A Thesis for the Degree of Master of Science

Excision of 5-hydroxymethylcytosine by DEMETER DNA glycosylase from *Arabidopsis*

애기장대 DEMETER DNA 글리코실라제에 의한

하이드록시메틸시토신 제거

FEBRUARY, 2013

HOSUNG JANG

MAJOR IN PLANT DEVELOPMENTAL GENETICS

DEPARTMENT OF PLANT SCIENCE

THE GRADUATE SCHOOL OF

SEOUL NATIONAL UNIVERSITY

Excision of 5-hydroxymethylcytosine by DEMETER DNA glycosylase from *Arabidopsis*

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

BY
HOSUNG JANG

MAJOR IN PLANT DEVELOPMENTAL GENETICS
DEPARTMENT OF PLANT SCIENCE
GRADUATE SCHOOL SEOUL NATIONAL UNIVERSITY

FEBRUARY, 2013

APPROVED AS A QUALIFIED THESIS OF HOSUNG JANG
FOR THE DEGREE OF MASTER OF SCIENCE
BY THE COMMITTE MEMBERS

CHAIRMAN _____

Byoung-Cheorl Kang, Ph.D.

VICE-CHAIRMAN _____

Jin Hoe Huh, Ph.D.

MEMBER _____

Doil Choi, Ph.D.

Excision of 5-hydroxymethylcytosine by DEMETER DNA glycosylase from *Arabidopsis*

HOSUNG JANG

DEPARTMENT OF PLANT SCIENCE

THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

ABSTRACT

In plants and animals, 5-methylcytosine (5mC) serves as an epigenetic mark to repress gene expression and plays important roles in gene imprinting and transposon silencing. The mammalian genome also contains 5-hydroxymethylcytosine (5hmC), formed by TET family-mediated oxidation of 5mC, which is referred to as the 6th base. 5hmC is abundant in mouse Purkinje neurons and embryonic stem cells, and regarded as an important intermediate during active DNA demethylation in mammals. However, it is not clearly demonstrated that 5hmC is present in plants. In *Arabidopsis*, DEMETER (DME) DNA glycosylase efficiently removes 5mC, which results in DNA demethylation and transcriptional activation of target genes. Here I showed that 5hmC was also excised by DME in vitro, raising the possibility that 5hmC may exist in plants and serve as an important intermediate for DNA demethylation. I performed a two dimensional thin layer chromatography (2D-TLC) analysis to verify whether 5hmC

is present in *Arabidopsis*. I detected no or very little, if any, 5hmC from the *Arabidopsis* genome, suggesting that it is very unlikely for plants to have 5hmC. In addition, I searched for DME mutants with altered substrate specificity. A DME mutant protein, Q1362R, was capable of excising mismatched thymine more readily than 5mC. Our findings suggest that the region upstream of the Fe-S cluster is crucial for shaping the recognition pocket of DME, providing the substrate specificity for this family of enzymes.

Key Words**:** DEMETER, DNA demethylation, glycosylase, 5mC, 5hmC recognition pocket

Student number: 2011-21210

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

5mC	5-methylcytosine
DME	DEMETER
BER	Base excision repair
ROS1	REPRESSOR OF SILENCING 1
DML2	DEMETER-LIKE 2
DML3	DEMETER-LIKE 3
MEA	MEDEA
5hmC	5-hydroxymethylcytosine
TET	Ten eleven translocate
5fC	5-formylcytosine
5caC	5-carboxylcytosine
DNMT	DNA methyltransferase
2D-TLC	Two dimensional thin layer chromatography
SRA	SET- or RING- associated
UHRF1	Ubiquitin-like plant homeodomain and RING finger domain 1
MET1	METHYLTRANSFERASE1
DRM	DOMAINS REARRANGED METHYLTRANSFERASE
CMT3	CHROMOMETHYLASE 3
VIM	VARIANT IN METHYLATIOIN
RdDM	RNA-directed DNA methylation

AGO	Argonaute
RNAi	RNA interference
ssRNA	single-stranded RNA
dsRNA	double-stranded RNA
RDR2	RNA-DEPENDENT RNA POLYMERASE 2
nt	nucleotide
siRNA	small interfering RNAs
DCL3	DICER-LIKE 3
IGN	Intergenic non-coding
SUVH4	SRA domain of the SUPPRESSOR OF VARIEGATION 3-9 HOMOLOGUE 4
AP	Apyrimidinic/apurinic
HhH	Helix-hairpin-helix
GPD	Glycine/proline-rich loop with a conserved aspartate
EndoIII	Endonuclease III
hOGG1	human 8-oxoguanine DNA glycosylase
FWA	FLOWERING WAGENINGEN
TDG	Thymine DNA glycosylase
AID/APOBEC	Activation-induced cytidine deaminase/apolipoprotein B mRNA-editing catalytic polypeptides
SMUG1	Single-strand-selective monofunctional uracil-DNA glycosylase 1
TBE	Tris/ Boric acid/ EDTA

INTRODUCTION

DNA methylation is a simple but important epigenetic modification that generally represses gene expression both in plants and animals (Law and Jacobsen, 2010; Huh et al., 2008). Cytosine methylation at C5 position is mediated by DNA methyltransferases, producing 5-methylcytosine (5mC) (Law and Jacobsen, 2010). In plants, CG, CHH, and CHG sites can be methylated by different types of methyltransferases (Chan et al., 2005).

Removal of DNA methylation, called DNA demethylation, is conducted by DNA demethylases unique to plants (Wu and Zhang, 2010). DEMETER (DME) family proteins, specifically remove 5mC, and in subsequent reactions, base excision repair (BER) pathway enzymes are required to complete active DNA demethylation (Law and Jacobsen, 2010). With REPRESSOR OF SILENCING 1 (ROS1), DME-LIKE 2 (DML2), and DML3, DME family proteins have three conserved domains (domain A, glycosylase domain, domain B) for 5mC excision activity (Choi et al., 2002; Gong et al., 2002; Penterman et al., 2007). DME is a bifunctional DNA glycosylase/AP lyase that hydrolyzes the N-glycosidic bond between a base and ribose. DME directly excises 5mC from double-stranded DNA to produce a nick and β-elimination product (Gehring et al., 2006). DME randomly removes 5mC *in vitro* and its catalytic mutant (D1304N or K1286Q do not excise 5mC (Gehring et al., 2006).

DME is most studied DNA demethylase. DME is essential for endosperm development and gene imprinting, expressed in the central cell of female gametophyte (Choi et al., 2002). DME is known to induce maternal allele-specific

activation of *MEDEA (MEA)* by removing 5mC, which leads to endosperm development (Gehring et al., 2006). In heterozygous dme mutant, 50% of *MEA* allele was not expressed in central cell, which resulted in 50% of defects in endosperm development and seed formation. This 50% of seed abortion phenotype was partially restored to 25% in transgenic plants that harbored functional DME (Choi et al., 2004).

In mammals, however, 5mC-specific glycosylase is not identified yet, which suggests alternative mechanisms are required for DNA demethylation. It is recently reported that 5-hydroxymethylcytosine (5hmC), a modified base of 5mC detected in the mouse brain, is oxidized from 5mC by Ten eleven translocate (TET) family proteins (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009).

The role of 5hmC is still elusive but recent studies reported some biological significance related with 5hmC. Impaired 5hmC pattern was detected in myeloid cancers in TET2 mutants, and 5hmC seems to play an important role in mouse embryonic stem cell self-renewal process (Ko et al., 2010; Ito et al., 2010). Misregulation of 5-hydroxymethylation also affected cell lineage specification in mouse embryonic stem cells (Koh et al., 2011).

TET proteins can further convert 5hmC to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Ito et al., 2011; He et al., 2011). And 5caC was found to be removed by TDG glycosylase (He et al., 2011), which suggests that 5hmC may serve as an important intermediate for DNA demethylation. A recent study shows that DNA methyltransferase 3 (DNMT3) family proteins, *de novo* DNA methyltransferases, are involved in active decarboxylation of 5caC, which raises

the possibility of 5hmC involvement in DNA demethylation independent of BER (Chen et al., 2012). Previous studies also revealed that not every 5mC-recognizing proteins can bind to 5hmC (Valinluck et al., 2004) demonstrating a functional distinction from 5mC. 5hmC is present in various organisms including mouse, rabbit, bovine, and amphibians (Kriaucionis and Heintz, 2009; Wossidlo et al., 2011; Almeida et al., 2012).

To date, however, the presence of 5hmC in plants is under question, and thus it would be of great interest to verify whether 5hmC is present in the plant genome. In this study, I tested whether 5hmC exists in plants by HPLC and two dimensional thin layer chromatography (2D-TLC) analysis. In addition, I showed that DME 5mC glycosylase was able to recognize and excise another modified cytosine bases including 5hmC and that several amino acids comprising the putative binding pocket is important for substrate specificity of DME.

LITERATURE REVIEWS

1. From genetics to epigenetics

A model of DNA double helix was designed by Watson and Crick in 1953 (Watson and Crick, 1953). DNA, a basic genetic material, consists of four types of bases – adenine , guanine, thymine and cytosine. Based on these four bases, the world of molecular genetics had opened and has been prosporous for over fifty years. And a central dogma, with a concept of "DNA makes RNA makes protein", accelerated the development of molecular biology (Crick, 1970). But genetics cannot explain every phenomenon of organisms. For example, the identical twins show slightly different phenotypes each other although they have exactly the same genomes, which implies that there are other mechanisms beyond genetics (Fraga et al., 2005). And this paradox could be explained in part by the discovery of the fifth base 5-methylcytosine (5mC) that possesses a methyl group on C5 position of cytosine. The 5mC generating process is also called DNA methylation that is present both in plants and animals (Law and Jacobsen, 2010). DNA methylation is a simple modification but regulates gene expression by altering chromatin structure without change of DNA sequences (Cedar and Bergman, 2012). This phenomenon is defined as a part of 'epigenetics', the study of heritable changes in phenotype that does not involve DNA mutation (Allis et al., 2007). Epigenetics mechanisms consist of three major factors – DNA methylation, histone modification and small RNAs (Conaway, 2012). The three factors are correlated and interact each other to regulate gene expression. Now epigenetics is regarded as a powerful concept to explain many biological events including X-chromosome inactivation, transposon silencing and cell specification (Bonasio et al., 2010).

2. DNA methylation

As previously mentioned, DNA methylation is essential for diverse biological processes in both plants and animals. In mammals, it occurs at cytosine in CpG sites and plays an important role in genome stability, cell differentiation, imprinting, and X-inactivation (Chan et al., 2005, Huh et al., 2008, Bonasio et al., 2010).

DNA methylation, achieved by DNA methyltransferases, is divided into two types according to the manner of methylation - *de novo* methylation and maintenance methylation (Law and Jacobsen, 2010). *De novo* methylation, established by DNMT3 family, generates a new methylation pattern in early embryo development (Cheng and Blumenthal, 2008). And these global patterns are maintained by maintenance DNA methyltransferase DNMT1 (Chen and Riggs, 2011). When cell division occurs, the doubled genome produces DNA strand including unmethylated cytosines, which need to be methylated by DNMT1 to maintain previous DNA methylation pattern (Law and Jacobsen, 2010). After DNA replication, DNMT1, recruited by SET- or RING- associated (SRA) domain of ubiquitin-like plant homeodomain and RING finger domain 1 (UHRF1) protein that recognizes 5mC of hemimethylated DNA, methylates the cytosine on the other strand (Arita et al., 2008).

A DNMT1 loss-of-function mutant mouse shows genome-wide loss of DNA methylation, embryonic lethality at embryonic day 9.5, abnormal expression of imprinted genes, ectopic X-chromosome inactivation and activation of silent retrotransposon (Allis et al., 2007). A DNMT3 homozygous mutant shows

postnatal lethality at 4-8 weeks, male sterility and failure to establish methylation imprints in both male and female germ cells, which implies that DNMT family proteins are crucial for various biological phenomena regulated by DNA methylation (Allis et al., 2007).

Other than this, actually more people become interested in DNA methylation due to its relation with cancer. In many cancer cells, the DNA methylation level of tumor suppressor genes increases, which means the genes were suppressed and could not suppress tumor growth (Rodriguez-Paredes and Esteller, 2011).

In plants, on the other hand, DNA methylation can occur at every contexts of cytosine: CG, CHG and CHH, where H represents any nucleotide except guanine (Chan et al., 2005). It is quite distinct from the mammalian DNA methylation system, and requires more types of DNA methyltransferases than mammals. In *Arabidopsis thaliana,* three types of DNA methyltransferases add methyl group on cytosine that include METHYLTRANSFERASE1 (MET1), DOMAINS REARRANGED METHYLTRASNSFERASE (DRM) family proteins, and CHROMOMETHYLASE 3 (CMT3) (Chan et al., 2005). MET1 and DRM family proteins, showing a significant homology to mammalian DNMT family proteins, maintain and establish *de novo* DNA methylation, respectively, whereas CMT3, unique to plants, maintains CHG methylation and does not seem to exist in mammals. (Law and Jacobsen, 2010)

The maintenance DNA methyltransferase MET1, a homolog of mammalian DNMT1, is in charge of methylating CG in Arabidopsis. MET1 and

DNMT1 appear to maintain CG methylation in a similar manner because plants also have a VARIANT IN METHYLATIOIN (VIM) family of SRA domain proteins, counterparts of UHRF1 (Woo et al., 2008; Law and Jacobsen, 2010). The MET1 homozygous mutant exibited a delayed flowering phenotype producing more rosette and cauline leaves than wild type *Arabidopsis* (Kankel et al., 2003). Also, the methylation level of methylated loci decreased and became transcriptionally activated during gametogenesis in the absence of MET1 (Saze et al., 2003).

DRM family proteins, homologs of mammalian DNMT3, catalyze *de novo* DNA methylation at asymmetric CHH sites (Chan et al., 2005). *De novo* DNA methylation in plants is mediated by small RNAs via RNA-directed DNA methylation (RdDM) (Wassenegger, 2000). In addition to Dicer and Argonaute (AGO) family proteins composing RNA interference (RNAi) machinery, RdDM requires Pol IV and Pol V, two plant-specific RNA polymerases (Lahmy et al., 2010; Law and Jacobsen, 2010). The model of RdDM suggests that single-stranded RNA (ssRNA) transcripts corresponding to transposons and repeat elements, generated by Pol IV, are used to synthesize double-stranded RNA (dsRNA) by RNA-DEPENDENT RNA POLYMERASE 2 (RDR2). This dsRNA is digested into 24-nucleotide (nt) -long small interfering RNAs (siRNAs) by DICER-LIKE 3 (DCL3), which are then bound to AGO4. The siRNA-loaded AGO4 interacts with subunits of Pol V that is thought to transcribe intergenic non-coding (IGN) regions in the genome. The nascent IGN transcript base pairs with siRNA in AGO4, which recruits several chromatin remodeling factors including de novo methyltransferase DRM2. Finally, DRM2 methylates targeted loci in a manner that may be aided by

those chromatin modifying factors (Law and Jacobsen, 2010).

These methylation patterns can be maintained by another maintance methyltransferase CMT3 that methylates on CHG sites. CHG methylation is also related with histone methylation. The chromodomain in CMT3 recognizes methylated H3 tails and the methyltrasferase domain of CMT3 methylates neighboring CHG sites (Lindroth et al., 2004). Methylated DNA is recognized by the SRA domain of the SUPPRESSOR OF VARIEGATION 3-9 HOMOLOGUE 4 (SUVH4) that catalyzes histone 3 lysine 9 dimethylation (H3K9me2), which suggests that CMT3 and SUVH4 consistently maintain methylation level (Johnson et al., 2007).

In plants, transposons and other repetitive DNA elements are methylated to be suppressed due to a risk to genome stability (Lisch, 2009). When transposons are activated, they jump in the genome, which may result in unexpected mutations of functional genes or cause genome instability (Le et al., 2000). During male gametogenesis in angiosperm, microspore develops into tricellular pollen containing two sperm cells and a vegetative nucleus (Huh et al., 2008). It is reported that transposons that are otherwise silenced in other tissues by DNA methylation are activated and mobilized in pollen (Slotkin et al., 2009). In fact, the activation of transposons was restricted to vegetative nucleus, which can be allowed in a sense that genetic information stored in the vegetative nucleus is not inherited to the zygote (Slotkin et al., 2009; Huh et al., 2008). It was also reported that RdDM components were downregulated in pollen, which means DNA demethylation might be facilitated by passive mechanisms (Pina et al., 2005; Slotkin et al., 2009).

3. DNA demethylation in plants

DNA methylation is actually a reversible process, defined as DNA demethylation (Chan et al., 2005). DNA demethylation can be divided into passive and active mechanisms. When maintenance DNA methyltransferases, such as MET1, and CMT3 are downregulated, DNA methylation patterns are diluted through cell divisions. This non-enzymatical decrease of DNA methylation is referred to as passive DNA demethylation (Wu and Zhang, 2010). On the other hand, active DNA demethylation, necessitating a certain enzymatic activity, is conducted by DNA demethylases.

In plants, instead of directly removing a methyl group, which is implausible hard to occur, DNA demethylases are proposed to excise '5mC' from DNA and in the following steps BER enzymes complete the active DNA demethylation (Gehring et al., 2009). The BER pathway, usually required for repairing damaged or mismatched bases, includes serial enzymatic processes. DNA glycosylase flips out a damaged base from DNA and excises it, producing an Apyrimidinic/apurinic (AP) site and sugar lesion. This lesion is cleaved by AP lyase generating a single strand break. And AP endonuclease removes the sugar lesion to provide a 3'-hydroxyl group to allow DNA polymerase to incorporate a complementary base. Finally, DNA ligase seals a nick to complete the repair (David et al., 2007).

In Arabidopsis, four DME family demethylases are reported, known as DNA glycosylases that hydrolyze the N-glycosidic bond between base and ribose (Choi et al., 2002; Gong et al., 2002; Penterman et al., 2007). DME is a

bifunctional DNA glycosylase/AP lyase that directly excises 5mC from double stranded DNA to produce a nick and β-elimination product (Gehring et al., 2006). This product is thought to be a substrate of AP endonuclease and subsequent DNA polymerase and DNA ligase complete the DNA demethylation with the insertion of unmethylated cytosine into the abasic site (Gehring et al., 2006; 2009).

DME possesses three conserved domains including domains A, and B, and a glycosylase domain (Mok et al., 2010). Domain A was likely involved in DNA binding which was confirmed by gel mobility shift assay but the function of domain B is largely unknown. Glycosylase domain includes a helix-hairpin-helix (HhH) motif and a glycine/proline-rich loop with a conserved aspartate (GPD) that are common in other glycosylases such as MutY and Endonuclease III (EndoIII) in *E.coli* and human 8-oxoguanine DNA glycosylase (hOGG1) (Guan et al., 1998; Kuo et al., 1992; Bruner et al., 2000). DME randomly removed 5mC *in vitro* and its catalytic mutant D1304N did not excise 5mC *in vitro* and *in vivo* (Gehring et al., 2006; Choi et al., 2004). ROS1, the other DME family protein is also a 5mC glycosylase that facilitates DNA demethylation in long substrates and involves random surface sliding for target searching (Ponferrada-Marin et al., 2010; 2012). Another DME family proteins DML2 and DML3 appeared to catalyze demethylation but their specific role is unknown yet (Penterman et al., 2007).

DME is most studied DNA demethylase. DME is essential for endosperm development and gene imprinting, expressed in the central cell of female gametophyte (Choi et al., 2002). DME is known to induce maternal allele-specific activation of *MEA* by removing 5mC, which leads to endosperm development (Gehring et al., 2006). In heterozygous dme mutant, 50% of *MEA* allele was not

expressed in central cell, which resulted in 50% of defects in endosperm development and seed formation. This 50% of seed abortion phenotype was partially restored to 25% in transgenic plants that harbored functional DME (Choi et al., 2004).

DME was also expressed in the vegetative nuclei in pollen and thought to activate transposons and induce the production of small RNAs (Schoft et al., 2011). And these small RNAs might be transported to sperm cell and contribute to its genome stability by silencing transposons in the zygote after fertilization (Feng et al., 2010; Kohler et al., 2012). In addition, DME is thought to target several genes in the genome including FLOWERING WAGENINGEN (FWA), a homeodomain containing transcription factor that is essential for flowering time (Kinoshita et al., 2004). In fwa mutants, the promoter of the gene is hypomethylated and expressed that was otherwise highly methylated and repressed in wild type, which resulted in late flowering phenotype (Soppe et al., 2000; Kinoshita et al., 2004).

As described above, active DNA demethylation of plants is largely dependent on DME 5mC glycosylase, but no 5mC glycosylase was reported in mammals to directly remove 5mC. Recently, however, instead of 5mC glycosylase, a novel ‘base’ is proposed as an important intermediate for active DNA demethylation process in mammals.

4. The 6th base, 5-hydroxymethylcytosine (5hmC) and DNA demethylation in mammals

The ‘base’ is not only an essential component of DNA and RNA but also provides a means of functional specificity for biological processes. Adenine, guanine, cytosine and thymine store and inherit genetic information and 5mC also delivers epigenetic information, and thus it is reasonable to regard 5mC as a 5th base.

5hmC, newly emerging the 6th base, was primarily identified in T-even bacteriophage in 1953 and detected in mouse brain and liver DNA in 1972, which has been unnoticed for last four decades (Wyatt and Cohen, 1953; Penn et al., 1972). In 2009, however, substantial amounts of 5hmC were rediscovered in the mouse brain and embryonic stem cells, respectively, by two research groups and it was also present at lower levels in various tissues including liver, lungs, kidney, and muscle (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009).

During reprogramming of DNA methylation in mammals, immediately after fertilization, a genome-wide loss of DNA methylation occurs in the paternal genome, which appears to be active DNA demethylation (Wu and Zhang, 2010). Surprisingly, active DNA demethylation was linked with 5hmC accumulation. Along with a loss of 5mC level, 5hmC level increased in the paternal genome by TET3 that was 1800-fold more expressed in oocyte than in other tissues (Wossidlo et al., 2011).

5hmC is oxidized from 5mC by TET family proteins, 2-oxoglutarate- and Fe(II)-dependent dioxygenases (Tahiliani et al., 2009). TET proteins can further

oxidize 5hmC to 5fC and 5caC that were referred to as the 7th and 8th base and detected in mammalian DNA (Ito et al., 2011; He et al., 2011). 5hmC, 5fC, and 5caC are thought to be an important intermediates during DNA demethylation in mammals, supported by the report that 5caC is specifically removed by thymine DNA glycosylase (TDG) and that may be replaced with unmethylated cytosine by BER (He et al., 2011).

Unlike in plants, direct removal of 5mC is not reported in mammals. 5mC in mammals is converted to thymine by Activation-induced cytidine deaminase/apolipoprotein B mRNA-editing catalytic polypeptides (AID/APOBEC) and thymine is thought to be replaced with cytosine by TDG and BER (Bhutani et al., 2011). Emergence of 5hmC suggests another two possible DNA demethylation pathways involving 5hmC and cytosine derivatives. First is that 5mC is oxidized to 5hmC and further converted to 5hmU by AID/APOBEC, which is removed by TDG and single-strand-selective monofunctional uracil-DNA glycosylase 1 (SMUG1) (Bhutani et al., 2011). The other is that 5mC is converted to 5hmC and further processed to 5caC, which is removed by TDG or by spontaneous decarboxylation (He et al., 2011). A recent study shows that DNMT3, a *de novo* DNA methyltransferase, is involved in active decarboxylation of 5caC, which raises the possibility of DNA demethylation independent of BER (Chen et al., 2012).

The role of 5hmC is little studied. Impaired 5hmC pattern was detected in myeloid cancers with TET2 mutant, and 5hmC seems to play an important role in mouse embryonic stem cell self-renewal (Ko et al., 2010; Ito et al., 2010). Abnormal 5-hydroxymethylation level affects cell lineage specification in mouse

embryonic stem cells (Koh et al., 2011). However, the function of 5hmC is still elusive and more studies on 5hmC will require the improvement of 5hmC sequencing techniques (Booth et al., 2012; Yu et al., 2012).

MATERIALS AND METHODS

Preparation of plasmids

The construct of DMEΔN677ΔIDR1::lnk was generated as previously reported (Mok et al., 2010). Binding pocket mutant DME fragments, isolated from random mutagenesis study previously reported (Mok et al., 2010), were introduced into the pBG102 vector. Eventually the constructs were designed to produce Sumo-DMEΔN677 Q1362R mutant proteins.

Protein expression and purification

Plasmids were transformed into *E.coli* Rosetta2 (DE3) strain and cells were grown at 28℃ in two liters of LB containing 50 μg/ml each of kanamycin and chloramphenicol until the OD600 reached 0.4. Expression of DME was induced with 100 μM of IPTG at 18℃ for 16 hr. Cells were harvested by centrifugation at 7,000 rpm for 20 min at 4℃ and resuspended in 30 ml of lysis buffer (50 mM Tris-HCl pH 7.4, 100 mM NaCl, 10% glycerol, 0.1 mM DTT). 0.1 mM PMSF was added to prevent protease activity and cells were disrupted by sonication for 5 min (0.5x duty cycle) on ice. Cell extracts were centrifuged at 12,000 rpm for 25 min at 4℃ and the supernatant was filtered through nylon membrane with 0.45 μm pore (Advantec). The lysate was sequentially purified by three different types of columns: affinity column (His trap FF 5ml, GE Healthcare), ion exchange column (Heparin HP 5ml, 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 DTT) at -80℃ until use.

DNA substrates for labeling

35 mer oligonucleotides containing C, 5mC, 5hmC, 5fC, and 5caC were purchased by Midland Certified (TX, USA) and ones containing tymine and uracil were purchased by IDT (Table 1). 40 pmol of oligonucleotides were labeled by T4 polynucleotide kinase (Takara) with 30 μCi of [γ-^{32}P]ATP. Labled oligonucleotides were purified by Qiaquick Nucleotide Removal Kit (Qiagen) and annealed with complementary oligonucleotides. The mixture was boiled in water for 10 min and slowly cooled down for annealing to room temperature for at least 4 hr.

***In vitro* DNA glycosylase assay**

25 nM of oligonucleotides were incubated with 200 ng of DMEΔN677ΔIDR1::lnk in the glycosylase reaction buffer (10 mM Tris-HCl pH 7.4, 50 mM NaCl, 0.5 mM DTT) at 37℃ for 1 hr. Reaction was terminated by adding an equal volume of stop solution (98% formamide, 10mM EDTA, 0.2% Xylen cyanol FF, 0.2% of bromophenol blue) and boiled at 95℃ for 10 min. Reaction products were seperated on a 15% denaturing polyacrylamide gel containing 7.5 M urea and 1X TBE (Tris/ Boric acid/ EDTA).

Kinetics analysis

25 nM of 35 mer oligonucleotides containing each 5mC and 5hmC were incubated with 200 ng of DME in a time course for: 0.5, 1, 2, 3, 5, 10, 25, 60 min, respectively. Reaction was terminated by boiling and addition of 1 N NaOH at 95℃ for 10 min. The reactants were boiled at 95℃ for another 10 min in the equal

volume of stop solution, then separated on a 15% denaturing polyacrylamide gel described above.

Plant materials

Arabidopsis thaliana Columbia and Landsberg ecotype was used to prepare genomic DNA for HPLC analysis and 5hmC detection assay. Genomic DNA of diverse tissues from 5 weak-old plants was isolated by the standard CTAB method (Tel-Zur et al., 1999).

HPLC analysis

100 μg of genomic DNA was digested by 1000 U of Benzonase (Sigma), 500 mU of Phosphodiesterase I (Sigma), and 80 U of Shirimp Alkaline phosphatase (Takara) to produce nucleosides in DNA digestion buffer (20 mM Tris-HCl pH 7.9, 100 mM NaCl, 20mM $MgCl_2$) for 37℃ at 30hr. Nucleosides containing various bases are purified by 10 kDa cut-off column (Millipore) and concentrated by Speed-vac (Operon). Samples were analyzed by HPLC by the procedure described in reference (Pomerantz and McCloskey, 1990).

TLC analysis for 5hmC detection

DNA hydrolysis and thin layer chromatography (TLC) analysis were performed with modifications from (Ito et al., 2011). Briefly, 2 μg of genomic DNA was digested with 50 U of $Taq^{\alpha}I$ (New England Biolabs, NEB) and 10 U of Calf Intestinal Alkaline Phosphatase (NEB) was added to remove exposed phosphate groups. After purification, dephosphorylated products were end-labled by T4 Polynucleotide kinase with 20 μCi of [γ-^{32}P]ATP. Both DNA products from

dephosphorylation and kination were purified with Qiaquick Nucleotide Removal Kit. Labeled genomic DNA was digested with 1 U of Bal-31 (NEB) single strand specific nuclease, which generates dNMPs including dCMP, dmCMP and dhmCMP. These fragments were spotted on PEIcellulose TLC plates (Merck) and seperated for 16 hr. Spots were seperated in isobutyric acid: NH4OH: H2O = 66:2:20 for the first-dimensional and separated again in isopropanol: HCl: H2O = 70:15:15 for the second-dimensional TLC. The plates were exposed to an X-ray film or analyzed by phosphorimager scanning. 1.2 kb of synthetic oligonucleotides containing five Taq$^{\alpha}$I sites were generated by PCR amplification in the presence of dCTP, 5-methyl dCTP or 5-hydroxymethyl dCTP (Zymo research). PCR products were hydrolyzed and seperated on TLC plates by the same protocol as a control.

Table 1. Oligonucleotides used in *in vitro* glycosylase assay

Name	Sequence
R35C	5' GTACTGTGTGATACTATCGAATTCAGTATGATCTG 3'
R35mC	5' GTACTGTGTGATACTATmCGAATTCAGTATGATCTG 3'
R35hmC	5' GTACTGTGTGATACTAThmCGAATTCAGTATGATCTG 3'
R35fC	5' GTACTGTGTGATACTATfCGAATTCAGTATGATCTG 3'
R35caC	5' GTACTGTGTGATACTATcaCGAATTCAGTATGATCTG 3'
R35T	5' GTACTGTGTGATACTATTGAATTCAGTATGATCTG 3'
R35U	5' GTACTGTGTGATACTATUGAATTCAGTATGATCTG 3'
Reverse	5'CAGATCATACTGAATTCGATAGTATCACACAGTAC3'

The sequences of 35 nt oligonucleotides including cytosine, 5mC, 5hmC, 5fC, 5caC, thymine, uracil and their complementary sequence including guanine are introduced.

RESULTS

5hmC excision by DME

The substrate for the HhH superfamily DNA glycosylases that possess [4Fe-4S] cluster such as MutY and Endo III, is not restricted to a single base. It was previously reported that DME from Arabidopsis contains a [4Fe-4S] cluster motif, and remove both 5mC and thymine (Mok et al., 2010: Gehring et al., 2006). Because glycosylases usually have a slightly broad range of substrate specificity, I tested substrates other than 5mC and thymine (Bulychev et al., 1996; Dizdaroglu et al., 1993). 5hmC, an oxidized form of 5mC, is known as an important intermediate for DNA demethylation in mammals, but no 5hmC glycosylase was yet reported. And further oxidized forms, 5caC are thought to be removed by TDG in mammals (He et al., 2011). Thus, I tested glycosylase activity on these three modified cytosines (5hmC, 5fC, and 5caC) with DME 5mC glycosylase.

MBP-DMEΔN677ΔIDR1::lnk (Figure 2), the small DME protein where N terminal 677 amino acids and IDR1 were deleted and tagged with an MBP, was incubated with 35 mer oligonucleotides including cytosine, 5mC, 5hmC, 5fC, or 5caC to test for substrate specificity (Figure 3B). As expected, DME could not remove cytosine but efficiently excised 5mC (Figure 3A). Surprisingly, DME also recognizes and excises 5hmC, producing β- and δ-elimination products, suggesting that DME has 5hmC glycosylase activity. Also, DME showed little activity on 5caC, whereas 5fC was not excised (Figure 3A).

To compare the base excision efficiency of DME between 5mC and 5hmC, a time-dependent kinetics analysis was performed. With 25 nM of substrate

oligonucleotides, the excessive amount of DME produced approximately 18 nM of products for 5mC whereas 9 nM of products on 5hmC, which means DME has a two-fold higher excision rate for 5mC over 5hmC (Figure 3C).

5hmC was not detected in Arabidopsis by HPLC

Although 5hmC was removed by DME, the presence of 5hmC was not reported in plants before. 5hmC excision of DME raised the possibility that 5hmC might be present in plants. As 5hmC was present in brain, embryonic stem cells, and early embryogenesis stage in mammals (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009), it was necessary to find appropriate tissues as candidates that 5hmC might exist. A brain-like tissue was easily excluded because plants have no brain or neuronal cells. Tissues at early embryonic development stage was appropriate, and accordingly, floral buds of *Arabidopsis* were selected as a primarily candidate.

5hmC was detected and confirmed by several methods in mammals – 2D TLC, HPLC and Mass spectrometry (Kriaucionis and Heintz, 2009). First, I tried to detect 5hmC with an HPLC analysis. To produce the standard peaks, 1 kb of synthetic oligonucleotides including cytosine, 5mC, and 5hmC were generated by PCR amplification and digested and separated by HPLC. The peak of deoxycytidine (dC) emerged at 18.5 min, 5-hydroxymethyl-deoxycytidine (5hm-dC) at 19 min, 5-methyl-deoxycytidine (5m-dC) at 22.5 min corresponded to the previously reported data (Kriaucionis and Heintz, 2009) (Figure 4A).

Then, 100 μg of genomic DNA from floral buds of *Arabidopsis* was digested in the same manner and separated by HPLC. Eight peaks were produced

(Figure 4B, number 1-8) and fine peaks number (2,5,6, and 8) matched with dC, deoxyguanosine (dG), deoxythymidine (dT), and deoxyadenosine (dA), respectively. Peaks number 4 seems to be 5m-dC but unclear. Peaks number 1, 3, and 7 were not matched with standard peaks, suggesting the possibility of RNA contamination. And no peak for 5hm-dC was detected in this analysis.

5hmC was not detected in *Arabidopsis* by TLC

As HPLC analysis required too much amount of genomic DNA and was not sensitive enough, I tried another method to detect 5hmC – TLC based 5hmC detection assay. As a control, synthetic DNA used in the HPLC analysis was digested with Taq$^{\alpha}$I that recognizes TCGA sites. Digested DNA fragments were end-labeled and further digested into single nucleotides generating dNMPs including cytosine, 5mC and 5hmC, respectively (Figure 5). These were separated on TLC plates. 5hm-dCMP for control was detected between the spots for dCMP and dTMP (Figure 6A).

2 µg of genomic DNA of floral buds, leaves from Columbia and a shoot from Landsberg were used to detect 5hmC in *Arabidopsis* by 1D TLC analysis (Figure 6B). No 5hmC signal was detected between dCMP and dTMP.

To confirm the 1D TLC data, 2D TLC analysis was performed. The same genomic DNA of floral buds and leaves for 1D TLC was sequentially separated with two different solvents on TLC plates (Figure 7A, 7B). However, no 5hmC was detected, which suggests that 5hmC is not present in plants or in an extremely small amount, if any, and unique to animals.

Q1362R mutant DME prefers T over 5mC

DNA glycosylases usually have an active site pocket that recognizes and determines substrate specificity, whereas the residues comprising the active site pocket of DME are not well known. The [4Fe-4S] cluster in the HhH superfamily glycosylases was thought to play an important role in binding and recognizing substrates (Porello et al., 1998). DME also has four cysteine residues composing [4Fe-4S] cluster motif, and their single amino acid substitution mutants were inactive, implying that the [4Fe-4S] cluster motif is necessary for enzymatic activity and/or stability of DME (Mok et al., 2010).

Notably, amino acid residues at the upstream of [4Fe-4S] cluster motif were highly conserved among DME family proteins (Figure 8A), and to identify residues necessary for active site pocket in the glycosylase domain, single substitution mutants were isolated by random mutagenesis as previously reported (Mok et al., 2010). Mutants including Q1362R in the background of DMEΔN677 were cloned into the pBG102 vector and expressed, producing Sumo-tagged DMEΔN677 mutants. Purified mutant proteins were tested their substrate specificities such as cytosine, 5mC, and thymine.

Remarkably, Q1362R mutant DME protein in which glutamine at 1362 was substituted with arginine, displayed altered substrate preference (Figure 8B). The DMEΔN677 (Q1362R) mutant protein showed significantly reduced activity on 5mC/G, but inceased activity on T at T/G mismatch. This implies that the Q1362 residue is an essential member in the 5mC recognition pocket.

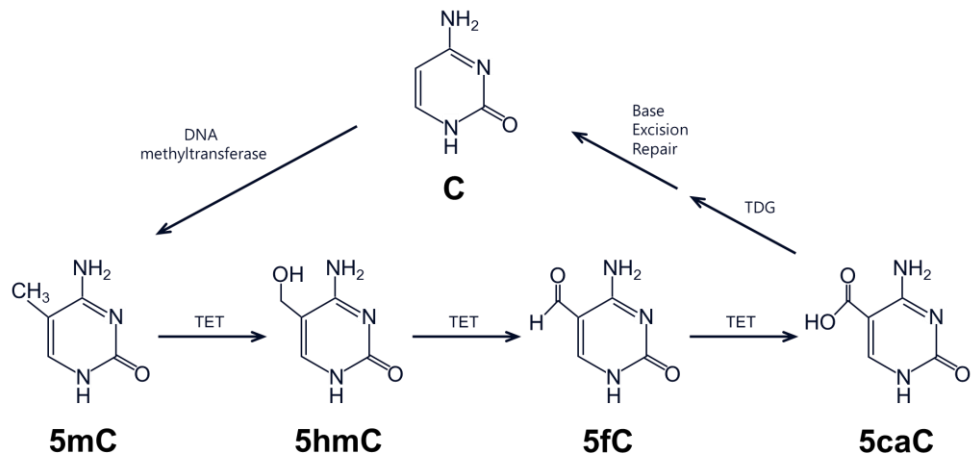

Figure 1. Structures of several modified cytosines.

5mC, generated by DNA methyltransferase, is oxidized to 5hmC by TET family proteins in mammals. TET proteins can also convert 5hmC to 5fC and 5caC. 5caC can be finally removed by TDG glycosylase with following base excision repair, which results in DNA demethylation. 5mC, 5-methylcytosine; 5hmC, 5-hydroxymethylcytosine; 5fC, 5-formylcytosine; 5caC, 5-carboxylcytosine.

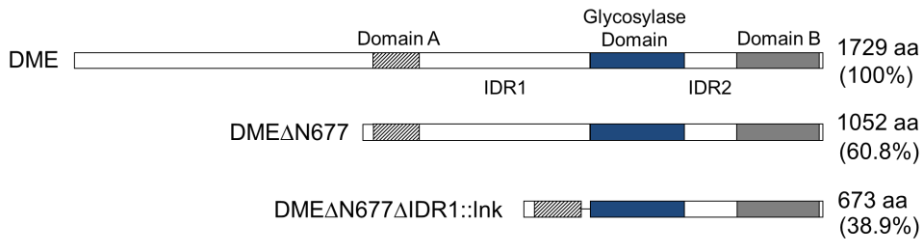

Figure 2. Construction of engineered DME.

Schematic diagrams of DME proteins with three conserved domains. N-terminal 677 amino acids were removed and the interdomain region IDR1 was replaced with a short linker sequence to produce DMEΔN677ΔIDR1::lnk where three conserved domains were still retained.

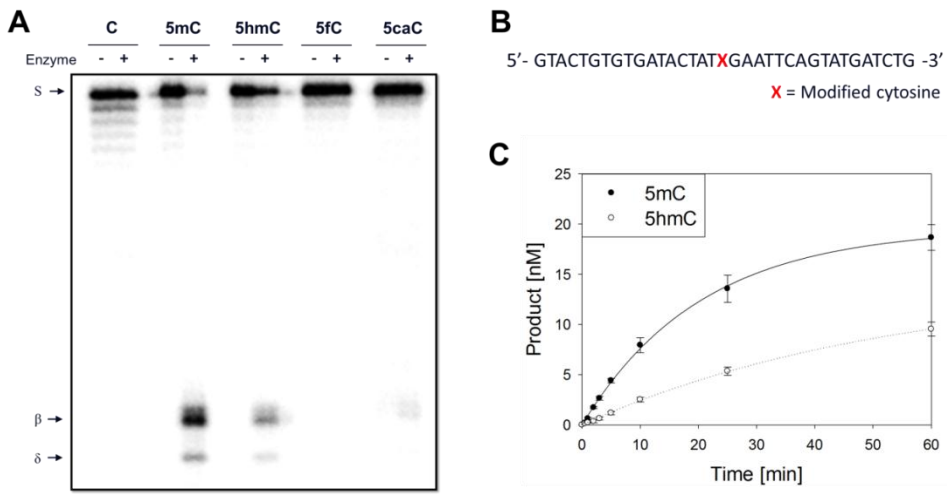

Figure 3. Excision of 5hmC by DME 5mC glycosylase.

(A) DME efficiently excised both 5mC and 5hmC, and showed little activity on 5caC, whereas 5fC was not excised by DME. (B) 35 nt oligonucleotide DNA substrates with a modified cytosine in the middle were labeled with ^{32}P. Position X indicates a modified cytosine paired with guanine. (C) Kinetic analysis was performed to compare the substrate specificity of DMEΔN677ΔIDR1::lnk. DME showed 2-fold higher preference for 5mC over 5hmC. Values are means ± standard deviation (SD) from three independent experiments.

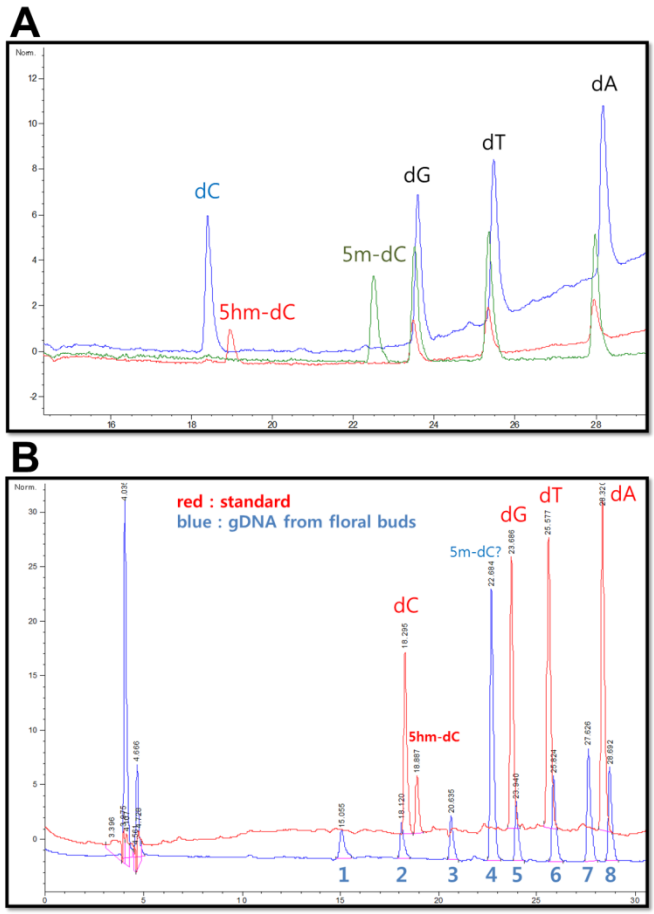

Figure 4. No 5hmC was detected by HPLC analysis.

(A) Nucleosides containing cytosine, 5hmC, 5mC, guanine, thymine, adenine were separated by HPLC. (B) Nucleosides of floral buds genomic DNA were separated by HPLC and compared with standard peaks. No 5hmC was detected.

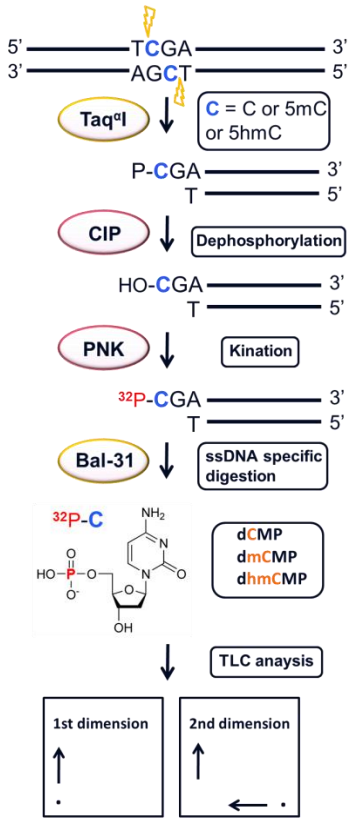

Figure 5. Schematic diagram of TLC analysis for 5hmC detection.

Genomic DNA from *Arabidopsis* was digested with Taq$^{\alpha}$I and exposed phosphate groups were susbstituted into labeled phosphates by Calf Intestinal Alkaline Phosphatase and Polynucleotide kinase. Labeled cytosines and cytosine derivatives were detached from genomic DNA by Bal-31 single strand specific nuclease, producing dCMP, methyl-dCMP and hydroxymethyl-dCMP. These fragments were seperated on TLC.

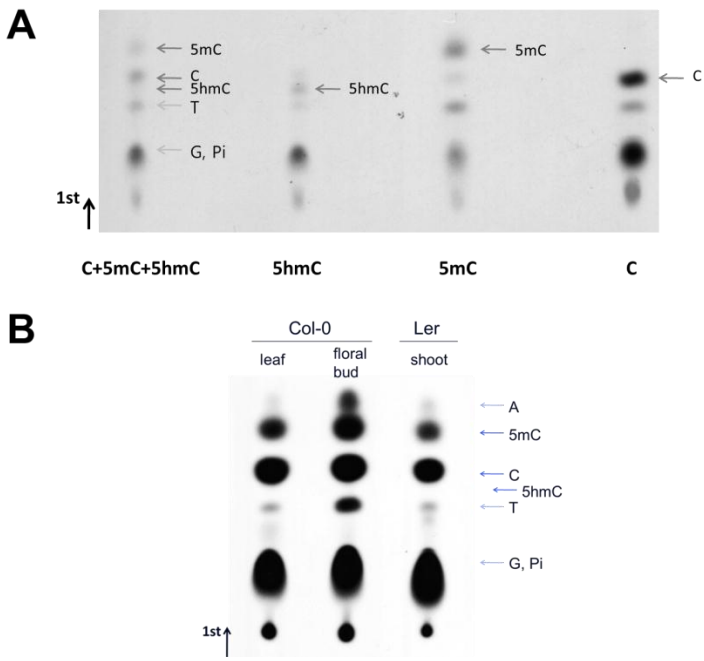

Figure 6. No 5hmC was detected from 1D TLC analysis.

(A) Synthetic DNA containing 5hmC, 5mC and cytosine were separated on TLC plates either respectively or as a mixture. (B) Genomic DNA from various tissue types (leaf, floral bud, shoot) was analyzed by TLC. No 5hmC was detected in any tissues.

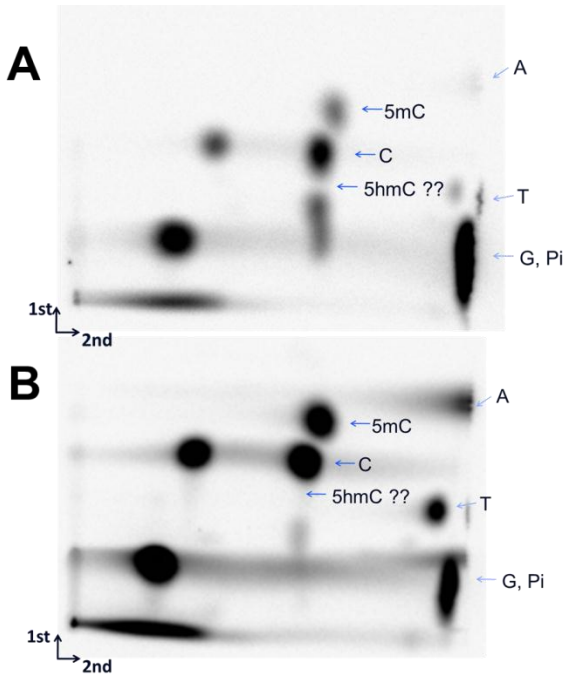

Figure 7. No 5hmC was detected from 2D TLC analysis.

To confirm the 1D TLC data, two-dimensional TLC (2D TLC) was performed with (A) Col-0 leaf and (B) Col-0 floral bud tissues, but no 5hmC was detected.

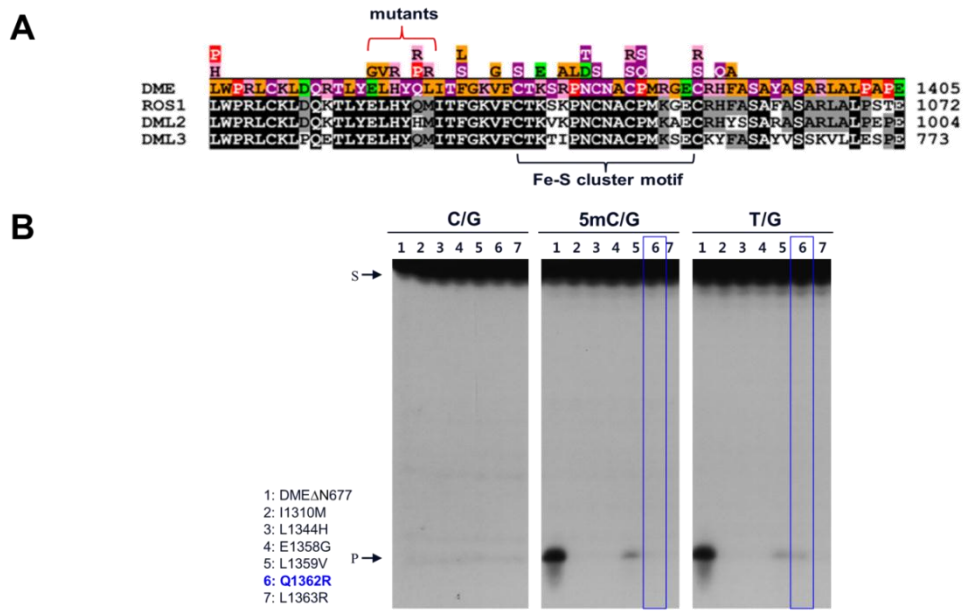

Figure 8. Substrate preference for T over 5mC of Q1362R mutant DME.

(A) To identify residues necessary for substrate specificity in the glycosylase domain, mutant DME proteins with single amino acid substitutions were generated by random mutagenesis. Several mutants with a substitution at the upstream of iron-sulfur cluster motif were selected and tested. (B) The DMEΔN677(Q1362R) mutant protein exhibited reduced activity on 5mC/G but increased activity on T/G mismatch.

DISCUSSION

5hmC excision by DME

DME is known to excise 5mC and thymine (Gehring et al., 2006), and other HhH superfamily DNA glycosylases possessing the [4Fe-4S] cluster also have a broad range of substrate specificity (Bulychev et al., 1996; Dizdaroglu et al., 1993). 5hmC, 5fC, and 5caC are thought to be intermediates for active DNA demethylation in mammals (Figure 1) (Ito et al., 2011; He et al., 2011). 5hmC was efficiently excised by DME 5mC glycosylase (Figure 3A), but the efficiency of enzymatic activity of DME on 5hmC was 2-fold lower than 5mC (Figure 3C). Concerning the electron cloud size of C5-cytosine, 5hmC is larger than 5mC (Rangam et al., 2012). It is reasonable that 5hmC might be slightly too large to fit in the recognition pocket, which may explain why DME displays 2-fold lower activity on 5hmC and much less activity on 5caC (Figure 3A). However, DME could not recognize or remove a formyl group of 5fC that concerns to be smaller than carboxyl group of 5caC, which implies that another factors may be important to determine the substrate specificity of DME such as a hydrogen bond between modified cytosine derivatives and amino acid residues in active site pocket of DME.

5hmC was not detected in *Arabidopsis* by HPLC

Four peaks (Figure 4B, number 2,5,6, and 8) matched with standard peaks, but the other peaks did not match with any DNA bases, which are likely to be contaminants of RNA according to the previous research (Gehrke et al., 1984). Peaks number 1 and 7 are thought to be a cytidine, and an adenosine, respectively, and peak number 3 might be a uridine. Peak number 4 could be 5-methyl-

deoxycytidine or a guanosine. But it might be the fusion of the 5-methyl-deoxycytidine and guanosine. No peak for 5hmC was detected.

5hmC was not detected in *Arabidopsis* by TLC

TLC is more sensitive than HPLC because signals can be detected with a small amount of genomic DNA. In leaf and floral bud tissues from *Arabidopsis*, no 5hmC signal was detected (Figure 6B). To confirm the data, the same tissue samples were analyzed by 2D TLC, but no 5hmC was detected in leaves and floral buds (Figure 7A, 7B). Taken together, 5hmC was not detected by HPLC and TLC, which implies that no or very little amount of 5hmC, if any, is present in *Arabidopsis* leaves and floral buds. In fact, in plants cytosine can be methylated on CG, CHG, and CHH sites and 55% of them are CG sites (Lister et al., 2008). However, experimental procedures cannot detect CHG or CHH sites, because Taq$^{\alpha}$I can only recognize and digest TCGA sequence (Figure 5). And the CG site is restricted to TCGA site that also restricts the detection level to a 1 in 16 chance, which means the target cytosine of this experiment was only 3% of total cytosines. In fact, 5hmC was detected in animals with only 5-6% because they used the same Taq$^{\alpha}$I restriction enzyme (Tahiliani et al., 2009). And they also detected lower amounts of 5fC and 5caC with the same technique (Ito et al., 2011). Thus, 5hmC does not appear to be present in *Arabidopsis* but unique to animals. In addition, homologs of TET proteins that catalyze the conversion of 5mC to 5hmC (Tahiliani et al., 2009), are not identified by BLAST search on TAIR (http://arabidopsis.org/) (Data not shown). Then why is 5hmC not present in plants? Both plants and animals have 5mC in common, which means they have evolved similar DNA methylation machineries (Law and Jacobsen, 2010). However, 5hmC appears to be

unique to animals, which implies that DNA demethylation might have evolved independently each other. Furthermore, unlike animals, plants have direct 5mC glycosylases (DME family proteins), and therefore, the intermediate of DNA demethylation, 5hmC, is not necessary.

Q1362R mutant DME prefers T over 5mC

A certain mutant DME showed altered substrate preference. The Q1362R mutant DME protein preferred thymine over 5mC (Figure 8), suggesting that Q1362R mutant may not recognize an NH_2 group of 5mC but oxygen at C=O on thymine. A recent study showed that ROS1 had altered preference for thymine over 5mC by replacing tyrosine with serine right beside the corresponding Q1362 residue (Ponferrada-Marin et al., 2010), implies that the residues around the Fe-S cluster is essential for recognition specificity in the active site pocket. DME excises both 5mC and thymine (Gehring et al., 2006). Manipulated DME protein with thymine preference can be used as a molecular biology tool to specifically remove thymine from T/G mismatch without affecting 5mC.

REFERENCES

Allis, C.D., Jenuwein, T., and Reinberg, D. (2007). Epigenetics (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press) p 2, 343

Almeida, R.D., Sottile, V., Loose, M., De Sousa, P.A., Johnson, A.D., and Ruzov, A. (2012). Semi-quantitative immunohistochemical detection of 5-hydroxymethyl-cytosine reveals conservation of its tissue distribution between amphibians and mammals. Epigenetics 7, 137-140.

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.

Bhutani, N., Burns, D.M., and Blau, H.M. (2011). DNA demethylation dynamics. Cell 146, 866-872.

Bonasio, R., Tu, S., and Reinberg, D. (2010). Molecular signals of epigenetic states. Science 330, 612-616.

Booth, M.J., Branco, M.R., Ficz, G., Oxley, D., Krueger, F., Reik, W., and Balasubramanian, S. (2012). Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. Science 336, 934-937.

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.

Bulychev, N.V., Varaprasad, C.V., Dorman, G., Miller, J.H., Eisenberg, M., Grollman, A.P., and Johnson, F. (1996). Substrate specificity of Escherichia coli MutY protein. Biochemistry 35, 13147-13156.

Cedar, H., and Bergman, Y. (2012). Programming of DNA methylation patterns. Annu Rev Biochem 81, 97-117.

Chan, S.W., Henderson, I.R., and Jacobsen, S.E. (2005). Gardening the genome: DNA methylation in Arabidopsis thaliana. Nat Rev Genet 6, 351-360.

Chen, C.C., Wang, K.Y., and Shen, C.K. (2012). The Mammalian de Novo DNA Methyltransferases DNMT3A and DNMT3B Are Also DNA 5-Hydroxymethylcytosine Dehydroxymethylases. J Biol Chem 287, 33116-33121.

Chen, Z.X., and Riggs, A.D. (2011). DNA methylation and demethylation in mammals. J Biol Chem 286, 18347-18353.

Cheng, X., and Blumenthal, R.M. (2008). Mammalian DNA methyltransferases: a structural perspective. Structure 16, 341-350.

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.

Choi, Y., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (2004). An invariant aspartic acid in the DNA glycosylase domain of DEMETER is necessary for transcriptional activation of the imprinted MEDEA gene. Proc Natl

Acad Sci U S A 101, 7481-7486.

Conaway, J.W. (2012). Introduction to theme "Chromatin, epigenetics, and transcription". Annu Rev Biochem 81, 61-64.

Crick, F. (1970). Central dogma of molecular biology. Nature 227, 561-563.

David, S.S., O'Shea, V.L., and Kundu, S. (2007). Base-excision repair of oxidative DNA damage. Nature 447, 941-950.

Dizdaroglu, M., Laval, J., and Boiteux, S. (1993). Substrate specificity of the Escherichia coli endonuclease III: excision of thymine- and cytosine-derived lesions in DNA produced by radiation-generated free radicals. Biochemistry 32, 12105-12111.

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

Fraga, M.F., Ballestar, E., Paz, M.F., Ropero, S., Setien, F., Ballestar, M.L., Heine-Suner, D., Cigudosa, J.C., Urioste, M., Benitez, J., Boix-Chornet, M., Sanchez-Aguilera, A., Ling, C., Carlsson, E., Poulsen, P., Vaag, A., Stephan, Z., Spector, T.D., Wu, Y.Z., Plass, C., Esteller, M. (2005). Epigenetic differences arise during the lifetime of monozygotic twins. Proc Natl Acad Sci U S A 102, 10604-10609.

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.

Gehring, M., Reik, W., and Henikoff, S. (2009). DNA demethylation by DNA repair. Trends Genet 25, 82-90.

Gehrke, C.W., McCune, R.A., Gama-Sosa, M.A., Ehrlich, M., and Kuo, K.C. (1984). Quantitative reversed-phase high-performance liquid chromatography of major and modified nucleosides in DNA. J Chromatogr 301, 199-219.

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, 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.

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., Xu, G.L. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science 333, 1303-1307.

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

Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C., and Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature 466, 1129-1133.

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.

Johnson, L.M., Bostick, M., Zhang, X., Kraft, E., Henderson, I., Callis, J., and Jacobsen, S.E. (2007). The SRA methyl-cytosine-binding domain links DNA and histone methylation. Curr Biol 17, 379-384.

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

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.

Ko, M., Huang, Y., Jankowska, A.M., Pape, U.J., Tahiliani, M., Bandukwala, H.S., An, J., Lamperti, E.D., Koh, K.P., Ganetzky, R., et al. (2010). Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature 468, 839-843.

Koh, K.P., Yabuuchi, A., Rao, S., Huang, Y., Cunniff, K., Nardone, J., Laiho, A., Tahiliani, M., Sommer, C.A., Mostoslavsky, G., Lahesmaa, R., Orkin, S.H., Rodig, S.J., Daley, G.Q., Rao, A. (2011). Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. Cell Stem Cell 8, 200-213.

Kohler, C., Wolff, P., and Spillane, C. (2012). Epigenetic mechanisms underlying

genomic imprinting in plants. Annu Rev Plant Biol 63, 331-352.

Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science 324, 929-930.

Kuo, C.F., McRee, D.E., Fisher, C.L., O'Handley, S.F., Cunningham, R.P., and Tainer, J.A. (1992). Atomic structure of the DNA repair [4Fe-4S] enzyme endonuclease III. Science 258, 434-440.

Lahmy, S., Bies-Etheve, N., and Lagrange, T. (2010). Plant-specific multisubunit RNA polymerase in gene silencing. Epigenetics 5, 4-8.

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.

Le, Q.H., Wright, S., Yu, Z., and Bureau, T. (2000). Transposon diversity in Arabidopsis thaliana. Proc Natl Acad Sci U S A 97, 7376-7381.

Lindroth, A.M., Shultis, D., Jasencakova, Z., Fuchs, J., Johnson, L., Schubert, D., Patnaik, D., Pradhan, S., Goodrich, J., Schubert, I., , Jenuwein, T., Khorasanizadeh, S., Jacobsen, S.E. (2004). Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE3. EMBO J 23, 4286-4296.

Lisch, D. (2009). Epigenetic regulation of transposable elements in plants. Annu Rev Plant Biol 60, 43-66.

Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., and Ecker, J.R. (2008). Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell 133, 523-536.

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 U S A 107, 19225-19230.

Penn, N.W., Suwalski, R., O'Riley, C., Bojanowski, K., and Yura, R. (1972). The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. Biochem J 126, 781-790.

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 U S A 104, 6752-6757.

Pina, C., Pinto, F., Feijo, J.A., and Becker, J.D. (2005). Gene family analysis of the Arabidopsis pollen transcriptome reveals biological implications for cell growth, division control, and gene expression regulation. Plant Physiol 138, 744-756.

Pomerantz, S.C., and McCloskey, J.A. (1990). Analysis of RNA hydrolyzates by liquid chromatography-mass spectrometry. Methods Enzymol 193, 796-824.

Ponferrada-Marin, M.I., Martinez-Macias, M.I., Morales-Ruiz, T., Roldan-Arjona, T., and Ariza, R.R. (2010). Methylation-independent DNA binding modulates specificity of Repressor of Silencing 1 (ROS1) and facilitates

demethylation in long substrates. J Biol Chem 285, 23032-23039.

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.

Ponferrada-Marin, M.I., Roldan-Arjona, T., and Ariza, R.R. (2012). Demethylation initiated by ROS1 glycosylase involves random sliding along DNA. Nucleic Acids Res 40, 11554-11562.

Porello, S.L., Cannon, M.J., and David, S.S. (1998). A substrate recognition role for the [4Fe-4S]2+ cluster of the DNA repair glycosylase MutY. Biochemistry 37, 6465-6475.

Rangam, G., Schmitz, K.M., Cobb, A.J., and Petersen-Mahrt, S.K. (2012). AID enzymatic activity is inversely proportional to the size of cytosine C5 orbital cloud. PLoS One 7, e43279.

Rodriguez-Paredes, M., and Esteller, M. (2011). Cancer epigenetics reaches mainstream oncology. Nat Med 17, 330-339.

Saze, H., Mittelsten Scheid, O., and Paszkowski, J. (2003). Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. Nat Genet 34, 65-69.

Schoft, V.K., Chumak, N., Choi, Y., Hannon, M., Garcia-Aguilar, M., Machlicova, A., Slusarz, L., Mosiolek, M., Park, J.S., Park, G.T., Fischer, R.L., Tamaru, H. (2011). Function of the DEMETER DNA glycosylase in the Arabidopsis thaliana male gametophyte. Proc Natl Acad Sci U S A 108,

8042-8047.

Slotkin, R.K., Vaughn, M., Borges, F., Tanurdzic, M., Becker, J.D., Feijo, J.A., and Martienssen, R.A. (2009). Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell 136, 461-472.

Soppe, W.J., Jacobsen, S.E., Alonso-Blanco, C., Jackson, J.P., Kakutani, T., Koornneef, M., and Peeters, A.J. (2000). The late flowering phenotype of fwa mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. Mol Cell 6, 791-802.

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

Tel-Zur, N., Abbo, S., Myslabodski, D., and Mizrahi, Y. (1999). Modified CTAB procedure for DNA isolation from epiphytic cacti of the genera Hylocereus and Selenicereus (Cactaceae). Plant Mol Biol Rep 17, 249-254.

Valinluck, V., Tsai, H.H., Rogstad, D.K., Burdzy, A., Bird, A., and Sowers, L.C. (2004). Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). Nucleic Acids Res 32, 4100-4108.

Wassenegger, M. (2000). RNA-directed DNA methylation. Plant Mol Biol 43, 203-220.

Watson, J.D., and Crick, F.H. (1953). The structure of DNA. Cold Spring Harbor

Symp Quant Biol 18, 123-131.

Woo, H.R., Dittmer, T.A., and Richards, E.J. (2008). Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in Arabidopsis. PLoS Genet 4, e1000156.

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.

Wyatt, G.R., and Cohen, S.S. (1953). The bases of the nucleic acids of some bacterial and animal viruses: the occurrence of 5-hydroxymethylcytosine. Biochem J 55, 774-782.

Yu, M., Hon, G.C., Szulwach, K.E., Song, C.X., Zhang, L., Kim, A., Li, X., Dai, Q., Shen, Y., Park, B., et al. (2012). Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. Cell 149, 1368-1380.

초 록

식물과 동물에서 5-메틸시토신은, 후생유전학적으로 유전자의 발현을 조절하여 유전자 각인, 트랜스포존 억제 등에 중요한 역할을 한다. 포유류 유전체에는 이 뿐 만이 아니라 5-하이드록시메틸시토신이 존재하는데 이는 TET 단백질의 산화작용에 의해 이루어지고 최근 6번째 염기로 일컬어지고 있다. 5-하이드록시메틸시토신은 쥐의 푸르키네 뉴런과 배아 줄기 세포에 많이 존재하고 포유류의 능동적DNA 탈메틸화 과정에서 중요한 중간산물의 역할을 한다. 그러나 이러한 5-하이드록시메틸시토신이 식물에도 존재하는지는 확실히 규명되지 않았다. 애기장대에서 DME는 5-메틸시토신을 직접 제거 하고 결과적으로 DNA 탈메틸화와 유전자 발현을 유도한다. 본 논문에서는 5-하이드록시메틸시토신이 시험관 내에서 DME에 의해 제거됨을 보이고 5-하이드록시메틸시토신이 식물에도 존재할 수 있다는 가능성을 얻어 그 여부를 확인하기 위해 2차원 얇은 층 크로마토그래피 실험을 수행하였다. 애기장대 유전체에서 5-하이드록시메틸시토신을 검출 하지 못하여 5-하이드록시메틸시토신이 매우 조금 존재하거나 존재하지 않는다는 추론을 했다. 또한,DME 단백질의 기질 특이성에 관여하는 Q1362R 돌연변이 단백질을 발견하였다. Q1362R은 5-메틸시토신보다 티민을 더 잘 제거 하였고 DME의 Fe-S 부분 상위에 존재하는 아미노산들이 기질 특이성에 관여할 것으로 생각된다.