



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

A DISSERTATION FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

**Modular Configuration of DEMETER DNA
Demethylase and Its Application to
Epigenome Editing**

**DEMETER DNA 탈메틸 효소의 모듈형 구성
개발과 후성유전체 편집 기술을 위한 응용 연구**

AUGUST 2019

HOSUNG JANG

MAJOR IN HORTICULTURAL SCIENCE AND BIOTECHNOLOGY

DEPARTMENT OF PLANT SCIENCE

COLLEGE OF AGRICULTURE AND LIFE SCIENCES

THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

**Modular configuration of DEMETER DNA
demethylase and its application to epigenome editing**

UNDER THE DIRECTION OF DR. JIN HOE HUH
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL OF
SEOUL NATIONAL UNIVERSITY

BY
HOSUNG JANG

MAJOR IN HORTICULTURAL SCIENCE AND BIOTECHNOLOGY
DEPARTMENT OF PLANT SCIENCE

AUGUST 2019

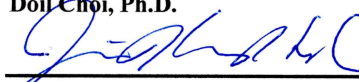
APPROVED AS A QUALIFIED DISSERTATION OF HOSUNG JANG
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
BY THE COMMITTEE MEMBERS

CHAIRMAN




Doil Choi, Ph.D.

VICE-CHAIRMAN



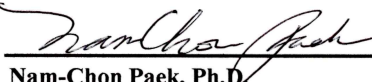
Jin Hoe Huh, Ph.D.

MEMBER



Cecile Segonzac, Ph.D.

MEMBER



Nam-Chon Paek, Ph.D.

MEMBER



Sangho Lee, Ph.D.

Modular Configuration of DEMETER DNA Demethylase and Its Application to Epigenome Editing

HOSUNG JANG

Department of Plant Science, Seoul National University

ABSTRACT

DNA methylation is a key epigenetic modification which regulates gene expression and chromatin structure in higher eukaryotes. DNA methylation is often mitotically heritable but can be removed by active mechanisms in an enzyme-dependent manner. In Arabidopsis, the DEMETER (DME) DNA glycosylase specifically excises 5-methylcytosine (5mC), generating harmful 3' blocking intermediates, which should be immediately processed by subsequent base excision repair machineries. DME and three other family members, REPRESSOR OF SILENCING 1 (ROS1), DEMETER-LIKE 2 (DML2) and DML3 share similar domain structures, but display distinct catalytic efficiencies. The DME family proteins are large, multi-domain DNA glycosylases with variable sequences

connecting conserved domains, but the contribution of these structures to the regulation of catalytic activity and the enzymes required for downstream demethylation pathway after 5mC excision are largely unknown. In this study, I extensively manipulated DME protein by reducing in size to identify the minimal regions necessary for 5mC excision activity. Domain swapping experiments revealed that the glycosylase domains of DME family members are functionally equivalent, and compatibility between conserved domains is critical for DNA demethylation *in vitro*. In addition, I demonstrated that ABASIC ENDONUCLEASE 1-LIKE (APE1L) and ABASIC ENDONUCLEASE-REDOX PROTEIN (ARP) are responsible for trimming unusual 3' end structures after 5mC excision, which proposes more complete DNA demethylation pathway in plants. Finally, I applied DME to epigenome editing by fusion with a transcription activator-like effector (TALE) DNA binding module. TALE-DME fusion protein showed delicate modulation of DNA methylation at specified genomic loci. Taken together, these studies will broaden our understanding of the fundamental regulatory mechanism of DNA methylation and transcription, and provide a promising avenue to produce various epigenetic traits by targeted DNA demethylation.

Keywords: DEMETER, DNA demethylation, domain structure, AP endonuclease, epigenome editing, TALE DNA binding module

Student number: 2013-30328

CONTENTS

ABSTRACT.....	i
CONTENTS.....	iii
LIST OF TABLES.....	vii
LIST OF FIGURES.....	viii
LIST OF ABBREVIATIONS.....	xi

GENERAL INTRODUCTION	1
-----------------------------------	----------

CHAPTER 1. Biochemical study of the domain structure of DNA demethylase

ABSTRACT.....	18
INTRODUCTION.....	20
MATERIALS AND METHODS.....	23

Construction of the minimized DME with catalytic core

Cloning of the glycosylase domain swapping constructs between DME family proteins

Cloning of domain swapping constructs between DME and DML2

Protein expression and purification

DNA substrate for radiolabeling

<i>In vitro</i> DNA glycosylase assay	
Single turnover kinetics of DME	
Electrophoretic mobility shift assay	
Sequence alignment and phylogenetic analysis	
RESULTS.....	35
Three conserved domains of DME comprise the minimal entity for 5mC excision <i>in vitro</i>	35
Conserved domains of DME family proteins are catalytically compatible <i>in vitro</i>	40
Functional motifs in the conserved domains of DME.....	49
DISCUSSION.....	52
REFERENCES.....	58

CHAPTER 2. Biochemical study of the DNA demethylation pathway

ABSTRACT.....	64
INTRODUCTION.....	65
MATERIALS AND METHODS.....	69
Expression and purification of the proteins	
<i>In vitro</i> 5mC excision assay	
<i>In vitro</i> 3' phosphodiesterase assay	

<i>In vitro</i> nucleotide incorporation assay	
<i>In vitro</i> 3' phosphatase assay	
<i>In vitro</i> AP endonuclease assay on AP sites	
Electrophoretic mobility shift assay	
RESULTS.....	78
DME and ROS1 5mC DNA glycosylases generate a 3'-PUA as a primary blocking intermediate in early 5mC excision.....	78
Enzyme-dependent δ -elimination process following 5mC excision.....	80
Arabidopsis encodes three putative AP endonucleases APE1L, APE2 and ARP.....	85
APE1L and ARP process major 5mC excision intermediate 3'-PUA to generate 3'-OH.....	86
3' phosphatase activity of ARP.....	90
Biochemical activities of MBP-free AP endonucleases.....	94
ARP has AP site incision activity.....	94
DNA binding activity of AP endonucleases.....	98
DISCUSSION.....	100
REFERENCES.....	108

CHAPTER 3. Application to epigenome editing by targeted DNA demethylation

ABSTRACT.....	115
INTRODUCTION.....	117
MATERIALS AND METHODS.....	121
Cloning of the TALE-DME fusion constructs	
Plant materials and growth conditions	
<i>Agrobacterium</i> -mediated Arabidopsis transformation	
Genotyping of the transgenic plants	
RNA isolation and gene expression analysis	
Quantitative real-time PCR (qRT-PCR) analysis	
Locus specific bisulfite sequencing	
RESULTS.....	130
Generation of transgenic plants expressing TALE-DME fusion proteins.....	130
DNA demethylation activity of TALE-DME in transgenic plants.....	132
DISCUSSION.....	140
REFERENCES.....	143
ABSTRACT IN KOREAN.....	148

LIST OF TABLES

CHAPTER 1

Table 1-1. Oligonucleotides used in this study.....	31
Table 1-2. Catalytic rate constants ($k_{\text{cat-st}}$) of purified proteins calculated from single turnover kinetics.....	39

CHAPTER 2

Table 2-1. List of oligonucleotides for biochemical assays.....	75
---	----

CHAPTER 3

Table 3-1. Oligonucleotides used in this study.....	127
Table 3-2. TALE-DME targeting sequences.....	128

LIST OF FIGURES

CHAPTER 1

Figure 1-1. Structures of manipulated DME fragments and chimeric proteins between DME and other family members.....	32
Figure 1-2. Sequences of the synthetic linkers and structures of chimeric fragments generated by domain swapping between DME and DML2	33
Figure 1-3. SDS-PAGE analysis of purified proteins used for <i>in vitro</i> 5mC excision assay.....	34
Figure 1-4. The <i>in vitro</i> 5mC excision activity of minimal DME catalytic core with three conserved domains.....	37
Figure 1-5. The conserved motifs of the DME family proteins in Arabidopsis.....	43
Figure 1-6. The <i>in vitro</i> 5mC excision activity of chimeric proteins generated by swapping glycosylase domains between DME family members.....	44
Figure 1-7. The <i>in vitro</i> 5mC excision activity of DME-DML2 chimeric proteins.....	46
Figure 1-8. Electrophoretic mobility shift assay of DME-DML2 chimeric proteins.....	48

Figure 1-9. Sequence alignment of the conserved domains of DME family proteins.....	51
---	----

CHAPTER 2

Figure 2-1. Schematic diagrams of manipulated DME and ROS1.....	76
Figure 2-2. SDS-PAGE analysis of purified proteins.....	77
Figure 2-3. The 5mC excision products generated by DME and ROS1 DNA glycosylases.....	82
Figure 2-4. Enzyme-dependent δ -elimination process during 5mC excision.....	84
Figure 2-5. <i>In vitro</i> reconstitution of DNA demethylation with Arabidopsis AP endonucleases.....	88
Figure 2-6. The 3' phosphatase activity of Arabidopsis AP endonucleases.....	92
Figure 2-7. <i>In vitro</i> AP endonuclease activities of Arabidopsis MBP-free AP endonucleases.....	96
Figure 2-8. Electrophoretic mobility shift assay of Arabidopsis AP endonucleases.....	99
Figure 2-9. Model of active DNA demethylation pathway in Arabidopsis.....	106

CHAPTER 3

Figure 3-1. Diagrams of the manipulated vectors for cloning.....	129
Figure 3-2. Schematic diagrams of TALE binding sites and TALE-DME constructs.....	131
Figure 3-3. Bisulfite conversion of genomic DNA extracted from each T1 transgenic plant.....	134
Figure 3-4. DNA demethylation activity of TD3 and TD5 TALE-DME fusion proteins in T1 transgenic plants.....	136
Figure 3-5. Expression level of <i>FWA</i> in TALE-DME T1 transgenic plants.....	138
Figure 3-6. Flowering time of TALE-DME T1 transgenic plants.....	139

LIST OF ABBREVIATION

3'-PUA	3'- phosphor- α , β -unsaturated aldehyde
5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
AP	Apurinic/aprimidinic
APE1L	Abasic endonuclease 1-like
APE2	Abasic endonuclease 2
ARP	Abasic endonuclease-redox protein
BER	Base excision repair
CGI	CpG island
CMT3	CHROMOMETHYLASE 3
CRISPR	Clustered, regularly interspaced, short palindromic repeats
DME	DEMETER
DML2	DME-LIKE 2
DML3	DME-LIKE 3
DNMT	DNA methyltransferase
DRM2	DOMAINS REARRANGED METHYLTRANSFERASE 2
EEP	Endonuclease/exonuclease/phosphatase
Endo III	Endonuclease III
Endo IV	Endonuclease IV
FACT	Facilitates chromatin transaction
FCL	4Fe-4S cluster loop
FIS2	FERTILIZATION-INDEPENDENT SEED2

FWA	FLOWERING WAGENINGEN
GMO	Genetically modified organisms
GPD	Glycine/proline-rich loop with a conserved aspartic acid
hAPE1	human AP endonuclease 1
HhH	Helix-hairpin-helix
hOGG1	human 8-oxoguanin DNA glycosylase
IDR	Interdomain region
LIG1	DNA ligase I
MEA	MEDEA
MET1	DNA METHYLTRANSFERASE 1
ncRNA	noncoding RNA
ROS1	REPRESSOR OF SILENCING 1
RRM	RNA recognition motif
SSB	Single-strand break
TALE	Transcription activator-like effector
TDG	Thymine-DNA glycosylase
TET	Ten-eleven translocation
UHRF1	Ubiquitin-like PHD and RING finger domains 1
ZDP	Zinc finger DNA 3' phosphoesterase
ZFP	Zinc finger proteins

GENERAL INTRODUCTION

DNA methylation

DNA methylation is a stable epigenetic modification crucial for gene imprinting, transposon silencing, and many developmental processes in higher eukaryotes (Huh et al., 2008; Smith and Meissner, 2013). DNA methylation is achieved by adding a methyl group on C5 position of cytosine (5-methylcytosine, 5mC), and usually occurs in the symmetric context of CG dinucleotides in mammals. The initial DNA methylation patterns are established by DNA methyltransferase 3 (DNMT3), and maintained by maintenance methyltransferase DNMT1 (Law and Jacobsen, 2010; Wu and Zhang, 2010). During cell division, DNMT1 interacts with ubiquitin-like PHD and RING finger domains 1 (UHRF1) which has strong preference for hemimethylated DNA, and methylates the cytosine on the complementary strand (Arita et al., 2008; Law and Jacobsen, 2010; Wu and Zhang, 2010; Lyco, 2018).

In plants, DNA methylation occurs at cytosines in all sequence contexts including the symmetric CG and CHG (H represents A, T or C), and the asymmetric CHH sites. Three different enzymes are responsible for catalyzing these DNA methylation patterns: CG and CHG methylation are

maintained by DNA METHYLTRANSFERASE 1 (MET1) and CHROMOMETHYLASE 3 (CMT3), respectively, whereas CHH methylation is achieved through *de novo* methylation by DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) (Law and Jacobsen, 2010; Wu and Zhang, 2010).

DNA demethylation in mammals

DNA demethylation is a reverse process of DNA methylation, which can be achieved through either passive or active mechanisms. Passive DNA demethylation occurs in a replication-dependent manner when DNA methyltransferases such as DNMT1 and MET1 are down-regulated or absent. During the early embryogenesis, both maternal and paternal genomes of the mammalian zygote undergo global DNA demethylation which is referred to as epigenetic reprogramming (Wu and Zhang, 2010; Iurlaro et al, 2017). As cell division progresses, the maternal genome experiences gradual and passive DNA demethylation with exclusion of DNMT1 from the nucleus (Carlson et al., 1992). However, paternal genome undergoes rapid genome-wide DNA demethylation immediately after fertilization, implying that this process requires an enzyme-dependent mechanism (Wu and Zhang, 2010; Guo et al., 2014).

Ten-eleven translocation (TET) family proteins are mammalian DNA demethylases which mediate the oxidation of 5mC to 5-hydroxymethylcytosine (5hmC) (Tahiliani et al., 2009). In the mammalian zygote, 5hmC accumulates in the paternal genome along with dramatic decrease of 5mC, which indicates that TET-mediated active DNA demethylation occurs in mammals (Wossidlo et al., 2011). TET proteins further catalyze the oxidation of 5hmC to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC), albeit TET proteins prefer 5mC to 5hmC or 5fC as a substrate (He et al., 2011; Ito et al., 2011; Hu et al., 2015; Wu and Zhang, 2017). The 5fC and 5caC bases can be removed by thymine-DNA glycosylase (TDG), and subsequent base excision repair (BER) enzymes such as AP endonuclease complete DNA demethylation (He et al., 2011; Maiti and Drohat, 2011; Kohli and Zhang, 2013; Wu and Zhang, 2017).

DNA demethylation in plants

Although plants and mammals have highly conserved DNA methylation systems, they have evolved distinct DNA demethylation machineries. In plants, the DEMETER (DME) DNA glycosylase family proteins catalyze direct excision of 5mC from DNA and initiate active DNA demethylation to permit transcription of target genes (Choi et al., 2002;

Gong et al., 2002; Agius et al., 2006; Gehring et al., 2006). As a bifunctional 5mC DNA glycosylase/lyase, DME catalyzes the cleavage of a sugar-phosphate backbone after 5mC excision via β - and δ -elimination reactions, producing harmful 3'-blocking intermediates such as 3'-phosphor- α , β -unsaturated aldehyde (3'-PUA) and 3'-phosphate (Gehring et al., 2006; Lee et al., 2014). These 5mC excision intermediates are processed to provide 3'-OH by BER enzymes such as AP endonuclease and 3' phosphatase for subsequent polymerization. Biochemical studies have revealed that Arabidopsis ABASIC ENDONUCLEASE 1-LIKE (APE1L) and ABASIC ENDONUCLEASE-REDOX PROTEIN (ARP) are capable of processing a major intermediate 3'-PUA, whereas ZINC FINGER DNA 3' PHOSPHOESTERASE (ZDP) and ARP are involved in processing 3'-phosphate (Martinez-Macias et al., 2012; Lee et al., 2014). Genetic studies showed that the *ape1l zdp* double mutant is lethal, suggesting their roles in DME-initiated DNA demethylation during seed development (Li et al., 2015b). Although the specific DNA polymerase involved in subsequent polymerization is not well-characterized, DNA polymerase λ presumably performs the incorporation of unmethylated cytosines to abasic sites (Uchiyama et al., 2004; Uchiyama et al., 2009), and DNA ligase I completes the final ligation step of BER (Córdoba-Cañero et al., 2011; Li et al., 2015a).

Function and structure of the plant DNA demethylase

Recent genome-wide DNA methylome analysis revealed that DME removes DNA methylation at numerous target loci in the Arabidopsis central cell, and the resulting changes in DNA methylation patterns are inherited by the fertilized endosperm (Park et al., 2016). During early endosperm development, DME establishes maternal-specific expression of *MEDEA (MEA)*, *FERTILIZATION-INDEPENDENT SEED2 (FIS2)* and *FLOWERING WAGENINGEN (FWA)*, while the paternal alleles remain methylated and silenced (Choi et al., 2002; Kinoshita et al., 2004; Gehring et al., 2006; Jullien et al., 2006). In particular, DME specifically removes 5mC of maternal *MEA* allele for gene activation, which is required for endosperm and seed development. Consequently, homozygous *dme* mutants are embryonic lethal, whereas the *DME/dme* heterozygote produces 50% of viable seeds (Choi et al., 2002). On the other hand, three additional DME family members are reported in Arabidopsis - *REPRESSOR OF SILENCING 1 (ROS1)*, *DME-LIKE 2 (DML2)* and *DML3*. ROS1 is required for regulation of transgene expression and production of stomatal lineage cells, whereas both DML2 and DML3 appear to facilitate DNA demethylation to regulate proper development in somatic tissues (Gong et al., 2002; Penterman et al., 2007; Yamamuro et al., 2014).

The DME family proteins have a modular structure, and share three conserved C-terminal domains (the glycosylase domain and flanking domains A and B), albeit they show distinct 5mC excision efficiencies *in vitro* (Penterman et al., 2007; Mok et al., 2010; Ponferrada-Marin et al., 2011). The DME family belongs to the helix-hairpin-helix (HhH) superfamily of DNA glycosylase, which contains an HhH motif and a glycine/proline-rich loop with a conserved aspartic acid (GPD) in the central glycosylase domain. The catalytic aspartic acid and lysine residues in the HhH-GPD motif are essential for DME activity (Choi et al., 2004; Gehring et al., 2006; Mok et al., 2010). Homology modeling of DME suggests that domain A and the glycosylase domain are involved in composition of the 5mC recognition pocket (Ponferrada-Marin et al., 2011; Brooks et al., 2014), and domain A is also required for non-specific DNA binding activity through its mixed charge cluster motif (Mok et al., 2010). Domain B contains a permuted CXXC and a divergent RNA recognition motif (RRM), but its function remains largely unknown (Iyer et al., 2011).

Targeting mechanism of the plant DNA demethylase

In contrast to the global DNA demethylation in mammals, no strong evidence exists for genome-wide demethylation in plants (Zhu, 2009). It is

plausible that certain guiding molecules are required for DME targeting. It is still largely unknown how DME family proteins are guided to target loci to initiate DNA demethylation and transcriptional activation, but several studies have provided some clues about targeting mechanism of DME. DME family proteins might be guided to target sequences by interacting with other chromatin binding modules. Indeed, there are numerous cases in which histone modifying enzymes such as histone acetyltransferases and histone methyltransferases are recruited to the target sites via the interaction with bromodomain or chromodomain proteins that bind to acetylated histones and methylated DNA, respectively (Marmorstein and Zhou, 2014; Patel, 2016). In *Arabidopsis*, histone H1.2 was identified as a DME-interacting protein via yeast two-hybrid and *in vitro* pull-down assays (Rea et al., 2012), suggesting that DME can also be guided by chromatin-associated proteins. In addition, DME family proteins can be recruited to target genomic regions by recognizing specific chromatin signatures unique to DNA demethylation targets. These signatures may include open chromatin structure with low histone density, or novel histone modifications required for active demethylation (Qian et al., 2012).

Alternatively, DME family proteins may utilize some transcription-associated patterns such as mRNA transcripts or noncoding RNAs

(ncRNAs). It is noteworthy that some Polycomb group complexes are guided to their targets using ncRNA as an adaptor molecule in animals and plants (Fitzpatrick et al., 2002; Mak et al., 2002; Rinn et al., 2007; Heo and Sung, 2011). Although the involvement of RNA molecules in DME targeting is largely hypothetical, the presence of CXXC and RRM motifs in domain B suggests their potential role as RNA binding modules for the guidance. In addition, RNA binding proteins may guide DME family proteins to target loci, as exemplified by ROS3 which is a small RNA binding protein that guides ROS1 for sequence-specific DNA demethylation (Zheng et al., 2008), albeit the direct interaction between ROS1 and ROS3 is still elusive.

DME is reported to possess two classes of target sites including GC-rich heterochromatin regions and accessible euchromatin regions (Frost et al., 2018). First, DME is known to target intergenic and heterochromatin regions for DNA demethylation (Ibarra et al., 2012). The Arabidopsis Facilitates chromatin transaction (FACT) histone chaperone complex is required for guidance of DME to heterochromatic targets and specific imprinted genes (Ikeda et al., 2011; Frost et al., 2018). The FACT complex is conserved in diverse organisms including yeast, mammals and plants, and is required for transcription of RNA polymerase II through nucleosomal

templates (Hondele and Ladurner, 2013; Hsieh et al., 2013). This suggests that FACT complex might also assist DME demethylation through nucleosome-rich heterochromatin regions (Frost et al., 2018). Second, DME appears to be preferentially guided to less compact euchromatin regions, and many of DME targets in the central cell are found near the promoters of imprinted genes such as *MEA*, *FWA* and *FIS2* (Choi et al, 2002; Kinoshita et al., 2004; Gehring et al., 2006; Jullien et al., 2006), and even euchromatic transposable elements (Ibarra et al., 2012). These studies reveal the function of DNA demethylase and enhance our understanding of DNA methylation dynamics in plants.

The thesis work focused on the biochemical study of the plant DNA demethylase and downstream events of the DNA demethylation pathway. This study also includes the application of plant DNA demethylase to epigenome editing and addresses the following three topics:

Chapter 1: Biochemical study of the domain structure of DNA demethylase

Chapter 2: Biochemical study of the DNA demethylation pathway

Chapter 3: Application to epigenome editing by targeted DNA demethylation

REFERENCES

- Agius, F., Kapoor, A., and Zhu, J.K.** (2006). Role of the Arabidopsis DNA glycosylase/lyase ROS1 in active DNA demethylation. *Proc. Natl. Acad. Sci. USA* **103**: 11796-11801.
- Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y., and Shirakawa, M.** (2008). Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature* **455**: 818-821.
- Brooks, S.C., Fischer, R.L., Huh, J.H., and Eichman, B.F.** (2014). 5-methylcytosine recognition by *Arabidopsis thaliana* DNA glycosylases DEMETER and DML3. *Biochemistry* **53**: 2525-2532.
- Carlson, L.L., Page, A.W., and Bestor, T.H.** (1992). Properties and localization of DNA methyltransferase in preimplantation mouse embryos: implications for genomic imprinting. *Genes Dev.* **6**: 2536-2541.
- Choi, Y., Gehring, M., Johnson, L., Hannon, M., Harada, J.J., Goldberg, R.B., Jacobsen, S.E., and Fischer, R.L.** (2002). DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in Arabidopsis. *Cell* **110**: 33-42.
- Cordoba-Canero, D., Roldan-Arjona, T., and Ariza, R.R.** (2011). Arabidopsis ARP endonuclease functions in a branched base excision DNA repair pathway completed by LIG1. *Plant J.* **68**: 693-702.
- Fitzpatrick, G.V., Soloway, P.D., and Higgins, M.J.** (2002). Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1. *Nat. Genet.* **32**: 426-431.

- Frost, J.M., Kim, M.Y., Park, G.T., Hsieh, P.H., Nakamura, M., Lin, S.J.H., Yoo, H., Choi, J., Ikeda, Y., Kinoshita, T., Choi, Y., Zilberman, D., and Fischer, R.L.** (2018). FACT complex is required for DNA demethylation at heterochromatin during reproduction in Arabidopsis. *Proc. Natl. Acad. Sci. USA* **115**: E4720-E4729.
- Gehring, M., Huh, J.H., Hsieh, T.F., Penterman, J., Choi, Y., Harada, J.J., Goldberg, R.B., and Fischer, R.L.** (2006). DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. *Cell* **124**: 495-506.
- Gong, Z., Morales-Ruiz, T., Ariza, R.R., Roldan-Arjona, T., David, L., and Zhu, J.-K.** (2002). ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase. *Cell* **111**: 803-814.
- Guo, F., Li, X., Liang, D., Li, T., Zhu, P., Guo, H., Wu, X., Wen, L., Gu, T.P., Hu, B., Walsh, C.P., Li, J., Tang, F., and Xu, G.L.** (2014). Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. *Cell Stem Cell* **15**: 447-459.
- He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C.X., Zhang, K., He, C., and Xu, G.L.** (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**: 1303-1307.
- Heo, J.B., and Sung, S.** (2011). Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science* **331**: 76-79.
- Hondele, M., Stuwe, T., Hassler, M., Halbach, F., Bowman, A., Zhang, E.T., Nijmeijer, B., Kotthoff, C., Rybin, V., Amlacher, S., Hurt, E.,**

- and Ladurner, A.G.** (2013). Structural basis of histone H2A-H2B recognition by the essential chaperone FACT. *Nature* **499**: 111-114.
- Hsieh, F.K., Kulaeva, O.I., Patel, S.S., Dyer, P.N., Luger, K., Reinberg, D., and Studitsky, V.M.** (2013). Histone chaperone FACT action during transcription through chromatin by RNA polymerase II. *Proc. Natl. Acad. Sci. USA* **110**: 7654-7659.
- Hu, L., Lu, J., Cheng, J., Rao, Q., Li, Z., Hou, H., Lou, Z., Zhang, L., Li, W., Gong, W., Liu, M., Sun, C., Yin, X., Li, J., Tan, X., Wang, P., Wang, Y., Fang, D., Cui, Q., Yang, P., He, C., Jiang, H., Luo, C., and Xu, Y.** (2015). Structural insight into substrate preference for TET-mediated oxidation. *Nature* **527**: 118-122.
- Huh, J.H., Bauer, M.J., Hsieh, T.F., and Fischer, R.L.** (2008). Cellular programming of plant gene imprinting. *Cell* **132**: 735-744.
- Ibarra, C.A., Feng, X., Schoft, V.K., Hsieh, T.F., Uzawa, R., Rodrigues, J.A., Zemach, A., Chumak, N., Machlicova, A., Nishimura, T., Rojas, D., Fischer, R.L., Tamaru, H., and Zilberman, D.** (2012). Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. *Science* **337**: 1360-1364.
- Ikeda, Y., Kinoshita, Y., Susaki, D., Ikeda, Y., Iwano, M., Takayama, S., Higashiyama, T., Kakutani, T., and Kinoshita, T.** (2011). HMG domain containing SSRP1 is required for DNA demethylation and genomic imprinting in Arabidopsis. *Dev. Cell* **21**: 589-596.
- Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y.** (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**: 1300-1303.
- Iurlaro, M., von Meyenn, F., and Reik, W.** (2017). DNA methylation homeostasis in human and mouse development. *Curr. Opin. Genet.*

Dev. **43**: 101-109.

- Iyer, L.M., Abhiman, S., and Aravind, L.** (2011). Natural history of eukaryotic DNA methylation systems. *Prog. Mol. Biol. Transl. Sci.* **101**: 25-104.
- Jullien, P.E., Kinoshita, T., Ohad, N., and Berger, F.** (2006). Maintenance of DNA methylation during the Arabidopsis life cycle is essential for parental imprinting. *Plant Cell* **18**: 1360-1372.
- Kinoshita, T., Miura, A., Choi, Y., Kinoshita, Y., Cao, X., Jacobsen, S.E., Fischer, R.L., and Kakutani, T.** (2004). One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation. *Science* **303**: 521-523.
- Kohli, R.M., and Zhang, Y.** (2013). TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* **502**: 472-479.
- Law, J.A., and Jacobsen, S.E.** (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **11**: 204-220.
- Lee, J., Jang, H., Shin, H., Choi, W.L., Mok, Y.G., and Huh, J.H.** (2014). AP endonucleases process 5-methylcytosine excision intermediates during active DNA demethylation in Arabidopsis. *Nucleic Acids Res.* **42**: 11408-11418.
- Li, Y., Duan, C.G., Zhu, X., Qian, W., and Zhu, J.K.** (2015a). A DNA ligase required for active DNA demethylation and genomic imprinting in Arabidopsis. *Cell Res.* **25**: 757-760.
- Li, Y., Cordoba-Canero, D., Qian, W., Zhu, X., Tang, K., Zhang, H., Ariza, R.R., Roldan-Arjona, T., and Zhu, J.K.** (2015b). An AP endonuclease functions in active DNA demethylation and gene imprinting in Arabidopsis. *PLoS Genet.* **11**: e1004905.

- Lyko, F.** (2018). The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nat. Rev. Genet.* **19**: 81-92.
- Maiti, A., and Drohat, A.C.** (2011). Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J. Biol. Chem.* **286**: 35334-35338.
- Mak, W., Baxter, J., Silva, J., Newall, A.E., Otte, A.P., and Brockdorff, N.** (2002). Mitotically stable association of polycomb group proteins *eed* and *enx1* with the inactive x chromosome in trophoblast stem cells. *Curr. Biol.* **12**: 1016-1020.
- Marmorstein, R., and Zhou, M.M.** (2014). Writers and readers of histone acetylation: structure, mechanism, and inhibition. *Cold Spring Harb. Perspect. Biol.* **6**: a018762.
- Martinez-Macias, M.I., Qian, W., Miki, D., Pontes, O., Liu, Y., Tang, K., Liu, R., Morales-Ruiz, T., Ariza, R.R., Roldan-Arjona, T., and Zhu, J.K.** (2012). A DNA 3' phosphatase functions in active DNA demethylation in *Arabidopsis*. *Mol. Cell* **45**: 357-370.
- Mok, Y.G., Uzawa, R., Lee, J., Weiner, G.M., Eichman, B.F., Fischer, R.L., and Huh, J.H.** (2010). Domain structure of the DEMETER 5-methylcytosine DNA glycosylase. *Proc. Natl. Acad. Sci. USA* **107**: 19225-19230.
- Park, K., Kim, M.Y., Vickers, M., Park, J.S., Hyun, Y., Okamoto, T., Zilberman, D., Fischer, R.L., Feng, X., Choi, Y., and Scholten, S.** (2016). DNA demethylation is initiated in the central cells of *Arabidopsis* and rice. *Proc. Natl. Acad. Sci. USA* **113**: 15138-15143.
- Patel, D.J.** (2016). A structural perspective on readout of epigenetic histone and DNA methylation marks. *Cold Spring Harb. Perspect. Biol.* **8**:

a018754.

- Penterman, J., Zilberman, D., Huh, J.H., Ballinger, T., Henikoff, S., and Fischer, R.L.** (2007). DNA demethylation in the Arabidopsis genome. *Proc. Natl. Acad. Sci. USA* **104**: 6752-6757.
- Ponferrada-Marin, M.I., Parrilla-Doblas, J.T., Roldan-Arjona, T., and Ariza, R.R.** (2011). A discontinuous DNA glycosylase domain in a family of enzymes that excise 5-methylcytosine. *Nucleic Acids Res.* **39**: 1473-1484.
- Qian, W., Miki, D., Zhang, H., Liu, Y., Zhang, X., Tang, K., Kan, Y., La, H., Li, X., Li, S., Zhu, X., Shi, X., Zhang, K., Pontes, O., Chen, X., Liu, R., Gong, Z., and Zhu, J.K.** (2012). A histone acetyltransferase regulates active DNA demethylation in Arabidopsis. *Science* **336**: 1445-1448.
- Rea, M., Zheng, W., Chen, M., Braud, C., Bhangu, D., Rognan, T.N., and Xiao, W.** (2012). Histone H1 affects gene imprinting and DNA methylation in Arabidopsis. *Plant J.* **71**: 776-786.
- Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Bruggmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., and Chang, H.Y.** (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**: 1311-1323.
- Smith, Z.D., and Meissner, A.** (2013). DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14**: 204-220.
- Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., and Rao, A.** (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1.

Science **324**: 930-935.

Uchiyama, Y., Takeuchi, R., Kodera, H., and Sakaguchi, K. (2009). Distribution and roles of X-family DNA polymerases in eukaryotes. *Biochimie* **91**: 165-170.

Uchiyama, Y., Kimura, S., Yamamoto, T., Ishibashi, T., and Sakaguchi, K. (2004). Plant DNA polymerase lambda, a DNA repair enzyme that functions in plant meristematic and meiotic tissues. *Eur. J. Biochem.* **271**: 2799-2807.

Wossidlo, M., Nakamura, T., Lepikhov, K., Marques, C.J., Zakhartchenko, V., Boiani, M., Arand, J., Nakano, T., Reik, W., and Walter, J. (2011). 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat. Commun.* **2**: 241.

Wu, S.C., and Zhang, Y. (2010). Active DNA demethylation: many roads lead to Rome. *Nat. Rev. Mol. Cell Biol.* **11**: 607-620.

Wu, X., and Zhang, Y. (2017). TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat. Rev. Genet.* **18**: 517-534.

Yamamuro, C., Miki, D., Zheng, Z., Ma, J., Wang, J., Yang, Z., Dong, J., and Zhu, J.K. (2014). Overproduction of stomatal lineage cells in *Arabidopsis* mutants defective in active DNA demethylation. *Nat. Commun.* **5**: 4062.

Zheng, X., Pontes, O., Zhu, J., Miki, D., Zhang, F., Li, W.X., Iida, K., Kapoor, A., Pikaard, C.S., and Zhu, J.K. (2008). ROS3 is an RNA-binding protein required for DNA demethylation in *Arabidopsis*. *Nature* **455**: 1259-1262.

Zhu, J.K. (2009). Active DNA Demethylation Mediated by DNA Glycosylases. *Annu. Rev. Genet.* **43**: 143-166.

CHAPTER 1

Biochemical study of the domain structure of DNA demethylase

ABSTRACT

DNA methylation is a key epigenetic mark which regulates gene expression and chromatin structure in eukaryotes. DNA methylation is often mitotically heritable but can be removed by active mechanisms in a replication-independent manner. In Arabidopsis, the DEMETER (DME) DNA glycosylase specifically removes 5-methylcytosine (5mC) from DNA and induces gene imprinting essential for seed viability in endosperm. There are three other DME family members present in the Arabidopsis genome; REPRESSOR OF SILENCING 1 (ROS1), DEMETER-LIKE 2 (DML2) and DML3, which share similar domain structures but display distinct catalytic efficiencies. The DME family proteins are large, multi-domain DNA glycosylases with variable sequences connecting conserved domains, but the contribution of these structures to the regulation of catalytic activity is largely unknown. In this study, I extensively manipulated DME protein by reducing in size to identify the minimal regions necessary for 5mC excision activity. I demonstrate the modular configuration of DME forming a cassette structure for following domain swapping experiments. The central glycosylase domain of minimal cassette of DME was replaced with that of other family members, producing three chimeric cassettes. Notably,

chimeric fusion of catalytic domains dramatically restored the 5mC excision activity of DML2 *in vitro*, and further domain swapping experiment between DME and DML2 revealed that DNA binding activity is associated with catalytic activity of these chimeric proteins. These results demonstrate that the glycosylase domains of DME family members have comparable 5mC excision activity, and suggest that compatible modular configuration among three conserved domains is critical for DNA demethylation.

INTRODUCTION

DNA methylation is a primary epigenetic modification, regulating chromatin structure and gene expression in many eukaryotes (Allis and Jenuwein, 2016). DNA methylation generally refers to the addition of a methyl group onto position 5 of the pyrimidine ring of cytosine, and is catalyzed by DNA methyltransferases (DNMTs) to form 5-methylcytosine (5mC). DNA methylation is often associated with transcriptional gene silencing and regulates many developmental processes such as cell differentiation, reproduction, gene imprinting and X-chromosome inactivation (Huh et al., 2008; Smith and Meissner, 2013). In response to developmental cues, DNA methylation needs to be dynamically regulated, and the reverse process of DNA methylation also plays an essential role for transcriptional control (Wu and Zhang, 2014). DNA demethylation can be achieved through a passive or active mechanism. Passive DNA demethylation is replication-dependent and often caused by inactivation of DNMTs over successive cell divisions. In contrast, active DNA demethylation requires enzyme activity to catalyze the removal of 5mC in a replication-independent manner. In plants, the DEMETER (DME) family proteins catalyze direct excision of 5mC from DNA and initiate active DNA

demethylation to induce transcription of target genes (Choi et al., 2002; Gong et al., 2002; Gehring et al., 2006).

The Arabidopsis genome encodes four *DME* family genes including *DME*, *REPRESSOR OF SILENCING 1 (ROS1)*, *DME-LIKE 2 (DML2)* and *DML3* (Choi et al., 2002). *DME* was initially identified as a regulator of gene imprinting essential for seed development in Arabidopsis (Choi et al., 2002). *DME* is primarily expressed in the central cell (Ibarra et al., 2012; Park et al., 2017) to induce the expression of target genes such as *MEDEA (MEA)*, *FERTILIZATION-INDEPENDENT SEED2 (FIS2)*, and *FLOWERING WAGENINGEN (FWA)*, all of which are imprinted in fertilized endosperm (Choi et al., 2002; Kinoshita et al., 2004; Gehring et al., 2006; Jullien et al., 2006). In contrast, *ROS1*, *DML2* and *DML3* facilitate DNA demethylation required for proper development in sporophytic tissues (Gong et al., 2002; Penterman et al., 2007; Zhu et al., 2014). The *DME* family proteins share conserved domain structures, notably a central catalytic glycosylase domain and flanking domains A and B (hereafter referred to as GD, AD and BD, respectively), important for the 5mC excision activity of this family of proteins (Agius et al., 2006; Gehring et al., 2006; Penterman et al., 2007). These three domains are tethered by the interdomain regions (IDRs) which are highly variable in sequence and

length among the family members (Penterman et al., 2007; Mok et al., 2010).

The DME family contains a core with multiple conserved domains, and except for the well-characterized glycosylase domain, very little is known about the function of the other domains. In this study, I defined the minimal catalytic sequences necessary for 5mC excision activity of DME, while revealing functional compatibility between the members of the family. Although the four *Arabidopsis* DME family proteins have distinct 5mC excision efficiencies *in vitro*, domain swapping experiments showed that GD of DME can be substituted by those of other members, which suggests that the conserved AD and BD of each member are likely important for member-specific functions. This study also reveals an interchangeable modular configuration of DME family proteins and catalytic core of domains necessary and sufficient for DNA demethylation.

MATERIALS AND METHODS

Construction of the minimized DME with catalytic core

Oligonucleotides used in this study are listed in Table 1-1. The pBG102-DME Δ N677 Δ IDR1::lnk vector (Mok et al., 2010) was digested with *Bam* HI and *Sal* I restriction enzymes to produce a DME Δ N677 Δ IDR1::lnk fragment (referred to as DME^{CTD} Δ 1 hereafter, Figure 1-1A). The restriction fragment was cloned into a pLM302 vector (obtained from the Center for Structural Biology at Vanderbilt University), resulting pLM302-DME^{CTD} Δ 1 (Figure 1-1A). The pLM302 vector is a pET27a (EMD Biosciences) derivative that encodes an N-terminal 6xHis and a Maltose Binding Protein (MBP) for fusion with a protein of interest. Constructs used for protein expression was tagged with a 6xHis+MBP in the pLM302 vector background.

To produce the minimized DME fragment comprising three core domains (domain A, glycosylase domain, and domain B) and two synthetic linkers, using pBG102-DME^{CTD} Δ 1 as a template, primers DG124 and DG125 containing a short linker sequence L2 (CGRGSSGNGSSGNPR) were extended to opposite directions to replace an IDR2. After the treatment

with *Dpn* I, the extension product was PCR-amplified with primers DG144 and DG044 and digested with *Bam* HI and *Sal* I. The restricted fragment was cloned into the pLM302 vector at the corresponding sites creating pLM302-DME^{CTD}Δ2 (referred to as DDD hereafter, Figure 1-1A). The same procedure was used to generate pLM302-DME^{CTD}Δ1ΔL1 with primers DG219 and DG220 for extension. To make a catalytically inactive DME, a DME fragment containing a single amino acid substitution (K1544Q) from pBG102-DME^{CTD}Δ1(K>Q) (Mok et al., 2010) was cloned into the pLM302-DDD, producing pLM302-DDD(K>Q).

Cloning of the glycosylase domain swapping constructs between DME family proteins

The glycosylase domains of ROS1, DML2 and DML3 were PCR-amplified from a cDNA library with primer pairs DG175 and DG176, DG177 and DG178, DG179 and DG180 respectively. Each amplified product was substituted with glycosylase domain of DDD in a pEGFP-C1 vector background (Takara), which generated three chimeric constructs consisting of the glycosylase domains of ROS1, DML2 and DML3 flanked by the domains A and B of DME (referred to as DRD, D2D and D3D, respectively, Figure 1-1B). Three chimeric constructs and DDD were cloned

into the *Bam* HI and *Sal* I sites of the pLM302 vector for protein purification.

Cloning of domain swapping constructs between DME and DML2

The linker L2 of pEGFP-C1-D2D was replaced with a new synthetic linker L3 (CGRASSGNGSSGNPR) containing a single *Sac* II site for following cloning procedure (Figure 1-2A). Oligonucleotides DG276 and DG277 was annealed and digested with *Not* I and *Sac* II restriction enzymes. The digested linker fragment was inserted into the corresponding sites of pEGFP-C1-D2D to produce pEGFP-C1-D2D with L3. A partial fragment of D2D was inserted into the *Pst* I and *Eco* RI of pBS-DME^{CTD}Δ1 (pBlueScript II KS) to produce pBS-D2D. Finally, the D2D fragment with L3 from pBS-D2D was cloned into the *Bam* HI and *Xho* I sites of the pLM302 vector to generate pLM302-D2D with L3. Note that all DME-DML2 swapping constructs contain L3 instead of L2 between the glycosylase domain and domain B.

Domain boundaries of DML2 were determined based on secondary structure prediction and amino acid sequence homology. The domain A and B of DML2 were amplified from Arabidopsis cDNA with primer pairs DG278-DG279, and DG280-DG281, respectively. Each fragment was

substituted into the domain A and B of pLM302-D2D to produce DME-DML2 swapping constructs, pLM302-22D and pLM302-D22, respectively. At the same time, the central glycosylase domain of DME was inserted into pLM302-D2D to generate pL302-DDD that harbors L3. In the same manner, the glycosylase domain of DME and domain B of DML2 were cloned into pLM302-22D, resulting pLM302-2DD and pLM302-222, respectively. Domain A of DML2 was cloned into the pLM302-DDD resulting pLM302-2DD. The glycosylase domain of DME was cloned into the pLM302-222 producing pLM302-2D2. Finally, the glycosylase domain of D22 was replaced with that of DME to produce pLM302-DD2. Taken together, 8 possible combinations of constructs between DME and DML2 were obtained (Figure 1-2B).

Protein expression and purification

Constructs were transformed into *E. coli* Rosetta2 (DE3) strain (EMD Millipore) and cells were grown at 28°C in LB medium containing 50 µg/mL each of kanamycin and chloramphenicol antibiotics until the OD₆₀₀ reached 0.4. Protein expression was induced with 0.1 mM IPTG at 18°C for 16 h. Cells were harvested by centrifugation at 7,000 rpm for 20 min at 4°C and resuspended in 30 mL of lysis buffer (50 mM Tris-HCl, pH

7.4, 100 mM NaCl, 10% glycerol, 0.1 mM dithiothreitol, 0.1 mM PMSF). Cells were sonicated (0.5x duty cycle) on ice for 5 min and cell extracts were clarified by centrifugation at 12,000 rpm for 25 min at 4°C. The supernatant was filtered through nylon membrane with 0.45 µm pore (Advantec) and the lysate was sequentially purified by three different types of columns: affinity column (His trap FF 5 mL, GE Healthcare), ion exchange column (Heparin HP 5 mL, GE Healthcare) and size exclusion (Superdex 200, GE Healthcare). The final eluted fractions were concentrated and aliquoted with 50% glycerol and stored in a storage buffer (20 mM Tris-HCl, pH 7.4, 40 mM NaCl, 4% glycerol, 0.1 mM dithiothreitol) at -80°C until use. Note that all the proteins used in this study are tagged with an MBP (Figure 1-3).

DNA substrate for radiolabeling

The 30-mer oligonucleotide containing 5mC in the middle of the sequence (5'- CTGTGTGATACTAT[5mC]GAATTCAGTATGATC -3') and its complementary strand were synthesized (Integrated DNA Technologies). Forty pmol of oligonucleotide was radiolabeled by T4 polynucleotide kinase (Takara) with 30 µCi of [γ -³²P]ATP (Perkin Elmer). The labeled oligonucleotide was purified by Qiaquick Nucleotide Removal

Kit (Qiagen) and annealed with complementary oligonucleotides to produce double-stranded DNA substrate. The mixture was boiled in water for 10 min and slowly cooled down for annealing to room temperature for at least 3 h.

***In vitro* DNA glycosylase assay**

Radiolabeled oligonucleotide substrate (25 nM) was incubated with purified enzymes (100 nM) in the glycosylase reaction buffer (10 mM Tris-HCl, pH 7.4, 50 mM NaCl, 0.5 mM dithiothreitol, 0.2 mg/mL Bovine serum albumine) at 37°C for 1 h. Reaction was terminated by adding an equal volume of stop solution (98% formamide, 10 mM ethylenediaminetetraacetic acid (EDTA), 0.2% Xylen cyanol FF, 0.2% of bromophenol blue) and boiled at 95°C for 10 min. Reaction products were separated on a 15% denaturing polyacrylamide gel containing 7.5 M urea and 1x TBE (Tris/ Borate/ EDTA) at 1,200 V for 4 h. The gel was exposed to X-ray film (Fujifilm) at -80°C.

Single turnover kinetics of DME

The same reaction condition described in the *In vitro* DNA glycosylase assay was used for kinetics study but reactions were terminated at a various time point by adding 100 mM NaOH and boiling for 10 min.

And then one volume of stop solution (98% formamide, 10 mM EDTA, 0.2% Xylen cyanol FF, 0.2% of bromophenol blue) was added and reactions were boiled at 95°C for additional 10 min. Reaction products were loaded on a 15% polyacrylamide gel containing 7.5 M Urea and 1x TBE. After electrophoresis at 1,200 V for 4 h, the gel was exposed to a phosphorimager screen (Fujifilm) and the radioactivity was measured using the Fujifilm BAS-5000 (Fujifilm). The intensity of the amount of products relative to that of substrate was measured using the Multi Gauge V2.2 (Fujifilm). Estimation of single turnover kinetics ($k_{\text{cat-st}}$) was accorded to Begley et al. (2003). Briefly, the catalytic rate constant was calculated using the following equation: $[\text{product}] = A[1 - \exp[(-k_{\text{cat-st}}) \times \text{min}]]$, where A represents the amplitude of the exponential phase.

Electrophoretic mobility shift assay

Radiolabeled oligonucleotide (100 nM) containing 5mC was incubated with increased amount of purified enzymes (0, 7, 35, 175 nM) in the DNA binding buffer (10 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.05% Triton X-100, 0.1 mg/mL BSA, 10% glycerol, 10 mM dithiothreitol) at 25°C for 15 min. Reactions were separated on a native polyacrylamide gel

(4% acrylamide, 2.5% glycerol and 0.5x TBE) at 50 V for 2 h. The gel was exposed to X-ray film at -80°C.

Sequence alignment and phylogenetic analysis

The glycosylase domain sequences of DME family proteins were obtained from TAIR (www.arabidopsis.org) and amino acid alignment was performed using ClustalX v2.1 (Larkin et al., 2007). The sequences were visualized by GeneDoc v2.7.00 (Nicholas and Nicholas, 1997), and the phylogenetic tree was constructed by the neighbor-joining method using MEGA v6 (Tamura et al., 2013) with 2,000 bootstrap replications.

Table 1-1. Oligonucleotides used in this study.

Name	Sequence (5'→3')
DG044	TTAAGTCGACTTAGGTTTTGTTGTTCTTC
DG124	CTGGCAACGGGAGTTCAGGTAATCCGCGGCCAACAATAAA ACTCAAC
DG125	GAACTCCCGTTGCCAGACGAACCGCGGCCGCACTCTGGTG CCGGTAAAG
DG144	AATTGGATCCTACAAAGGAGATGGTGCAC
DG175	AATTGGTACCAGCCAGTGGGATTGTTAAGAAGAG
DG176	AATTGCGGCCGCACTCTGTACTTGGTAAAGCAAG
DG177	AATTGGTACCAGCCAGTGGGATAGTTTGAGAAAGG
DG178	AATTGCGGCCGCACTCTGGTTCTGGTAAAGCAAG
DG179	AATTGGTACCAGCCAGTGAACAATCTTAGAAGG
DG180	AATTGCGGCCGCACTCTGGACTCTCGAGAAGAAC
DG219	GATTCCCTGACTATGAAGCAATAAGACGTGC
DG220	CATAGTCAGGGAATCGAGCAGCTAG
DG276	AATTGCGGCCGCGCTTCGTCTGGCAACGGGAGTTCAGGTA ATCCGCGGTAA
DG277	TTAACC GCGGATTACCTGAACTCCCGTTGCCAGACGAAGCG CGGCCGCAATT
DG278	AATTAGATCTTACAAAAAGTCGTATGAAG
DG279	AATTCTGCAGGAGGAACTCAGCAGCTAAATCC
DG280	AATTCCGCGGCCTACCATCATCCTCAACAAGG
DG281	AATTGTCGACTCATTCTCTGTCTTCTTTAG

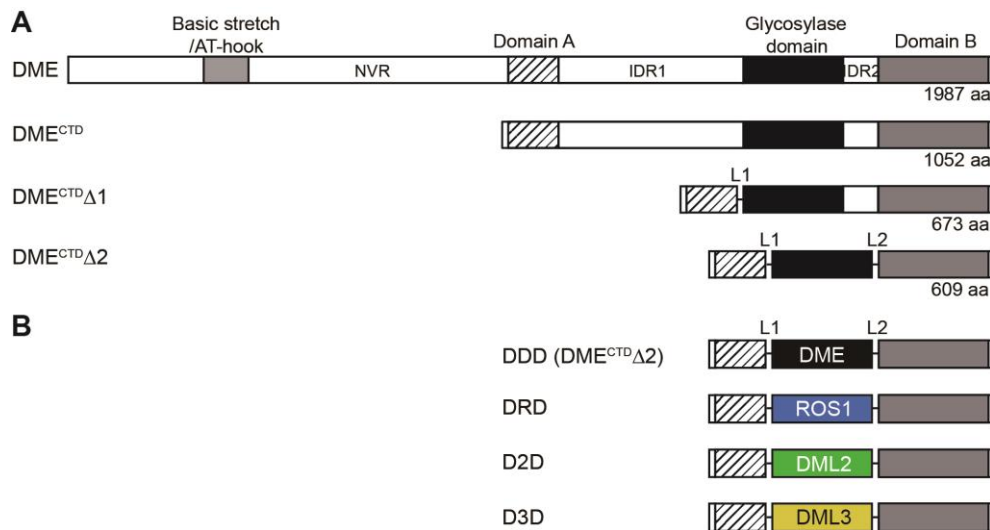


Figure 1-1. Structures of manipulated DME fragments and chimeric proteins between DME and other family members.

(A) Structures of full-length and manipulated DME fragments. DME^{CTD} has C-terminal domains without N-terminal 935 amino acids. DME^{CTD} Δ 1 has deletions of N-terminal 935 amino acids and interdomain region IDR1, with the latter being replaced by a synthetic linker L1. DME^{CTD} Δ 2 consists of three core domains tethered with linkers L1 and L2, instead of IDR1 and IDR2. NVR, N-terminal variable region. (B) The DDD (DME^{CTD} Δ 2) cassette of the minimal catalytic core and its chimeric constructs, DRD, D2D and D3D.

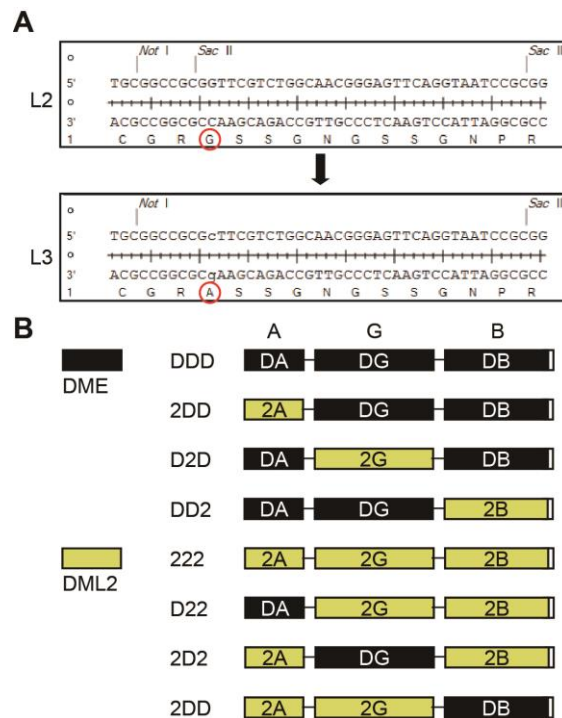


Figure 1-2. Sequences of the synthetic linkers and structures of chimeric fragments generated by domain swapping between DME and DML2.

(A) Sequences of the synthetic linkers L2 and L3. A single amino acid substitution of L2 generated L3, used for following construction of DME-DML2 chimeric fragments. (B) Structures of possible combinations of chimeric fragments generated by domain swapping between DME and DML2. Domains of DME and DML2 were represented with black and yellow boxes, respectively.

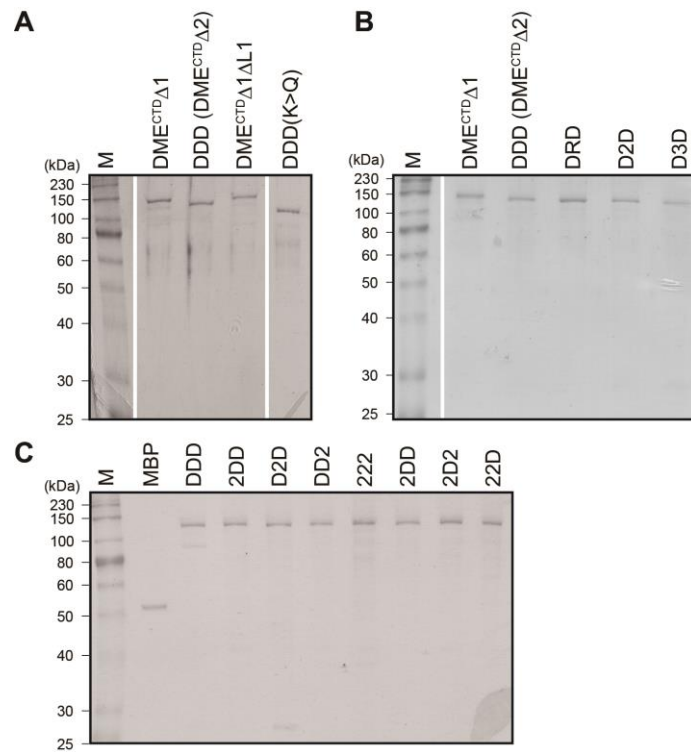


Figure 1-3. SDS-PAGE analysis of purified proteins used for *in vitro* 5mC excision assay.

N-terminal MBP tagged purified proteins (200 ng) for experiment in (A) Figure 1-4, (B) Figure 1-6, (C) Figure 1-7 were electrophoresed on a 10% SDS-PAGE gel. The sizes of the protein marker (kDa) were indicated to the left. M, size marker.

RESULTS

Three conserved domains of DME comprise the minimal entity for 5mC excision *in vitro*.

It was previously shown that three conserved domains of DME and ROS1 are essential, but most of the other variable regions are dispensable for 5mC excision activity *in vitro* (Mok et al., 2010; Hong et al., 2014). Deletion studies of DME showed that both DME without the entire N-terminal region (designated as DME^{CTD} in Figure 1-1A) and the manipulated DME fragment mostly comprising the three conserved C-terminal domains but still harboring interdomain region IDR2 (designated as DME^{CTD}Δ1 in Figure 1-1A) were able to remove 5mC *in vitro* (Mok et al., 2010). These results indicate that the core structure of DME mainly comprises the three domains as discrete modules. Homology modeling analysis supports this idea by predicting a model, in which the 5mC binding pocket of DME is composed of amino acid residues derived from domain A (AD) and the glycosylase domain (GD) (Brooks et al., 2014). To verify this interdomain 5mC binding pocket model, I directly tethered AD and GD without a synthetic linker L1 to produce DME^{CTD}Δ1ΔL1 (Figure 1-4A). The *in vitro* 5mC excision analysis showed that DME^{CTD}Δ1ΔL1 completely lost

5mC excision activity (Figure 1-4B), whereas DME^{CTD}Δ1 with a synthetic dodecapeptide linker sequence between AD and GD excised 5mC from hemimethylated oligonucleotide substrates *in vitro* (Figure 1-4C). This indicates that the flexibility between AD and GD is required to configure a catalytic core structure suitable for 5mC excision, thus the variable region between AD and GD is required for function *in vitro*, and its sequence must at least confer flexibility.

To further test the modular structure of DME with interdomain flexibility, all variable regions including both interdomain regions IDR1 and IDR2 were removed and connected with flexible synthetic linkers L1 and L2 to produce a DME^{CTD}Δ2 fragment consisting only of AD, GD, and domain B (BD) of DME (Figure 1-4A). *In vitro* analysis revealed that DME^{CTD}Δ2 excised 5mC with the single turnover rate constant ($k_{cat-st} = 0.0783 \pm 0.0027$) comparable to that of DME^{CTD}Δ1 ($k_{cat-st} = 0.0617 \pm 0.0010$) (Figures 1-4D and Table 1-2), suggesting that the three conserved domains comprise the minimal entity required for 5mC excision *in vitro*, but that a flexible sequence between AD and GD, and likely also BD and GD, are required for function.

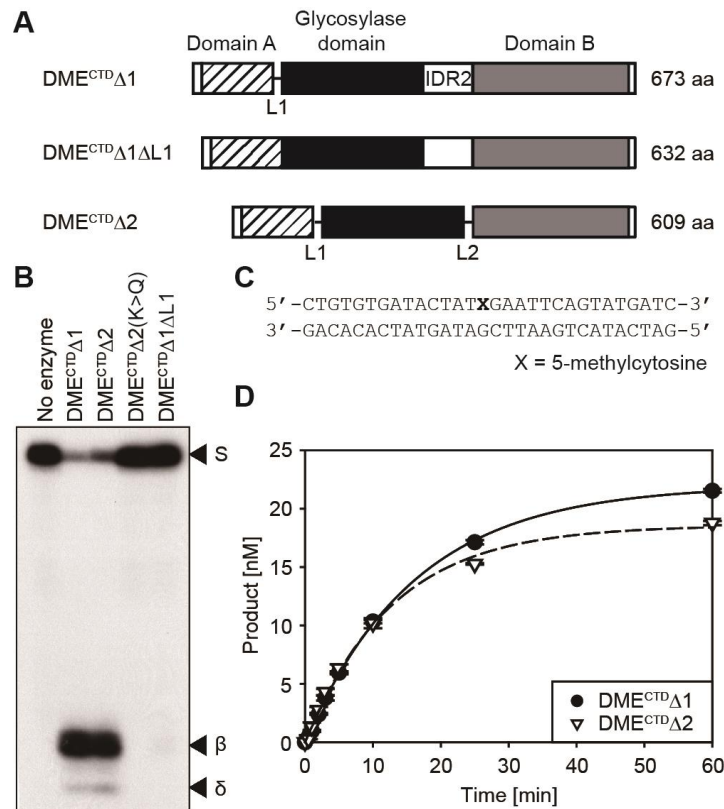


Figure 1-4. The *in vitro* 5mC excision activity of minimal DME catalytic core with three conserved domains.

(A) Diagrams of compact DME fragments consisting of three conserved domains. DME^{CTD}Δ1 has a synthetic linker L1 replacing IDR1, whereas DME^{CTD}Δ1ΔL1 has the first two domains directly tethered together. DME^{CTD}Δ2 has synthetic linkers L1 and L2 replacing IDR1 and IDR2, respectively. (B) *In vitro* 5mC glycosylase assay with DME catalytic core fragments. Positions of substrate (S) and β- and δ-elimination products (β, δ) are indicated to the right of the panel. K>Q, a catalytic mutant with a K1544Q substitution. (C) Oligonucleotide substrate used in (B). (D)

Kinetics analysis with DME^{CTD} Δ 1 and DME^{CTD} Δ 2 under single turnover conditions. Standard deviations were calculated from three independent experiments and plotted with error bars.

Table 1-2. Catalytic rate constants (k_{cat-st}) of purified proteins calculated from single turnover kinetics.

Protein	k_{cat-st} (min ⁻¹)
DME ^{CTD} Δ 1	0.0617 \pm 0.0010
DDD	0.0783 \pm 0.0027
DRD	0.0997 \pm 0.0044
D2D	0.1533 \pm 0.0191
D3D	0.0683 \pm 0.0024
D22	0.0644 \pm 0.0210
22D	0.1116 \pm 0.0101

Conserved domains of DME family proteins are catalytically compatible *in vitro*.

The DME family members in Arabidopsis share highly conserved domain structures, but each possesses distinct 5mC excision activity *in vitro* (Figure 1-5). In particular, DML2 displayed no discernable 5mC excision product unlike the other DME family members (Penterman et al., 2007), which raised the possibility that the activity of GD of DML2 might be restricted by its flanking domains AD and BD. In order to investigate whether these domains have functional compatibility between family members, I performed domain swapping experiments among the family members in a cassette configuration, in which the GD of the DME^{CTD}Δ2 platform (referred to as DDD hereafter) was replaced with that of ROS1, DML2 and DML3, and the resulting chimeric fragments were designated as DRD, D2D and D3D, respectively (Figure 1-6A). All chimeric recombinant proteins were found to efficiently excise 5mC *in vitro* with a substantial amount of reaction products (Figure 1-6B), albeit they were reported to have different catalytic efficiencies in the native configuration (Penterman et al., 2007). Notably, the catalytic rate constant of D2D ($k_{cat-st} = 0.1533 \pm 0.0191$) was about 2-fold higher than that of DDD ($k_{cat-st} = 0.0783 \pm 0.0027$), which demonstrates that domains AD and BD of DME dramatically restored the

5mC excision activity of GD of DML2, supporting the hypothesis that intrinsic glycosylase activity of DML2 could be restricted by flanking domains AD and BD (Figure 1-6C and Table 1-2). These results not only indicate that the central GDs of DME family proteins retain equivalent structures and activities, but also suggest that DNA glycosylase domains of DME family proteins are catalytically compatible *in vitro*.

Besides the well-known central GD, the function of AD and BD is poorly understood, even though the previous random mutagenesis study revealed that a number of single amino acid substitutions abolishing the catalytic activity of DME *in vitro* are largely confined to the conserved domains AD and BD (Mok et al., 2010). To further investigate the function of flanking domains AD and BD, I performed domain swapping experiments between functionally distinguishable DME and DML2. Using the previously produced minimal DDD cassette configuration, 8 possible constructs of DME-DML2 swapping combination were created, and all chimeric proteins were purified for biochemical assay. Consistent with the previous report (Penterman et al., 2007), the 222 protein which consists of three domains of DML2 with flexible linkers, displayed no 5mC excision activity as expected (Figures 1-7A and 1-7B). Replacement of AD or BD of DDD cassette with that of DML2 produced 2DD or DD2 chimeric proteins,

both of which showed extremely reduced DME activity (Figures 1-7A and 1-7B). This result suggests that DML2 has impaired AD and BD, which inhibit intrinsic DNA glycosylase activity of central GD of DML2. Conversely, introduction of DME domains into the 222 cassette generated D22, 2D2 and 22D chimeric proteins. Except for 2D2, both D22 and 22D proteins restored the 5mC excision activity of DML2. Notably, 22D more efficiently catalyzed 5mC excision compared to D22, which is supported by the kinetics study that the catalytic rate constant of 22D ($k_{cat-st} = 0.1116 \pm 0.0101$) was 1.7-fold higher than that of D22 ($k_{cat-st} = 0.0644 \pm 0.0210$) (Figure 1-7C and Table 1-2). Furthermore, the electrophoretic mobility shift assay revealed that different biochemical efficiencies of these chimeric proteins are correlated with DNA binding activity (Figure 1-8). These results demonstrate that the BD appears to be more critical to DME function than AD, which probably binds to DNA directly or serves as a structural role to facilitate DNA binding activity of DME.

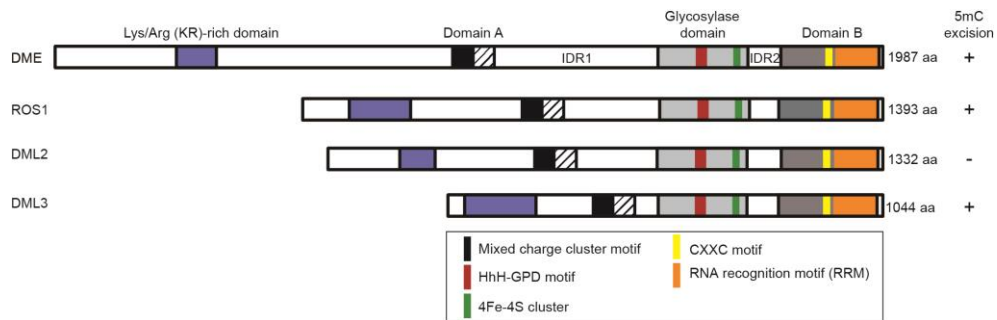


Figure 1-5. The conserved motifs of the DME family proteins in Arabidopsis.

Schematic diagrams of four DME family members in Arabidopsis. Conserved motifs are denoted in colored boxes. *In vitro* 5mC excision activity of DME family proteins estimated from the previous study (Penterman et al., 2007) is indicated to the right.

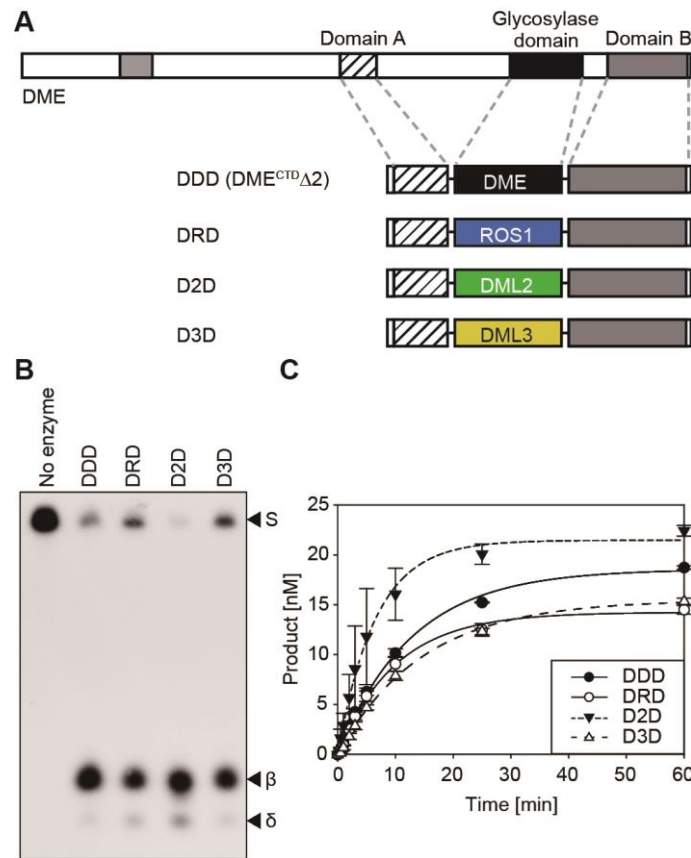


Figure 1-6. The *in vitro* 5mC excision activity of chimeric proteins generated by swapping glycosylase domains between DME family members.

(A) Structures of chimeric proteins between DME and other family members in a minimal cassette configuration. The central glycosylase domain of DME^{CTD}Δ₂ (shown as DDD) was replaced with that of ROS1, DML2 or DML3 to produce chimeric proteins DRD, D2D and D3D, respectively. (B) *In vitro* 5mC glycosylase assay with the chimeric proteins. The radiolabeled DNA substrate (25 nM) containing 5mC was incubated with 100 nM each of MBP-DDD, DRD, D2D and D3D at 37°C for 1 h.

Positions of the oligonucleotide substrate (S), and the reaction products (β , δ) were indicated to the right of the panel. (C) Kinetics analysis with the chimeric proteins under single turnover conditions. Reactions were terminated at various time points (0.5, 1, 2, 3, 5, 10, 25 and 60 min), and the amounts of reaction products were plotted over time. Standard deviations were calculated from three independent experiments and plotted with error bars.

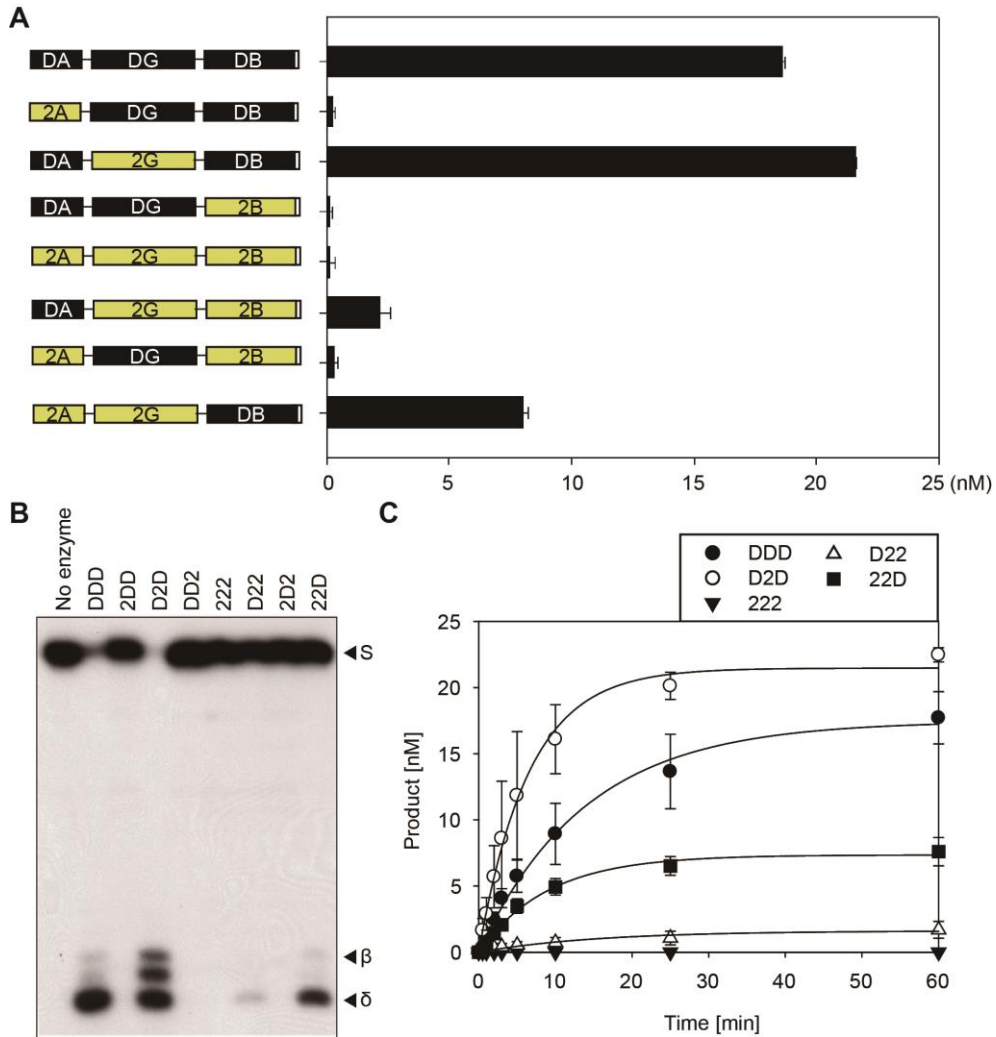


Figure 1-7. The *in vitro* 5mC excision activity of DME-DML2 chimeric proteins.

(A) *In vitro* 5mC glycosylase assay with the DME-DML2 chimeric proteins. Structures of chimeric proteins are represented to the left of the panel. The amount of the reaction products from experiment (B) was estimated and plotted as a bar graph. Standard deviations were calculated from three independent experiments and plotted with error bars. (B) *In vitro* 5mC

excision activity of the chimeric proteins. The radiolabeled DNA substrate (25 nM) containing 5mC was incubated with 100 nM each of chimeric proteins at 37°C for 1 h. Reactions were terminated by adding 100 mM NaOH with heat denaturing. Positions of the oligonucleotide substrate (S), and the reaction products (β , δ) were indicated to the right of the panel. (C) Kinetics analysis with the chimeric proteins under single turnover conditions. Reactions were terminated at various time points (0.5, 1, 2, 3, 5, 10, 25 and 60 min), and the amounts of reaction products were plotted over time. Standard deviations were calculated from three independent experiments and plotted with error bars.

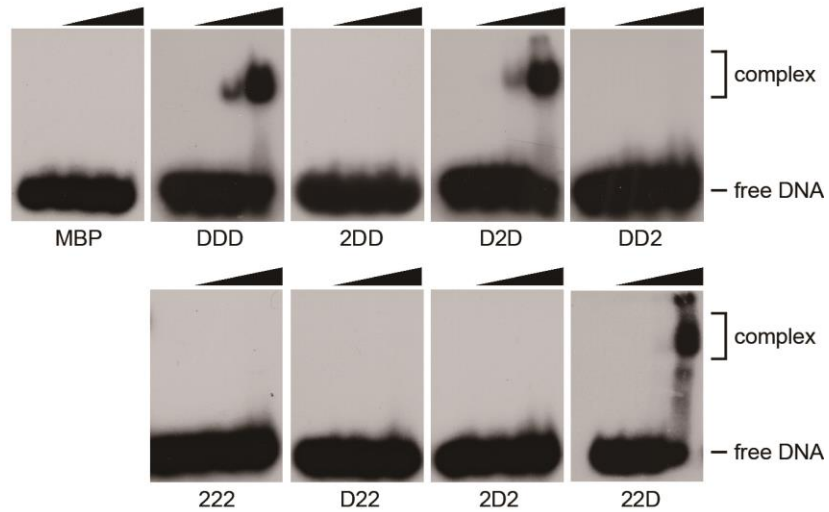


Figure 1-8. Electrophoretic mobility shift assay of DME-DML2 chimeric proteins.

The radiolabeled oligonucleotide substrate containing 5mC (100 nM) was incubated with increased amount (0, 7, 35, 175 nM) of MBP or MBP-tagged DME-DML2 chimeric proteins at 25°C for 15 min. Free DNA substrate and protein-DNA complex are indicated to the right.

Functional motifs in the conserved domains of DME

The DME family proteins are variable in size but share similar domain structures, notably for the conserved three domains AD, GD and BD tethered by highly variable unstructured regions (Figures 1-5 and 1-9). Especially, the central GD is known to be catalytically essential for 5mC excision, which is supported by the observation that GD showed high sequence similarities among the DME family members (Figure 1-9A). Although phylogenetic analysis on GD implies that DME and ROS1 are the closest to each other, while DML3 is less related to the other members (Figure 1-9C), DME family members are predicted to have diverged from a common ancestor, and they all share several motifs in the conserved domains (Figure 1-5). The core of DME family proteins comprises the helix-hairpin-helix (HhH) motif and a glycine/proline-rich loop with a conserved aspartic acid (GPD) followed by the 4Fe-4S cluster loop (FCL) motif, a permuted CXXC motif, and a divergent version of an RNA Recognition Motif (RRM) fold (Figures 1-9A and 1-9B). The catalytic pocket containing HhH and FCL motifs is predicted to span AD and GD (Brooks et al., 2014), whereas the CXXC motif and RRM fold are present in the C-terminal half of BD (Figure 1-9B). In particular, the CXXC motif between the FCL motif and RRM fold is found in many chromatin

modifiers in mammals such as DNMT1, Methyl-CpG-binding domain protein 1, and notably, mammalian DNA demethylase, ten-eleven translocation 1 (TET1) (Rhee et al., 2002; Tahiliani et al., 2009; Long et al., 2013). Recent study also reported that the CXXC motif is required for DNA binding and target selection (Xu et al., 2018), which allows us to presume a biological role of the CXXC motif in plants.

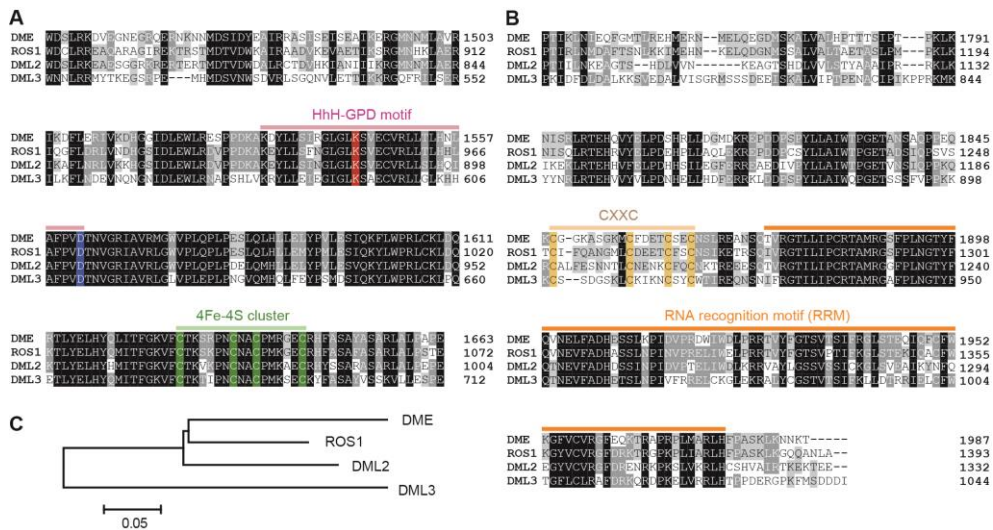


Figure 1-9. Sequence alignment of the conserved domains of DME family proteins.

(A) Amino acid sequence alignment of the glycosylase domain of the DME family proteins. Catalytic residues (K1544 and D1562 for DME) in the HhH-GPD motif and four cysteine residues of the 4Fe-4S cluster are denoted as red, blue and green, respectively. (B) Amino acid sequence alignment of the domain B of DME family proteins. The CXXC motif and the RNA recognition motif are denoted above the sequence. Four cysteine residues comprising the CXXC motif are colored in yellow. (C) The phylogenetic tree was constructed with the Neighbor-Joining method using the sequences of the glycosylase domains of the DME family proteins.

DISCUSSION

DME is a bifunctional 5mC DNA glycosylase/lyase required for active DNA demethylation in the central cell and the establishment of endosperm gene imprinting in Arabidopsis (Gehring et al., 2006; Park et al., 2017). The DME family has three conserved domains AD, GD and BD, and the core structures and catalytic mechanisms of this family of enzymes appear to be conserved in a wide range of organisms. However, it is known that the evolutionary conservation of DNA glycosylases is largely limited to the enzymatic core domains (Krokan and Bjoras, 2013). The sizes and sequence contexts of the other regions such as N-terminus and IDRs among DME family genes are various and not similar to one another in Arabidopsis (Figure 1-5). Previous deletion studies of DME demonstrated that N-terminal variable regions and IDR1 are not essential for the DME biochemical activity (Mok et al., 2010). In this study, I performed extensive manipulation of DME protein by deletion of both IDR1 and IDR2 to create the minimalized DME cassette (DDD) composed of the three core domains (Figure 1-4). The compact form of DDD, in which the flexible interdomain hinges were replaced with short polypeptide linkers, was sufficient to excise

5mC *in vitro*, attesting to the fact that both IDR1 and IDR2 are not critical for *in vitro* 5mC excision activity. However, consistent with the earlier studies (Gehring et al., 2006; Brooks et al., 2014; Hong et al., 2014), I also demonstrated that IDRs serve as flexible linkers to allow intra-molecular interaction between the three conserved domains and can be replaced with short synthetic polypeptide linkers (Figure 1-4B). Homology modeling supports such evolution of interdomain structure and predicts that the 5mC recognition pocket of DME is composed of amino acid residues derived from both AD and GD (Ponferrada-Marin et al., 2011; Brooks et al., 2014). Furthermore, *in vitro* biochemical study of ROS1 showed that addition of the purified BD protein restored base excision activity of the inactive recombinant AD-GD fusion peptide (Hong et al., 2014). Taken together, these studies suggest the modular nature of three conserved domains necessary and sufficient for biochemical function of DME as a catalytic core.

The DME family proteins in Arabidopsis share central GD with highly conserved motifs, but showed distinct biochemical efficiencies (Penterman et al., 2007; Figure 1-5). The central GD contains HhH-GPD motif and catalytic lysine and aspartate residues essential for nucleophilic attack at C-1 of the ribose ring to catalyze a breakage of N-glycosylic bond between the ribose sugar and 5mC (Jacobs and Schar, 2012; Figure 1-9A).

This HhH-GPD DNA glycosylase superfamily includes human 8-oxoguanin DNA glycosylase (hOGG1), *E. coli* adenine DNA glycosylase (MutY) and endonuclease III (Endo III) sharing conserved motifs (Kuo et al., 1992; Guan et al., 1998; Bruner et al., 2000). A canonical catalytic residue such as Asp1562 in DME is essential for the HhH-GPD glycosylase superfamily (Gehring et al., 2006; Mok et al., 2010), and four cysteine residues probably function to hold a 4Fe-4S cluster in place as in MutY and Endo III (Kuo et al., 1992; Guan et al., 1998). I initially assumed that differences in 5mC excision efficiency among DME family proteins might be due to the structural differences, but domain swapping experiments demonstrate that all GDs retain equivalent structure/function for 5mC excision (Figure 1-6B). Especially, D2D chimeric protein showed stronger 5mC excision activity than DME (Figure 1-6C), which suggests that the intrinsic 5mC glycosylase activity of GD in DML2 might be restricted by flanking AD and BD of DML2.

From a mechanistic point of view, the additional domains AD and BD may also provide some essential features for excision of 5mC. Unlike typical DNA glycosylases that recognize damaged or modified bases by detecting subtle changes in DNA helical structure, the DME family proteins remove 5mC from a normal Watson-Crick base pair. Therefore, a more

sophisticated mechanism will be required for initial target selection, and both AD and BD may help this process. For instance, nonspecific DNA binding activity provided by the mixed charge cluster motif of AD (Mok et al, 2010; Figure 1-5) may facilitate sliding of DME along the DNA duplex for target scanning (Jacobs and Schar, 2012). In addition, the permuted CXXC motif and RRM fold of BD may perform specific functions to scan target bases or stabilize the intermediate protein-DNA complex during 5mC excision, albeit their exact functions are still elusive. Domain swapping experiments between functionally distinguishable DME and DML2 provided important clues about the function of these domains. Arabidopsis DML2 appears to have impaired AD and BD, and replacing the BD of DML2 with that of DME (22D chimeric protein, Figure 1-7) successfully restored the 5mC excision activity of DML2, which implies that BD is more essential for DME function compared to AD. In addition, restoration of biochemical activity of these chimeric proteins was dependent on DNA binding activity (Figure 1-8), suggesting a structural role of BD in DNA binding.

So far very little is known about the function of BD, albeit its necessity for 5mC excision was already demonstrated by deletion and mutagenesis experiments (Gehring et al., 2006; Mok et al., 2010). Notably,

the second half of BD contains a permuted CXXC motif and the RRM fold (Iyer et al., 2011) (Figures 1-5 and 1-9B). The CXXC motif is also present in mammalian DNA demethylase, TET family proteins that convert 5mC into 5hmC (Long et al., 2013). TET1 contains a catalytic domain for 5mC dioxygenase activity at the C-terminus and the CXXC motif at the N-terminus (Pastor et al., 2013). The CXXC motif of TET1 is important for DNA binding to unmethylated CpG islands (CGIs) (Xu et al., 2011), which should prevent DNA methylation spreading into CGIs while maintaining the hypomethylated state of CGIs and transcriptional activity of the associated genes. Although it is still elusive whether the CXXC motif of DME is necessary for 5mC excision, the motif may involve in binding to genomic target or stabilization of DME-DNA structure during 5mC excision. Similarly, the role of the divergent RRM fold is also not fully understood. Although the involvement of RNA species in the active DNA demethylation process has not been firmly established, an RRM protein ROS3 required for ROS1 demethylation suggests a potential role of non-coding RNAs in the active DNA demethylation pathway in Arabidopsis (Zheng et al., 2008). Considering the general property of RRM that binds to single-stranded RNA/DNA, it is plausible that the RRM fold plays an important role for guiding DME to proper genomic targets for DNA demethylation while

utilizing RNA as a guiding cue. Taken together, member-specific BD is assumed to not only dictate target specificity by affecting protein conformation and influencing interacting partners, but play a structural role for DME catalysis, such as binding single-stranded DNA upon base interrogation to facilitate 5mC base excision.

REFERENCES

- Agius, F., Kapoor, A., and Zhu, J.K.** (2006). Role of the Arabidopsis DNA glycosylase/lyase ROS1 in active DNA demethylation. *Proc. Natl. Acad. Sci. USA* **103**: 11796-11801.
- Allis, C.D., and Jenuwein, T.** (2016). The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.* **17**: 487-500.
- Begley, T.J., Haas, B.J., Morales, J.C., Kool, E.T., and Cunningham, R.P.** (2003). Kinetics and binding of the thymine-DNA mismatch glycosylase, Mig-Mth, with mismatch-containing DNA substrates. *DNA Repair* **2**: 107-120.
- Brooks, S.C., Fischer, R.L., Huh, J.H., and Eichman, B.F.** (2014). 5-methylcytosine recognition by *Arabidopsis thaliana* DNA glycosylases DEMETER and DML3. *Biochemistry* **53**: 2525-2532.
- Bruner, S.D., Norman, D.P., and Verdine, G.L.** (2000). Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature* **403**: 859-866.
- Choi, Y., Gehring, M., Johnson, L., Hannon, M., Harada, J.J., Goldberg, R.B., Jacobsen, S.E., and Fischer, R.L.** (2002). DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in Arabidopsis. *Cell* **110**: 33-42.
- Gehring, M., Huh, J.H., Hsieh, T.F., Penterman, J., Choi, Y., Harada, J.J., Goldberg, R.B., and Fischer, R.L.** (2006). DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. *Cell* **124**: 495-506.
- Gong, Z., Morales-Ruiz, T., Ariza, R.R., Roldan-Arjona, T., David, L.,**

- and Zhu, J.-K.** (2002). ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase. *Cell* **111**: 803-814.
- Guan, Y., Manuel, R.C., Arvai, A.S., Parikh, S.S., Mol, C.D., Miller, J.H., Lloyd, R.S., and Tainer, J.A.** (1998). MutY catalytic core, mutant and bound adenine structures define specificity for DNA repair enzyme superfamily. *Nat. Struct. Biol.* **5**: 1058-1064.
- Hong, S., Hashimoto, H., Kow, Y.W., Zhang, X., and Cheng, X.** (2014). The carboxy-terminal domain of ROS1 is essential for 5-methylcytosine DNA glycosylase activity. *J. Mol. Biol.* **426**: 3703-3712.
- Huh, J.H., Bauer, M.J., Hsieh, T.F., and Fischer, R.L.** (2008). Cellular programming of plant gene imprinting. *Cell* **132**: 735-744.
- Ibarra, C.A., Feng, X., Schoft, V.K., Hsieh, T.F., Uzawa, R., Rodrigues, J.A., Zemach, A., Chumak, N., Machlicova, A., Nishimura, T., Rojas, D., Fischer, R.L., Tamaru, H., and Zilberman, D.** (2012). Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. *Science* **337**: 1360-1364.
- Iyer, L.M., Abhiman, S., and Aravind, L.** (2011). Natural history of eukaryotic DNA methylation systems. *Prog. Mol. Biol. Transl. Sci.* **101**: 25-104.
- Jacobs, A.L., and Schar, P.** (2012). DNA glycosylases: in DNA repair and beyond. *Chromosoma* **121**: 1-20.
- Jullien, P.E., Kinoshita, T., Ohad, N., and Berger, F.** (2006). Maintenance of DNA methylation during the Arabidopsis life cycle is essential for parental imprinting. *Plant Cell* **18**: 1360-1372.
- Kinoshita, T., Miura, A., Choi, Y., Kinoshita, Y., Cao, X., Jacobsen, S.E.,**

- Fischer, R.L., and Kakutani, T.** (2004). One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation. *Science* **303**: 521-523.
- Krokan, H.E., and Bjoras, M.** (2013). Base excision repair. *Cold Spring Harb. Perspect. Biol.* **5**: a012583.
- Kuo, C.F., McRee, D.E., Cunningham, R.P., and Tainer, J.A.** (1992). Crystallization and crystallographic characterization of the iron-sulfur-containing DNA-repair enzyme endonuclease III from *Escherichia coli*. *J. Mol. Biol.* **227**: 347-351.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., and Higgins, D.G.** (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947-2948.
- Long, H.K., Blackledge, N.P., and Klose, R.J.** (2013). ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. *Biochem. Soc. Trans.* **41**: 727-740.
- Mok, Y.G., Uzawa, R., Lee, J., Weiner, G.M., Eichman, B.F., Fischer, R.L., and Huh, J.H.** (2010). Domain structure of the DEMETER 5-methylcytosine DNA glycosylase. *Proc. Natl. Acad. Sci. USA* **107**: 19225-19230.
- Nicholas, K.B., Nicholas, H.B., and Deerfield, D.W.** (1997). GeneDoc: Analysis and visualization of genetic variation. *EMBnet news* **4**: 14.
- Park, J.S., Frost, J.M., Park, K., Ohr, H., Park, G.T., Kim, S., Eom, H., Lee, I., Brooks, J.S., Fischer, R.L., and Choi, Y.** (2017). Control of DEMETER DNA demethylase gene transcription in male and female gamete companion cells in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **114**: 2078-2083.

- Pastor, W.A., Aravind, L., and Rao, A.** (2013). TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nat. Rev. Mol. Cell Biol.* **14**: 341-356.
- Penterman, J., Zilberman, D., Huh, J.H., Ballinger, T., Henikoff, S., and Fischer, R.L.** (2007). DNA demethylation in the Arabidopsis genome. *Proc. Natl. Acad. Sci. USA* **104**: 6752-6757.
- Ponferrada-Marin, M.I., Parrilla-Doblas, J.T., Roldan-Arjona, T., and Ariza, R.R.** (2011). A discontinuous DNA glycosylase domain in a family of enzymes that excise 5-methylcytosine. *Nucleic Acids Res.* **39**: 1473-1484.
- Rhee, I., Bachman, K.E., Park, B.H., Jair, K.W., Yen, R.W., Schuebel, K.E., Cui, H., Feinberg, A.P., Lengauer, C., Kinzler, K.W., Baylin, S.B., and Vogelstein, B.** (2002). DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. *Nature* **416**: 552-556.
- Smith, Z.D., and Meissner, A.** (2013). DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14**: 204-220.
- Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., and Rao, A.** (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**: 930-935.
- Tamura, K., Stecher, G., Peterson, D., Filipowski, A., and Kumar, S.** (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30**: 2725-2729.
- Wu, H., and Zhang, Y.** (2014). Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell* **156**: 45-68.
- Xu, C., Liu, K., Lei, M., Yang, A., Li, Y., Hughes, T.R., and Min, J.**

- (2018). DNA Sequence Recognition of Human CXXC Domains and Their Structural Determinants. *Structure* **26**: 85-95 e83.
- Xu, Y., Wu, F., Tan, L., Kong, L., Xiong, L., Deng, J., Barbera, A.J., Zheng, L., Zhang, H., Huang, S., Min, J., Nicholson, T., Chen, T., Xu, G., Shi, Y., Zhang, K., and Shi, Y.G.** (2011). Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Mol. Cell* **42**: 451-464.
- Yamamuro, C., Miki, D., Zheng, Z., Ma, J., Wang, J., Yang, Z., Dong, J., and Zhu, J.K.** (2014). Overproduction of stomatal lineage cells in Arabidopsis mutants defective in active DNA demethylation. *Nat. Commun.* **5**: 4062.
- Zheng, X., Pontes, O., Zhu, J., Miki, D., Zhang, F., Li, W.X., Iida, K., Kapoor, A., Pikaard, C.S., and Zhu, J.K.** (2008). ROS3 is an RNA-binding protein required for DNA demethylation in Arabidopsis. *Nature* **455**: 1259-1262.

CHAPTER 2

Biochemical study of the DNA demethylation pathway

The research described in this chapter has been published in *Nucleic Acids Research*, <https://doi.org/10.1093/nar/gku834>

ABSTRACT

DNA methylation is the primary epigenetic modification important for gene regulation in plants and animals. DNA methylation is reversible and the base excision repair pathway is central to active DNA demethylation. In plants, DNA demethylation is initiated by the DEMETER (DME) family of 5-methylcytosine (5mC) DNA glycosylases which specifically recognizes and excises 5mC, whereas the downstream base excision repair events are largely unknown. During 5mC excision, DME generates harmful 3'-phosphor- α , β -unsaturated aldehyde (3'-PUA) and 3'-phosphate structures. These blocking lesions must be immediately processed by AP endonuclease in living cells to allow subsequent nucleotide extension. The kinetic studies revealed that these 3'-blocking lesions persist for a significant amount of time and at least two different enzyme activities are required to immediately process them. Here, I report that Arabidopsis AP endonucleases ABASIC ENDONUCLEASE 1-LIKE (APE1L) and ABASIC ENDONUCLEASE-REDOX PROTEIN (ARP) are responsible for trimming unusual 3' end structures after 5mC excision. Based on the present study, I suggest Arabidopsis has a branched pathway of active DNA demethylation, for which both AP endonuclease and 3' phosphatase activities are required.

INTRODUCTION

DNA methylation is the primary epigenetic modification that regulates gene expression and chromatin structure (Law and Jacobsen, 2010; Smith and Meissner, 2013; Allis and Jenuwein, 2016). Tight control of DNA methylation is important for diverse developmental processes in plants including gene imprinting and transposon silencing (Huh et al., 2008; Law and Jacobsen, 2010). In eukaryotes, DNA methylation is catalyzed by DNA methyltransferases to convert cytosine to 5-methylcytosine (5mC) (Law and Jacobsen, 2010). Similar to most epigenetic modifications, DNA methylation can be reversed. DNA demethylation, the reverse process of DNA methylation, can be classified into two different mechanisms. Passive DNA demethylation occurs in a replication-dependent manner which involves gradual decrease of the 5mC level by inactivation or down-regulation of maintenance DNA methyltransferases such as DNMT1 and MET1 in mammals and plants, respectively. In contrast, active DNA demethylation is enzymatically induced by DNA demethylases in a replication-independent manner (Wu and Zhang, 2010).

It is previously known that base excision repair (BER) machineries are employed to allow active DNA demethylation in plants and mammals.

According to the current models of active DNA demethylation in plants, 5mC is directly excised by specific DNA glycosylases and following BER enzymes complete DNA demethylation by inserting unmethylated cytosine into abasic site. (Agius et al., 2006; Gehring et al., 2006; Morales-Ruiz et al., 2006; Gehring et al., 2009; Zhu, 2009). However, DNA demethylation in animals is unlikely to involve direct removal of 5mC, but instead begins with conversion of 5mC to other bases such as thymine or 5-hydroxymethylcytosine (5hmC). The 5hmC base can be further converted to 5-formylcytosine and 5-carboxylcytosine, which are then excised by mismatched DNA glycosylases such as thymine DNA glycosylase (Tahiliani et al., 2009; Cortellino et al., 2011; Guo et al., 2011; Williams et al., 2011; Xu et al., 2011; Song and He, 2013).

DEMETER (DME) is a member of the plant-specific DNA demethylase family which was first identified in *Arabidopsis* (Choi et al., 2002). DME and family members such as REPRESSOR OF SILENCING 1 (ROS1), DEMETER-LIKE 2 (DML2) and DML3 exert 5mC glycosylase activity that specifically recognizes and excises 5mC from DNA (Agius et al., 2006; Gehring et al., 2006; Morales-Ruiz et al., 2006; Penterman et al., 2007; Ortega-Galisteo et al., 2008). As bifunctional DNA glycosylases with additional AP-lyase activity, the DME family proteins catalyze both 5mC

excision and the cleavage of a sugar-phosphate backbone via β - and δ -elimination reactions, which generate 3'-phosphor- α , β -unsaturated aldehyde (3'-PUA) and 3'-phosphate, respectively (Gehring et al., 2006). These intermediates must be processed to provide 3'-OH for subsequent polymerization by DNA polymerase, thus further demethylation steps may require BER machineries. In particular, AP endonucleases that act immediately downstream of DNA glycosylase are likely indispensable for processing such harmful lesions.

Recently, ZINC FINGER DNA 3' PHOSPHOESTERASE (ZDP) was proven necessary for ROS1-mediated DNA demethylation in Arabidopsis (Martinez-Macias et al., 2012). ZDP preferentially removed the δ -elimination product 3'-phosphate, providing a 3'-OH for subsequent polymerization and ligation (Martinez-Macias et al., 2012). However, generation of the 3'-phosphate by δ -elimination is a very slow process, which results in an inevitable open chromatin at 5mC excision site until the BER is completed (Ponferrada-Marin et al., 2009; Martinez-Macias et al., 2012). This is extremely harmful to living organisms and usually prevents DNA replication and transcription (Caldecott, 2008).

Here I showed that both DME and ROS1 5mC glycosylases produced 3'-PUA as a primary 5mC excision intermediate, requiring

immediate recruitment of DNA repair enzymes. To investigate the functional roles during 5mC excision, thorough biochemical analysis was performed with three AP endonucleases ABASIC ENDONUCLEASE 1-LIKE (APE1L), ABASIC ENDONUCLEASE 2 (APE2) and ABASIC ENDONUCLEASE-REDOX PROTEIN (ARP) present in the Arabidopsis genome. I demonstrated that both APE1L and ARP are able to process the 3'-PUA, and that ARP has additional 3' phosphatase activity to process 3'-phosphate generated by DME. These data suggest that active DNA demethylation processes in plants require two distinct, coordinated enzymatic activities for complete removal of 5mC excision intermediates.

MATERIALS AND METHODS

Expression and purification of the proteins

The DME Δ N677 Δ IDR1::lnk (DME Δ) and ROS1 Δ N509 (ROS1 Δ) fragments (Figure 2-1; Jang et al., 2014) were introduced into the pLM302 vector (Center for Structural Biology, Vanderbilt University), a pET-27a (EMD Millipore) derivative with an N-terminal 6xHis-MBP and a PreScission Protease (GE Healthcare) cleavage site. To produce the pLM302-APE1L, -APE2 and -ARP constructs, full-length sequences of APE1L, APE2 and ARP obtained from Arabidopsis cDNA were cloned into the pLM302 vector.

To express the proteins, all the cloned constructs were transformed into *E. coli* Rosetta2 (DE3) strain (EMD Millipore), respectively. A single colony was inoculated in 5 mL of LB medium containing kanamycin (50 μ g/mL) and chloramphenicol (50 μ g/mL). The culture was incubated at 37°C overnight. An aliquot of overnight culture was inoculated into 2 L LB medium with the same antibiotics and incubated at 30°C until OD₆₀₀ reached 0.4. Expression was induced with 0.1 mM IPTG at 16°C overnight with shaking. Cells were harvested by centrifugation at 7,000 rpm for 20 min at 4°C, and the pellet was resuspended in 30 mL of lysis buffer (50 mM

Tris-HCl, pH 7.4, 100 mM NaCl, 10% glycerol, 0.1 mM dithiothreitol, 0.1 mM PMSF). The lysate was sonicated on ice and clarified by centrifugation. The supernatant was filtered through nylon membrane with 0.45 μ m pore (Adventec), and loaded onto the HisTrap FF column (GE Healthcare) and eluted with a step gradient of imidazole (50, 75, 100, 150 and 250 mM). Fractions were concentrated and aliquoted with 50% glycerol and stored at -80°C until use (Figures 2-2A and 2-2B).

For the purification of the MBP-free AP endonucleases, the MBP-tagged APE1L, APE2, and ARP proteins purified through the HisTrap FF column, were incubated with PreScission Protease (GE Healthcare) at 4°C overnight to cleave off the MBP fragments. The Heparin HP column (GE Healthcare) was used to separate MBP from AP endonuclease fragments. The collected MBP-free AP endonuclease fractions were concentrated and stored with 50% glycerol at -80°C until use (Figure 2-2C).

***In vitro* 5mC excision assay**

Oligonucleotides used in this study were synthesized by Integrated DNA Technologies (Table 2-1). Forty pmol of F35[5mC] oligonucleotide was radiolabeled with [γ -³²P]ATP (6,000 Ci/mmol, Perkin Elmer) using T4 polynucleotide kinase (NEB) and then annealed to a complementary

oligonucleotide to produce double-stranded DNA substrate. For the *in vitro* 5mC excision assay, 25 nM each of radiolabeled oligonucleotide substrate was incubated with 100 nM MBP-DME Δ or 85 nM MBP-ROS1 Δ (Figure 2-2A) in the glycosylase reaction buffer (10 mM Tris-HCl, pH 7.4, 50 mM NaCl, 0.5 mM DTT, 200 μ g/mL BSA) at 37°C for 1 h. For the experiments in Figure 2-4, two separate reactions were performed at 37°C for 25 min and paused on ice or heat-inactivated at 65°C for 15 min, respectively, and then resumed at 37°C for additional 15 h and 35 min. Reactions were terminated at each time point indicated in Figure 2-4, by adding an equal volume of stop solution (98% formamide, 10 mM EDTA, 0.2% xylen cyanol FF, 0.2% bromophenol blue) and heat-denaturing at 95°C for 10 min. Reaction products were separated on a 15% denaturing polyacrylamide gel containing 7.5 M urea and 1x TBE.

***In vitro* 3' phosphodiesterase assay**

Radiolabeled 35-mer oligonucleotide (F35[5mC]) containing 5mC was annealed to the complementary strand to prepare the methylated DNA substrate; 25 nM of radiolabeled methylated oligonucleotide duplex was incubated with 100 nM MBP-DME Δ in the glycosylase reaction buffer (10 mM Tris-HCl, pH 7.4, 50 mM NaCl, 0.5 mM DTT, 200 μ g/mL BSA) at

37°C for 1 h. In the same reaction tube, 5 nM each of AP endonucleases was added and reaction was performed at 37°C for 20 min in the presence of 2.5 mM MgCl₂.

***In vitro* nucleotide incorporation assay**

The 25 nM radiolabeled oligonucleotide substrate containing 5mC was reacted with 100 nM MBP-DMEΔ in the glycosylase reaction buffer, and subsequently incubated with 5 nM AP endonucleases in the presence of 2.5 mM MgCl₂ at 37°C for 20 min. Following heat-inactivation at 65°C for 15 min, reactions was subjected to nucleotide incorporation with either 0.1 mM dCTP or 0.025 mM Cy3-dCTP using 5 units of Klenow fragment (3'→5' exo-) (NEB) at 25°C for 25 min.

***In vitro* 3' phosphatase assay**

The 17-mer oligonucleotide with a 3'-phosphate was 5'-end-labeled with [γ -³²P]ATP using T4 Polynucleotide Kinase (3' phosphatase minus, NEB). This radiolabeled upstream oligonucleotide and another 17-mer downstream oligonucleotide with a 5'-phosphate were annealed together to the 35-mer complementary strand to produce DNA substrate with a single nucleotide gap. The 25 nM DNA substrate was incubated with 5 nM each of

AP endonucleases in the AP endonuclease reaction buffer (10 mM Tris-HCl, pH 7.4, 50 mM NaCl, 200 µg/mL BSA, 2.5 mM MgCl₂, 1 mM DTT) at 37 °C for 1 h. For kinetics analysis, 5 nM MBP-ARP was incubated with 25 nM radiolabeled oligonucleotide duplex in a time course manner (2, 5, 10, 30, 60, 120, 240, 480 and 960 min).

***In vitro* AP endonuclease assay on AP sites**

The radiolabeled F35[AP] oligonucleotide containing a tetrahydrofuran (THF) was annealed to the complementary strand to generate the AP site analog substrate. Oligonucleotide substrate (25 nM) was incubated with 5 nM each of APE1L, APE2 or ARP in the AP endonuclease reaction buffer (10 mM Tris-HCl, pH 7.4, 50 mM NaCl, 200 µg/mL BSA, 2.5 mM MgCl₂, 1 mM DTT) at 37°C for 30 min. As a reaction control, the same oligonucleotide was reacted with 0.4 unit of hAPE1 (NEB).

Electrophoretic mobility shift assay

Radiolabeled 35-mer oligonucleotides (50 nM) containing a THF or 5mC were incubated with each purified MBP-APE1L, -APE2, or -ARP on ice for 10 min. The increasing amounts of protein (0, 100 and 300 nM) were

added to measure DNA binding activity in the DNA binding buffer (10 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.05% Triton X-100, 0.1 mg/mL BSA, 10% glycerol, 10 mM DTT). The reactions were separated on the 8% native polyacrylamide gel at 50 V for 2 h and the gel was exposed to an X-ray film.

Table 2-1. List of oligonucleotides for biochemical assays.

name	oligonucleotide sequence
F35[AP]	5'- GTACTGTGTGATACTAT[THF]GAATTCAGTATGATCTG
F35[5mC]	5'- GTACTGTGTGATACTAT[5mC]GAATTCAGTATGATCTG
R35	5'- CAGATCATACTGAATTCGATAGTATCACACAGTAC
F17F[3P]	5'- GTACTGTGTGATACTAT[3Phos]
[5P]F17B	5'- [5Phos]GAATTCAGTATGATCTG

*THF, tetrahydrofuran; 5mC, 5-methylcytosine; Phos, phosphorylation



Figure 2-1. Schematic diagrams of manipulated DME and ROS1.

DMEΔ has deletions of N-terminal 935 amino acids and interdomain region IDR1, with the latter being replaced by a synthetic linker. ROS1Δ has C-terminal domains without N-terminal 509 amino acids.

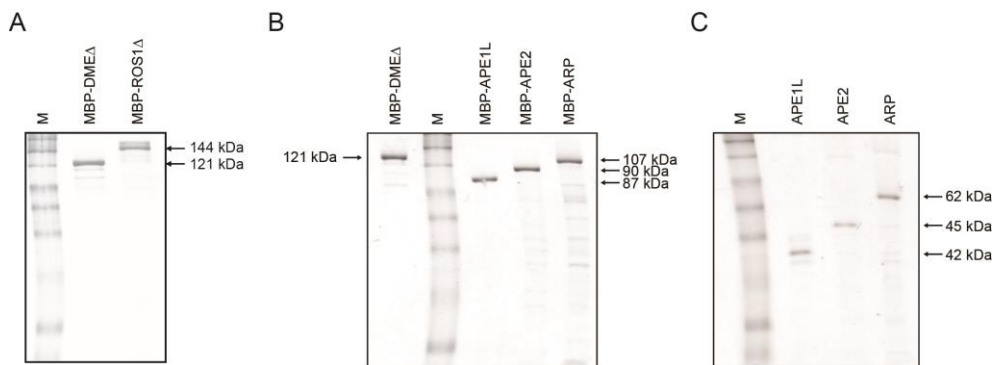


Figure 2-2. SDS-PAGE analysis of purified proteins.

The purified DME family proteins (A), MBP-tagged AP endonucleases (B), and MBP-free AP endonucleases (C) were electrophoresed on a 10% SDS-PAGE gel. The sizes of the proteins were indicated to the right. M, size marker.

RESULTS

DME and ROS1 5mC DNA glycosylases generate a 3'-PUA as a primary blocking intermediate in early 5mC excision

The first step of BER involves base removal by DNA glycosylase, which cleaves an N-glycosydic bond between the base and a ribose sugar creating an apurinic/apyrimidinic (AP) or abasic site (Figure 2-3A). Depending on the mode of base excision and the ability to execute strand cleavage, DNA glycosylases can be categorized into monofunctional and bifunctional enzymes (Scharer and Jiricny, 2001). After base excision, bifunctional DNA glycosylases catalyze the scission of the sugar phosphate backbone leaving the 3'-PUA by the process called β -elimination; a further δ -elimination process generates a 3'-phosphate (Figure 2-3A). Such β - and δ -elimination products need to be processed immediately to provide a 3'-OH for subsequent polymerization by DNA polymerase.

Previous studies showed that the DME family catalyzes both 5mC excision and a strand cleavage using a bifunctional mechanism, producing 3'-PUA and 3'-phosphate by β - and δ -elimination, respectively (Agius et al., 2006; Gehring et al., 2006; Morales-Ruiz et al., 2006; Penterman et al., 2007; Ortega-Galisteo et al., 2008). To carefully monitor the changes in

relative abundance of β - and δ -elimination processes, I performed time-course experiments on 5mC excision by DME and ROS1. When reacted with a 35-mer duplex DNA substrate that contained a 5mC residue at position 18 in the 5'-end-labeled strand, both DME and ROS1 produced 3'-PUA via β -elimination as a primary 5mC excision intermediate in early stages (< 5 min) (Figures 2-3B and 2-3C). As the reaction progressed, 3'-phosphate started to appear but in significantly lower amounts than 3'-PUA (Figure 2-3B). Prolonged reaction generated more 3'-phosphate while the amount of 3'-PUA gradually decreased, and after the reaction reached near completion, the 3'-phosphate became predominant (Figures 2-4A and 2-4C), implying the 3'-phosphate was a final 5mC excision product via δ -elimination of 3'-PUA.

These observations suggest that β -elimination precedes and δ -elimination follows as a relatively slow process. This may pose a serious problem to the DNA strands on which DNA demethylation occurs, particularly when a strand gap remains open for a long time. Importantly, two distinct enzyme activities may be required to completely clear blocking structures at the 3' of the strand cleavage site because 3'-PUA and 3'-phosphate are predominantly produced at different time windows by β - and δ -elimination, respectively.

Enzyme-dependent δ -elimination process following 5mC excision

To ensure that sequential β - and δ -elimination reactions are intrinsic to DME/ROS1 glycosylases, I tested whether the latter δ -elimination process is dependent on active enzyme activity. I incubated purified MBP-DME Δ protein with a 35-mer oligonucleotide duplex containing a 5mC residue in the middle of the 5'-end-labeled top strand. Twenty-five minutes after incubation, two separate reactions were either stalled on ice or terminated at 65°C for 15 min, and then both returned to 37°C for an additional 15 h and 35 min. I compared the rate of δ -elimination processes between the two experiments by measuring the amount of 3'-phosphate produced after reincubation (Figure 2-4). When the reaction was paused on ice for 15 min and then resumed, 3'-phosphate was continuously produced while the remaining 3'-PUA gradually disappeared (Figures 2-4A and 2-4C). By contrast, when DME was denatured by heating at 65°C for 15 min, production of 3'-phosphate was no longer accumulated after the reaction resumed (Figures 2-4B and 2-4D), suggesting that δ -elimination is an enzyme-dependent process. This result also implies that successive β - and δ -eliminations are intrinsically coupled processes.

I have shown that DME and ROS1 sequentially produce 3'-PUA and 3'-phosphate via β - and δ -elimination reactions, respectively (Figures 2-3B and 2-3C). Persistence of these two products in DNA would create a serious problem in plant cells, because they are part of single-strand break (SSB) damage and harmful blocking lesions in DNA replication and transcription (Caldecott, 2008). Although ZDP was proposed to play a central role in active DNA demethylation acting downstream of ROS1 (Martinez-Macias et al., 2012), the extremely slow turnover rate of 3'-PUA to 3'-phosphate by DME-mediated δ -elimination and early formation of 3'-PUA as a major intermediate (Figures 2-3B and 2-3C) imply that an additional mechanism is required to process 3'-PUA. Thus, I focused on the activities of plant AP endonucleases that are possibly involved in processing base excision intermediates.

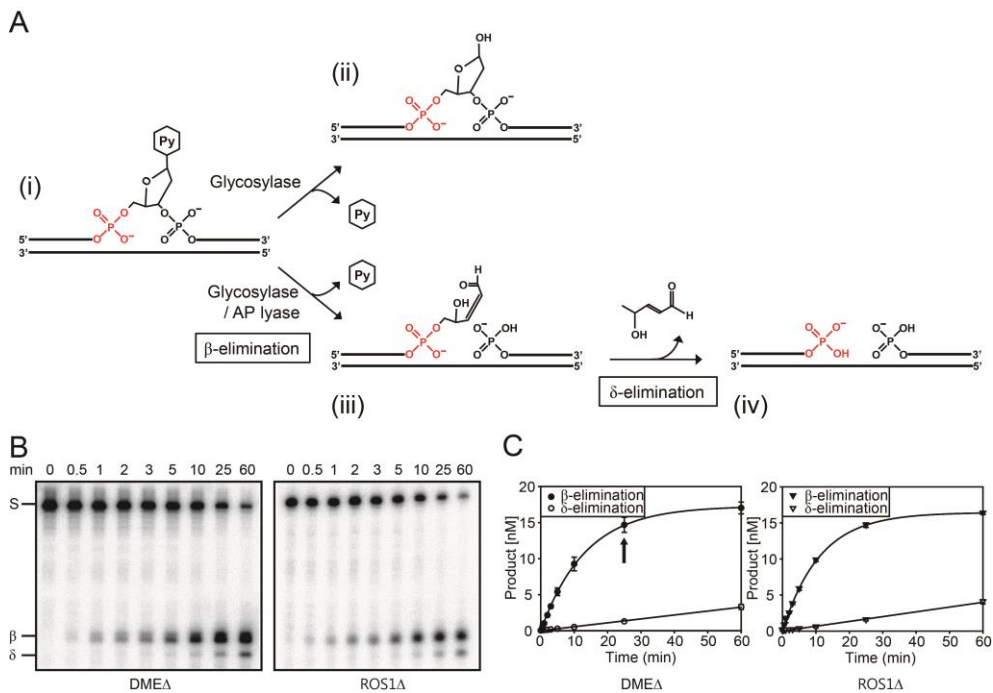


Figure 2-3. The 5mC excision products generated by DME and ROS1 DNA glycosylases.

(A) The base excision by monofunctional or bifunctional DNA glycosylase. Upon encountering a base to remove (i), monofunctional DNA glycosylases catalyze the hydrolysis of an N-glycosidic bond between the base and a ribose sugar creating an AP site (ii). However, bifunctional DNA glycosylases possess additional AP lyase activity that catalyzes the scission of the sugar phosphate backbone leaving the 3'-phosphor- α , β -unsaturated aldehyde (iii) by the β -elimination process. Further δ -elimination process generates a 3'-phosphate (iv), which is a blocking lesion for subsequent polymerization. (B) The 5mC excision products generated by DME and ROS1. Radiolabeled oligonucleotide substrate containing 5mC was incubated with purified MBP-DME Δ or -ROS1 Δ protein. Both 3'-PUA and

3'-phosphate were generated in the β - and δ -elimination processes by DME and ROS1. The major intermediate formed in the early reaction was 3'-PUA (β -elimination product) and as the reaction proceeded, 3'-phosphate (δ -elimination product) began to accumulate. (C) Relative amounts of β - and δ -elimination product accumulation. The amounts of every β - and δ -elimination product from (B) were quantitated using the phosphorimager and plotted over time. An arrow indicates the time point (25 min) when different temperatures (4°C or 65°C) were treated to the experiments in Figure 2-4. Error bars represent standard deviations from three independent experiments.

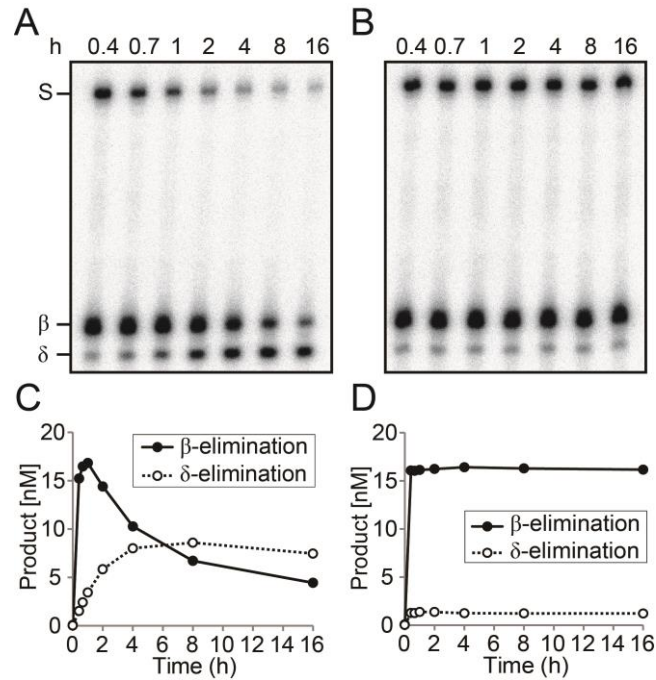


Figure 2-4. Enzyme-dependent δ -elimination process during 5mC excision

(A, B) Radiolabeled oligonucleotide duplex containing 5mC (25 nM) was reacted with purified MBP-DME Δ (100 nM) at 37°C for 25 min. Reactions were separated and paused on ice (A) or heat-inactivated at 65°C (B) for 15 min and then resumed at 37°C for additional 15 h and 35 min. Reactions were terminated at indicated time points and separated on a polyacrylamide gel. Oligonucleotide substrate (S) and β - and δ -elimination products (β , δ) are indicated to the left. (C, D) Relative amounts of β - and δ -elimination products from (A, B) were quantitated and plotted over time.

Arabidopsis encodes three putative AP endonucleases APE1L, APE2 and ARP

AP endonucleases are highly conserved among diverse species (Robertson et al., 2009) and the structures of *E. coli* exonuclease III and human AP endonuclease 1 (hAPE1) suggest a comparable mechanism for catalytic activity of AP endonucleases (Mol et al., 1995; Mol et al., 2000). The Arabidopsis genome contains three putative AP endonuclease genes *APE1L*, *APE2* and *ARP* (Figure 2-5A). All three AP endonucleases have a common endonuclease/exonuclease/phosphatase (EEP) domain that displays a significant homology to *E. coli* exonuclease III and hAPE1. The EEP domain comprises most protein structures of APE1L, whereas APE2 and ARP have an additional zinc-finger domain and SAP DNA binding domain, respectively (Figure 2-5A).

To date, ARP is the only plant AP endonuclease whose biochemical activity was characterized using purified protein or cell extract (Babiychuk et al., 1994; Córdoba-Cañero et al., 2011). Therefore, for comprehensive understanding of their biochemical characteristics related to BER and active DNA demethylation, I cloned all three putative AP endonuclease genes from Arabidopsis and prepared recombinant proteins from *E. coli* for *in vitro* assays.

APE1L and ARP process major 5mC excision intermediate 3'-PUA to generate 3'-OH

During the course of active DNA demethylation, DME and ROS1 catalyzed successive β - and δ -elimination processes after 5mC excision, generating 3'-PUA and 3'-phosphate blocking lesions, which should be immediately processed by following AP endonucleases (Figure 2-3B). To test whether Arabidopsis AP endonucleases can process 3'-PUA and 3'-phosphate acting downstream of DME-initiated DNA demethylation, I prepared a radiolabeled oligonucleotide duplex with 5mC on the top strand which was first reacted with DME for 1 h and then heat-inactivated to prevent additional enzymatic base excision. The DME reaction products, in which 3'-PUA was predominant in β -elimination and 3'-phosphate was minor in δ -elimination (Lane 2 in Figure 2-5B), were further reacted with MBP-APE1L, -APE2 or -ARP. Similar to hAPE1 (Lane 4 in Figure 2-5B), Arabidopsis APE1L and ARP catalyzed the conversion of 3'-PUA to 3'-OH, whereas APE2 displayed no discernable activity (Lanes 5-7 in Figure 2-5B). These results suggest that APE1L and ARP can remove harmful 3'-PUA, allowing nucleotide extension by DNA polymerase.

I hypothesized that following 5mC excision by DME family proteins, subsequent BER enzymes participate in the DNA demethylation

process by incorporating unmethylated cytidine in place of excised 5mC. I reconstituted the *in vitro* BER-mediated DNA demethylation pathway by demonstrating replacement of 5mC with unmethylated cytidine after DME base excision. After the conversion of 3'-PUA to 3'-OH by AP endonucleases (Lanes 3-7 in Figure 2-5B), the dCTP incorporation at the site of base excision was produced by Klenow DNA polymerase (Lanes 9-13 in Figure 2-5B). However, because incorporation of regular dCTP (18-nt) was hardly distinguishable from the spot corresponding to the 3'-PUA (β) on an acrylamide gel, I used Cy3-dCTP for better separation due to its high molecular weight. Accordingly, Cy3-dCTP was incorporated at the site of 5mC excision by Klenow fragment (Figure 2-5C). I observed that Cy3-dCTP was successfully inserted in the gap after the treatment of DME reaction products with APE1L or ARP (lanes 11 and 13 in Figure 2-5C), but no extension took place when treated with APE2 (lane 12 in Figure 2-5C). This demonstrates that both APE1L and ARP, but not APE2, successfully generate 3'-OH after 5mC excision, which is utilized for nucleotide extension by DNA polymerase during DNA demethylation.

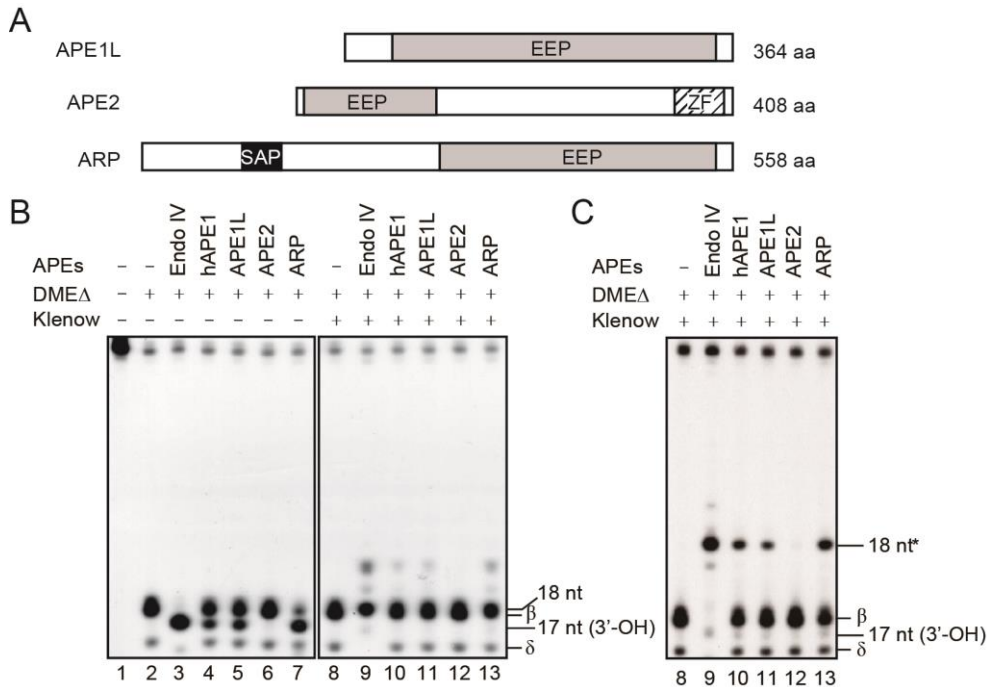


Figure 2-5. *In vitro* reconstitution of DNA demethylation with Arabidopsis AP endonucleases.

(A) Schematic diagrams of APE1L, APE2 and ARP proteins in Arabidopsis. EEP, endonuclease-exonuclease-phosphatase; SAP, SAF-A/B, Acinus and PIAS; ZF, GRF-type zinc-finger motif. (B) *In vitro* reconstitution of DNA demethylation with Arabidopsis AP endonucleases. Radiolabeled 35-mer DNA duplex with 5mC were reacted with MBP-DMEΔ (lane 2), and the reaction product was further incubated with each purified Arabidopsis AP endonuclease in the presence of 2.5 mM MgCl₂ (lanes 5–7). Subsequent dCTP incorporation (18-nt) was achieved by Klenow DNA polymerase to fill the gap generated by AP endonuclease (lanes 11–13). Reactions with *E. coli* Endonuclease IV (lanes 3 and 9) and human hAPE1 (lanes 4 and 10) were used as controls. The sizes of 3' end-processed (17-nt) and cytidine-

incorporated fragments (18-nt) relative to DME products (β and δ) were indicated to the right of the panel. Endo IV, Endonuclease IV. (C) *In vitro* reconstitution of DNA demethylation with Cy3-dCTP. As the size of cytidine-incorporated fragment is indistinguishable from β -elimination product, a larger molecule Cy3-dCTP was incorporated into the AP site after AP endonuclease reaction. Cy3-dCTP incorporation (18-nt) was denoted with an asterisk.

3' phosphatase activity of ARP

DME is shown to catalyze the β -elimination first to generate a 3'-PUA and a subsequent enzyme-dependent δ -elimination process produced a 3'-phosphate. This implies that an efficient removal of 3'-phosphate is a critical step towards nucleotide extension after 5mC excision. Since the 3' phosphatase activity of *E. coli* Endonuclease IV was previously reported (Bailey and Verly, 1989), I investigated whether Arabidopsis AP endonucleases are also able to process a 3'-phosphate generated by δ -elimination. Flanking a 1-nt central gap, both a 5' end-labeled 17-nt oligonucleotide with a 3'-phosphate and an unlabeled 17-nt oligonucleotide with a 5'-phosphate were annealed to the complementary strand to prepare a 35-nt oligonucleotide duplex mimicking a δ -elimination product (Figure 2-6C). This oligonucleotide substrate was incubated with each Arabidopsis AP endonuclease. APE1L or APE2 did not convert the 3'-phosphate of δ -elimination product to 3'-OH, whereas ARP displayed significant 3' phosphatase activity to generate 3'-OH (Figures 2-6A and 2-6B). The time-course study showed that ARP processed the 3'-phosphate at a slow rate reaching its plateau after 8 h of reaction (Figure 2-6D). This result strongly suggests that ARP itself is capable of processing both DME-catalyzed 5mC

excision intermediates 3'-PUA and 3'-phosphate with no support from other enzymes such as a 3' phosphatase.

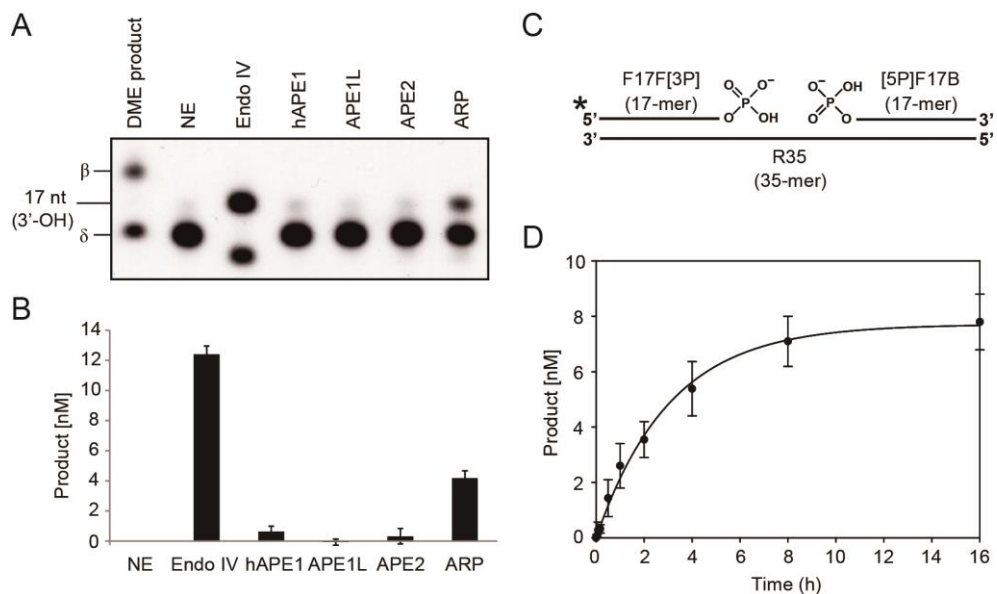


Figure 2-6. The 3' phosphatase activity of Arabidopsis AP endonucleases.

(A) The 3' phosphatase activity of Arabidopsis AP endonucleases. DNA substrate depicted in (C) was reacted with purified MBP-APE1L, -APE2 or -ARP at 37°C for 1 h. Only MBP-ARP protein converted a δ -elimination product analog to 3'-OH, like *E. coli* Endonuclease IV (Endo IV). Either MBP-APE1L or -APE2 did not catalyze such conversion, like human APE1 (hAPE1). Methylated DNA substrate used for DNA glycosylase assay was reacted with MBP-DME Δ and loaded alongside for size comparison. NE, no enzyme control. (B) Relative amounts of 3' phosphatase reaction products of Arabidopsis AP endonucleases. The amounts of 17-nt reaction products were measured by phosphorimager. (C) Structure of 35-mer oligonucleotide duplex that mimics δ -elimination product catalyzed by DME for 3' phosphatase assay. The radiolabeled upstream 17-mer oligonucleotide with a 3'-phosphate (F17F[3P]) and the downstream 17-mer with a 5'-phosphate

([5P]F17B) are annealed together to the complementary 35-mer strand (R35) to produce DNA substrate with a 1-nt gap in the middle. (D) Kinetics analysis of ARP 3' phosphatase activity. The above DNA substrate (25 nM) was reacted with MBP-ARP (5nM) at 37°C in a time-course manner. The amounts of 3' phosphatase reaction products at indicated time point were measured and plotted over time. Error bars represent standard deviations from three independent experiments.

Biochemical activities of MBP-free AP endonucleases

Although an N-terminal or C-terminal tagged protein has been typically used in biochemical studies, there is a possibility that the MBP tag might affect the activity of Arabidopsis AP endonucleases. Since all Arabidopsis AP endonucleases in this study were expressed with an N-terminal MBP tag, MBP was removed by PreScission Protease and the MBP-free AP endonucleases were repurified (Figure 2-2C). Removal of MBP had no effect on enzyme activity compared to MBP fusion but somewhat decreased the protein stability (Figure 2-7).

ARP has AP site incision activity

Besides the 3' phosphodiesterase activity and 3' phosphatase activity, AP endonuclease plays an essential role in processing AP sites generated either spontaneously or by monofunctional DNA glycosylases (Kim and Wilson III, 2012; Robertson et al., 2009). AP endonuclease catalyzes the incision of the DNA-sugar phosphate backbone at 5' of AP sites to prime DNA repair synthesis. To determine whether the Arabidopsis AP endonucleases have the canonical AP site incision activity, each MBP-free AP endonuclease was incubated with an end-labeled 35-mer oligonucleotide duplex containing a single AP site analogue THF. Consistent with previous

reports (Babiychuk et al., 1994; Córdoba-Cañero et al., 2011), purified ARP showed incision activity at AP sites generating 3'-OH, whereas neither APE1L nor APE2 displayed activity (Figure 2-7A). This result implies that ARP is the primary AP endonuclease in Arabidopsis incising AP sites in the BER pathway acting immediately downstream of monofunctional DNA glycosylases.

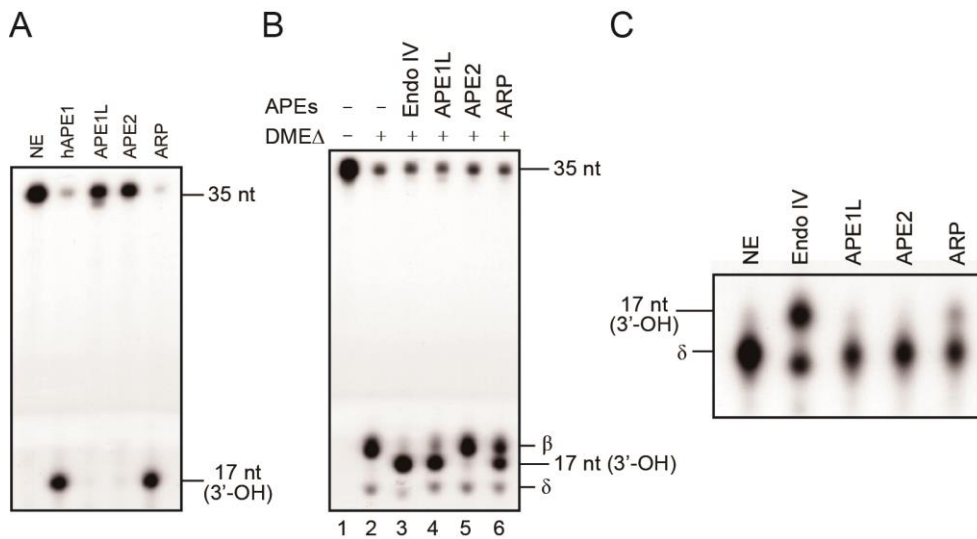


Figure 2-7. *In vitro* AP endonuclease activities of Arabidopsis MBP-free AP endonucleases.

(A) AP site incision activity of Arabidopsis AP endonucleases. Radiolabeled 35-mer oligonucleotides (25 nM) containing a THF, an AP site analog, at position 18 (R35AP) were reacted with 5 nM each of purified MBP-free APE1L, APE2 and ARP in the AP endonuclease reaction buffer at 37°C for 30 min. Sizes of the substrate (35-nt) and the product (17-nt) were indicated to the right. NE, no enzyme; hAPE1, human APE1. (B) The 3' phosphodiesterase activity on DME-treated products. Radiolabeled 35-mer oligonucleotides (25 nM) containing 5mC were reacted with DME (lane 2), and the reaction product was further incubated with each purified Arabidopsis AP endonuclease in the presence of 2.5 mM MgCl₂ (lanes 4-6). The 3' end-processed fragment (17-nt) was indicated relative to DME treated products (β, δ) to the right of the panel. Endo IV, Endonuclease IV. (C) The 3'-phosphatase activity of purified Arabidopsis AP endonucleases. The 35-mer oligonucleotide duplex (25 nM) that mimics a DME-catalyzed

δ -elimination product was prepared and reacted with 5 nM each of APE1L, APE2 and ARP in the AP endonuclease buffer at 37°C for 60 min. The DNA substrate (δ) and the product (17-nt) were indicated to the left.

DNA binding activity of AP endonucleases

It is previously reported that some AP endonucleases have additional functions besides the AP site processing including transcription factor stimulation and checkpoint signaling (Babiychuk et al., 1994; Willis et al., 2013; Xanthoudakis et al., 1992). For recruiting other proteins or modulating their activities, AP endonuclease should have a strong affinity to DNA where the events occur regardless of enzyme activity. Thus, I examined DNA binding activity of Arabidopsis AP endonucleases using the electrophoretic mobility shift assay (Figure 2-8). All Arabidopsis AP endonucleases were found to bind DNA substrate containing THF or 5mC. Interestingly, even enzymatically inactive APE2 displayed strong affinity to DNA, suggesting that APE2 DNA binding is independent of enzyme activity. Also, considering methylated DNA is not a direct substrate of AP endonuclease, all Arabidopsis AP endonucleases appear to have non-specific DNA binding properties. This suggests that Arabidopsis AP endonucleases, including biochemically inactive APE2, may have some unknown biological functions.

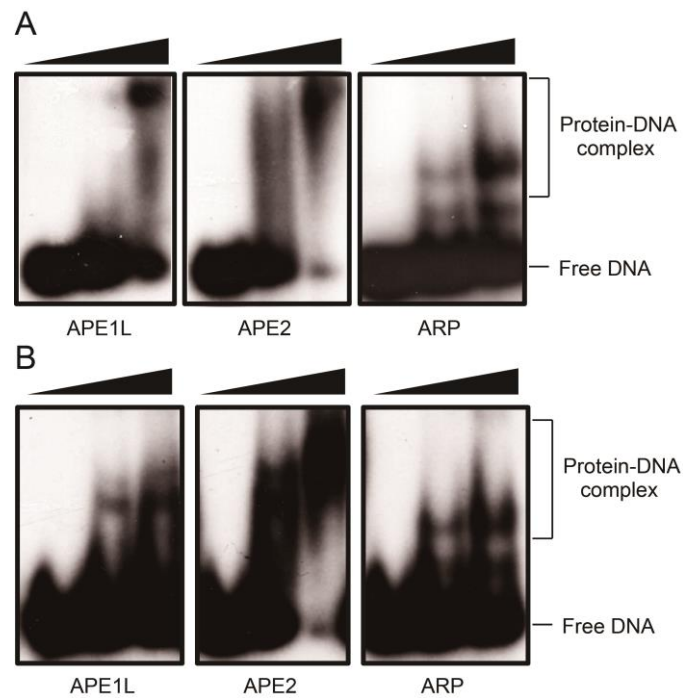


Figure 2-8. Electrophoretic mobility shift assay of Arabidopsis AP endonucleases.

Electrophoretic mobility shift assay was performed with radiolabeled 35-mer oligonucleotides containing a THF (A) or 5mC (B). Oligonucleotide substrates (50 nM) were incubated with increasing amounts (0, 100 and 300 nM) of each MBP-APE1L, -APE2 or -ARP on ice for 10 min. The protein-DNA complex and free DNA are indicated to the right.

DISCUSSION

Although DNA methylation systems are highly conserved between plants and mammals, DNA demethylation systems are evolutionarily divergent. The most striking difference is the presence of enzymes that remove 5mC directly from DNA. In plants, the 5mC DNA glycosylase family, also known as the DME family proteins, are capable of catalyzing excision of 5mC at target DNA sequences for gene activation (Choi et al., 2002; Gong et al., 2002; Gehring et al., 2006; Hsieh et al., 2009). DME 5mC DNA glycosylases produce 3'-PUA and 3'-phosphate by successive β - and δ -elimination processes during the course of 5mC excision (Gehring et al., 2006). These 3' structures remain as blocking lesions for subsequent nucleotide extension, which should be immediately processed to form 3'-OH and the two enzyme activities are likely engaged. One involves 3' phosphatase activity, for which ZDP was recently proposed to process 3'-phosphate after 5mC excision (Martinez-Macias et al., 2012), and the other enzyme may employ 3' phosphodiesterase activity provided by AP endonucleases. Since DME family proteins have a relatively slow turnover rate for base excision compared to other conventional DNA glycosylases

(Ponferrada-Marin et al., 2009), the production of 3'-PUA and 3'-phosphate requires a significant amount of time for reaction completion (Figure 2-4A). In addition, DME/ROS1 proteins produce 3'-phosphate as an end product, and the responsible δ -elimination reaction is an enzyme-dependent and extremely slow process (Figure 2-4). This implies that ZDP 3' phosphatase by itself cannot process 3'-phosphate immediately upon 5mC excision, which should induce critical damage on DNA due to persisting SSBs and harmful 3' blocking lesions. Thus, for plants to maintain genome stability during DNA demethylation, additional AP endonuclease activities must exist for the removal of 3'-PUA and/or 3'-phosphate shortly after 5mC excision.

Arabidopsis encodes three AP endonucleases APE1L, APE2 and ARP (Figure 2-5A), but only ARP has been characterized for its biochemical activity (Babiychuk et al., 1994; Córdoba-Cañero et al., 2011). Previous reports showed that Arabidopsis ARP has an AP site incision activity toward acid-depurinated DNA, and plays an important role in progression of short- and long-patch BER (Babiychuk et al., 1994; Córdoba-Cañero et al., 2011). However, the *in vitro* enzymatic activities of APE1L and APE2 were not comparatively examined under the same reaction conditions. The comprehensive biochemical analysis of three AP

endonucleases reveals that only ARP has the AP site incision activity (Figure 2-7A), and both APE1L and ARP are capable of processing 3'-PUA to generate 3'-OH allowing subsequent nucleotide incorporation (Figures 2-5B and 2-5C). By contrast, APE2 displayed no discernable biochemical activity for any substrate examined in this study. Interestingly, APE1L showed significant activity for 5mC excision intermediates but not for an AP site analog THF (Figure 2-5B), suggesting that APE1L functions specifically in the DNA demethylation pathway and the 3'-end trimming process requires mechanisms distinct from those of AP site incision. In addition, ARP showed 3' phosphatase activity like *E. coli* Endonuclease IV (Figure 2-6), which implies that ARP itself might be sufficient to process diverse base excision intermediates such as an AP site, 3'-PUA and 3'-phosphate. Notably, the enzymatic activity of ARP to process both 3'-PUA and 3'-phosphate suggests the possibility that ARP can effectively remove 5mC excision intermediates without the participation of another enzymes.

It was previously shown that no *ape1l ape2* double homozygous mutant was retrieved from genetic crosses among AP endonuclease mutants, suggesting that at least one of intact *APE1L* or *APE2* gene is necessary for seed development in Arabidopsis (Murphy et al., 2009). This indicates a functional redundancy between *APE1L* and *APE2*, although *ARP* can be

dispensable for reproductive development. This is quite surprising not only in that ARP was shown to process most harmful lesions generated by base excision, but also in the fact that APE2 showed no catalytic activity considering its incomplete catalytic domain. However, I still cannot rule out the possibility that some enzymes are capable of replacing the function of ARP, and that APE2 has additional functions besides canonical AP endonuclease activity. Supporting this idea, several lines of evidence showed that AP endonuclease plays an essential role for signal transduction and activation of transcription factors in diverse organisms. For example, *Xenopus* APE2 is required to activate checkpoint kinase 1 (Chk1) in response to oxidative stress using its Chk1-binding motif (Willis et al., 2013). Human APE1 also enhances the DNA binding activity of heterodimers of Fos-Jun transcription factors (Xanthoudakis et al., 1992), and the similar effect was reported for *Arabidopsis* ARP (Babiychuk et al., 1994). In addition, *Arabidopsis* APE2 still has strong DNA binding activity (Figure 2-8), suggesting that APE2 may have some distinct functions regardless of canonical AP endonuclease activity.

Considering proper DNA demethylation is important for female gametophyte-specific gene imprinting (Choi et al., 2002; Gehring et al., 2006; Huh et al., 2008), it is possible that mutations in AP endonucleases

likely induce gametophyte lethality by dysregulation of DNA demethylation by DME in the central cell before fertilization. However, I cannot exclude the possibility that there is an unknown mechanism that allows DNA demethylation independently of AP endonuclease activity in the female gametophyte. Another study proposed that ZDP 3' phosphatase is essential for DNA demethylation and its mutation caused DNA hypermethylation and transcriptional gene silencing of a reporter gene (Martínez-Macías et al., 2012). Although *zdp* mutant did not exhibit any conspicuous developmental defects, *ape11 zdp* double mutant showed embryonic lethal phenotype with DNA hypermethylation and down-regulation of imprinted genes in the endosperm (Martínez-Macías et al., 2012; Li et al., 2015b). These results suggest a distinct function of APE1L for the establishment of female gametophyte-specific DNA demethylation and gene activation by DME, which can be partly compensated by ZDP.

Biochemical studies on DNA demethylases and AP endonucleases allowed me to propose a model of plant-specific DNA demethylation pathways involving BER (Figure 2-9). A damaged, mismatched or modified (5mC in this case) base is recognized by monofunctional or bifunctional DNA glycosylases. Monofunctional DNA glycosylases such as MAG, MYH and UNG catalyze only the cleavage of an N-glycosidic bond

between a base and a ribose sugar leaving an AP site (the pathway on the left in Figure 2-9), whereas bifunctional enzymes such as FPG, OGG1, NTH, MBD4L, and importantly, DME/ROS1 family DNA demethylases cleave the phosphodiester bond 3' to the AP site, concomitantly with base removal using an intrinsic AP-lyase activity (the pathway in the middle in Figure 2-9). The primary base excision intermediates have abnormal 3'-end structures – 3'-PUA and 3'-phosphate generated by β - and δ -elimination processes, respectively. The 3'-PUA is removed by APE1L or ARP, and δ -elimination product 3'-phosphate can be processed by ARP or ZDP. After the generation of 3'-OH, DNA polymerase will insert an unmethylated cytosine and DNA ligase seal the gap. Specific DNA polymerase is still unclear, but DNA polymerase λ is presumably thought to perform this function (García-Díaz et al., 2000; Uchiyama et al., 2004; Uchiyama et al., 2009). The final ligation step of BER in Arabidopsis is likely performed by DNA ligase I (LIG1) (Córdoba-Cañero et al., 2011; Li et al., 2015a). Given the fact that active DNA demethylation utilizes many of DNA repair systems, it is important to investigate the repair machineries acting downstream of 5mC excision to better understand the molecular dynamics of DNA demethylation processes. This will also provide an insight into the active DNA demethylation systems for epigenetic gene regulation in plants.

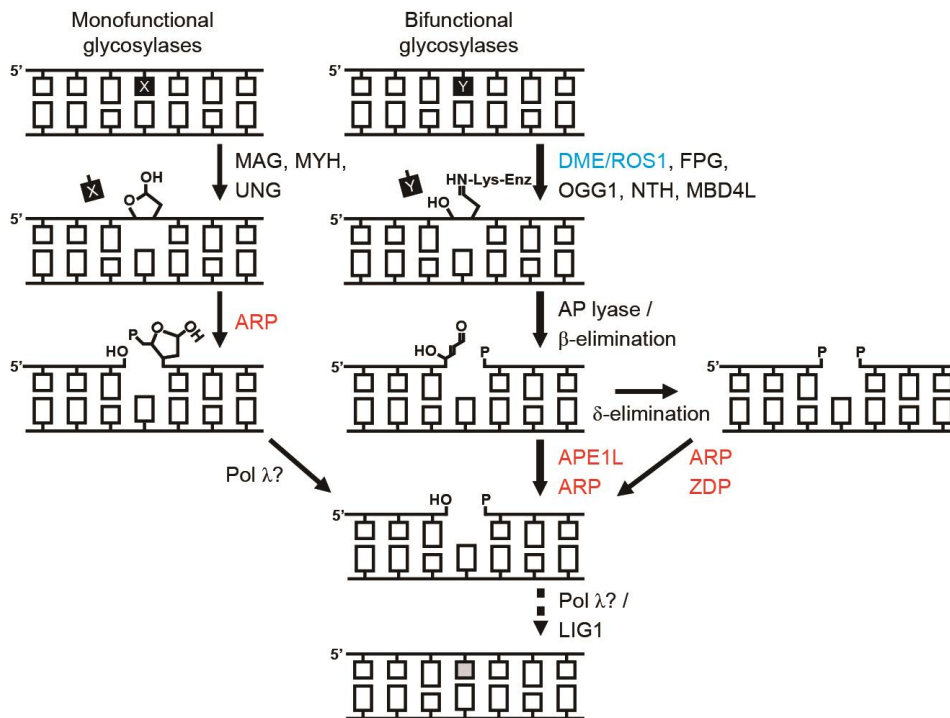


Figure 2-9. Model of active DNA demethylation pathway in Arabidopsis.

Various forms of bases are addressed by monofunctional or bifunctional DNA glycosylases. After base excision by monofunctional DNA glycosylases, the AP endonuclease ARP primarily catalyzes the cleavage of the phosphodiester bond, leaving a 3'-OH and 5'-dRP (shown in the left). The resulting 5'-dRP is presumably processed by DNA polymerase λ to generate 5'-phosphate at the nick. For bifunctional DNA glycosylases including the DME/ROS1 family of 5mC DNA glycosylases, base excision and a strand break simultaneously occur due to associated AP-lyase activity (shown in the middle). The resulting 3'-PUA is either directly processed by AP endonucleases APE1L and ARP, or further converted to 3'-phosphate via δ -elimination. This 3'-phosphate can be processed by ARP or ZDP 3' phosphatase (shown in the right). Consequently, DNA polymerase λ and

LIG1 will complete BER. The proposed pathway of active DNA demethylation in Arabidopsis is denoted by thick arrows in the middle.

REFERENCES

- Agius, F., Kapoor, A., and Zhu, J.K.** (2006). Role of the Arabidopsis DNA glycosylase/lyase ROS1 in active DNA demethylation. *Proc. Natl. Acad. Sci. USA* **103**: 11796-11801.
- Allis, C.D., and Jenuwein, T.** (2016). The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.* **17**: 487-500.
- Babiychuk, E., Kushnir, S., Van Montagu, M., and Inze, D.** (1994). The *Arabidopsis thaliana* apurinic endonuclease Arp reduces human transcription factors Fos and Jun. *Proc. Natl. Acad. Sci. USA* **91**: 3299-3303.
- Bailly, V., and Verly, W.G.** (1989). The multiple activities of Escherichia coli endonuclease IV and the extreme lability of 5'-terminal base-free deoxyribose 5-phosphates. *Biochem. J.* **259**: 761-768.
- Caldecott, K.W.** (2008). Single-strand break repair and genetic disease. *Nat. Rev. Genet.* **9**: 619-631.
- Choi, Y., Gehring, M., Johnson, L., Hannon, M., Harada, J.J., Goldberg, R.B., Jacobsen, S.E., and Fischer, R.L.** (2002). DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in Arabidopsis. *Cell* **110**: 33-42.
- Cordoba-Canero, D., Roldan-Arjona, T., and Ariza, R.R.** (2011). Arabidopsis ARP endonuclease functions in a branched base excision DNA repair pathway completed by LIG1. *Plant J.* **68**: 693-702.
- Cordoba-Canero, D., Morales-Ruiz, T., Roldan-Arjona, T., and Ariza, R.R.** (2009). Single-nucleotide and long-patch base excision repair

of DNA damage in plants. *Plant J.* **60**: 716-728.

- Cortellino, S., Xu, J.F., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D., Abramowitz, L.K., Bartolomei, M.S., Rambow, F., Bassi, M.R., Bruno, T., Fanciulli, M., Renner, C., Klein-Szanto, A.J., Matsumoto, Y., Kobi, D., Davidson, I., Alberti, C., Larue, L., and Bellacosa, A.** (2011). Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* **146**: 67-79.
- Garcia-Diaz, M., Dominguez, O., Lopez-Fernandez, L.A., de Lera, L.T., Saniger, M.L., Ruiz, J.F., Parraga, M., Garcia-Ortiz, M.J., Kirchhoff, T., del Mazo, J., Bernad, A., and Blanco, L.** (2000). DNA polymerase lambda (Pol lambda), a novel eukaryotic DNA polymerase with a potential role in meiosis. *J. Mol. Biol.* **301**: 851-867.
- Gehring, M., Reik, W., and Henikoff, S.** (2009). DNA demethylation by DNA repair. *Trends Genet.* **25**: 82-90.
- Gehring, M., Huh, J.H., Hsieh, T.F., Penterman, J., Choi, Y., Harada, J.J., Goldberg, R.B., and Fischer, R.L.** (2006). DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. *Cell* **124**: 495-506.
- Gong, Z., Morales-Ruiz, T., Ariza, R.R., Roldan-Arjona, T., David, L., and Zhu, J.K.** (2002). ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase. *Cell* **111**: 803-814.
- Guo, J.U., Su, Y., Zhong, C., Ming, G.L., and Song, H.** (2011). Hydroxylation of 5-methylcytosine by TET1 promotes active DNA

demethylation in the adult brain. *Cell* **145**: 423-434.

- Hsieh, T.F., Ibarra, C.A., Silva, P., Zemach, A., Eshed-Williams, L., Fischer, R.L., and Zilberman, D.** (2009). Genome-wide demethylation of Arabidopsis endosperm. *Science* **324**: 1451-1454.
- Huh, J.H., Bauer, M.J., Hsieh, T.F., and Fischer, R.L.** (2008). Cellular programming of plant gene imprinting. *Cell* **132**: 735-744.
- Jang, H., Shin, H., Eichman, B.F., and Huh, J.H.** (2014). Excision of 5-hydroxymethylcytosine by DEMETER family DNA glycosylases. *Biochem. Biophys. Res. Commun.* **446**: 1067-1072.
- Jullien, P.E., Kinoshita, T., Ohad, N., and Berger, F.** (2006). Maintenance of DNA methylation during the Arabidopsis life cycle is essential for parental imprinting. *Plant Cell* **18**: 1360-1372.
- Kim, Y.J., and Wilson, D.M., 3rd.** (2012). Overview of base excision repair biochemistry. *Curr. Mol. Pharmacol.* **5**: 3-13.
- Law, J.A., and Jacobsen, S.E.** (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **11**: 204-220.
- Li, Y., Duan, C.G., Zhu, X., Qian, W., and Zhu, J.K.** (2015a). A DNA ligase required for active DNA demethylation and genomic imprinting in Arabidopsis. *Cell Res.* **25**: 757-760.
- Li, Y., Cordoba-Canero, D., Qian, W., Zhu, X., Tang, K., Zhang, H., Ariza, R.R., Roldan-Arjona, T., and Zhu, J.K.** (2015b). An AP endonuclease functions in active DNA demethylation and gene imprinting in Arabidopsis. *PLoS Genet.* **11**: e1004905.
- Martinez-Macias, M.I., Qian, W., Miki, D., Pontes, O., Liu, Y., Tang, K., Liu, R., Morales-Ruiz, T., Ariza, R.R., Roldan-Arjona, T., and Zhu, J.K.** (2012). A DNA 3' phosphatase functions in active DNA

demethylation in Arabidopsis. *Mol. Cell* **45**: 357-370.

- Mol, C.D., Izumi, T., Mitra, S., and Tainer, J.A.** (2000). DNA-bound structures and mutants reveal abasic DNA binding by APE1 and DNA repair coordination [corrected]. *Nature* **403**: 451-456.
- Mol, C.D., Kuo, C.F., Thayer, M.M., Cunningham, R.P., and Tainer, J.A.** (1995). Structure and function of the multifunctional DNA-repair enzyme exonuclease III. *Nature* **374**: 381-386.
- Morales-Ruiz, T., Ortega-Galisteo, A.P., Ponferrada-Marin, M.I., Martinez-Macias, M.I., Ariza, R.R., and Roldan-Arjona, T.** (2006). DEMETER and REPRESSOR OF SILENCING 1 encode 5-methylcytosine DNA glycosylases. *Proc. Natl. Acad. Sci. USA* **103**: 6853-6858.
- Murphy, T.M., Belmonte, M., Shu, S., Britt, A.B., and Hatteroth, J.** (2009). Requirement for abasic endonuclease gene homologues in Arabidopsis seed development. *PLoS One* **4**: e4297.
- Ortega-Galisteo, A.P., Morales-Ruiz, T., Ariza, R.R., and Roldan-Arjona, T.** (2008). Arabidopsis DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks. *Plant Mol. Biol.* **67**: 671-681.
- Penterman, J., Zilberman, D., Huh, J.H., Ballinger, T., Henikoff, S., and Fischer, R.L.** (2007). DNA demethylation in the Arabidopsis genome. *Proc. Natl. Acad. Sci. USA* **104**: 6752-6757.
- Ponferrada-Marin, M.I., Roldan-Arjona, T., and Ariza, R.R.** (2009). ROS1 5-methylcytosine DNA glycosylase is a slow-turnover catalyst that initiates DNA demethylation in a distributive fashion. *Nucleic Acids Res.* **37**: 4264-4274.
- Robertson, A.B., Klungland, A., Rognes, T., and Leiros, I.** (2009). DNA

repair in mammalian cells: Base excision repair: the long and short of it. *Cell. Mol. Life Sci.* **66**: 981-993.

Scharer, O.D., and Jiricny, J. (2001). Recent progress in the biology, chemistry and structural biology of DNA glycosylases. *Bioessays* **23**: 270-281.

Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14**: 204-220.

Song, C.X., and He, C. (2013). Potential functional roles of DNA demethylation intermediates. *Trends Biochem. Sci.* **38**: 480-484.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., and Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**: 930-935.

Uchiyama, Y., Takeuchi, R., Kodera, H., and Sakaguchi, K. (2009). Distribution and roles of X-family DNA polymerases in eukaryotes. *Biochimie* **91**: 165-170.

Uchiyama, Y., Kimura, S., Yamamoto, T., Ishibashi, T., and Sakaguchi, K. (2004). Plant DNA polymerase lambda, a DNA repair enzyme that functions in plant meristematic and meiotic tissues. *Eur. J. Biochem.* **271**: 2799-2807.

Williams, K., Christensen, J., Pedersen, M.T., Johansen, J.V., Cloos, P.A., Rappsilber, J., and Helin, K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**: 343-348.

Willis, J., Patel, Y., Lentz, B.L., and Yan, S. (2013). APE2 is required for ATR-Chk1 checkpoint activation in response to oxidative stress.

Proc. Natl. Acad. Sci. USA **110**: 10592-10597.

Wu, S.C., and Zhang, Y. (2010). Active DNA demethylation: many roads lead to Rome. *Nat. Rev. Mol. Cell Biol.* **11**: 607-620.

Xanthoudakis, S., Miao, G., Wang, F., Pan, Y.C., and Curran, T. (1992). Redox activation of Fos-Jun DNA binding activity is mediated by a DNA repair enzyme. *EMBO J.* **11**: 3323-3335.

Xu, Y., Wu, F., Tan, L., Kong, L., Xiong, L., Deng, J., Barbera, A.J., Zheng, L., Zhang, H., Huang, S., Min, J., Nicholson, T., Chen, T., Xu, G., Shi, Y., Zhang, K., and Shi, Y.G. (2011). Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Mol. Cell* **42**: 451-464.

Zhu, J.K. (2009). Active DNA Demethylation Mediated by DNA Glycosylases. *Annu. Rev. Genet.* **43**: 143-166.

CHAPTER 3

Application to epigenome editing by targeted DNA demethylation

ABSTRACT

DNA methylation is a stable epigenetic mark implicated in diverse biological processes including gene imprinting and transposon silencing by regulating chromatin structure and gene expression. DNA methylation is antagonistically regulated by enzymatic chromatin modifiers such as DNA methyltransferase and demethylase. To precisely modify epigenetic marks, DNA binding modules including zinc finger protein, transcription activator-like effector (TALE) and the clustered, regularly interspaced, short palindromic repeats (CRISPR)/dCas9 system can be combined with diverse chromatin modifiers. Recent studies report successful modulation of DNA methylation with DNA methyltransferase or mammalian demethylase, whereas implication of plant DNA demethylase in epigenome editing is still elusive. In this study, I designed programmable TALE DNA binding modules targeting *FLOWERING OF WAGENINGEN (FWA)* locus in *Arabidopsis*, one of the well-characterized epialleles regulated by DNA methylation. The TALE modules were fused with plant-derived DNA demethylase DEMETER (DME) to produce TALE-DME fusion proteins. TALE-DME displayed significant DNA demethylation activity on *FWA* locus in T1 transgenic plants, and one of the transgenic lines exhibited

dramatic increase of *FWA* expression level along with late flowering phenotype. This study not only provides a powerful tool for regulating gene expression by epigenome editing, but an opportunity for comprehensive understanding of the causal links between transcription and DNA demethylation in plants.

INTRODUCTION

DNA methylation is a stable epigenetic mark that is essential for gene imprinting, transposon silencing, and many developmental processes in higher eukaryotes (Huh et al., 2008; Smith and Meissner, 2013). DNA methylation usually occurs in symmetric CG sequence context in animals, which is established by DNA methyltransferase 3 (DNMT3) and maintained by maintenance methyltransferase DNMT1. In plants, however, DNA methylation occurs at cytosines in both symmetric (CG, CHG) and asymmetric (CHH) sequence contexts. DNA methylation patterns of CG, CHG and CHH are achieved by DNA METHYLTRANSFERASE 1 (MET1), CHROMOMETHYLASE 3 (CMT3) and DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2), respectively (Law and Jacobsen, 2010; Wu and Zhang, 2010; Lyco, 2018).

Although plants and animals have significantly conserved DNA methylation machineries, their DNA demethylation systems appear to have independently evolved. In animals, Ten-eleven translocation (TET) family proteins are responsible for catalyzing the oxidation of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC). TET proteins further convert

5hmC to 5-formylcytosine and 5-carboxylcytosine, both of which are known to be removed by thymine-DNA glycosylase (TDG). After base excision by TDG, subsequent base excision repair enzymes complete DNA demethylation (Kohli and Zhang, 2013; Wu and Zhang, 2017). Unlike animals that require a series of enzymatic modifications of 5mC for DNA demethylation, plants utilize the DEMETER (DME) DNA glycosylase family proteins which directly excise 5mC from DNA, replacing it with unmethylated cytosine via the base excision repair pathway (Huh et al., 2008; Law and Jacobsen, 2010). Through DNA demethylation, the expression of several target genes can be reactivated (Huh et al., 2008).

The *FLOWERING OF WAGENINGEN (FWA)* locus is a representative epiallele regulated by DNA methylation. In wild-type *Arabidopsis*, the expression of *FWA* is repressed with hypermethylation at the promoter region throughout the life cycle, except for its temporary expression in endosperm. However, demethylation of this locus in *fwa-1* mutant led to ectopic expression of *FWA* with late flowering phenotype (Kinoshita et al., 2006; Weigel and Colot, 2012). Recent studies about epigenome editing focused on *FWA* locus because of the obvious phenotypic difference and stable maintenance of epigenetic pattern in *FWA*

locus (Johnson et al., 2014; Gallego-Bartolomé et al., 2018; Papikian et al., 2019).

Epigenome editing directly alters chromatin marks including DNA methylation and histone modifications at specific genomic loci, which regulates transcription of downstream target gene (Kungulovski and Jeltsch, 2015). Targeted epigenome manipulation is typically based on enzymatic chromatin modifiers and programmable DNA binding modules. The DNA binding modules for specific DNA recognition, commonly include zinc finger proteins (ZFP), transcription activator-like effector (TALE), and clustered, regularly interspaced, short palindromic repeats (CRISPR)/Cas9 system. ZFP and TALE have been extensively used as targeting platforms, which recognize DNA by sequence-specific protein-DNA interaction, whereas the CRISPR/Cas9 system is based on Watson-Crick base pairing between a guide RNA and target DNA strand (Sander and Joung, 2014; Kungulovski and Jeltsch, 2015; Thakore et al., 2016). Due to intrinsic nuclease activity of Cas9 protein that probably disturbs epigenome editing, mutagenesis of catalytic residues in Cas9 DNA cleavage domains was performed, generating a dead Cas9 (dCas9). The dCas9 protein is catalytically inactive for DNA cleavage, but still retains its DNA binding

activity, which can be widely used for epigenome editing (Cheng et al., 2013; Mali et al., 2013).

Previous studies have reported that programmable DNA binding modules fused with diverse chromatin modifiers including DNA methyltransferases and mammalian DNA demethylases are capable of regulating the expression of target gene by epigenome editing (Morita et al., 2016; Huang et al., 2017; Gallego-Bartolomé et al., 2018; Papikian et al., 2019). However, it is still elusive whether plant DNA demethylase can be used for precise epigenome editing. In this study, plant-specific DNA demethylase DME was fused with the TALE DNA binding module to produce TALE-DME fusion protein. Several lines of T1 transgenic plants that overexpress *TALE-DME* were obtained, which induced DNA demethylation in target *FWA* promoter region. In addition, one of the T1 transgenic lines displayed dramatic increase of *FWA* expression along with late flowering phenotype. This study provides a powerful tool for targeted DNA demethylation by application of plant-derived DNA demethylase to epigenome editing.

MATERIALS AND METHODS

Cloning of the TALE-DME fusion constructs

Oligonucleotides used in this study are listed in Table 3-1. The CaMV 35S promoter (p35S) sequence was introduced into the GUS site of the pBI101.1 binary vector, producing pBI101.1-p35S without a GUS tag. The multiple cloning sites (MCS) of pBI101.1-p35S were manipulated using oligonucleotides DG2338 and DG2339 to provide proper restriction enzyme sites, generating a pBJ101 vector (Figure 3-1A). The pBJ101 binary vector was subjected to further cloning steps below, and used for *Arabidopsis* transformation. For construction of the TALE-DME fusion fragments, the MCS of p326-RFP (Choi et al., 2005) was replaced with newly synthesized MCS (Figure 3-1B) using oligonucleotides DG2332 and DG2333 to produce a pJ326 vector. The pJ326 vector was used for following TALE-DME cloning.

A 3xFLAG sequence containing *Xba* I and *Xho* I restriction sites was PCR-amplified with primers DG2334 and DG2335 and introduced into the pJ326 vector. Subsequently, TALE fragments containing an N-terminal HA and an NLS were cloned into the *Pme* I and *Rsr* II sites of the pJ326-3xFlag vector. Diverse TALE modules targeting *FWA* promoter region were

designed and synthesized by ToolGen (Table 3-2). Each TALE module was PCR-amplified with primers DG2388 and DG2389 that harbors *Pme* I and *Rsr* II restriction sites. Finally, the DME Δ fragment (Figure 2-1; Jang et al., 2014) was cloned into the *Bam* HI and *Sal* I sites located downstream of the TALE sequence to produce pJ326-TALE-DME. The TALE-DME fragments fused with both N-terminal 3xFLAG and HA tags were digested with *Pml* I and *Avr* II, and then cloned into the corresponding sites of the previously produced pBJ101 vector. The pBJ101-TALE-DME and pBJ101 vector constructs were used for following *Agrobacterium*-mediated *Arabidopsis* transformation.

Plant materials and growth conditions

Arabidopsis thaliana ecotype Columbia-0 (Col-0) was subjected to *Agrobacterium*-mediated transformation. The *fwa-1* mutant was adopted from the previous report (Koornneef et al., 1991). Seeds were sterilized with 30% bleach solution and stratified at 4°C for 2 days, and plated on a 0.5x MS nutrient medium with 1% sucrose and 0.8% plant agar. Germinated seedlings were transferred to soil and grown in the growth room under 16 h of light and 8 h of dark cycles at 23°C. To measure flowering time, the

number of rosette leaves was counted when the visible bolt of each plant was appeared.

***Agrobacterium*-mediated *Arabidopsis* transformation**

Agrobacterium GV3101 cells containing each plasmid construct were prepared with heat-shock transformation. Cells were incubated in LB liquid media at 30°C for 2 h and then on LB solid media with 50 µg/mL of kanamycin and 25 µg/mL of rifampicin at 30°C for 2 days. A floral dipping method was performed for *Agrobacterium*-mediated infiltration on 6 week-old plants. Cell suspension was prepared in 500 mL of LB medium with 50 µg/mL of kanamycin and 25 µg/ml of rifampicin at 30°C for 24 h. Cells were harvested by centrifugation at 7,000 rpm for 20 min at 4°C, and resuspended in infiltration media (0.5x MS salts, 5% sucrose, 0.025% Silwet-77, pH 5.0). Floral buds of plants were dipped in *Agrobacterium* suspension for 1.5 min, kept from the light overnight and grown under normal conditions (Clough and Bent, 1998).

Genotyping of the transgenic plants

The T1 transgenic plants were selected on 0.5x MS medium with 50 µg/mL of kanamycin. For genomic DNA (gDNA) extraction, cotyledons

from 14-day-old seedlings were grinded and incubated in the gDNA extraction buffer (200 mM Tris-HCl, pH 7.4, 250 mM NaCl, 25 mM EDTA, 0.5% SDS) at 65°C for 1 h. After centrifugation at 13,000 rpm for 15 min at 4°C, an equal volume of isopropanol was added to the supernatant. The pellet was obtained by isopropanol precipitation and washed with 70% ethanol. After drying the pellet, gDNA was eluted with distilled water. Using the eluted gDNA as a template, the presence of the *TALE-DME* transgene and the pBJ101 empty vector control was confirmed by PCR-amplification with primer pairs DG991-DG998 and DG2404-DG2363, respectively.

RNA isolation and gene expression analysis

Total RNA was isolated from the rosette leaves of six-week-old T1 transgenic plants using TRIzol (Ambion). The first-strand cDNA was synthesized using Oligo(dT) and SuperScript II reverse transcriptase (Invitrogen) with manufacturer's instruction. The expression level of target genes was estimated by PCR using gene specific primers: DG1231 and DG1334 for endogenous *FWA* expression, DG135 and DG026 for *TALE-DME* transgene, and DG244 and DG245 for *ACTIN11*.

Quantitative real-time PCR (qRT-PCR) analysis

To quantitate the expression level of the target genes, quantitative real-time PCR was performed using Rotor-Gene Q cycler (Qiagen) with SYBR Green Q master mix (Genet Bio). Using 60 ng of the synthesized cDNA as a template, the expression level of target genes were examined by qRT-PCR with corresponding primer pairs: DG2380 and DG2381 for *FWA* expression, DG772 and DG773 for *DME* expression, and DG1261 and DG1262 for *UBQ10* expression. Thermal cycling reaction was performed at 95°C for 10 min followed by 40-50 cycles of 95°C for 10 sec, 60°C for 15 sec, and 72°C for 35 sec. The expression level of *FWA* or *DME* relative to *UBQ10* was calculated by previously described (Schmittgen and Livak, 2008). Standard deviations were calculated from two independent technical repeats.

Locus specific bisulfite sequencing

The gDNA was isolated from the rosette leaves of six-week-old T1 transgenic plants using standard Cetyl-trimethylammonium bromide (CTAB) extraction method. Bisulfite conversion of gDNA was performed using EpiTect Bisulfite Kits (Qiagen) with manufacturer's instruction. After the bisulfite conversion reaction, both target and control regions were PCR-

amplified with specific primers. The *ATP Sulfurylase Arabidopsis 1 (ASA1)* gene, hypomethylated in the Arabidopsis genome was used as an unmethylated control to confirm the complete bisulfite conversion of gDNA. The *ASA1* region was PCR-amplified with primers DG662 and DG663, whereas the promoter region of *FWA* was PCR-amplified with primers DG344 and DG345. Each PCR product was cloned into the RBC T&A cloning kit (Real Biotech Corporation) and 10-15 clones were sequenced. The sequenced reads were aligned and the methylation patterns were analyzed by CyMATE (<http://www.cymate.org>).

Table 3-1. Oligonucleotides used in this study.

Name	Sequence (5'→3')
DG026	CAGTGTTTCGTTGATCGAGTTTGC
DG135	CCTGCAGGATCAAGTGGTAATG
DG244	AACTTCAACACTCCTGCCATG
DG245	CTGCAAGGTCCAAACGCAGA
DG344	GGTTYTATAYTAATATYAAAGAGTTATGGGYGAAG
DG345	CAAARTACTTTACACATAARCRAAAAACARACAAATC
DG662	TAAGAGAATTAATAATGYAAAATTTGYAAATA
DG663	CCTAACAAATRCAATCTRRTTTCRAAAAATRAC
DG772	TCGTCTCCTTGATGGTATGGA
DG773	GTGCCGAATTCGCTGTTT
DG991	AATACATGTAGGGATCCGAATTCAAGATCTACGC
DG998	GGGGGCACCCGTCAGTG
DG1231	CTAGGTGCAAAGAGATGGCTCG
DG1261	CGTTGACTGGGAAAACACTACT
DG1262	GTCCTGGATCTTGGCTTTCA
DG1334	GCAGTTGGATTGATGCCACC
DG2332	CTAGCACGTGTCTAGACTCGAGGGTTTAAACCGGACCGCCGGA TCCGTCGACCCTAGGCATTTAAATGAGCT
DG2333	CATTTAAATGCCTAGGGTCGACGGATCCGGCGGTCCGGTTTAA ACCTCGAGTCTAGACACGTG
DG2334	AATTTCTAGAATGGACTACAAAGACCATGACG
DG2335	AATTTCTGAGGTCATCGTCATCCTTGTAATCG
DG2338	CCACGTGTCTAGACTCGAGGGATCCGTCGACCCTAGGCATTTA AATGAGCT
DG2339	CATTTAAATGCCTAGGGTCGACGGATCCCTCGAGTCTAGACAC GTGGGTAC
DG2363	CCAGTCACGACGTTGTAAAACG
DG2380	TCCAAGTCTGCACTCATTC
DG2381	AACTGGAGCTGCTGATGGTT
DG2388	AATTTGTTTAAACATGGTGTACCCCTACGACGTG
DG2389	ATTCGGTCCGTTCACTTTTGACTAGCAACGCG
DG2404	CCCACTATCCTTCGCAAGAC
DG2876	ATTGATTTTTGTTGTTAAAAATAAAATYYATGTGAAGG
DG2878	CTTCRATAAARAATATATRARATTCTCRACRRAAA

Table 3-2. TALE-DME targeting sequences.

Name	Sequence (5'→3')
TD3	TTCATATATTCTTTATCGA
TD5	TAGTGTTTACTTGTTTAAGG

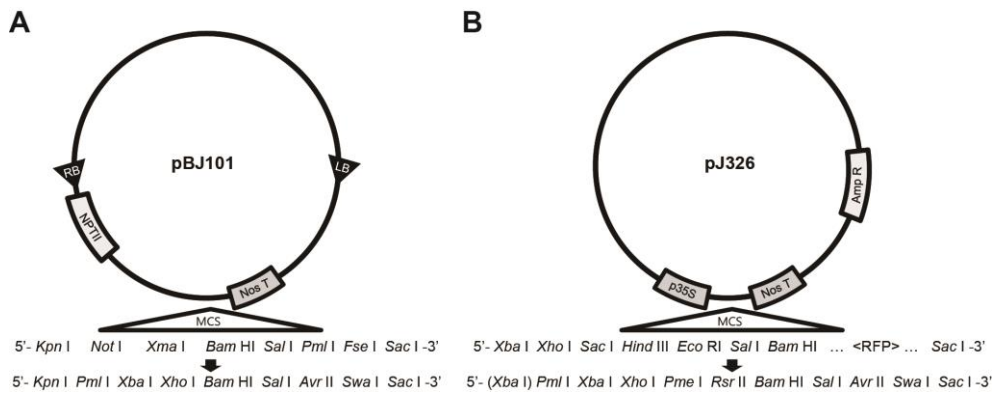


Figure 3-1. Diagrams of the manipulated vectors for cloning.

(A) The pBJ101 binary vector used for Arabidopsis transformation. The sequences resided between *Kpn* I and *Sac* I of the pBI101.1-p35S vector was replaced with newly synthesized MCS sequence including *Pml* I and *Avr* II. (B) The pJ326 vector used for TALE-DME cloning. The restriction enzymes sites and RFP sequence between *Xba* I and *Sac* I of the p326 vector were replaced with proper MSC depicted below the arrow.

RESULTS

Generation of transgenic plants expressing TALE-DME fusion proteins.

The DME family is a plant-specific DNA demethylase that directly recognizes and removes 5mC from DNA (Agius et al., 2006; Gehring et al., 2006; Penterman et al., 2007). Since DME binds to DNA in a non-sequence specific manner (Mok et al., 2010), a programmable DNA binding module is required for guiding DME to precise genomic loci. I utilized plant pathogen-derived TALE as a DNA binding module to combine with DME DNA demethylase. Two programmable TALEs targeting *FWA* promoter were designed (TD3 and TD5 in Figure 3-2A), and fused with DME to produce TALE-DME (Figure 3-2B). As *FWA* expression is normally repressed by DNA methylation in wild-type Col-0, introduction of TALE-DME into Col-0 plant is presumed to reactivate the expression of *FWA* by targeted DNA demethylation. Thus, several lines of T1 transgenic plants that express TALE-DMEs (TD3 and TD5) and two lines of empty vector control were produced and analyzed.

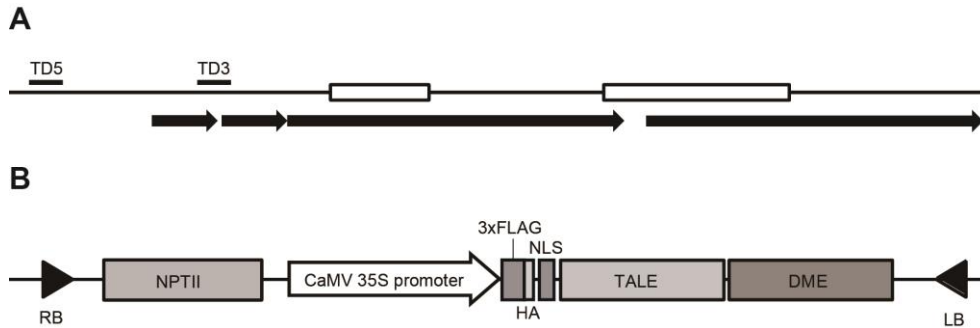


Figure 3-2. Schematic diagrams of TALE binding sites and TALE-DME constructs.

(A) A diagram of Arabidopsis *FWA* promoter with TALE target sites. Open boxes represent the first and second exons of *FWA*, and black arrows indicate the tandem repeat regions. TALE-DME binding sites (TD3 and TD5) are shown above the *FWA* promoter with black bars. (B) Constructs of pBJ101-TALE-DME used for Arabidopsis transformation. The *TALE-DME* transgene fused with N-terminal 3xFLAG and HA tags was driven under CaMV 35S promoter. The *NPTII* gene was used as a selection marker. The right border (RB) and left border (LB) sequences are depicted with triangles.

DNA demethylation activity of TALE-DME in transgenic plants

Two independent lines of each transgenic plant expressing TD3, TD5 and vector control were selected and DNA methylation levels of those plants were analyzed by bisulfite sequencing. Complete bisulfite conversion of cytosine to thymine bases was confirmed by amplification of *ASAI* region (Figure 3-3) which is reported to be unmethylated in *Arabidopsis* (Jeddeloh et al., 1998). Local bisulfite sequencing analysis showed that most of transgenic lines including TD3-1, TD3-2 and TD5-3 displayed significant decrease of DNA methylation level of *FWA* locus compared to vector control (Figures 3-4A and 3-4C), whereas TD5-2 showed similar DNA methylation pattern with vector (Figure 3-4B). TD5-3 and two TD3 transgenic lines showed a broad range of DNA demethylation (up to 400 bp from the TALE binding site) with 10-40% efficiency. These results imply that TALE-DME fusion protein successfully induces DNA demethylation at the specific target loci in plants.

To find a causal link between targeted DNA demethylation and transcription of target gene, RT-PCR analysis was performed. Although *TALE-DME* transgene was expressed in all transgenic plants, transcription of *FWA* was only detected in TD5-2 and TD5-3 lines (Figure 3-5A). Further quantitation of *FWA* and *DME* mRNA level revealed that significant up-

regulation of *FWA* and *DME* expression compared to vector control was detected in the TD5-3 line (Figure 3-5B). This result was also supported by the late flowering phenotype of TD5-3 transgenic line (Figure 3-6). These data suggest that TALE-DME appears to reactivate *FWA* expression through targeted DNA demethylation, which leads to late flowering phenotype.

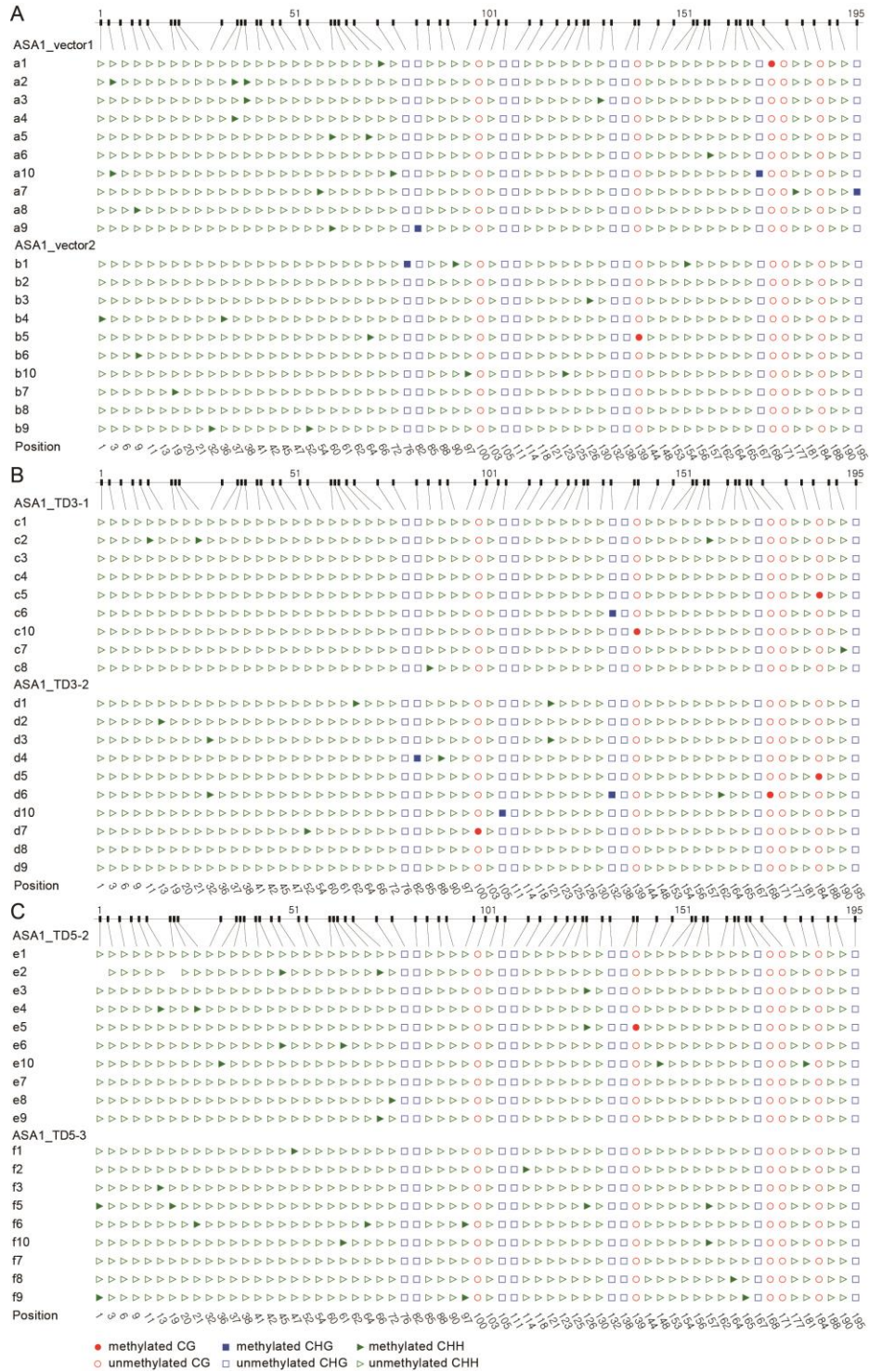


Figure 3-3. Bisulfite conversion of genomic DNA extracted from each T1 transgenic plant.

DNA methylation profiles of *ASA1* locus in T1 transgenic plants that express empty vector (A), TD3 (B) and TD5 (C). The sites of CG, CHG and CHH methylation are colored with red, blue and green, respectively.

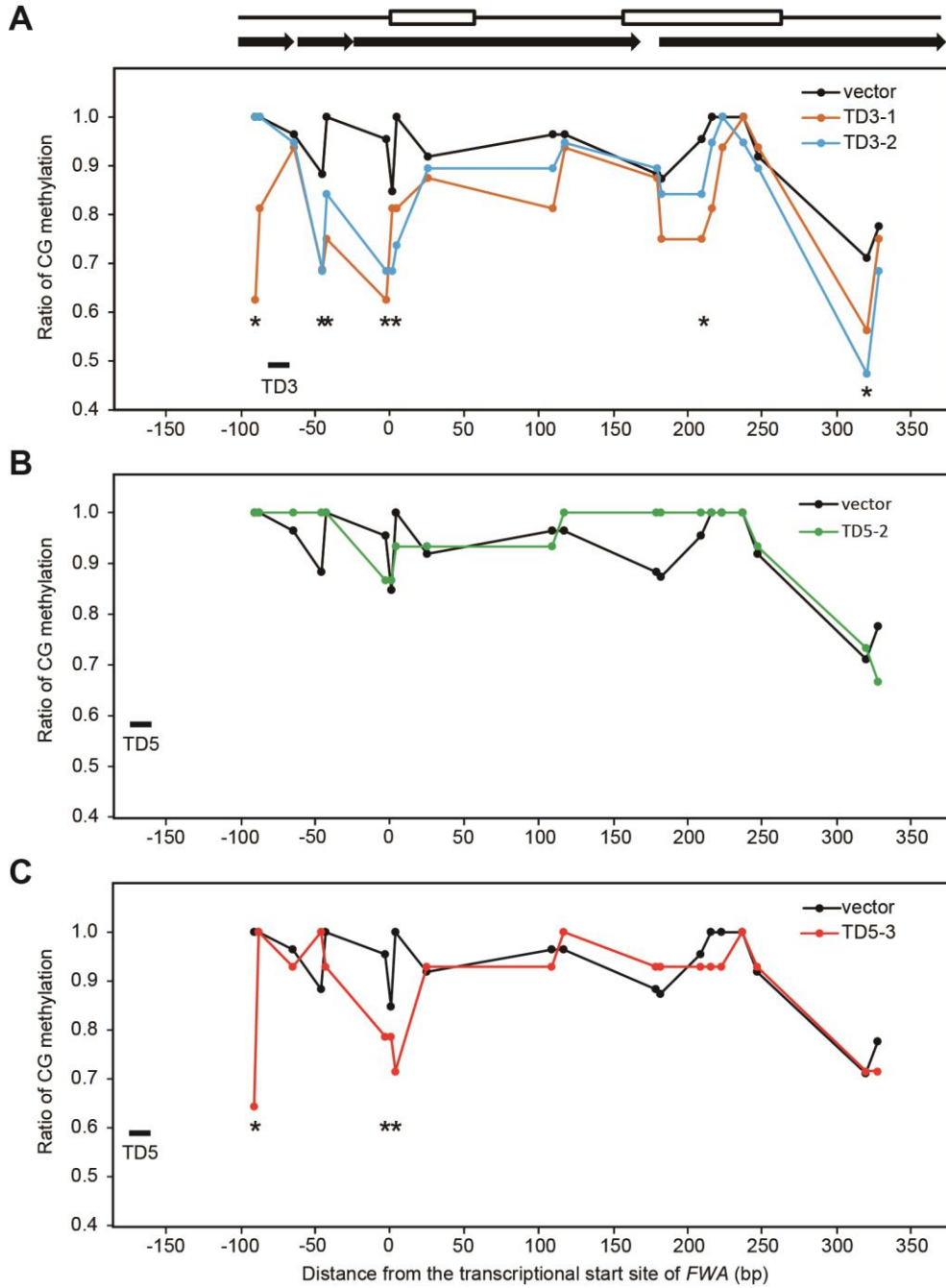


Figure 3-4. DNA demethylation activity of TD3 and TD5 TALE-DME fusion proteins in T1 transgenic plants.

Ratio of CG methylation levels in TD3 (A), TD5-2 (B) and TD5-3 (C) TALE-DME T1 transgenic plants. TALE-DME binding sites are depicted with black bars. Genomic DNA structure including exons (open boxes) and tandem repeat regions (black arrows) of the *FWA* locus is represented above the graph. The different CG positions along the length of the *FWA* promoter are numbered relative to transcription start site (x-axis). Asterisks indicate differentially methylated cytosines in TALE-DME transgenic lines compared to vector control lines. Note that the mean values of DNA methylation levels of two independent vector control plants were plotted with black lines (Vector).

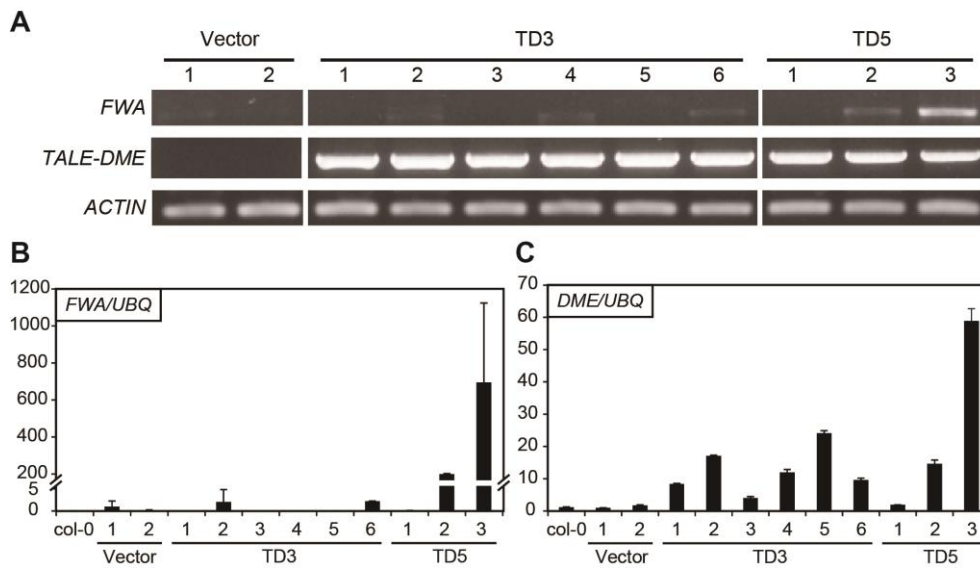


Figure 3-5. Expression level of *FWA* in TALE-DME T1 transgenic plants.

(A) Expression of *FWA* in T1 transgenic plants expressing *TALE-DME*. RT-PCR was performed with cDNA prepared from 6 lines of TD3 and 3 lines of TD5 transgenic plants. Two independent control lines harboring pBJ101 empty vector were also analyzed. *ACTIN* was used as a control. (B, C) Relative expression levels of *FWA* (B) and *DME* (C) compared to *UBQ10* mRNA levels in TALE-DME T1 transgenic plants were analyzed by qRT-PCR. Expression levels of each gene was normalized to Vector-1 control and plotted as bar graphs. Error bars represent standard deviations from two technical repeats.

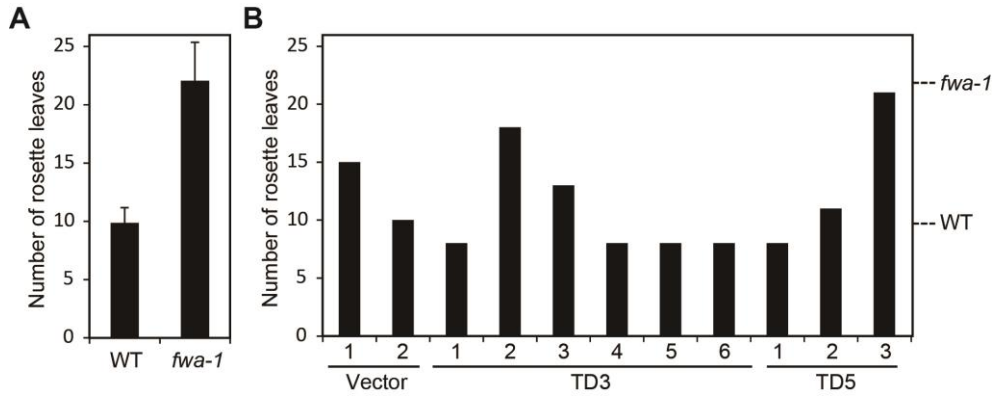


Figure 3-6. Flowering time of TALE-DME T1 transgenic plants.

(A) The number of rosette leaves was counted and plotted as a bar graph when the visible bolts of WT (Col-0) and *fwa-1* were appeared. Error bars represent standard deviations from biological repeats (n = 15, 16) (B) Flowering time of TD3 and TD5 TALE-DME T1 transgenic plants was plotted as a bar graph. Two independent transgenic lines expressing empty vector were analyzed to estimate flowering time as controls. Average values of flowering time calculated from (A) are indicated to the right of the panel.

DISCUSSION

The classical ZFP has been widely used as a DNA binding module over the past decades, whereas methylated DNA attenuated the DNA binding affinity of ZFP, which limited its epigenetic application (Lei et al., 2018). Later, TALE derived from plant pathogenic bacteria was reported as a more improved DNA binding module, which specifically recognizes single nucleotide by its repeat domains (Gaj et al., 2013). By virtue of great design flexibility of TALE, several previous studies have reported TALE-directed epigenome editing in mammals (Gaj et al., 2013; Thakore et al., 2016; Lei et al., 2018). A chimeric TALE module involving DNA methyltransferases DNMT3A and DNMT3L induced hypermethylation of the *CDKN2A* locus in human fibroblasts (Bernstein et al., 2015). In addition, TALE was also fused with TET1 mammalian DNA demethylase to transcriptionally activate target genes by site-specific DNA demethylation in HEK293 and HeLa cells (Maeder et al., 2013).

In this study, I propose a TALE-based targeted DNA demethylation system for plants using plant-specific DNA demethylase. I showed that TALE-DME introduced into *Arabidopsis* displayed significant DNA

demethylation activity on *FWA* locus in T1 transgenic plants (Figures 3-4A and 3-4C). The differentially methylated cytosines compared to vector controls (denoted with asterisks in Figure 3-4) in transgenic plants were clustered around the short tandem repeat region of *FWA* known to be highly methylated in wild-type (Weigel and Colot, 2012). It is reasonable that this region is located near the TALE-DME binding sites, and may be important for regulating transcription of *FWA*. Furthermore, TD5-3 transgenic plant showed a causal relationship between *FWA* expression and DNA demethylation along with late flowering phenotype (Figures 3-5 and 3-6), which indicates that delicate modulation of DNA methylation in the specific genomic loci is required for target gene expression. On the contrary, DNA demethylation in two TD3 lines did not lead to *FWA* expression (Figures 3-4A and 3-5), suggesting that an additional mechanism may exist for *FWA* activation other than DNA demethylation of the *FWA* promoter.

Due to a slow turnover rate of DME catalysis, DME tightly binds to an abasic site after 5mC excision (Ponferrada-Marin et al., 2009). This feature of DME is distinguished from other chromatin modifiers, which can prevent efficient DNA demethylation at the target loci during epigenome editing. To improve the efficiency of DNA demethylation and broaden the range of the target loci, the CRISPR-SunTag system recruiting multiple

copies of chromatin modifier is required for efficient targeted DNA demethylation. After engineering catalytically inactive dCas9, numerous researchers have focused on CRISPR/dCas9-mediated epigenome editing. Using CRISPR or CRISPR-SunTag system, dCas9 protein was fused with DNMT3A or TET1 to manipulate site-specific DNA methylation, leading to successful regulation of target genes (Liu et al., 2016; Morita et al., 2016; Huang et al., 2017). Recent studies also reported that epigenetic features established by epigenome editing persisted in T2 generation even if the transgene was already removed from plants (Gallego-Bartolomé et al., 2018; Papikian et al., 2019). These results suggest that epigenome editing alters heritable epigenetic marks without sequence change, which can be applied to create genetically modified organisms (GMO)-free crops. Taken together, epigenome editing including this research will provide a powerful tool for regulating dynamics of epigenetic modifications, which leads to a promising avenue to produce various epigenetic traits regulated by DNA methylation.

REFERENCES

- Agius, F., Kapoor, A., and Zhu, J.K.** (2006). Role of the Arabidopsis DNA glycosylase/lyase ROS1 in active DNA demethylation. *Proc. Natl. Acad. Sci. USA* **103**: 11796-11801.
- Bernstein, D.L., Le Lay, J.E., Ruano, E.G., and Kaestner, K.H.** (2015). TALE-mediated epigenetic suppression of CDKN2A increases replication in human fibroblasts. *J. Clin. Invest.* **125**: 1998-2006.
- Cheng, A.W., Wang, H., Yang, H., Shi, L., Katz, Y., Theunissen, T.W., Rangarajan, S., Shivalila, C.S., Dadon, D.B., and Jaenisch, R.** (2013). Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res.* **23**: 1163-1171.
- Choi, K., Kim, S., Kim, S.Y., Kim, M., Hyun, Y., Lee, H., Choe, S., Kim, S.G., Michaels, S., and Lee, I.** (2005). SUPPRESSOR OF FRIGIDA3 encodes a nuclear ACTIN-RELATED PROTEIN6 required for floral repression in Arabidopsis. *Plant Cell* **17**: 2647-2660.
- Clough, S.J., and Bent, A.F.** (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**: 735-743.
- Gaj, T., Gersbach, C.A., and Barbas, C.F., 3rd.** (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* **31**: 397-405.
- Gallego-Bartolome, J., Gardiner, J., Liu, W., Papikian, A., Ghoshal, B., Kuo, H.Y., Zhao, J.M., Segal, D.J., and Jacobsen, S.E.** (2018).

Targeted DNA demethylation of the Arabidopsis genome using the human TET1 catalytic domain. Proc. Natl. Acad. Sci. USA **115**: E2125-E2134.

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

Huang, Y.H., Su, J., Lei, Y., Brunetti, L., Gundry, M.C., Zhang, X., Jeong, M., Li, W., and Goodell, M.A. (2017). DNA epigenome editing using CRISPR-Cas SunTag-directed DNMT3A. Genome Biol. **18**: 176.

Huh, J.H., Bauer, M.J., Hsieh, T.F., and Fischer, R.L. (2008). Cellular programming of plant gene imprinting. Cell **132**: 735-744.

Jang, H., Shin, H., Eichman, B.F., and Huh, J.H. (2014). Excision of 5-hydroxymethylcytosine by DEMETER family DNA glycosylases. Biochem. Biophys. Res. Commun.

Jeddeloh, J.A., Bender, J., and Richards, E.J. (1998). The DNA methylation locus DDM1 is required for maintenance of gene silencing in Arabidopsis. Genes Dev. **12**: 1714-1725.

Johnson, L.M., Du, J., Hale, C.J., Bischof, S., Feng, S., Chodavarapu, R.K., Zhong, X., Marson, G., Pellegrini, M., Segal, D.J., Patel, D.J., and Jacobsen, S.E. (2014). SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. Nature **507**: 124-128.

Kinoshita, Y., Saze, H., Kinoshita, T., Miura, A., Soppe, W.J., Koornneef, M., and Kakutani, T. (2007). Control of FWA gene silencing in *Arabidopsis thaliana* by SINE-related direct repeats. Plant J. **49**: 38-

45.

- Kohli, R.M., and Zhang, Y.** (2013). TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* **502**: 472-479.
- Koornneef, M., Hanhart, C.J., and van der Veen, J.H.** (1991). A genetic and physiological analysis of late flowering mutants in *Arabidopsis thaliana*. *Mol. Gen. Genet.* **229**: 57-66.
- Kungulovski, G., and Jeltsch, A.** (2016). Epigenome Editing: State of the Art, Concepts, and Perspectives. *Trends Genet.* **32**: 101-113.
- Law, J.A., and Jacobsen, S.E.** (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **11**: 204-220.
- Lei, Y., Huang, Y.H., and Goodell, M.A.** (2018). DNA methylation and demethylation using hybrid site-targeting proteins. *Genome Biol.* **19**: 187.
- Liu, X.S., Wu, H., Ji, X., Stelzer, Y., Wu, X., Czauderna, S., Shu, J., Dadon, D., Young, R.A., and Jaenisch, R.** (2016). Editing DNA methylation in the mammalian genome. *Cell* **167**: 233-247 e217.
- Lyko, F.** (2018). The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nat. Rev. Genet.* **19**: 81-92.
- Maeder, M.L., Angstman, J.F., Richardson, M.E., Linder, S.J., Cascio, V.M., Tsai, S.Q., Ho, Q.H., Sander, J.D., Reyon, D., Bernstein, B.E., Costello, J.F., Wilkinson, M.F., and Joung, J.K.** (2013). Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. *Nat. Biotechnol.* **31**: 1137-1142.
- Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L., and Church, G.M.** (2013). CAS9 transcriptional

- activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* **31**: 833-838.
- Mok, Y.G., Uzawa, R., Lee, J., Weiner, G.M., Eichman, B.F., Fischer, R.L., and Huh, J.H.** (2010). Domain structure of the DEMETER 5-methylcytosine DNA glycosylase. *Proc. Natl. Acad. Sci. USA* **107**: 19225-19230.
- Morita, S., Noguchi, H., Horii, T., Nakabayashi, K., Kimura, M., Okamura, K., Sakai, A., Nakashima, H., Hata, K., Nakashima, K., and Hatada, I.** (2016). Targeted DNA demethylation in vivo using dCas9-peptide repeat and scFv-TET1 catalytic domain fusions. *Nat. Biotechnol.* **34**: 1060-1065.
- Papikian, A., Liu, W., Gallego-Bartolome, J., and Jacobsen, S.E.** (2019). Site-specific manipulation of Arabidopsis loci using CRISPR-Cas9 SunTag systems. *Nat. Commun.* **10**: 729.
- Penterman, J., Zilberman, D., Huh, J.H., Ballinger, T., Henikoff, S., and Fischer, R.L.** (2007). DNA demethylation in the Arabidopsis genome. *Proc. Natl. Acad. Sci. USA* **104**: 6752-6757.
- Ponferrada-Marin, M.I., Roldan-Arjona, T., and Ariza, R.R.** (2009). ROS1 5-methylcytosine DNA glycosylase is a slow-turnover catalyst that initiates DNA demethylation in a distributive fashion. *Nucleic Acids Res.* **37**: 4264-4274.
- Sander, J.D., and Joung, J.K.** (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* **32**: 347-355.
- Schmittgen, T.D., and Livak, K.J.** (2008). Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protoc.* **3**: 1101-1108.
- Smith, Z.D., and Meissner, A.** (2013). DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14**: 204-220.

- Thakore, P.I., Black, J.B., Hilton, I.B., and Gersbach, C.A.** (2016). Editing the epigenome: technologies for programmable transcription and epigenetic modulation. *Nat. Methods* **13**: 127-137.
- Weigel, D., and Colot, V.** (2012). Epialleles in plant evolution. *Genome Biol.* **13**: 249.
- Wu, S.C., and Zhang, Y.** (2010). Active DNA demethylation: many roads lead to Rome. *Nat. Rev. Mol. Cell Biol.* **11**: 607-620.
- Wu, X., and Zhang, Y.** (2017). TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat. Rev. Genet.* **18**: 517-534.

ABSTRACT IN KOREAN

DNA 메틸화는 중요한 후성유전학적 표지로, 고등생물에서 염색질의 구조와 유전자 발현을 조절한다. 일반적으로 DNA 메틸화는 세포 분열 이후에도 유지가 되지만 필요한 경우 특정 효소에 의해 제거될 수 있고, 이를 'DNA 탈메틸화'라고 한다. 애기장대의 DEMETER (DME) DNA 글리코실라제는 DNA에 존재하는 5-메틸시토신(5mC) 염기를 특이적으로 인지하여 제거하고 생물체에 해로운 방해생성물을 만들어내는데, 이는 AP 엔도뉴클리아제 효소에 의해 안전하게 제거될 수 있다. 이후, DNA 중합효소가 메틸기가 없는 시토신을 DNA에 넣어주게 되면 비로소 DNA 탈메틸화가 완성된다. DME 유전자군에 속하는 4개의 유전자인 *DME*, *REPRESSOR OF SILENCING 1 (ROS1)*, *DEMETER-LIKE 2 (DML2)* 그리고 *DML3*는 모두 비슷한 도메인 구조를 가졌지만, 서로 다른 5mC 제거 효율을 보인다. DME 유전자군에 속하는 단백질들은 크기가 크고 진화적으로 잘 보존된 여러 개의 도메인으로 이루어져 있으며, 이 도메인들은 변이가 많은 시퀀스를 통해 연결되어 있다. 하지만 이러한 구조적 특징이 DME의 효소 활성 조절에 미치는 영향이나, DME의 5mC 제거 반응 이후에 일어나는 DNA 탈메틸화 하위 경로에 대한 연구는 많이 미흡한 실정이다. 본 연구에서는 DME 단백질의 크기를 극단적

으로 줄여서 5mC 제거 활성화에 필요한 최소한의 도메인을 탐색하고자 했고, DNA 탈메틸 효소의 모듈형 구성 개발에 성공하였다. 이 후, 도메인 치환 실험을 통해, DME 유전자군의 글리코실라제 도메인들이 모두 비슷한 잠재적 5mC 제거 능력을 가지고 있고, 도메인들 사이의 호환성이 DME의 효소 활성화에 매우 중요하게 작용한다는 것을 밝혔다. 또한, 염기 절단 수리를 담당하는 애기장대 AP 엔도뉴클리아제에 속하는 ABASIC ENDONUCLEASE 1-LIKE (APE1L)과 ABASIC ENDONUCLEASE-REDOX PROTEIN (ARP)가 DME의 효소 활성화 반응 이후 생성되는 비정상적인 방해생성물을 안전하게 제거할 수 있다는 것을 증명했고, 이를 통해 식물 내 DNA 탈메틸화 경로에 대한 더욱 완성된 모델을 제시할 수 있었다. 마지막으로 DME를 후성유전체 편집 기술에 응용하기 위해 DNA 부착 모듈인 transcription activator-like effector (TALE)와 융합하는 연구를 진행하였는데, 이 과정에서 만들어진 TALE-DME 융합 단백질은 원하는 유전체 부위에서 정교하게 DNA 메틸화를 조절하는 것을 확인할 수 있었다. 종합하면, 본 연구는 DNA 탈메틸 효소의 모듈형 구성을 개발하여 도메인 기능 연구와 DNA 탈메틸화 경로 연구를 위한 기초지식을 제공하였고, 이를 후성유전체 편집 기술에 응용하여 농업적으로 유용한 다양한 후성유전학적 형질 창출에 기여할 수 있을 것으로 기대된다.