 2.53. Working model of MBD7 function in anti-silencing.
Table 2.1.
Results of immunoprecipitation against MBD7-4xMYC followed by LC-MS.

<table>
<thead>
<tr>
<th>Accession</th>
<th>Gene ID</th>
<th>Gene name</th>
<th>Description</th>
<th>Unique peptides</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q9FJF4</td>
<td>AT5G59800</td>
<td>MBD7</td>
<td>Methyl-CpG-binding domain-containing protein 7</td>
<td>9</td>
</tr>
<tr>
<td>Q9SYQ0</td>
<td>AT1G20870</td>
<td>IDM3</td>
<td>Alpha-crystallin domain of heat shock protein-containing protein</td>
<td>12</td>
</tr>
<tr>
<td>F4HYL2</td>
<td>AT1G54840</td>
<td>IDM2</td>
<td>Alpha-crystallin domain of heat shock protein-containing protein</td>
<td>1</td>
</tr>
<tr>
<td>F4IXE7</td>
<td>AT3G14980</td>
<td>IDM1</td>
<td>Acyl-CoA N-acyltransferase with RING/FYVE/PHD-type zinc finger domain</td>
<td>1</td>
</tr>
</tbody>
</table>
3. IDENTIFICATION OF A MOLECULAR REGULATOR INVOLVED IN THE ACTIVE DNA DEMETHYLATION PATHWAY BASED ON GENETIC SCREENING IN *ARABIDOPSIS*

3.1 Abstract

DNA methylation is an important epigenetic mark in diverse biological processes. DNA methylation status can be dynamically controlled by methylation and active demethylation processes. In plants, active DNA demethylation relies on a base excision repair pathway. Recently, a subfamily of DNA glycosylases has been identified as DNA demethylation enzymes. However, the mechanism by which demethylation is targeted to specific loci is poorly understood. To date, the only identified regulatory component in the active DNA demethylation pathway is ROS3 (REPRESSOR OF SILENCING 3), a small RNA-binding protein. ROS3 is colocalized with ROS1 (REPRESSOR OF SILENCING 1) and is capable of binding single-stranded RNA. The current working model is that ROS3 is involved in targeting demethylase enzymes to specific loci through its RNA-binding activity. Here, we describe an enhancer mutant screen in the *ros3* background and propose to characterize a mutant, *rte1* (ros three enhancer 1), from the screen in order to identify a novel component involved in the regulation of active DNA demethylation.

3.2 Introduction

Active DNA demethylation is critical for many plant processes such as gene imprinting, preventing the spread of DNA methylation from repetitive sequences to neighboring genes, and maintaining a proper dynamic state of transposons [1].
In plants, researchers have identified a family of DNA glycosylase/lyases, including ROS1, DME, and DMLs, whose enzymatic activities result in the replacement of a methylated cytosine with an unmethylated cytosine [2–4]. A single-stranded RNA-binding protein, ROS3, has been identified and to date is the only regulatory factor known to be involved in the active DNA demethylation pathway in plants [5].

In mammals, almost all components identified in active DNA demethylation are enzymes that can be classified into two categories. Enzymes in one category have methylcytosine modification activities, and these include AID (activation-induced deaminase) 127 with deaminase activity and TET1 with hydroxylase activity. Enzymes in the second category are related to the base excision repair (BER) pathway, and these include TDG (thymine-DNA glycosylase) and MBD4 (methylcytosine-binding protein), both of which have glycosylase activity [3,6–9]. In addition, recent studies showed that an ELP3-containing elongator complex may play an indirect role in active DNA demethylation [10].

Although enzymes functioning in active DNA demethylation have been identified and studied, the mechanism by which demethylation is targeted to specific loci remains unclear in both animals and in plants. We hypothesize that plants contain regulatory factors that work together with ROS3 or function in a parallel pathway with ROS3 to direct demethylases to specific loci. Therefore, research with the Arabidopsis model system may increase our understanding of how active DNA demethylation occurs in plants and may also provide clues concerning the regulatory mechanisms of active DNA demethylation in mammals.

In our mutant screening system, a transgene was introduced into C24 ecotype plants of Arabidopsis. The transgene consists of the firefly luciferase reporter (LUC) under control of the ABA-, drought-, salt-, and cold stress-responsive RD29A promoter [2]. Expression of the proRD29A-LUC transgene in our transgenic Arabidopsis line (renamed C24-luc) has been very stable for many generations over the last 17 years. This system has been used to screen for mutants with higher or lower expression of the proRD29A-LUC transgene. Many novel components have been
identified from mutants with an enhanced expression of the proRD29A-LUC transgene, but only two mutants have been identified so far with a repressed expression of proRD29A-LUC. The two mutants with a repressed expression of proRD29A-LUC are ros1 and ros3 [2,5]. In ros1 and ros3 mutants, the transgene is silenced because of the hypermethylation of the transgenic RD29A promoter. To identify new components involved in the DNA demethylation pathway, we performed a genetic screen for the ros3 enhancer mutants. We found that after many generations of inbreeding, the ros3 mutant in C24-luc shows abnormal bioluminescence in mature plants but not in seedlings. Therefore, 10-day-old seedlings could be used to screen for ros3 enhancer mutants. In this study, rte1 (ros three enhancer mutant 1) was identified and characterized.

3.3 Results

3.3.1 Isolation of ros3 enhancer mutants

To obtain ros3 enhancer mutants, ros3 seeds were mutagenized with the chemical mutagen ethyl methanesulfonate (EMS). The M2 population was used for enhancer mutant screens. Expression of the transgene proRD29A::LUC in 10-day-old seedlings of ros3 is induced by cold treatment (Fig.3.1 A). We screened for mutants in which the stress-induced luminescence signal was weaker than in the ros3 mutant. The phenotype of one enhancer mutant, referred to as rte1, was confirmed (Fig.3.1 A). After a 24-h cold treatment, the C24-luc wild type and ros3 mutants displayed strong luminescence signals, suggesting that the proRD29A-LUC transgene was well induced (Fig.3.1A). The luminescence signal, however, was completely repressed in the rte1 mutant and in ros1-1, which served as a positive control (Fig.3.1A). When we treated 3-week-old seedlings with both cold and salt stress (300 mM NaCl solution) for 6 h, similar luminescence phenotypes were observed (Fig.3.1B). In mature leaves, LUC was not strongly induced by cold or salt treatment in ros3, but the luminescence signal was weaker in rte1 than in ros3 (Fig.3.1C).
Transcript levels of proRD29A:: LUC were assessed by real-time PCR. The results were consistent with the luminescence phenotypes that the LUC transgene was well induced in C24-luc and ros3 but not in ros1-1 or rte1 by cold treatment (Fig.3.2A). Because the transgene driven by the RD29A promoter was repressed, we tested the expression of the endogenous RD29A gene. Like the transgene, the endogenous RD29A gene was not well induced in rte1 relative to C24-luc and ros3, but its transcript level was higher in rte1 than in ros1 (Fig.3.2B). COR15A was well induced in all tested plants by cold treatment (Fig.3.2C), indicating that the cold induction was effective and also suggesting that the repressed expression of the transgene or endogenous gene in rte1 and ros1 was not due to abolishment of the stress signaling pathway.

Another transgene, the 35S promoter- driven NPTII transgene, was also in our screening system. The C24-luc WT has a kanamycin-tolerant phenotype, while ros3 has moderate kanamycin sensitivity (Fig.3.3A). rte1 has stronger kanamycin sensitivity than ros3, suggesting that NPTII is more repressed in rte1 than in ros3. This inference was supported by assessment of the expression level of NPTII (Fig.3.3B).

3.3.2 rte1 is a DNA hypermethylated mutant

Because rte1 has a stronger transgene and endogenous gene silencing phenotype than ros3, we wanted to determine whether the DNA methylation phenotype is enhanced in rte1 relative to ros3. Four Chop-PCR markers were used to check the methylation status of rte1. Two markers were used to detect loci in transgenic and endogenous RD29A promoters, and the other two markers are known to be ros1 endogenous DNA hypermethylation targets. In Chop-PCR, methylation-sensitive enzymes were used to digest genomic DNA before PCR amplification was performed with primers that covered the digested sites. The results can be analyzed by comparing the intensities of the amplified PCR products: a higher PCR product intensity indicates a higher methylation level of the genomic DNA. C24-luc and ros3 served as negative controls, and ros1-1 served as a positive control. The results show that
the DNA methylation level at transgenic, endogenous RD29A promoter loci is higher in \textit{rte1} than in \textit{ros3} (Fig.3.4). At \textit{AT4G18650} promoter and \textit{Pm36} LOCUS, \textit{rte1} showed a little DNA hypermethylation phenotype (Fig.3.4).

Chop-PCR could only check the methylation level of specific cytosines. To confirm the hypermethylation level of the whole \textit{RD29A} promoter region, we performed bisulfite sequencing of the transgenic \textit{RD29A} promoter region. Bisulfite sequencing is used to determine the methylation level at specific regions by treating genomic DNA with sodium bisulfite. The bisulfite treatment can convert unmethylated cytosine to uracil (C to U), while leaving 5-methylcytosine unaffected. When a PCR reaction is carried out using the converted DNA as template, adenine (A) instead of guanine (G) will be paired with uracil (U). Therefore, unmethylated cytosine will convert to thymidine (C to T), and methylated cytosine (mC) will remain cytosine (Fig.3.5A). The bisulfite sequencing results showed that, compared to the \textit{ros3} mutant, the \textit{rte1} mutant is hypermethylated in all three contexts (CG, CHG, and CHH) (Fig.3.5B). The C24-luc and \textit{ros3} mutants were used as negative controls, while the \textit{ros1-1} mutant was used as a positive control.

3.3.3 Isolation of the \textit{rte1} single mutant

The \textit{rte1} mutant was screened from mutant pools generated in the \textit{ros3} mutant background. The \textit{rte1} mutant is actually a double mutant with both the \textit{rte1} mutation and the \textit{ros3} mutation. To facilitate further research on the functions of the enhancer gene, we isolated the single enhancer mutant (hereafter referred to as the \textit{rte1} single mutant).

The \textit{rte1} mutant was back crossed to C24-luc, and the F2 population was used to screen for the \textit{rte1} single mutant. F2 seeds were plated on MS medium and grown for 12 days. After a 24-h cold treatment, F2 plants were imaged with a CCD camera. Seedlings with a lower luminescence signal than C24-luc were selected and transferred to soil. One week later, genomic DNA was extracted from each selected
plant, and genotyping was carried out to test for the presence of the \textit{ros3} mutation. Plants without the \textit{ros3} mutation but with a low luminescence signal were isolated as the single mutants, and their seeds were collected. Two individual lines of the single mutant were isolated, and their seeds were collected separately. Their low luminescence signal phenotype was confirmed in mature leaves (Fig.3.6A) and was reconfirmed in their progeny (Fig.3.6B). The luminescence phenotype is stronger in \textit{rte1/ros3} than in either \textit{ros3} or \textit{rte1} single mutants. The additive effect indicates that RTE1 may work independently or in parallel with ROS3. We measured the kanamycin resistance of these two single mutants. The \textit{rte1} single mutant #12 was very sensitive to kanamycin (Fig.3.7), while the \textit{rte1} single mutant #19 was less sensitive (data not shown).

### 3.3.4 The methylome of the \textit{rte1} single mutant

Based on its kanamycin phenotype, \textit{rte1} single mutant #12 was selected to further characterize the function of the \textit{RTE1} gene. Chop-PCR marker in endogenous RD29A promoter was used to check the methylation status of the \textit{rte1} single mutant #12. Compared to C24-luc plants, \textit{rte1} single mutant #12 showed a hypermethylation phenotype (Fig.3.8A). To further characterize the influence of \textit{RTE1} on DNA methylation, we performed whole-genome bisulfite sequencing using 12-day-old seedlings of the \textit{rte1} single mutant #12. Consistent with Chop-PCR results, the \textit{RD29A} promoter region in \textit{rte1} single mutant #12 is hypermethylated (Fig.3.8B).

By comparison with C24-luc bisulfite sequencing results, we identified 498 differentially methylated regions (DMRs) in the \textit{rte1} single mutant #12. Most DMRs are hypermethylated regions (Fig.3.9A), suggesting that \textit{RTE1} functions in DNA demethylation. The composition of hypermethylated DMRs indicated that \textit{RTE1} does not have preference to TE, gene, or intergenic regions (Fig.3.9B). Four regions were selected to illustrate the endogenous hypermethylated regions of \textit{rte1} single mutant #12 (Fig.3.10).
3.3.5 Genetic mapping of *rte1*

Genetic mapping was used to roughly locate the *rte1* mutation. For genetic mapping, the *rte1* mutant was crossed to the wild type of the Columbia ecotype (Col-0) without the *proRD29A-LUC* transgene. The F2 population was screened for the enhancer mutant phenotype based on luminescence imaging and PCR genotyping for the presence of *proRD29A-LUC*. Plants with the *proRD29A-LUC* but without cold-induced *LUC* expression were selected as the mapping population for further experiments.

Col-0 and C24 are two ecotypes of *Arabidopsis*, and SSLP (simple sequence length polymorphism) markers can be used to distinguish the two types of alleles in the F2 mapping population. The recombination frequency (RF) can reflect the genetic distance between two genes. The RF of unlinked genes is roughly 50%. Linked genes cannot segregate independently during germ cell meiosis so that their RF is <50%. In our case, the loci close to the enhancer mutation should show linkage with the phenotype. Twenty SSLP markers, which are evenly distributed in the five *Arabidopsis* chromosomes, were selected and examined in the mapping population. According to the mapping table, 6 Mb to 8 Mb of Chromosome 4 showed linkage to the phenotype of the *rte1* mutant (Table 3.1).

3.4 Discussion

The C24-luc system is an important genetic screening system. The first identified anti-silencing factor, ROS1, was obtained with this system [2]. The other important anti-silencing factor obtained with this system was ROS3 [5]. To date, these are the only anti-silencing factors screened from this system. Although obtaining additional anti-silencing factors with this system is difficult, in the current study we used the C24-luc system and *ros3* enhancer screening to identify another factor involved in anti-silencing, RTE1. Genetic mapping was performed, and 6 Mb to 8 Mb of Chromosome 4 showed linkage to the phenotype of the *rte1* mutant. For unknown reasons,
the stress-induced transgene expression is not stable, so that the luminescence phenotype is not fully linked with the mutation in the \textit{rte1} mutant (data not shown). As a result, we could not use genetic mapping to better define the region containing the \textit{rte1} mutation.

Although the gene responsible for the anti-silencing phenotype of the \textit{rte1} mutant remains unknown, the anti-silencing role of \textit{RTE1} was supported and confirmed by the analysis of the \textit{rte1} mutant and \textit{rte1} single mutants. In the \textit{rte1} mutant, both the transgene and endogenous gene driven by the RD29A promoter were silenced because of DNA hypermethylation at the RD29A promoter region (Fig.3.2). In \textit{rte1} single mutants, the luciferase transgene was silenced (Fig.3.6), and hundreds of DNA hypermethylated regions were identified through whole-genome bisulfide sequencing (Fig.3.9). To exclude the possibility that RTE1 is a known gene, we sequenced all known anti-silencing genes but found no homozygous mutation (data not shown), indicating that RTE1 is a new component involved in anti-silencing and in preventing DNA hypermethylation. Considering its novelty and its confirmed anti-silencing phenotype, additional research is warranted to identify and characterize the mutation in the \textit{rte1} mutant.

Although \textit{RTE1} was found by \textit{ros3} enhancer screening, the anti-silencing role of \textit{RTE1} does not depend on ROS3. That inference is supported by the repressed transgene expression and DNA hypermethylation phenotype in the \textit{rte1} single mutants. Because \textit{rte1} mutant was obtained from \textit{ros3} enhancer screening, it is an \textit{rte1} and \textit{ros3} double mutant. Comparing the luminescence phenotypes of the \textit{rte1} mutant and the \textit{rte1} single mutant, we found that the luciferase transgene was more silenced in the \textit{rte1} and \textit{ros3} double mutant than in the \textit{rte1} single mutant, indicating that \textit{RTE1} and \textit{ROS3} have additive effects. We therefore suspect that RTE1 may function in parallel with or downstream of ROS3.

ROS3 is a small, RNA-binding protein and is co-localized with ROS1 [5]. A previous report suggested that DNA demethylation by ROS1 may be guided by RNAs bound to ROS3 [5]. Therefore, if RTE1 functions in a parallel pathway of ROS3,
RTE1 could be a protein that interprets epigenetic markers, including specific histone modifications or cytosine modification. By recognizing specific epigenetic marks, RTE1 may guide DNA demethylases, like ROS1, to their target loci. ROS1 removes methylcytosine from the DNA backbone through its glycosylase/lyase activity. This leaves a nucleotide gap, which must be refilled with an unmethylated cytosine by DNA repair-related enzyme activities. If RTE1 functions downstream of ROS3, RTE1 could be a DNA repair-related enzyme that works with ROS1 to perform DNA demethylation. However, the \textit{rte1} mutant was not sensitive to MMS (methyl methanesulfonate) (data not shown), which is a chemical that damages DNA. It follows that if RTE1 is an enzyme, it is probably a methyl-cytosine-specific rather than a general DNA repair factor.

As noted, \textit{rte1} was isolated through an enhancer genetic screening for anti-silencing factors. Our results suggest that RTE1 is a previously unknown component involved in anti-silencing and in preventing DNA hypermethylation. Additional research is required to determine which gene is responsible for the silencing phenotype of the \textit{rte1} mutant. The further characterization of the \textit{rte1} mutant should increase our understanding of how active DNA demethylation is regulated in plants, and may also provide clues regarding the regulatory mechanisms of active DNA demethylation in mammals.

### 3.5 Material and methods

#### 3.5.1 Plant materials and growth conditions

All plants were \textit{Arabidopsis} in the C24 genetic background. C24-luc plants contain transgenes, including 35S promoter-driven \textit{NPTII} (\textit{pro35S: NPTII}) and \textit{RD29A} promoter-driven \textit{LUC} (\textit{proRD29A:LUC}). Seeds were disinfested in 50% (vol/vol) bleach for 15 minutes and then washed five times with sterilized water. After stratification at 4C for 48 h, the seeds were plated on $\frac{1}{2}$MS medium. Plants were grown at 23C in a growth chamber with 16 h of light and 8 h of darkness. Kanamycin
sensitivity was tested on \( \frac{1}{2} \) MS with 50 mM kanamycin for 2 weeks. Cold stress was imposed by exposure to 4C for 24 h. Salt stress was imposed by exposure to a 300 mM NaCl solution.

### 3.5.2 Mutant screen

For screening \( \text{ros3} \) enhancer mutants, \( \text{ros3} \) seeds were mutagenized with ethyl methanesulfonate (EMS). Roughly 20000 individual lines from the M2 population were used for screening \( \text{ros3} \) enhancer mutants. The seeds were plated on MS medium with C24-LUC and \( \text{ros3} \) as controls. Seedlings were grown for 12 days, and exposed to cold treatment (4C for 24 h) to induce LUC expression. After cold treatment, 1 mM luciferin, the substrate of luciferase, was sprayed on the surface of the 12-day-old seedlings. Then the seedlings were placed in a dark chamber equipped with a CCD camera (charged couple device). Five minutes of exposure time was used to detect the emission of light from the seedlings. With the above screening, about 500 seedlings with reduced LUC emission compared to \( \text{ros3} \) were chosen and transferred from MS plates to soil. The seeds of individual lines were harvested separately. Lower induced luminescence emission of these putative mutant lines was confirmed by using the same method described earlier in the paragraph. After one round of confirmation, the number of putative lines was reduced to 62. For these 62 putative lines, another round of confirmation was performed. Through two rounds of tests, the phenotype of one enhancer mutant, referred to as \( \text{rte1} \), was confirmed.

### 3.5.3 Real-Time PCR

Total RNA was extracted with Trizol reagent (Ambion, cat#15596-026) from 12-day-old seedlings following the manufacturer’s instructions, DNA contamination was removed from the total RNA using the Turbo DNase (Ambion, cat#AM2238) kit. A 10-\( \mu \)g quantity of total RNA was treated with 1 \( \mu l \) of DNase I in a 20-\( \mu l \) reaction for 30 min at 37C, and then the reaction was stopped by treatment with 2 \( \mu l \) of inactivation
reagent for 5 min at room temperature. Reverse transcription was performed with 1 μg of RNA and oligo dT primers using the qScript cDNA SuperMix kit (Quanta, cat# 95048-025). Then, 1 μl of cDNA was used as template in each qPCR reaction. Real-time PCR were carried out on a BIO-RAD machine and using PerfeCTa SYBR Green FastMix (Quanta, cat# 95072-250). ACTIN was used as an internal control.

3.5.4 Chop-PCR

Genomic DNA was extracted from 12-day-old mutants and wild-type seedlings. After RNase (invitrogen) treatment at 37°C for 1 h, DNA was purified with phenochloroform-isopropyl (Thermo, cat# 17909) and precipitated by 2V ethanol and 1/10V NaOAC (3 M, pH 5.2). In a 30-μl reaction, methylation-sensitive enzymes were used to digest 500 ng of DNA for 16 h, and 2 μl was used for PCR amplification.

3.5.5 Individual-locus bisulfite sequencing

The DNeasy Plant Mini Kit (QIAGEN) was used to extract genomic DNA from 3-week-old seedlings. According to the manufacturer’s protocol, the BisulFlash DNA Modification Kit (Epigentek) was used to convert 200 ng of genomic DNA with sodium bisulfite. For each 20-μl PCR reaction, 2 μl of bisulfite-treated DNA was used. The Wizard SV Gel and PCR Clean-Up system (Promega) were used to purify PCR products, which were later subcloned into the T-easy vector (Promega) following the manufacturer’s protocol. Twenty independent clones were sequenced from each genetic background to calculate the methylation level.

3.5.6 Whole-genome bisulfite sequencing

Genomic DNA was extracted from 12-day-old seedlings of the rte1 single mutant #12 and wild-type plants. The bisulfite-sequencing library was constructed and sequenced by the Purdue Genomics Core Facility.
3.6 References


3.7 Figures and tables

Fig. 3.1. Luminescence phenotype of the *rte1* mutant.
(A) Luminescence phenotype of 12-day-old seedlings. After 12-day-old seedlings were treated at 4C for 24 h, their luminescence signals were assessed. A luminescence signal was not observed in *rte1* or in the positive control, *ros1-1*. The diagram on the right indicates the position of the different seedlings.
(B) Luminescence phenotype of 3-week-old seedlings. Three-week-old seedlings were placed on a plate containing a 300 mM NaCl solution and then were kept at 4C for 6 h. The luminescence signal was weaker for *rte1* than for *ros3*. The diagram on the right indicates the position of the different seedlings.
(C) Luminescence phenotype of mature leaves following both cold and salt treatment. For C24-luc or each mutant, three pieces of leaves were taken from mature plants. The stress treatment was same as described in (B). The luminescence phenotype of mature leaves was consistent with that of 12-day-old and 3-week-old seedlings.
Fig. 3.2. Silencing of RD29A promoter-driven transgenic and endogenous genes in the rte1 mutant.

(A, B) Transcript levels of LUC and RD29A in the wild type, ros1-1, ros3, and rte1. RNA from 12-day-old seedlings was used for qRT-PCR. Plants was treated at 4C for 24 h to induce the expression of transgenic and endogenous genes. Both transgenic (LUC) and endogenous (RD29A) genes driven by the RD29A promoter were suppressed in the rte1 mutant.

(C) Cold-induced expression level of the control gene, COR15A. RNA extraction and cold treatment were the same as described in (A, B). COR15A was well induced in all tested plants, indicating an effective cold treatment.
Fig. 3.3. Silencing of the 35S promoter-driven NPTII transgene in the rte1 mutant.

(A) Kanamycin resistance phenotype of rte1 mutant. Kanamycin resistance was assessed on 1/2 MS medium containing 50 μM kanamycin. Mutant and wild-type seedlings grew normally on 1/2 MS. Kanamycin sensitivity was much greater in rte1 than in ros3 or C24-luc.

(B) Expression level of NPTII. RNA from 12-day-old seedlings was used for qRT-PCR. The expression of NPTII was completely repressed in rte1 and ros1, while some transcripts were detected in ros3. The results were consistent with the kanamycin-resistant phenotype.
Fig. 3.4. Examination of the *rte1* methylation status using Chop-PCR.
Chop-PCR was performed for each mutant and the wild type with four biological replicates. For each reaction, an equal amount of genomic DNA was digested by methylation-sensitive restriction enzymes. BsmAI was used for the endogenous *RD29A* promoter marker. BstUI was used for markers at transgenic *RD29A* promoter and *At4g18650* promoter. HhaI was used for the Pm36 marker. *ACTIN2* served as a non-digestion control.
Fig. 3.5. Methylation level of the *RD29A* promoter region in the *rte1* mutant.

(A) A diagram showing the bisulfite sequencing mechanism. “C” refers to unmethylated cytosine, while “mC” refers to methylated cytosine. The unmethylated DNA is displayed above, and the methylated DNA is displayed below. “U” in red represents the uracil converted from unmethylated cytosine. In the PCR reaction, adenine (A) will be paired with uracil (U).

(B) Methylation level of the RD29A promoter region. Endogenous and transgenic *RD29A* promoters were not differentiated in this analysis. *ros1-1* served as a positive control.
Fig. 3.6. Isolation of *rte1* single mutants. 
(A) Luminescence phenotype of mature leaves of *rte1* single mutants. Two to three leaves are displayed for each mutant or the wild type. Two individual lines of *rte1* single mutants were selected, including *rte1* single mutant #12 and *rte1* single mutant #19. *LUC* gene expression was induced by a 6-h exposure to a 300 mM NaCl solution at 4°C. 
(B) Luminescence phenotype of seedlings of the *rte1* single mutant. Twelve-day-old seedlings were used to check the luminescence phenotype. *LUC* gene expression was induced by a 24-h cold treatment (at 4°C).
Fig. 3.7. Kanamycin-resistance phenotype of the \textit{rte1} single mutant. C24-luc served as a negative control, and \textit{ros1-1} served as a positive control. Kanamycin resistance was tested on $\frac{1}{2}$ MS medium containing 50 mM kanamycin. Fourteen-day-old seedlings of the \textit{rte1} single mutant #12 were more sensitive to kanamycin than C24-luc.
Fig. 3.8. Methylation status of the RD29A promoter in the "rte1" single mutant #12.

(A) Examination of the Chop-PCR marker at the endogenous RD29A promoter in the "rte1" single mutant #12. Chop-PCR for the mutant and wild type were performed with two biological replicates. BstUI was used for endogenous RD29A promoter marker. ACTIN2 served as a non-digestion control.

(B) Methylation status of the RD29A gene in the wild type, the "ros1-1" mutant, and the "rte1" single mutant #12. The bisulfite sequencing result is displayed as a screenshot of genome browser tracks. The upper and middle lanes indicate the methylation status in C24-luc and ros1-1, respectively. The lower lane indicates the methylation status in the "rte1" single mutant #12. Each orange bar represents a methyl cytosine, and the height of the bar indicates the methylation level of the single cytosine. RD29A gene was diagramed as above the screenshot.
Fig. 3.9. Analysis of the genome-wide DNA methylation status in the \emph{rte1} single mutant #12.
(A) Hyper- and hypo-methylated DMRs in the \emph{rte1} single mutant #12. Whole-genome bisulfide sequencing was performed using 12-day-old seedlings. Using wild type plants as control, hyper- and hypo methylated regions in \emph{rte1} single mutant #12 were determined.
(B) The composition of hypermethylated regions in the \emph{rte1} single mutant #12. The percentage of each type (either TE, gene, or intergenic region) is indicated.
Fig. 3.10. Screenshots showing endogenous hypermethylated regions in the \textit{rte1} single mutant #12.
Total DNA methylation is labeled with an orange bar. CG methylation is labeled with a yellow bar. CHG methylation is labeled with a blue bar. CHH methylation is labeled with a pale pink bar. Each bar represents a methyl cytosine, and the height of the bar indicates the methylation level of the single cytosine. For each type of methylation, the methylation status of C24-luc is displayed in the upper lane, and the methylation status of the \textit{rte1} single mutant #12 is displayed in the lower lane. The position of each region is labeled above the screenshot, and the corresponding gene is diagramed at the bottom of the screenshot.
Table 3.1.

Rough mapping of the rte1 mutant.

*rte1* was crossed with Col-0. In the F2 population, seedlings with an abnormal luminescence phenotype were selected and were genotyped for the presence of *proRD29A-LUC*. Ninety-six seedlings were used for the rough mapping of *rte1* mutants. The SSLP markers were named with three digits indicating the position of the marker (*e.g.* 416 is the SSLP marker near 16Mbp on Chr.4).

<table>
<thead>
<tr>
<th>Chr.</th>
<th>102</th>
<th>108</th>
<th>113</th>
<th>119</th>
<th>127</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio of Col-0 alleles to total alleles</td>
<td>52%</td>
<td>49%</td>
<td>53%</td>
<td>52%</td>
<td>46%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chr.</th>
<th>203</th>
<th>210</th>
<th>214</th>
<th>218</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio of Col-0 alleles to total alleles</td>
<td>48%</td>
<td>50%</td>
<td>54%</td>
<td>57%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chr.</th>
<th>301</th>
<th>312</th>
<th>319</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio of Col-0 alleles to total alleles</td>
<td>46%</td>
<td>50%</td>
<td>43%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chr.</th>
<th>401</th>
<th>406</th>
<th>408</th>
<th>409</th>
<th>416</th>
<th>418</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio of Col-0 alleles to total alleles</td>
<td>25%</td>
<td>20%</td>
<td>20%</td>
<td>20%</td>
<td>28%</td>
<td>25%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chr.</th>
<th>509</th>
<th>515</th>
<th>522</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio of Col-0 alleles to total alleles</td>
<td>51%</td>
<td>50%</td>
<td>61%</td>
</tr>
</tbody>
</table>
4. SUMMARY

In this study, we designed two genetic screens to identify new anti-silencing factors which may be involved in DNA demethylation and prevention of transcriptional gene silencing. One genetic screen is based on the stress inducible \textit{RD29A} promoter fused with the firefly luciferase reporter gene (\textit{RD29A::LUC}), and the second genetic screen is based on cauliflower mosaic virus \textit{35S} promoter-driven \textit{SUC2} (\textit{35S::SUC2}) transgene. Using these two systems, we identified two new anti-silencing factors, MBD7 and RTE1. MBD7 is a methyl-CpG-binding protein. MBD7 is required for the proper expression of the \textit{35S::SUC2} transgene by preventing DNA hypermethylation at \textit{35S} promoter region. In addition, thousands of DMRs have been identified in \textit{mbd7-1} mutants through whole genome bisulfite sequencing, supporting the role of MBD7 in preventing genome-wide DNA hypermethylation. Using ChIP-seq assays, we found that MBD7 binding is correlated with CG methylation density. The MBD7 binding to its genomic targets is associated with its anti-silencing role.

From the \textit{RD29A::LUC} system, \textit{RTE} was identified as a putative new anti-silencing factor. RTE1 is required for the proper expression of the \textit{RD29A::LUC} transgene by preventing DNA hypermethylation at the \textit{RD29A} promoter region. Whole genome bisulfite sequencing results revealed that RTE1 is required for preventing DNA hypermethylation at hundreds of genomic loci. Due to difficulties associated with the complex and sometimes unstable silencing of the \textit{RD29A::LUC} in the mapping population generated from crosses between C24 \textit{rte1} and Col wild type, we have not been able to precisely map the \textit{rte1} mutation to identify the responsible gene for the silencing of \textit{RD29A::LUC} in the \textit{rte1} mutant. Further efforts are needed in order to identify the \textit{RTE1} gene in the future.

Two types of anti-silencing factors can be found from these two genetic screens: components required for prevention of silencing of both \textit{RD29A} and \textit{35S} promoters
and components required for prevention of silencing of only the 35S promoter. For examples, the rosl mutant was first identified using the RD29A::LUC system, and recently more alleles of rosl mutants were isolated from the 35S-SUC2 systems. Therefore, ROS1 is required for the anti-silencing of both the RD29A::LUC and 35S::SUC2 transgenes. Unlike ROS1, IDM1 and IDM2 are only required for the anti-silencing of 35S::SUC2 and 35S::NPTII, but not required for the anti-silencing of RD29A::LUC, suggesting that different mechanisms exist for the anti-silencing of the transgenic RD29A promoter and transgenic 35S promoter.

Based on this study and previous studies, the two genetic screening systems appear to have the following similarities. First, in the RD29A::LUC system, multiple copies of the RD29A:LUC transgenes are inserted into the genome. In the 35S::SUC2 system, there are three transgenes which are 35S::SUC2, 2x35S::NPTII, and 2x35S::HPTII. Therefore, multiple copies of the 35S promoter exist in the WT plants. Second, at the promoter regions of both RD29A::LUC and 35S::SUC2, there are high levels of DNA methylation in wild type plants, although the methylation is still insufficient to cause transgene silencing. In the anti-silencing mutants, the high level DNA methylation spreads to the neighboring regions and the spreading causes the silencing of RD29A::LUC and 35S::SUC2.

The two systems also have important differences. This difference may help us understand why different factors are required for the silencing of transgenic RD29A and 35S promoters, respectively. Although high levels of DNA methylation exist at both the RD29A and 35S promoter regions in wild type plants, the types of DNA methylation at these two promoters regions are different. At the RD29A promoter, DNA methylation mainly occurs in the CHG and CHH sequence contexts. In contrast, the DNA methylation at the 35S promoter region is mostly CG methylation. The MBD7 binding to the target loci is correlated with the CG methylation density but not with CHG and CHH methylation densities as shown in Chapter 2. MBD7 binding to the 35S promoter appears to be due to the high CG methylation density at the 35S promoter region, whereas the transgenic RD29A promoter does not seem to have
enough CG density to allow MBD7 binding. In summary, through the genetic screens we found two new anti-silencing factors, and our characterization of the mbd7 and rte1 mutants has contributed to a better understanding of active DNA demethylation and anti-silencing mechanisms in plants.
VITA
VITA

Zhaobo Lang
1314 Neil Armstrong Dr. Apt 8
West Lafayette, IN 47906
(650)308-6971
email: lang.katrina@gmail.com

Education
September 2005-June 2009: B.S. in Biotechnology, Beijing Normal University, Beijing, China
September 2006-June 2009: Minor in International Economic and Trade, Beijing Normal University, Beijing, China
September 2009-December 2010: Graduate student in Plant Science Department of Botany and Plant Sciences, University of California, Riverside
January 2011-present: Ph.D student in Plant Molecular Genetics, Department of Horticulture and Landscape Architecture, Purdue University

Research Experience
September 2008-June 2009: Visiting student, National Institute of Biological Sciences, Beijing. Advisor: Dr. Yan Guo
September 2009-present: Graduate Research Assistantship, UC Riverside and Purdue University. Advisor: Dr. Jian-Kang Zhu